



SHI 2022

Proceedings of the 18th Scandinavian
Conference on Health Informatics

August 22-24, 2022
Tromsø, Norway

SHI 2022

Proceedings of the 18th Scandinavian
Conference on Health Informatics

Editors

Dr. André Henriksen,
UiT The Arctic University of Norway

Associate Professor Elia Gabarron,
Østfold University College, Norway

Professor Vivian Vimarlund,
Linköping University, Sweden

August 22-24, 2022
Tromsø, Norway

Published by:

Linköping University Electronic Press, Sweden
Linköping Electronic Conference Proceedings No. 187
ISBN 978-91-7929-344-4 (PDF)
URL <https://doi.org/10.3384/9789179293444>
ISSN 1650-3686 (print)
ISSN 1650-3740 (online)

© The Authors, 2022

Preface

This proceeding presents the papers presented at the 18th Scandinavian Conference on Health Informatics - SHI 2022 in Tromsø, Norway on August 22-24, 2022.

The SHI conference is an annual scientific event attended by scientist and practitioners working in the field of Health Informatics.

The area of Health informatics is driven by development in technologies and informatics research that are advancing in parallel and for the integration of information and communication of health and social care. The area includes issues related to EHR, eHealth services and systems, intelligent systems, mobile health applications, telemedicine, assistive technology, artificial intelligence, and Internet of Things.

The field of Health Informatics in this conference is examined from a very broad perspective with participants presenting research outcomes and with focus on eHealth, informatics, Assistive technologies, machine learning, gaming, Internet of the Things, Implementation of eHealth services, EHR, Artificial Intelligence, management and organizational aspects, legal and social issues.

More than 40 manuscripts were received with a total of 32 accepted as full papers and 11 as extended abstracts after peer-review. The Editorial Board selected publications with relevance and quality of the field to provide a state-of-the-art of the area. Authors of these communications are researchers of 35 different affiliations (see below table), and 13 countries (in alphabetic order: Czech Republic, Denmark, Finland, Germany, India, Italy, Luxembourg, Netherlands, Norway, Pakistan, Sweden, Switzerland, and USA).



Affiliations	Country	Authors
UiT The Arctic University of Norway	Norway	42
Aalborg University	Denmark	37
Norwegian Centre for E-health Research	Norway	30
NTNU - Norwegian University of Science and Technology	Norway	12
Steno Diabetes Center North Denmark	Denmark	11
University of Agder	Norway	8
Arcada University of Applied Sciences	Finland	7
University Hospital Düsseldorf	Germany	7
Østfold University College	Norway	6
Aalborg University Hospital	Denmark	5
University of Oslo	Norway	5
Stockholm University	Sweden	5
Technical University of Munich	Germany	4
Novo Nordisk	Denmark	3
VTT Technical Research Centre	Finland	3
Islamia College University Peshawar	Pakistan	3
University of Luxembourg	Luxembourg	2
Oslo University Hospital	Norway	2
St. Olav's University Hospital	Norway	2
Illinois Institute of Technology	USA	2
Prague University of Economics and Business	Czech Republic	1
University College of Northern Denmark	Denmark	1
DIOPTIC GmbH	Germany	1
JADENT GmbH	Germany	1
Medical University of Brandenburg	Germany	1
Health Information Systems Programme	India	1
University of Pisa	Italy	1
Leiden University	Netherlands	1
DIPS ASA	Norway	1
Kristiania University College	Norway	1
Norwegian Quality Improvement of Laboratory Examinations	Norway	1
University of Bergen	Norway	1
University Hospital of North Norway	Norway	1
Uppsala University	Sweden	1
Geneva University Hospital	Switzerland	1

All contributions are peer reviewed. The Editorial board expect this proceeding will be of interest for researchers and practitioners working in the field of Health Informatics.

Among many other things, 2022 made us aware of the value of networking. The past years, due to the Covid pandemic, demonstrated the necessity for access to information to improve health and healthcare, and the need for professionals to use and apply that information to transform healthcare. At the same time, the past years made us aware of different approaches and of the importance of Health informatics for renew and innovation of health and social care to improve the delivery of health and social care services to benefit citizens and patients, as we were forced to adapt to a “new normal” in many areas.

The Editors would like to thank the members of the Scientific Program Committee, the Organizing Committee, and all reviewers, who carried out the very professional, through and objective refereeing of the scientific work in order to achieve a high-quality publishing achievement for this scientific event.

The Editors would like to thank Linköpings University press, Sweden for the publication of the proceedings as an Open access Book.

We expect the SHI 2022 provides a forum for research and researchers in Health Informatics, as well as to facilitate scientific discussion, share experiences, and promote collaboration and networking between researchers and practitioners across all countries.

In addition to the delegates presentations, SHI 2022 had the following keynote speakers: Prof. Sabine Koch from Karolinska Institutet (Sweden), Dr. Morten Hasselstrøm Jensen from Aalborg University (Denmark), Prof. and chief physician Audny Anke from the University Hospital of North-Norway (Norway), and research scholar & psychologist Henriette Michalsen, University Hospital of North-Norway (Norway).

Tromsø, Norway, 22-24 August 2022

André Henriksen, Elia Gabarron, Vivian Vimarlund

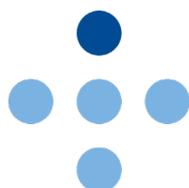
SPONSORS



UiT The Arctic
University of Norway



**The Research Council
of Norway**



HELSE NORD IKT



Editors

- André Henriksen, UiT The Arctic University of Norway
- Elia Gabarron, Østfold University College
- Vivian Vimarlund, Linköping University

Scientific Program Committee

- Stefano Bonacina, Karolinska Institutet
- Meghan Bradway, Norwegian Centre for E-health Research
- Jeppe Eriksen, Aalborg University
- Mariann Fossum, University of Agder
- Martin Gerdes, University of Agder
- Marte Hoff Hagen, NTNU - Norwegian University of Science and Technology
- Stine Hangaard, Aalborg University
- Gunnar Hartvigsen, UiT The Arctic University of Norway
- Ole Hejlesen, Aalborg University
- Lisa Korsbakke Emtekær Hæssum, Aalborg University Hospital
- David-Zacharie Issom, University of Geneva
- Morten Hasselstrøm Jensen, Aalborg University
- Erlend Johannessen, UiT The Arctic University of Norway
- Jonas Johansson, UiT The Arctic University of Norway
- Hege Mari Johnsen, University of Agder
- Lasse Krejberg, Aalborg University
- Dillys Larbi, Norwegian Centre for E-health Research
- Unn Manskow, Norwegian Centre for E-health Research
- Santiago Martinez, University of Agder
- Miroslav Muzny, Norwegian Centre for E-health Research

- Phuong Ngo, Norwegian Centre for E-health Research
- Monica Patrascu, University of Bergen
- Susanna Pelagatti, University of Pisa
- Keiichi Sato, Illinois Institute of Technology
- Gro-Hilde Severinsen, Norwegian Centre for E-health Research
- Line Silsand, Norwegian Centre for E-health Research
- Terje Solvoll, Norwegian Centre for E-health Research
- Juan Carlos Torrado Vidal, University of Bergen
- Ashenafi Zebene Woldaregay, Norwegian Center for Clinical Artificial Intelligence
- Eirik Årsand, UiT The Arctic University of Norway

Local Organizing Committee

- André Henriksen, UiT the Arctic University of Norway
- Gunnar Hartvigsen, UiT the Arctic University of Norway
- Julia Alexandra Sen, UiT the Arctic University of Norway
- Pietro Randine, UiT the Arctic University of Norway
- Erlend Johannessen, UiT the Arctic University of Norway
- Tina Rishaug, UiT the Arctic University of Norway
- Eirik Årsand, UiT the Arctic University of Norway

Table of contents

Articles

Session 1: Health Information Systems and central components

01. Hong Li, Torbjørg Meum and Miria Grisot.....2
Exploring Early-Stage Implementation of Digitally Enabled Remote Care: Case Studies from Norway and China
02. Olga Golburean, Ove Lintvedt, Espen Sobbakken Nordheim, Rune Pedersen, Luis Marco-Ruiz and Arild Faxvaag.....9
Does your EHR support a Learning Healthcare System? An exploration of possible indicators
03. Clara Bender, Marie Germund Nielsen, Ole Hejlesen and Stine Hangaard.....14
Assessing the use of telemedicine among people with diabetes: A Danish translation and cross-cultural adaptation of the Telemedicine Usability Questionnaire
04. Thomas Hellstén, Jonny Karlsson, Christian Häggblom and Jyrki Kettunen.....21
Towards Accurate Computer Vision-Based Marker Less Human Joint Localization for Rehabilitation Purposes
05. Anette Jøsendal and Trine Bergmo.....26
From paper-based to electronic prescribing of multidose drug dispensing — effects on pharmacy workload

Session 2: Human Factors in Health Informatics design

06. Yogita Thakral.....33
Digital Monitoring of Antibiotic Resistance (ABR) in Low- and Middle-Income Countries: A Narrative Literature Review
07. Andre Henriksen, Gerit Pfuhl, Ashenafi Zebene Woldaregay, David-Zacharie Issom, Eirik Årsand, Keiichi Sato and Gunnar Hartvigsen.....41
Expectations of users and non-users of wearable sensors and mobile health applications

08. Artur Liebrecht, Roman Bibo, Bastian Dewitz, Sebastian Kalkhoff, Sobhan Moazemi, Markus Rollinger, Jean-Michel Asfour, Klaus-Jürgen Janik, Artur Lichtenberg, Hug Aubin and Falko Schmid.....46
ARMAGNI: Augmented Reality Enhanced Surgical Magnifying Glasses
09. Erlend Johannessen, André Henriksen, Gunnar Hartvigsen, Alexander Horsch, Eirik Årsand and Jonas Johansson.....52
Ubiquitous digital health-related data: clarification of concepts
10. Julie Egmo, Katrine Holmstrup Sørensen, Tanja Fredensborg Holm, Ole Hejlesen, Morten Hasselstrøm Jensen and Stine Hangaard.....59
The Perceived Usefulness of a Contactless Sleep Monitor for Adults with Type 1 Diabetes: A Qualitative Study

Session 3: eHealth & mHealth Solutions

11. Sarala Ghimire, Santiago Martine and Martin Gerdes64
Self-imperative Care of Pregnancy using IoT Solutions
12. Michal Dolezel and Åsa Cajander.....69
Towards Mapping of Information Technology-Induced Alterations in Online Physicians' Professional Identities: A Conceptual Framework and Empirical Illustrations from Sweden
13. Tina Rishaug, André Henriksen, Anne-Marie Aas, Gunnar Hartvigsen, Kåre Inge Birkeland and Eirik Årsand.....77
Designing an e-Health Program for Lifestyle Changes in Diabetes Care
14. Pietro Randine, John Graham Cooper, Gunnar Hartvigsen and Eirik Årsand.....82
Towards a New Model for Chronic Disease Consultations
15. Magnus Stellander, André Henriksen, Henriette Michalsen, Audny Anke, Daniel Ursin, Santiago Martinez, Susanna Pelagatti, Keiichi Sato, Vebjørn Haugland, Erlend Johannessen, Juan Carlos Torrado and Gunnar Hartvigsen.....88
Sorterius - An augmented reality app for encouraging outdoor physical activity for people with intellectual disabilities

16. Hege Mari Johnsen and Jannicke Rabben..... 92
Student-staff Co-creation of Serious Games - Lessons Learned
17. Dillys Larbi, Henriette Lauvhaug Nybakke, Karianne Lind, Eirin Rødseth and Elia Gabarron..... 97
A scoping review of diabetes telemedicine research in Norway

Session 4: Citizen-centric Health Informatics

18. Lisa Korsbakke Emtekær Hæsum and Ole Kristian Hejlesen.....103
Development of an Interactive Communication Model with Integrated Teach-Back – using a web-based IT solution to create synergy between research and practice
19. Elia Gabarron, André Henriksen and Anders Nordahl-Hansen....111
Social media, physical activity and autism: better or bitter together? A scoping review
20. Dillys Larbi, Helle Sandsdalen, Elia Gabarron, Eirik Årsand and André Henriksen.....118
User preferences for a physical activity chatbot connected to an activity tracker and integrated into a social media platform
21. Shaedul Islam, Hüseyin Demirci and Gabriele Lenzini.....124
Privacy-preserving Polygenic Risk Scoring using Homomorphic Encryption
22. Jeppe Eriksen, Kristina Tornbjerg Eriksen and Pernille Bertelsen.....128
Citizens' use of Health Information Technology between 2013-2021 in Denmark: A longitudinal study

Session 5: Security, Privacy, and Ethics

23. Henrik Skaug Sætra, Anders Nordahl-Hansen, Eduard Fosch-Villaronga and Christine Dahl136
Normativity assumptions in the design and application of social robots for autistic children

Session 6: AI and ML in eHealth applications and services

24. Jaakko Lähteenmäki, Juha Pajula and Emmi Antikainen.....141
Development of medical applications based on AI models and register data – regulatory considerations
25. Jonas Tana, Bao Cai and Leonardo Espinosa-Leal.....147
A temporal analysis of depression related tweets - a case study in Finland
26. Seshathiri Dhanasekaran, Anders Andersen, Randi Karlsen, Anne Håkansson and André Henriksen.....154
Data collection and smart nudging to promote physical activity and a healthy lifestyle using wearable devices
27. Casper Knudsen and Pernille Bertelsen.....159
Maintaining Data Quality at the hospital department level – The data work of medical secretaries
28. Alexander Dolk, Hjalmar Davidsen, Hercules Dalianis and Thomas Vakili.....166
Evaluation of LIME and SHAP in Explaining Automatic ICD-10 Classifications of Swedish Gastrointestinal Discharge Summaries
29. Andrius Budrionis, Taridzo Chomutare, Therese Olsen Svenning and Hercules Dalianis174
The Influence of NegEx on ICD-10 Code Prediction in Swedish: How is the Performance of BERT and SVM Models Affected by Negations?
30. Jo Inge Arnes, Alexander Hapfelmeier and Alexander Horsch.....179
Autostrata: Improved Automatic Stratification for Coarsened Exact Matching
31. Miroslav Muzny, Meghan Bradway, Håvard Blixgård and Eirik Årsand.....187
Designing, implementing, and testing a modern electronic clinical study management system – the HUBRO system

32. Meghan Bradway, Henriette Nybakke, Stine Ingebrigsten and Kari Dyb.....195
Who are the “Hard-to-Reach” groups in chronic-health and health technology research? – A scoping review

Abstracts

Abstract session

33. Siranush Karapetyan and Alexander Hapfelmeier.....204
The concept of conditional method agreement trees with single measurements per subject
34. Camilla Heisel Nyholm Thomsen, Stine Hangaard, Thomas Kronborg, Peter Vestergaard, Ole Hejlesen and Morten Hasselstrøm Jensen.....206
Time for using Machine Learning for Basal Insulin Dose Guidance for People with Type 2 Diabetes? Preliminary Results from a Systematic Review
35. Nynne Sophie Holdt-Caspersen, Stine Hangaard Casper, Peter Vestergaard, Claus Dethlefsen and Morten Hasselstrøm Jensen.....208
The Adherence to Oral Second Line Antidiabetic Medication in People with Type 2 Diabetes - A Protocol for a Systematic Review
36. Gilmini Dantanarayana and Prasad Jayaweera.....210
Three Dimensional Ontological Perspective for Designing Healthcare Services
37. Lovisa Jäderlund Hagstedt, Maria Hägglund, Göran Petersson, Nadia Davoody and Helena Hvitfeldt.....212
Impact of the Covid-19 pandemic on use of Video consultations in a Swedish Primary care setting
38. Tanja Fredensborg Holm, Julie Egmoose, Katrine Holmstrup Sørensen, Ole Hejlesen, Morten Hasselstrøm Jensen and Stine Hangaard.....214
Exploring the validity of a contactless monitor used to measure vital parameters during sleep: A pilot study

39. Molly Gibson, Magne Halvorsen, Mikkel Angelo Anchissi Joner, Mari Lehne, Elise Rabassa Stautland, Gustaf Svensson, Hanna Thevik, Georgia Fehler, and Stefan Hochwarter.....216
Pilot Testing and Evaluation of Participatory Patient Record in Psychiatric Care
40. Jannie Toft Damsgaard Nørlev, Morten Hasselstrøm Jensen, Ole Hejlesen, Tinna Björk Aradóttir, Nicholas William Ciccone and Stine Hangaard.....218
Quantification of insulin adherence in adults with insulin-treated type 2 diabetes: a protocol for a systematic review
41. Katrine Holmstrup Sørensen, Tanja Fredensborg Holm, Julie Egmose, Ole Hejlesen, Morten Hasselstrøm Jensen and Stine Hangaard.....220
Exploring the user-friendliness of a contactless monitoring system used for sleep monitoring: A usability study
42. Marte Hoff Hagen.....222
Exploring Digital Psychosocial Follow-up for Survivors of Childhood Critical Illness
43. Muhammad Kamran, Mohib Ullah, Ali Shariq Imran and Muhammad Sajjad.....224
Automatic Report Generation for Medical Images

Articles

Exploring Early-Stage Implementation of Digitally Enabled Remote Care Case Studies from Norway and China

Hong Li*, Torbjørg Træland Meum*, Miria Grisot
University of Oslo, Department of Informatics, Norway hong.li@ifi.uio.no

*Equally contributed authors

Abstract

The onset of COVID-19 has the accelerated adoption and diffusion of remote care services to address the challenges of postoperative follow-up and support patients receiving care at home. Despite the great importance of this emerging technological solution, the field is still in its infancy. This paper presents findings based on feedback from an early-stage implementation of digitally enabled remote care in Norway and China, respectively. The same solution has been implemented in the two countries with a particular focus on patients with chronic conditions and postoperative rehabilitation.

Keywords

Remote care, digital technology, case study, early-stage implementation

1 INTRODUCTION

In this paper, we report from ongoing research on the implementation of digitally enabled remote care. We refer digitally enabled remote care to the integration of digital technologies into work practices of remote care that not only focuses on monitoring an individual's vital signs, mobility and general safety in the home [5], but also puts value on training an individual to utilise relevant skills to control or reduce the impact of disease on physical health status [7]. Despite the realm represents an exciting empirical field and it is becoming of the utmost importance to the world at large, digitally enabled remote care has proved to be challenging to successfully implement and use by healthcare organisations [6]. Digital technologies for remote care are collaborative tools supporting the communication and coordination of patients and health professionals. Their use fundamentally changes the traditional interactions between patients and health professionals [3][4], and should be designed to support a variety of user needs [32]. They are also tools that transform the role of patients from care receivers to active data generators [10] co-producing meaningful data [19]. Current literature on digitally enabled remote care has studied the design and use of these technologies for specific group of patients with specific health conditions. However, we still have limited insights into the challenges that emerge from the transformation of organisational practices and technology in use in remote care.

This work encompasses an existing digitally enabled remote care solution (see Figure 1), which was initially developed in Norway to offer remote primary care for patients with chronic diseases. It consists of a web-based patient monitoring portal for health professionals to handle digital home follow-up, and a mobile application

specifically designed for patients with a set of integrated measuring devices. The same solution has been implemented in Norway and China. Several pilots are currently carried out in the two countries in specialised care offering remote care services to hospital patients. The aim of our research is to understand what are the lessons learned from the early-stage implementation of digitally enabled remote care in Norway and China. Thus, we address the following research question: what challenges emerge when implementing the same digital solution for remote care services in Norway and China?

This paper is structured as follows. Firstly, we position the study in relation to relevant literature on digital technologies for remote care. Secondly, we present our conceptual lens of sociotechnical perspectives on remote care and argue how our study can address the research gaps. Thirdly, we describe the research methodology, followed by the study setup in Norway and China, respectively. Then, we reveal findings based on some early feedback we have received from the early-stage implementation in the two countries. Lastly, we conclude the paper with a brief summary and our future work.

2 SOCIOTECHNICAL PERSPECTIVES ON REMOTE CARE

Digital solutions are considered as a key enabler to promote health, prevent diseases, and provide patient-centred care that meet citizens' needs [15]. International strategies have called for a paradigm shift in the way healthcare is organised [33]. In particular, the need for new models that enable patient-centred services and a shift from hospital-centred systems to more integrated care structures have been highlighted. The use of remote care technologies is considered a promising solution for person-centred health services and has increasingly been used for people with

The 18th Scandinavian Conference on Health informatics, Tromsø, Norway, August 22-24, 2022. Organized by UiT The Arctic University of Norway. Conference Proceedings published by Linköping University Electronic Press at <https://doi.org/10.3384/ecp187>. © The Author(s). This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>

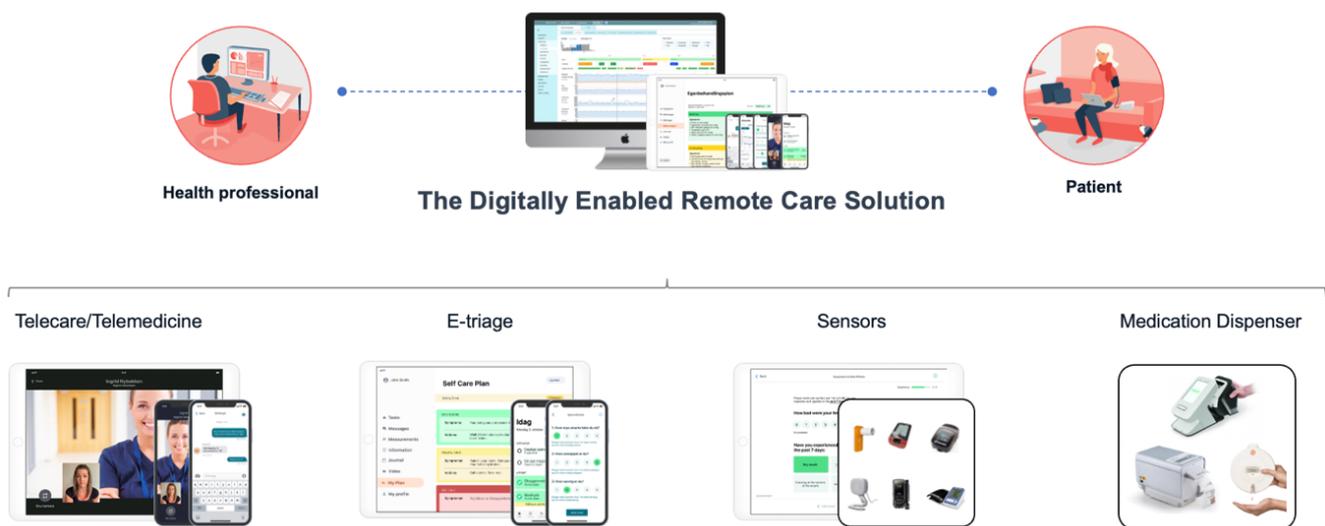


Figure 1. Overview of the digitally enabled remote care solution (image source: [12]).

chronic diseases. Studies on remote care are diverse and have increased considerably in recent years. A systematic review and meta-analyses have showed promising results in effectiveness and cost-effectiveness of e-health interventions to patients with somatic diseases [14]. However, several of the studies included in the review also showed inconsistent evidence. Similar studies have pointed out the challenges of implementing telehealth programmes in daily practice [31] and the need to theorising the distinct processes required to achieve widespread adoption [22]. It is beyond doubt that the use of digital technologies has the potential to improve health services for patients with chronic diseases. Nevertheless, several studies have identified that the uptake of digital remote care in daily practice is slow [6][13][31] and highlighted the need for a better understanding of the mutual relationship between technology and organisational practice [17].

Furthermore, current health data infrastructures are typically segregated in silos, differing in purposes, data sharing methods, regulatory compliance practice, and the users' roles. The trend of preventive and personalised healthcare implies that health data increasingly can be sourced from nonconventional health data sources, such as patients' personal devices, living environment, and Internet industry. Remote care solutions represent early examples of this. Even though some research have validated the potential significance of using patient-generated health data (PGHD) for improving healthcare performance [11] [26], most of these PGHD have not yet been widely used in a well-organised and structured way.

In this study, we draw on the sociotechnical approach to Health Informatics that has emphasised the contextual nature of health information [8] and consider design and implementation as an iterative, incremental change process [9]. Several studies in this tradition have highlighted the contextual aspect of translating new interventions in health care [24], co-creation of health services [16], and cultivation strategies that includes process-orientation, user mobilisation, and learning [18]. Moreover, Greenhalgh et al. have argued that we need to study the mutual relationship between human actions and the wider

organisational and system context [17]. They have further developed a theory-informed framework that can be used to identify the degree of complexity of a technology programme and to identify potential challenges of adoption and implementation in real-world settings [17].

Overall, there are substantial gaps between the potential capabilities of digital remote care technologies to provide more efficient healthcare and the current adoption in daily practice. Previous studies have identified some of the challenges of implementing digital remote care and highlighted the need to explore the dynamic interaction between organisational practices and technologies in use. We attempt to address these gaps by unfolding lessons learned during the early-stage implementation of digitally enabled remote care in relatively large-scale pilot studies in Norway and China.

3 RESEARCH DESIGN

3.1 Research methodology

The research is designed as multiple case studies over three years (2021-2024) to investigate how digital technologies can be scaled, adapted, and evaluated to ensure high-quality remote care in Norway and China. The intention is to use the cases for contrasting key decisions about the sociotechnical set up of the pilots in the different hospitals in Norway and China. In this paper we focus on the early stage of this process. Overall, both the pilot implementations and our research activities have been delayed due to the COVID-19 pandemic.

In the early stage, empirical data have been collected from documents and interviews. As illustrated in Table 1, we have reviewed relevant documents such as publicly available documents, internal project reports, policies, and strategies at national level. In addition, we have also carried out ten interviews so far with our key stakeholders, for instance, health professionals from the hospitals and staff from the vendor organisation. In the next phase of the research, we will carry out observation in the hospitals to unfold how the solution is used by the health professionals and patients in the two countries

Country	Case study	Level of care	Patient type	Source of the empirical data so far
China	A hospital	Tertiary care	Postoperative thoracic patients	<ul style="list-style-type: none"> • Internal project report written by the hospitals (2) • Semi-structured interview with a project coordinator from the vendor organisation (1)
	B hospital		Postoperative cardiac patients	
Norway	1 municipality	Primary care	Patients with chronic conditions	<ul style="list-style-type: none"> • Publicly available documents (3) and videos (4) • Internal project documents (2) • Semi structured interview with nurses (3), project manager (1), vendor (2), IT-department (3)
	3 hospitals	Secondary care		

Table 1: Current status of the ongoing case studies.

Notes and recordings from the interviews have been reviewed and transcribed to identify key patterns and themes. We have followed the general qualitative coding principles [27] and the guidelines of thematic analysis [20] for data analysis. Ethical approval has been granted by the Norwegian Centre for Research Data (Ref. 600335). Participation is voluntary and written informed consent has been obtained from all research participants. All data have been anonymised and securely stored.

3.2 Ongoing case studies

Table 1 shows our case studies from Norway and China that represent different cultural, organisational, and clinical settings. For instance, the Chinese case has a particular focus on postoperative remote care and rehabilitation of thoracic and cardiac surgery patients. On the other hand, the Norwegian case has a broader scope and includes patients with chronic conditions who need follow-up by health professionals.

3.2.1 The Norwegian case studies

Digitally enabled remote care in Norway has mainly been driven by national strategies and the need for sustainable care services in the future. The use of digital technologies is considered a key measure for realising more coordinated and person-centred services [30]. A national programme for the use of welfare technologies was established and a total of six municipalities in Norway participated in the programme for the use of digital remote care. The participation involved the design of a new digital service that was carried out in collaboration amongst municipalities, suppliers, hospitals, service designers, and project managers from the Norwegian Directorate of Health. A cooperation agreement with GPs and staff at the hospitals was established in the municipalities. During the project period, workshops were organised with local actors to develop guidelines, procedures, and adjustment of the digital solution.

The overall goal of the national programme was to gain knowledge about the use of digital remote care to provide national recommendations for further implementation of the service. The target group for digital remote care has mainly been patients suffering from chronic diseases, such as chronic obstructive pulmonary disease, heart disease, diabetes, mental disorders, and cancer. The aim of the national programme was to gain more knowledge about the

effects and benefits of the use of digital remote care and the evaluation was carried out as randomised clinical trials. A total of 735 patients were recruited for the study and the evaluation showed promising findings such as increased safety and coping for patients, as well as increased insight into the course of the disease. Overall, the experience gained from the national programme has provided useful knowledge and experience that will be continued in further scaling of digital remote care to new user groups and health services across organisational units in Norway.

3.2.2 The Chinese case studies

The Healthy China 2030 blueprint has been established as a national strategy since 2016 [28], which has fast-tracked the development of digitally enabled remote care. Against the backdrop of COVID-19, a digital health care boom has been sparked in China. The solution for remote care is considered a timely proposal and has been implemented in two hospitals in Shanghai, China since 2021. The goal of the pilots was to enrol 1,000 patients who are undergoing thoracic and cardiac surgery rehabilitation, and establish a corresponding database for patient follow-up. By November 2021, A hospital has enrolled a total of 190 postoperative thoracic surgery patients, of which 79 cases in the control group and 111 cases in the experimental group, accounted for 38% of the overall plan. B hospital has enrolled a total of 194 postoperative cardiac surgery patients, of which 103 cases in the control group and 91 cases in the experimental group, accounted for 38.8% of the overall plan. According to the types of disease, there were 108 cases (55.8%) of coronary artery bypass graft, 77 cases (39.5%) of various aortic replacement operations, 9 cases (4.8%) of aortic endovascular repair operations, and 65 cases (33.3 %) of combined heart valve surgery. All the above-mentioned patients have completed the informed consent before discharge. In terms of the project personnel, A hospital invested twelve medical doctors, specifically, two of them had senior professional titles, one had a junior professional title, and the rest of them were medical interns, namely, six master students, two doctoral students, and one postdoctoral resident doctor. The medical interns participated in the data collection, follow-up, data analysis, equipment management, and other matters of the patients enrolled in the project. B hospital also invested twelve personnel, including five doctors and seven nurses. Additionally, there were one doctor, three nurses, and three

technicians invested by its cooperative unit to help with professional interpretation of the electrocardiogram results.

4 FINDINGS AND DISCUSSION

In this section we present the findings about the early-stage implementation of the solution. This is a critical stage where key decisions shaping the long term of the implementation process are taken.

4.1 Early feedback from the Chinese cases

4.1.1 The experience so far from the Chinese hospitals' perspectives

Our findings show that there are differences pertaining to the feedback from the two hospitals in China where the solution had been implemented. We noted several factors that might have influenced their respective user experience. The factors are related to organisational and leadership level, which is in line with the factors that have been reported in the previous studies [1][6]. The purpose of enabling the solution for remote care is to bring patients and clinicians closer together. Nevertheless, the digital technology itself is merely a tool to support and enhance the two hospitals' existing healthcare services. What is ultimately provided to their patients is health care service per se, which needs their own clinical teams to deliver. Therefore, it is crucial that the two hospitals equip themselves with the competencies needed for delivering high-quality health care services to their patients and are proactive in investing time and effort to learn, use, and implement the digital technology.

B hospital had a relatively stronger clinical team and was acting proactively in implementing the solution. The hospital had invested sufficient manpower and time to learn the technology, and consequently mastered it and managed to use it in a flexible and creative way. As a result, their experience and feedback with regard to the implementation of and experience with the solution turned out to be positive. For example, it was remarked by their clinical team that a healthy and harmonious doctor-patient communication model was built while accumulating experience in patient rehabilitation. The solution allowed their patients to enjoy health care without having to leave their homes, which not only saved the cost of inpatient treatment, but also saved time for patients who were using the outpatient services; and hence it enhanced their patients' user experience. From B hospital's perspective, the medical resources can be saved for patients who are more urgently in need of hospitalisation, which ultimately saves the country's medical expenditures. On the other hand, the user experience of and feedback from A hospital tended to be less satisfying. One representative of the vendor organisation commented that this was partly due to having a weaker clinical team on the A hospital side. The solution was used by medical interns, who did not have rich clinical experience, rather than the senior doctors in A hospital. Moreover, in comparison to B hospital, they did not invest enough manpower and time to cultivate their staff to use the solution properly.

4.1.2 The experience so far from the Chinese patients' perspectives

According to an internal report written by the Chinese hospitals [29], the majority of the patients were satisfied

with the solution and several patients showed their interest in extending the duration to use the application. With the popularisation of smartphones, most patients—even some elderly patients—could master the use of the remote monitoring mobile application after getting an easy hands-on training. For postoperative cardiac surgery patients, some of them would choose to be transferred to secondary hospitals or rehabilitation hospitals to continue treatment for a period of time after being discharged, which might add another cost to them. Using the remote monitoring application could give the patients a sense of security, making it possible for patients who did not unnecessarily need to continue treatment in the hospital to be able to recover at home. Meanwhile, this can also reduce the risk of cross-infection in the hospital. For postoperative thoracic surgery patients, by pushing notifications of the rehabilitation plan via patients' end device and urging them to complete it through daily exercise tasks, a result of lung function recheck conducted by B hospital indicated a better recovery for patients. This shows that there is a need for remote monitoring, and it is positive in saving medical expenses and helping patients to recover.

Interestingly, one key challenge regarding the use of end devices was revealed. In the Norwegian context, a tablet with one application specifically designed for patients would be handed over to them. In this case, patients would have the same end device with a single application mode. However, patients in the Chinese context had to download the application for patients on their own end device, which was often a smartphone with multiple applications. This can be problematic, for example, for senior patients who are less tech savvy, especially when the application is not supported on the smartphone operating system version on their end devices. Furthermore, when the application for patients is not the only application on an end device, it will not function in a purpose-driven mode where nonessential applications and device settings are rendered inaccessible. This can cause distractions and increase data security risks which ultimately minimise the optimum use of the application for patients.

4.1.3 The major challenges that the vendor is facing in the Chinese market

Developing suitable business models that can adapt to a variety of market needs in different contexts has been suggested to be one of the key factors to succeed in scaling up digital remote care technologies [1][6]. Similarly, the digital technology solution in question needs a business model to make it commercialisable and sustainable in China. The project coordinator from the vendor organisation explained that the architecture of the system was standardised except offering different operating languages. It requires customisation and localisation to scale the infrastructure and service of the solution in China and make it adaptable to the Chinese market.

Another challenge was that the digital technology supplier in China was short-staffed due to a lack of financing. When a technical issue is discovered, it often cannot be fixed in a timely manner. The issue must be reported to the head quarter first, and then the staff based in Norway will try to solve the problem. The time zone difference between Norway and China makes it longer to solve a problem, and the user experience is therefore influenced by this.

Furthermore, the staff in the head quarter mainly come from Norway or other European countries, and hence might have difficulty understanding the Chinese needs due to the cultural differences.

4.2 Early feedback from the Norwegian cases

4.2.1 The experience so far from the Norwegian patients' perspectives

Digital remote care has been introduced into municipal health services as part of a national trial in Norway. The target group in the national trial has been patients with chronic diseases, such as diabetes, chronic obstetric diseases, and heart diseases. The evaluation of the national trial shows promising results, for instance, increased patients' safety and abilities to manage their chronic diseases; furthermore, the digital solution is perceived to be user-friendly and easy-to-use by the enrolled patients after receiving a short training [30]. The digital platform has built-in functionality that facilitates self-management of patients' chronic diseases, that is, 'the individual's ability to manage the symptoms, treatment, physical and psychosocial consequences and life style changes inherent in living with a chronic condition' [7] (p. 178). Some key features in the mobile application used by patients are a set of integrated medical measurement devices, a questionnaire to report symptoms, and a self-treatment plan that provides an overview of measures that the patient needs to perform based on vital measurements and symptoms. Self-reported data in the patient app becomes available in the digital platform and the nurses employed at the follow-up centre provide feedback to patients based on self-reported data. However, sharing information among the actors involved in the follow-up of patients remains a challenge that needs to be addressed in the further implementation of digital remote care. GPs and specialists at the hospital usually do not have access to the digital platform for remote care, and this interaction can be demanding as commented by an employee in the municipal health service: 'For the patients, it is often very demanding that they have to retell everything, the whole history since the last time they were at a check-up, so it is difficult for them to remember everything (...) Many patients bring their own tablet to the GP or to the hospital to show the history (...) It would have been very beneficial if the GP and the hospital could have access to patients' measurement history in a graphical representation'. Consequently, there are ongoing activities to facilitate more seamless interactions among the actors involved in the service.

4.2.2 The experiences so far from the Norwegian primary and secondary healthcare services

Overall, the healthcare professionals in the Norwegian cases are satisfied with the digital solution. Scaling the solution to more users and user areas has been part of the ongoing activities. Telemonitoring centres, who had been involved in the national trial, have been established in the municipality. Nurses employed at the telemonitoring centres are responsible for the daily follow-up of the patients. Furthermore, a collaboration agreement was made with some GPs and specialists at the hospital who had provided professional advice on the follow-up of the patients during the national trial. The new service thus facilitated improved coordination of the service and a nurse

at a telemonitoring centre stated that digital remote care has filled a gap that previously existed among the home care service, the GP and the hospital. Nevertheless, it is still a challenge to share information among all actors involved in the digital follow-up of patients. The digital remote care platform works well as a standalone system, but the integration with existing systems used in the health service is a challenge. First and foremost, there is a lack of integration between the digital solution for remote care and EPRs used in the municipality. All information related to the patient's health care must be available in the EPR and the lack of integration means that information from the digital remote care solution needs to be transferred manually into the EPR. This entails a risk of incorrect registration as well as duplication of work for nurses at the telemonitoring centre. The digital remote care platform is not available to all employees in the home care services and relevant information about patient's health condition needs to be available in the EPR to provide justifiable health services. Patients with chronic diseases often need follow-up from the specialist health service and lack of integration with EPRs used at hospitals is also a challenge. Thus, seamless interaction between different actors has become a challenge and a nurse stated that 'the benefits are not achieved if only the municipality works on this and if not doctors and outpatient clinics and all are connected'.

Some of the issues outlined above include interdisciplinary collaboration and ongoing work on integration and reconfiguration of the digital solution has been initiated to implement digital remote care both in the municipality and at hospitals. For example, there is ongoing work to develop standardised interfaces between the digital remote care solution and the EPR systems. However, this is not only a technological issue, but also to decide what information is appropriate to share with whom and in what situation. For example, patient generated information can be valuable to clinicians in the hospital; however, they do not have the ability to follow up daily measurements for all patients. There is thus ongoing work to adapt the digital solution to different needs, roles, and responsibilities in a new integrated solution.

Lessons learned from participation in small scale projects have shown how digital remote care can be useful for several user groups and during the pandemic the digital solution was adapted to digital follow-up of COVID-19 patients. Furthermore, several hospitals have ambitions to offer home-based services and digital follow-up of patients have also been scaled to new care settings such as digital outpatient clinics. Ongoing activities involve digital follow-up of patients who need long-term follow-up, such as patients with diabetes, lung diseases, epilepsy, cancer and so on. These patient groups need regular follow-up by health professionals and are traditionally performed by physical attendance at the nearest hospital. Digital remote care is considered an opportunity for more person-centred follow-up for these patients, and it is expected that self-monitoring of vital sign and symptoms will provide more continuity in the follow-up of the patients. However, the digital platform needs to be reconfigured to meet needs and requirements in a new clinical context.

The case briefly outlined above illustrates the sociotechnical complexity of digital remote care and the

need for different levels of integration between health and social care depending on the type of services being offered [6]. Several studies have shown how digital technologies such as digital remote care enable innovation in the context of patient-centred care [2][23]. The notion of recombining has been highlighted by [23] to shed light on how digital resources can be combined and recombined in various ways to enable flexibility and adaptability in digital infrastructures. We will use these theoretical perspectives as analytical lenses by following the further implementation of digital remote care in different clinical settings.

4.2.3 Sharing experience amongst the Norwegian health communities

Implementation of digital solutions for remote care in Norway in general are still at an early stage, and the processes so far have been a step by step, iterative process. Local projects have been established to facilitate further implementation of digital remote care. National and local communities for the development of remote care have also been established where regular meetings and/or webinars are organised to share experience and expertise. A variety of projects have enrolled motivated users who have gradually improved the solution based on the experience gained with regard to how digital technologies can enable better health care services. This strategy can be characterised as bootstrapping which emphasises the importance of identifying motivated users, learning from simpler areas before moving into more complex one, and motivating more users through successful demonstrations [21]. Moreover, the strategy for scaling the solution has been based on experiences from small-scale pilot projects. As stated by [25], ‘the temporariness of pilot implementation is central to their role in infrastructure evolution because it enables experimentation and learning’ (p. 446).

5 CONCLUSION

We have reported lessons learned from the early-stage implementation of digitally enabled remote care in Norway and China. Our findings revealed the key stakeholders’ feedback regarding their experience so far with the solution for remote care in the two countries. We have illustrated some challenges that need to be solved in the further scaling of digital remote care to new patient groups and user areas. However, seamless interaction across organisational units remains a challenge and requires integration among different systems and measuring devices. We further argue that this is a socio-technical issue, where the digital solution needs to be adapted to different needs, roles, and responsibilities in the digital service. Investigating the early stages of the implementation process in Norway and China to understand how early decisions on the sociotechnical setup affect scaling the solution and the services at a later stage in the two contexts is considered as our future work. We hope our ongoing case studies in the two countries will inspire diverse research communities that are keen on developing digital technologies for remote care in different contexts, as we believe that digitally enabled remote care will open a new era of health care which is underway.

6 REFERENCES

- [1] Aaen, J. “Competing Concerns on Emerging Welfare Technologies. A review of eight prevailing debates in current literature” in *Scandinavian Journal of Information Systems* 33, no. 1: 7. 2021.
- [2] Aanestad, M., Vassilakopoulou, P., Øvrelid, E. “Collaborative innovation in healthcare: Boundary resources for peripheral actors” in *ICIS 2019 Proceedings*. 24. 2019.
- [3] Andersen, T. O., Bansler, J. P., Kensing, F., Moll, J., Mønsted, T., Nielsen, K. D., Nielsen, O. W., Petersen, H. H., Svendsen, J. H. “Aligning Concerns in Telecare: Three Concepts to Guide the Design of Patient-Centred E-Health” in *Computer Supported Cooperative Work (CSCW)*, 27(3–6), 1181–1214. 2018.
- [4] Bardram, J. E., Bossen, C., Thomsen, A. “Designing for transformations in collaboration: a study of the deployment of homecare technology” in *Proceedings of the 2005 international ACM SIGGROUP conference on Supporting group work*, pp. 294–303. 2005.
- [5] Barlow, J., Curry, R., Chrysanthaki, T., Hendy, J., Taher, N. “Developing the capacity of the remote care industry to supply Britain’s future needs” in *HaCIRIC* November 2012, pp. 1–25. 2012.
- [6] Barlow, J., Hendy, J., Chrysanthaki, T. “Scaling-up remote care in the United Kingdom: Lessons from a decade of policy intervention” in *Essential Lessons for the Success of Telehomecare*, pp. 223–236. Ios Press, 2012.
- [7] Barlow, J., Wright, C., Sheasby, J., Turner, A., Hainsworth, J. “Self-management approaches for people with chronic conditions: a review” in *Patient education and counseling* 48.2 (2002): 177–187. 2002.
- [8] Berg, M. “Implementing information systems in health care organizations: myths and challenges” in *International journal of medical informatics* 64, no. 2–3: 143–156. 2001.
- [9] Berg, M., Aarts, J., Van Der Lei, J. “ICT in health care: sociotechnical approaches” in *Methods of information in medicine* 42, no. 04: 297–301. 2003.
- [10] Cerna, K., Grisot, M., Islind, A. S., Lindroth, T., Lundin, J., Steineck, G. “Changing Categorical Work in Healthcare: the Use of Patient-Generated Health Data in Cancer Rehabilitation” in *Computer Supported Cooperative Work (CSCW)* 29, no. 5: 563–586. 2020.
- [11] Cohen, D. J., Keller, S. R., Hayes, G. R., Dorr, D. A., Ash, J. S., Sittig, D. F. “Integrating patient-generated health data into clinical care settings or clinical decision-making: lessons learned from project healthdesign” in *JMIR human factors* 3, no. 2: e5919. 2016.
- [12] Dignio. “Dignio Connected Care”. <https://dignio.com/no/product/>, (accessed 24 June 2022).
- [13] Dyb, K., Berntsen, G. R., Kvam, L. “Adopt, adapt, or abandon technology-supported person-centred care initiatives: healthcare providers’ beliefs matter” in *BMC Health Services Research* 21, no. 1: 1–13. 2021.

- [14] Elbert, N. J., Van Os-Medendorp, H., Van Renselaar, W., Ekland, A. G., Hakkaart-Van Roijen, L., Raat, H., Nijsten, T. E. C., Pasmans, S. G. M. A. “Effectiveness and Cost-Effectiveness of Ehealth Interventions in Somatic Diseases: A Systematic Review of Systematic Reviews and Meta-Analyses” in *Journal of medical Internet research* 16, no. 4: e2790. 2014.
- [15] European Commission. “Communication From the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee Of the Regions Empty on enabling the digital transformation of health and care in the Digital Single Market; empowering citizens and building a healthier society”. <https://digital-strategy.ec.europa.eu/en/library/communication-enabling-digital-transformation-health-and-care-digital-single-market-empowering>, (accessed 24 June 2022).
- [16] Greenhalgh, T., Jackson, C., Shaw, S., Janamian, T. “Achieving research impact through co-creation in community-based health services: literature review and case study” in *The Milbank Quarterly* 94, no. 2: 392-429. 2016.
- [17] Greenhalgh, T., Wherton, J., Papoutsi, C., Lynch, J., Hughes, G., A’Court, C., Hinder, S., Fahy, N., Procter, R., Shaw, S. “Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies” in *Journal of medical Internet research* 19, no. 11: e8775. 2017.
- [18] Grisot, M., Hanseth, O., Thorseng, A. A. “Innovation of, in, on infrastructures: articulating the role of architecture in information infrastructure evolution” in *Journal of the Association for Information Systems* 15, no. 4: 2. 2014.
- [19] Grisot, M., Moltubakk Kempton, A., Hagen, L., Aanestad, M. “Data-work for personalized care: Examining nurses’ practices in remote monitoring of chronic patients” in *Health informatics journal* 25, no. 3: 608-616. 2019.
- [20] Guest, G., MacQueen, K. M., Namey, E. E. *Applied thematic analysis*. Sage publications, 2012.
- [21] Hanseth, O., Aanestad, M. “Design as bootstrapping. On the evolution of ICT networks in health care” in *Methods of information in medicine* 42, no. 04: 385-391. 2003.
- [22] James, H. M., Papoutsi, C., Wherton, J., Greenhalgh, T., Shaw, S. E. “Spread, scale-up, and sustainability of video consulting in health care: systematic review and synthesis guided by the NASSS framework” in *Journal of medical Internet research* 23, no. 1: e23775. 2021.
- [23] Lindroth, T., Grisot, M. “How Lightweight Technologies Support Digital Innovation in the Context of Patient-Centered Care” in *Eleventh Scandinavian Conference on Information Systems (SCIS2020)*. 2020.
- [24] May, C. R., Finch, T. L., Cornford, J., Exley, C., Gately, C., Kirk, S., Jenkins, K. N., Osbourne, J., Robinson, a L., Rogers, A., Wilson, R., Mair, F. S. “Integrating telecare for chronic disease management in the community: what needs to be done? ” in *BMC health services research* 11, no. 1: 1-11. 2011.
- [25] Mønsted, T., Hertzum, M., Søndergaard, J. “A socio-temporal perspective on pilot implementation: Bootstrapping preventive care” in *Computer Supported Cooperative Work (CSCW)* 29, no. 4: 419-449. 2020.
- [26] Park, Y. R., Lee, Y., Kim, J. Y., Kim, J., Kim, H. R., Kim, Y. H., Kim, W. S., Lee, J. H. “Managing patient-generated health data through mobile personal health records: analysis of usage data” in *JMIR mHealth and uHealth* 6, no. 4: e9620. 2018.
- [27] Saldaña, J. *The Coding Manual for Qualitative Researchers*. Sage publications, 2015.
- [28] Tan, X., Liu, X., Shao, H. “Healthy China 2030: a vision for health care” in *Value in health regional issues* 12: 112-114. 2017.
- [29] The Chinese hospitals in the research project. *Project annual progress review*. Unpublished report, 2021.
- [30] The Norwegian Directorate of Health. “*Digital hjemmeoppfølging-sluttrapport fra nasjonal utprøving*”. <https://www.helsedirektoratet.no/rapporter/digital-hjemmeoppfølging-sluttrapport-fra-nasjonal-utprøving-2018-2021>, (accessed 24 June 2022).
- [31] Varsi, C., Solberg Nes, L., Kristjansdottir, O. B., Kelders, S. M., Stenberg, U., Zangi, H. A., Børøund, E., Weiss, K. E., Stubhaug, A., Asbjørnsen, R. A., Westeng, M., Ødegaard, M., Eide, H. “Implementation strategies to enhance the implementation of eHealth programs for patients with chronic illnesses: realist systematic review” in *Journal of medical internet research* 21, no. 9: e14255. 2019.
- [32] Woll, A., Bratteteig, T. “A trajectory for technology-supported elderly care work” in *Computer Supported Cooperative Work (CSCW)* 28, no. 1: 127-168. 2019.
- [33] World Health Organization. “*WHO global strategy on integrated people-centred health services 2016-2026*”. <http://www.who.int/servicedeliverysafety/areas/people-centred-care/en>, (accessed 24 June 2022).

7 ACKNOWLEDGEMENT

The DigiRemote project (No. 310137) has received funding from an IKTPLUSS-IKT and digital innovation grant from the Research Council of Norway.

Does your EHR support a Learning Healthcare System?

An exploration of possible indicators

Olga Golburean¹, Ove Lintvedt², Espen Solbakken Nordheim², Rune Pedersen², Luis Marco-Ruiz² and Arild Faxvaag¹

¹Faculty of medicine and health sciences, NTNU and Clinic of Orthopaedics, Rheumatology and Dermatology, St.Olavs hospital, Trondheim, Norway olga.olburean@ntnu.no. ²Norwegian Centre for E-health Research, Tromsø, Norway.

Abstract

Background: The Learning Health System is a concept born from the Institute of Medicine roundtable in 2006 which aims for overcoming the limitations of evidence-based medicine. Since its conception several nations and organizations have actively proposed health policies, research plans, and innovation strategies for its implementation. The effective realization of practices enabling a Learning health system hinge on the availability of Information system functionality and tools.

Objective: To describe the EHR functionality that is needed and corresponding indicators for use in evaluation.

Methods: Analysis of the content of influential scientific publications on the topic.

Results and discussion: The key enabling practices are linked to the life-cycle of biomedical knowledge as well as to knowledge on how to implement evidence-based care.

Keywords

Learning health system; Learning healthcare system; Learning health care system; Electronic health record; Evaluation of EHR;

1 INTRODUCTION

1.1 Learning health system

Learning Healthcare System (LHS) is a US health policy initiative and research agenda that was conceptualized by the Institute of medicine (now the National Academy of medicine) in 2006 to identify and address the need for ‘reengineering clinical research and healthcare delivery’ to advance the systematic development and implementation of evidence-based medicine [25]. The initiative has developed into a broad research agenda that is supported by a worldwide community of researchers.

According to Friedman [13], learning refers to having access to, and analyzing data to continuously seek to improve care by developing and implementing new knowledge. Health is both a valuable objective and an imperative for high quality care. System relates to the components that must act in unison to “achieve goals not attainable by any subset of the components”.

The Learning health system (LHS) initiative has been developed against the backdrop of accelerating healthcare costs, unequal access to care, and a plethora of quality and safety issues in the US and elsewhere. Although the possible benefits of LHS have been ascertained, for instance, in PEDSnet, and TRANSFoRm project, there has been little adoption in practice [9], [11], [26], [10]. A wide range of barriers can be identified when implementing LHS, such as: ethical issues, the complexity of clinical decision-making, missing data (patients data are dispersed across multiple organizations), and the ability of clinicians to routinely used, maintain, and monitor routinely produced data extracted from EHR [11],[5]. In line with the discussion aforementioned,

Nordic healthcare has established initiatives to promote and accelerate high-quality care by promoting a transparent and knowledge-based healthcare system, capable of sharing data across sectors and secondary use of data gathered in everyday practice [8]. Furthermore, since 2012 the Nordic council of ministers has supported a Nordic network cooperation on indicator development [17].

Most of the literature on LHSs is mainly theoretical [27]. Several models and frameworks have been developed in the last decade [2]. A framework for value-creating learning health systems [22] seems to fit with the Nordic healthcare system. Their framework has four elements that characterize an LHS: core values, pillars and accelerators, processes, and outcomes. Their core values are also valued in Nordic healthcare: participatory leadership, inclusiveness, scientific rigor, person-centeredness, equity, and solidarity.

1.2 Electronic health records

Electronic health records (EHR) are health information systems that manage and integrate: a) clinical notes; b) medical tests such as laboratory, pathology etc.; c) PACS of medical image systems; and other ancillary systems such as Clinical Decision Support modules [16]. An EHR can be considered an information system for the programming, delivery, documentation and assessment of knowledge-, competence- and skills- based care in a safe and effective manner. The scientific literature describes a Learning health system as a socio-technical system [14], and the performance of the system hinges on the comprehensive and effective interaction between the human worker and the system. Etheredge [9] defines a LHS as a system that generates and applies data from EHR to increase the value of healthcare.

The 18th Scandinavian Conference on Health informatics, Tromsø, Norway, August 22-24, 2022. Organized by UiT The Arctic University of Norway. Conference Proceedings published by Linköping University Electronic Press at <https://doi.org/10.3384/ecp187>. © The Author(s). This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>

9 The 18th Scandinavian Conference on Health informatics, Tromsø, Norway, August 22-24, 2022.

1.3 Evaluation science

LHSs architectural framework comprises of five dimensions: scientific, technical, ethical, social, and goals. Lessard's framework suggested six decision layers that adjust these dimensions: The performance layer models goals, the scientific layer, the organizational layer, the data, and information technology layer, and the ethics and security layer [18]. Our focus in the present paper is on the data and information technology layer, more specifically on LHS which generates and applies data derived from EHR.

Like any other healthcare technology, an EHR is a product of design and engineering. Just like biomedical knowledge, health information technologies are conceptualized, designed, developed, taken into use, assessed, evaluated, phased out and, in the end, replaced by something newer and hopefully better.

The evaluation and assessment of EHRs and other health information system tools both serve to validate the chosen designs and to assess the true outcomes of using the technology. The science of evaluation of health information systems is the key to accelerate learning, both in the organizations that design, engineer and sell such systems as well as among their customers.

An LHS, with its socio-technical nature, makes them difficult to evaluate [14]. In addition, the very nature of an LHS is a system that is constantly evolving. The outcome measures and available data may also evolve.

One of the biggest challenges in evaluating research on LHS is defining relevant indicators, outcomes, and levels of measurement [4]. LHS and EHR share similar barriers and facilitators [20], [21].

As a means to assess and foster good designs as well as killing bad designs, evaluation science is the key to quick but lifelong learning in the health informatics community. Considering that EHR is an essential prerequisite for implementation and expansion of LHS, it is imperative to identify indicators for evaluating the contribution of EHR to a learning healthcare system.

1.4 Objective

To describe the EHR functionality that is needed to support a Learning healthcare system and to develop candidate indicators that can be used as a starting point in the evaluation of EHRs with respect to the ability to support Learning health systems.

1.5 Research question:

What are the potential indicators and which data can say something about their value?

2 METHODS

Focussed review on the literature on learning healthcare systems. Sampling strategy: A comprehensive search was conducted in June 2022, using the PubMed (MEDLINE) and Scopus databases. The search was limited to articles published between 2007 and 2022. The keywords used were: "learning health systems" OR "learning healthcare system" OR "learning health care system" AND "electronic health record".

Inclusion criteria: We included papers that complied with any of the following criteria: (a) explained the concept of the Learning Healthcare System; (b) covered the implementation of an LHS; (c) described the evaluation of a LHS reporting on key features to evaluate it. Articles that fulfilled inclusion criteria, and that present a significant theoretical cornerstone of key features of the LHS system and articles that describe the implication of EHR to accelerate the development of an LHS were revised and included in the paper..

Exclusion criteria: papers covering specific implementations such as AI projects reusing data from EHR but not clearly framed within a LHS were excluded.

Database search resulted in 2859 articles.

Firstly, papers were checked on title only, by reviewer (OG), excluding 2734 studies. Secondly, title and abstract were checked excluding 12 studies. Thirdly, full text review was performed by reviewers (AF and OG) excluding 71 studies. Articles meeting inclusion criteria but containing information that is covered in the already cited sources were excluded due to the page number limitation, which was set by the conference committee. The full flowchart for the screening of articles is depicted in Figure 1

We did not use any specific extraction form. For each publication we identified all the aspects that were likely to contribute to the implementation of LHS. Whenever possible we mapped LHS implementation contributors to potential indicators of success in the LHS adoption.

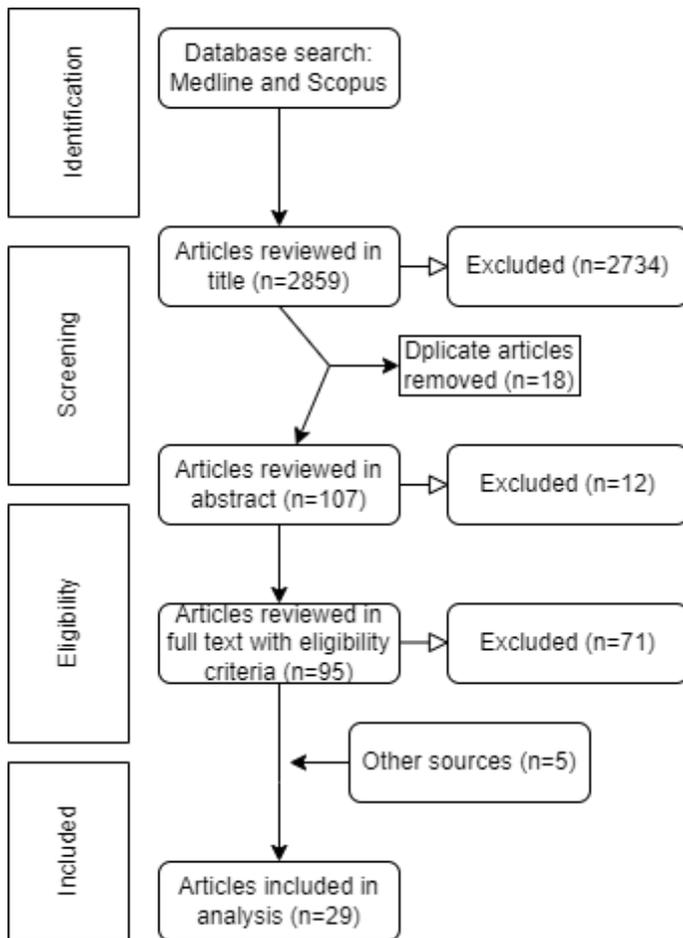


Figure 1. Flowchart followed in the review.

2.0 Results

2.1 Features of a learning health system

Since a health system is a socio-technical system, this means that the technology alone does not suffice to provide services that are of value to the patient. In such a socio-technical system, people that are engaged in the key enabling practices are supported by data and effective tools [13]. Greene [15] describes an evolving learning health system and a model for implementing learning health system practices in a healthcare organization. In the UK, McLachlan and co-workers have developed a conceptual framework to characterize learning health systems [19].

Taken together, the key enabling practices are related to insight- and knowledge-building as well as implementation and dissemination of care delivery practices that are data-driven and based upon the developed insights and knowledge.

2.2 Health information system functionality to support a learning health system

A list of EHR / Health information system back-end functionalities and human-facing front-end tools are presented in table 1. The functionality encompasses the automatic sampling of patient-controlled data, analysis of data for knowledge development, risk modeling, programming, delivery, documentation, and assessment of care.

2.3 Possible indicators

[A list of possible indicators is also presented in table 1.](#)

Learning health system feature or functionality	Possible Indicator
Functionality for automatic sampling and storage of data about the health problem, the care that is provided and the care outcomes [3], [20], [21].	Data that are sampled by the patient are made accessible for the decision-maker via the EHR Outcome data that are recorded by the patient are available and can be integrated and analyzed together with data about the processes that contributed to the outcome.
Access to a shared health data infrastructure [11], [1], [7].	The EHR can be used to share and aggregate data about patients with rare diseases.
Access to a shared knowledge development infrastructure [7], [24].	The EHR can be used to enact the participation in multicenter clinical trials
Functionality to define and characterize patient cohorts with use of data from multiple sources [7], [20], [21], [29].	The EHR can be used to define and characterize patient cohorts
Functionality for system-wide surveillance of safety of health operations [15], [20], [21].	The EHR can be used to monitor the safety of health operations across the healthcare system The EHR can be used to describe and assess the quality of the care that has been provided to patients in the cohorts
Functionality to characterize and assess the quality of care that has been provided to the patients that constitute the cohort [15], [23].	Data from the EHR can be used to train a machine with the objective to model and assess risk (Predictive patient risk modeling)
Tools to apply insights and experiences of what works into doable programs of care that are ready to be implemented in clinical workflows across the health system [15], [1].	Data from the EHR can be used to train a machine with the objective to model and assess risk (Predictive patient risk modeling) Availability of EHR functionality that supports the implementation, monitoring and evaluation of care pathways, documentation templates and order sets.

Tools to make clinically relevant knowledge available at point of care across the entire healthcare system [15], [16].

Tools that engage patients in decision-making [28].

Tools to disseminate learning health system practices in wider networks [15], [20], [21].

Availability of EHR functionality that supports the implementation, monitoring and evaluation of Clinical decision-support.

Patients can use a GUI to enact in shared decision-making

Availability of EHR - functionality to share and disseminate knowledge artifacts by using open, standardized formats (e.g Open CDS) and shared knowledge repositories.

Table 1 Features of a Learning health system and the corresponding indicators.

2.4 Data sets that can inform the value of those indicators

The data should preferably be obtained from the data sets that directly arise from clinicians', application analysts' or business intelligence developers' use of the EHR-system. The use of survey constructs is an alternative when system- or log data cannot be obtained.

3 DISCUSSION

According to McLachlan's framework [19], the operability of an LHS relies on the functionality of the EHR. In order to deliver patient-centered/ personalized care, data within EHR should be available, capable of transferring to a different context, and possible to share across multiple organizations. LHS can be seen as an expansion of the existing health care system into a system that is adept at learning from patient data [12]. Furthermore, routinely collected data can give a more accurate and detailed picture of the clinical care, contrary to the data collected specifically for the purpose of research [6].

To our knowledge, there is no data available about candidate indicators that can be used as a starting point in the evaluation of EHRs with respect to the ability to support Learning health systems.

In this report, we have motivated the development of indicators for assessing EHR-system back-end functionality and clinician-facing front-end tools that aid in the realization of Learning health systems. The list of indicators can inform the search for the data points that can be used to populate the indicators.

4 CONCLUSION

The scientific literature describes a Learning health system as a socio-technical system. In such a system, people that are engaged in the key enabling practices are supported by data and effective tools. In a Learning health system, key enabling practices are related to insight- and knowledge-building as well as implementation and dissemination of care delivery practices that are data-driven and based upon the developed insights and knowledges. We have drafted and presented a list of possible indicators for assessing HER functionality that are to support the key enabling practices.

5 REFERENCES

- [1] Ainsworth, J., og I. Buchan. «Combining Health Data Uses to Ignite Health System Learning». *Methods of Information in Medicine* 54, nr. 6 (2015): 479–87.
- [2] Allen, Claire, Katie Coleman, Kayne Mettert, Cara Lewis, Emily Westbrook, og Paula Lozano. «A Roadmap to Operationalize and Evaluate Impact in a Learning Health System». *Learning Health Systems* 5, nr. 4 (2021): e10258.
- [3] Austin, Elizabeth J., Cynthia LeRouge, Jenney R. Lee, Courtney Segal, Savitha Sangameswaran, Joseph Heim, William B. Lober, Andrea L. Hartzler, and Danielle C. Lavalley. «A Learning Health Systems Approach to Integrating Electronic patient-reported Outcomes across the Health Care Organization.» *Learning Health Systems* 5, no. 4 (October 2021).
- [4] Atsma, Femke, Glyn Elwyn, og Gert Westert. «Understanding unwarranted variation in clinical practice: a focus on network effects, reflective medicine and learning health systems». *International Journal for Quality in Health Care* 32, nr. 4 (4. juni 2020): 271–74. <https://doi.org/10.1093/intqhc/mzaa023>.
- [5] Deans, Katherine J., Sara Sabihi, and Christopher B. Forrest. «Learning Health Systems». *Seminars in Pediatric Surgery* 27, nr. 6 (desember 2018): 375–78.
- [6] Deeny, Sarah R, and Adam Steventon. «Making Sense of the Shadows: Priorities for Creating a Learning Healthcare System Based on Routinely Collected Data.» *BMJ Quality & Safety* 24, no. 8 (August 1, 2015): 505.
- [7] Easterling, Douglas, Anna C. Perry, Rachel Woodside, Tanha Patel, and Sabina B. Gesell. «Clarifying the Concept of a Learning Health System for Healthcare Delivery Organizations: Implications from a Qualitative Analysis of the Scientific Literature.» *Learning Health Systems* 6, no. 2 (April 2022).
- [8] Einhorn, E.S., 2019. Healthcare in the Nordics. Nordics.info [Healthcare in the Nordics](https://nordics.info/Healthcare-in-the-Nordics)
- [9] Etheredge, Lynn M. «A Rapid-Learning Health System». *Health Affairs* 26, nr. Supplement 1 (januar 2007): w107–18..
- [10] Ethier, Jean-François, Mark McGilchrist, Adrien Barton, Anne-Marie Cloutier, Vasa Curcin, Brendan C. Delaney, and Anita Burgun. «The TRANSFoRM Project: Experience and Lessons Learned Regarding Functional and Interoperability Requirements to Support Primary Care.» *Learning Health Systems* 2, no. 2 (April 2018): e10037.
- [11] Forrest, Christopher B., Peter Margolis, Michael Seid, and Richard B. Colletti. «PESNet: How a Prototype Pediatric Learning Health System Is Being Expanded into a National Network». *Health Affairs (Project Hope)* 33, nr. 7 (juli 2014): 1171–77.
- [12] Foley, Tom, and Luke Vale. «A Framework for Understanding, Designing, Developing and Evaluating Learning Health Systems.» *Learning Health Systems*, May 20, 2022.
- [13] Friedman, Charles P., Adam K. Wong, and David Blumenthal. «Achieving a Nationwide Learning Health System». *Science Translational Medicine* 2, nr. 57 (10. november 2010): 57cm29.
- [14] Friedman, Charles, Joshua Rubin, Jeffrey Brown, Melinda Buntin, Milton Corn, Lynn Etheredge, Carl

- Gunter, mfl. «Toward a Science of Learning Systems: A Research Agenda for the High-Functioning Learning Health System». *Journal of the American Medical Informatics Association: JAMIA* 22, nr. 1 (januar 2015): 43–50.
- [15] Greene, Sarah M., Robert J. Reid, and Eric B. Larson. «Implementing the Learning Health System: From Concept to Action». *Annals of Internal Medicine* 157, nr. 3 (7. august 2012): 207–10.
- [16] Hoffmann, Mikael, Robert Vander Stichele, David W Bates, Johanna Björklund, Steve Alexander, Marine L Andersson, Ane Auraaen, et al. «Guiding Principles for the Use of Knowledge Bases and Real-World Data in Clinical Decision Support Systems: Report by an International Expert Workshop at Karolinska Institutet.» *Expert Review of Clinical Pharmacology* 13, no. 9 (September 1, 2020): 925–34.
- [17] Hyppönen, Hannele, Arild Faxvaag, Heidi Gilstad, Gudrun Audur Hardardottir, Lars Jerlvall, Maarit Kangas, Sabine Koch, et al. «Nordic EHealth Indicators: Organisation of Research, First Results and Plan for the Future.» *Studies in Health Technology and Informatics* 192 (2013): 273–77.
- [18] Lessard, Lysanne, Wojtek Michalowski, Michael Fung-Kee-Fung, Lori Jones, and Agnes Grudniewicz. «Architectural Frameworks: Defining the Structures for Implementing Learning Health Systems.» *Implementation Science* 12, no. 1 (December 2017): 78.
- [19] McLachlan, Scott, Henry W. W. Potts, Kudakwashe Dube, Derek Buchanan, Stephen Lean, Thomas Gallagher, Owen Johnson, Bridget Daley, William Marsh, og Norman Fenton. «The Heimdall Framework for Supporting Characterisation of Learning Health Systems». *BMJ Health & Care Informatics* 25, nr. 2 (1. april 2018).
- [20] McLachlan, Scott, Kudakwashe Dube, Evangelia Kyrimi, and Norman Fenton. «LAGOS: Learning Health Systems and How They Can Integrate with Patient Care.» *BMJ Health & Care Informatics* 26, no. 1 (October 2019): e100037.
- [21] McLachlan, Scott, Kudakwashe Dube, Owen Johnson, Derek Buchanan, Henry W.W. Potts, Thomas Gallagher, og Norman Fenton. «A framework for analysing learning health systems: Are we removing the most impactful barriers?» *Learning Health Systems* 3, nr. 4 (21. mars 2019): e10189.
- [22] Menear, Matthew, Marc-André Blanchette, Olivier Demers-Payette, og Denis Roy. «A framework for value-creating learning health systems». *Health Research Policy and Systems* 17, nr. 1 (9. august 2019): 79.
- [23] Miriovsky, Benjamin J., Lawrence N. Shulman, and Amy P. Abernethy. «Importance of Health Information Technology, Electronic Health Records, and Continuously Aggregating Data to Comparative Effectiveness Research and Learning Health Care.» *Journal of Clinical Oncology* 30, no. 34 (December 1, 2012): 4243–48.
- [24] Nordo, Amy Harris, Hugh P. Levaux, Lauren B. Becnel, Jose Galvez, Prasanna Rao, Komathi Stem, Era Prakash, and Rebecca Daniels Kush. «Use of EHRs Data for Clinical Research: Historical Progress and Current Applications.» *Learning Health Systems* 3, no. 1 (January 2019): e10076.
- [25] Olsen, Aisner D, and McGinnis Jm. «The Learning Healthcare System: Workshop Summary». PubMed, 2007.
- [26] Platt, Richard, Susan S. Huang, and Jonathan B. Perlin. «A Win for the Learning Health System». *NAM Perspectives*, 29. mai 2013.
- [27] Pomare, Chiara, Zeyad Mahmoud, Alex Vedovi, Louise A. Ellis, Gilbert Knaggs, Carolyn L. Smith, Yvonne Zurynski, and Jeffrey Braithwaite. «Learning Health Systems: A Review of Key Topic Areas and Bibliometric Trends». *Learning Health Systems* 6, nr. 1 (2022): e10265.
- [28] Wagle, Anjali A., Nino Isakadze, Khurram Nasir, og Seth Shay Martin. «Strengthening the Learning Health System in Cardiovascular Disease Prevention: Time to Leverage Big Data and Digital Solutions». *Current Atherosclerosis Reports* 23, nr. 5 (10. mars 2021): 19.
- [29] Wouters, Roel H.P., Rieke Graaf, Emile E. Voest, and Annelien L. Bredenoord. «Learning Health Care Systems: Highly Needed but Challenging.» *Learning Health Systems* 4, no. 3 (July 2020).

6 ACKNOWLEDGEMENTS

This work was supported by a research grant from the Liaison Committee between the Central Norway Regional Health Authority (RHA) and the Norwegian University of Science and Technology (NTNU) and by the National Centre for E-health research.

Assessing the use of telemedicine among people with diabetes: A Danish translation and cross-cultural adaptation of the Telehealth Usability Questionnaire

Clara Bender*¹, Marie Germund Nielsen², Ole Hejlesen¹, Stine Hangaard^{1,3}

¹Department of Health Science and Technology Aalborg University, Denmark *csch@hst.aau.dk

²Department of Clinical Medicine, Aalborg University, Denmark

³Steno Diabetes Centre North Denmark, Denmark

Abstract

The aim was to translate and cross-culturally adapt the original Telehealth Usability Questionnaire (TUQ) into Danish and pre-test the translation on a sample of Danish patients with diabetes who received telemedicine. Participants with diabetes (n = 34) completed the Danish TUQ and participated in semi-structured interviews. The overall internal consistency was 0.857. The internal consistency for the five sub-groups ranged from 0.241 (sub-group four) to 0.857 (sub-group five). The study demonstrates an overall accepted statistical internal consistency compared with the original TUQ, which makes it a reliable tool to assess patients' perceptions after using a telemedicine service.

Keywords

Telemedicine, questionnaire, adaption, forward-backwards translation, reliability

1 INTRODUCTION

Telemedicine solutions may use many different types of communication technology to support patients remotely. Various types of telemedicine solutions have been proposed for a wide range of patients [1–5], including patients with diabetes, pulmonary diseases, heart disease, cancer, and so on [6–10]. Moreover, a meta-review from 2017 focusing on telehealth interventions to support self-management of long-term conditions of diabetes, heart failure, chronic obstructive pulmonary disease, and cancer suggested that telemedicine is a safe way to deliver self-management support and showed that users have a high degree of acceptance of telemedicine. However, telemedicine solutions were not consistently superior to usual care [6]. This lack of consistent effect may be explained by the shortcomings of telemedicine studies, as their quality has been questioned [11]. One of the shortcomings of telemedicine studies is that they fail to explore patients' experiences with the telemedicine solution provided [11]. However, it is important to increase knowledge about telemedicine solutions and their usability to understand which solution is appropriate for which patient [12]. Considering the COVID-19 pandemic, this challenge has become even more relevant as the use of telemedicine has increased [13–17]. This highlights the need for tools to evaluate patients' perceptions of the quality of the telemedicine solutions provided.

Surveys have been commonly used to assess perceptions and outcomes in telemedicine studies among both patients and healthcare professionals [16,18].

The quality (i.e., consistency and transferability) of telemedicine studies has been questioned, and improved quality and reporting of telemedicine survey studies have been called for [12]. Moreover, to the best of our knowledge, the availability of surveys in the Scandinavian languages, including Danish, are very limited, underlining the need for a Danish validated and cross-culturally adapted survey that evaluates patients' perceptions of usefulness and satisfaction with a provided telemedicine solution.

Several of the existing telemedicine-specific questionnaires are lacking in their ability to evaluate more than one telehealth system. Furthermore, most of the questionnaires are only intended for either clinicians or patients and not both at the same time [1,18]. The Telehealth Usability Questionnaire (TUQ), however, has several advantages since it is intended for use with various types of telehealth systems and can be used both for clinicians and patients. The final TUQ consists of 21 items assessing computer usability, telemedicine solutions, and their quality. This assessment focuses on usefulness, ease of use, effectiveness, reliability, and satisfaction [19]. The original development of TUQ was based on four phases: 1) literature review, 2) construct development, 3) item development, and 4) examination of reliability.

To test the reliability of the content of the originally developed TUQ, 53 participants (21 males and 32 females) were included in a three-month study, where participants who regularly used telehealth technologies were asked to complete the TUQ based on their recent interaction with a selected system. At the same time, participants who had never used or did not regularly use telehealth technologies

The 18th Scandinavian Conference on Health informatics, Tromsø, Norway, August 22-24, 2022. Organized by UiT The Arctic University of Norway. Conference Proceedings published by Linköping University Electronic Press at <https://doi.org/10.3384/ecp187>. © The Author(s). This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>

were asked to take part in a simulated telehealth session and afterwards complete the TUQ based on their interaction [19]. After the three months were over, statistical analysis was conducted and showed a solid, robust, and versatile TUQ instrument [19]. Since the results from the study by Parmanto et al. showed that TUQ is a relevant questionnaire to evaluate telehealth systems, a Danish translation and cross-cultural adaptation of it is highly needed due to the increased use of telemedicine solutions in the Danish healthcare sector.

Thus, the aim of the present study was to translate the TUQ into Danish, pre-test it, and thereby cross-culturally adapt the translation using a sample of Danish patients with diabetes who had recently received a telemedicine service in the form of a video consultation.

2 METHODS AND MATERIALS

2.1 Procedures

The English version of the TUQ was translated and adapted for the Danish culture in accordance with the guidelines for the process of cross-cultural adaptation of self-report measures provided by Beaton et al. [20]. The term cross-cultural adaptation refers to “a process that looks at both language (translation) and cultural adaption issues in the process of preparing a questionnaire for use in other settings” [20]. The English version of the TUQ includes 21 items with the possibility to respond using a seven-point Likert scale, where one represents total disagree and seven represents fully agree or N/A. Besides the 21 items, a space for more elaborate comments about the telemedicine system is provided in the TUQ.

The cross-cultural adaptation of the TUQ was performed in line with the process given by Beaton et al. [20] by using forward-backwards translation. Table 1 gives an overview of the steps in the process, which is described in detail below the table. The letters in the brackets say which authors did which of the tasks in the validation process.

Stage	Action
I	Forward translation by two translators
II	Synthesis of the two translators into one (T-12)
III	Backwards translation by two translators
IV	Expert committee review
V	Pre-testing and completing

Table 1 an overview of Beaton et al.’s five-stage cross cultural adaption process [20].

Stage I: Forward translation

Two forward translators translated the English version of the TUQ into Danish and produced two individual versions (T1 and T2) of the Danish TUQ. Both translators were bilingual, with Danish as their mother tongue. One translator had a clinical background and was familiar with the terms: telemedicine and usability. The other translator did not have a clinical background and was not familiar with the concepts of the TUQ. Both forward translators were selected based on the criteria given by Beaton et al. [20].

Time estimation: The number of working hours that each of the forward translators spent varied between two and three hours. Their time was spent on literal translation work

and on how to include cultural differences in the translation of the original TUQ (i.e., how the TUQ could be implemented in a Danish context).

Stage II: Synthesis of the initial translations (T-12)

Working from the original TUQ and the two independent initial Danish translations (T1 and T2), the two forward translators reached a consensus on a new version of the Danish TUQ (T-12). The out-come (i.e., the issues addressed by T1 and T2 in this synthesis step and how the issues were resolved) was described in a written report.

Time estimation: The number of working hours for the second stage was approximately one working day (eight hours) per translator.

Stage III: Backwards translation

To ensure the validity of the first two steps in the translation process, a third step, back-wards translation, was performed. Two independent backwards translators with English as their mother tongue translated the T-12 version of TUQ back into English and thus produced two different back translations (BT1 and BT2). Both backwards translators were familiar with the concepts of the TUQ, but neither of them had a medical background. The backwards translators were selected based on the criteria given by Beaton et al. [20].

Time estimation: Each backwards translator spent approximately half to one working day (four to eight hours) on the translation.

Stage IV: Expert committee review

The expert committee was composed of the two forward translators, the two backwards translators, two health professionals, a methodologist, and a language professional. During their meeting, the expert committee reviewed all the produced material from the previous stages (stages I, II, and III, i.e., the English TUQ and each translation [T1, T2, T-12, BT1, and BT2]) together with the corresponding written report. The expert committee discussed the translations and developed a pre-final version of the Danish TUQ. Due to the COVID-19 pandemic, the expert committee meeting was held digitally using Microsoft Teams.

Time estimation: the expert committee used approximately half a working day (three hours). The working hours did not include the preparation time that each of the committee members spent prior to the expert committee meeting.

Stage V: Pre-testing and completing

Participants with diabetes (n = 34) completed the Danish TUQ. Subsequently, they participated in semi-structured interviews concerning their basic demographics, their responses to the TUQ, and the relevance of each item within the Danish TUQ. This step was conducted to ensure the reliability and consistency of the Danish TUQ.

After the last stage, the authors had a series of meetings and went through the comments given by the interviewed subjects. Again, a consensus was reached, and the authors agreed on a final version of the Danish translation of the TUQ.

Time estimation: The researchers spent approximately 30 minutes per interview (plus additional preparation time before each interview and post-processing of the data). The

duration of the physical meeting between the authors lasted approximately two hours.

2.2 Participants in the Pre-test

All included participants (n = 34) were diagnosed with diabetes and had received a telemedicine service in the form of a video consultation within the last two months. The included participants were recruited from the four Steno Diabetes Centres located in Denmark (Steno Diabetes Centre North Denmark, Steno Diabetes Centre Aarhus, Steno Diabetes Centre Zealand, and Steno Diabetes Centre Odense). The aim of the Steno Diabetes Centres is to improve and prolong the lives of people with diabetes in Denmark [21].

The participants were selected using consecutive sampling, which refers to the inclusion of all accessible participants at multiple data collection sites - in this study, the four Steno Diabetes Centres. From each of the four sites, a manager provided a list with the names of potential participants. The potential participants had each received a short description of the aim of the study by healthcare professionals from one of the Steno Diabetes Centres. The participants had given written informed consent, so researchers (two assistant professors with experience within telemedicine and questionnaire translation) from Aalborg University were permitted to contact them by telephone (due to the COVID-19 pandemic, the interviews were performed by telephone). Each participant received a phone call during which they received a detailed description of the aim of the study and the study procedure. If the potential participants agreed to participate in the study, a time and date for a new phone call was planned. After the call ended, each subject received an email including the Danish TUQ, written information about the study, and their rights as a participant in the study. Moreover, the participants were encouraged to fill out the Danish TUQ based on their recent experience with teleconsultation.

Approximately one week after receiving the email, each participant received a phone call from the researchers. First, the participants were asked to provide basic demographics (sex, age, civil status, level of employment, level of education, comorbidities, and diagnosis/reason they had used the tele-medicine service). Second, each participant was asked to provide their response for each questionnaire item and their understanding of the meaning and relevance of each of the items. In closing, after going through all 21 items, the researchers asked if there was anything the participants wanted to add further to the interview. During and after all 34 semi-structured interviews, the researchers took notes.

The following criteria were outlined for the inclusion of participants in the pre-test of the Danish TUQ:

Inclusion criteria: diagnosed with diabetes (either type 1 or type 2 diabetes), have received a telemedicine service related to their diabetes, > 18 years old (both men and women were included), and able to read and understand Danish.

Exclusion criteria: blindness, too ill to participate, dementia, or other cognitive impairment (judged by the healthcare professionals at the Steno Diabetes Centres), unable to read or understand Danish and thereby fill out the Danish TUQ.

2.3 Statistical Analysis

Descriptive statistics were presented with mean and standard deviation (\pm SD) or percentage of the participants. The overall internal consistency of the Danish TUQ was measured using Cronbach's alpha, based on the data collected from the semi-structured interviews [22]. For each of the five sub-groups within the TUQ, the internal consistency was also calculated. The sub-groups were designed by Parmanto et al. based on covering all usability factors (i.e., usefulness, ease of use, effectiveness, reliability, and satisfaction) [19]. These sub-groups were sub-group one (items 1–3), sub-group two (items 4–6), sub-group three (items 7–14), sub-group four (items 15–17), and sub-group five (items 18–21).

The descriptive and reliability analyses were conducted in IBM SPSS Statistics, version 27 [23]. The significance level was set at 0.05.

2.4 Ethical Approval

The study was conducted in accordance with the Helsinki Declaration [24] and Danish legislation; questionnaires and qualitative studies do not require ethical approval since they are based on written consent [25]. Likewise, it was not required to have an approval from a data protection officer. Each participant received time for deliberation before giving informed written consent to participate in the study. The data analysis was executed anonymously. The AGREE checklist was followed.

3 RESULTS

Basic information on the participants		
Number of participants	34	
Male (sex)	18	53%
Age (y)	50.4 (\pm 13.5)	
Living situation		
In the city (defined as > 20,000 inhabitants)	21	62%
In rural area	13	38%
Civil status		
Married or living with a partner	28	82%
Living alone	6	18%
Level of employment		
Full-time employment	15	44%
Less than 37 hours per week	5	15%
No job (including those who have retired)	14	41%
Level of education		
9 th or 10 th grade or less (some only completed 7 th grade)	2	6%
High school	5	15%
Higher education	12	35%
Skilled worker (trade, industry, office, etc.)	15	44%
Comorbidities		
Comorbidities (yes)	14	41%

Comorbidities (no)	11	32%
Comorbidities (not answered)	9	26%

Table 2: Basic demographics of the 34 participants in the study.

From table 2 an overview of the demographic information of the participants included in the study is given. The average age of the participants was 50.4 (SD: 13.5), and most of the participants (62%) lived in the city (defined as > 20,000 inhabitants).

Diagnose/reason why the participant received the telemedicine service		
Type 1 diabetes	17	50%
Type 2 diabetes	4	11.8%
Type 1.5 diabetes	2	5.9%
Screening for diabetes	2	5.9%
Follow up on a knee operation	1	2.9%
Dercum's disease	1	2.9%
Follow up after a Gastric bypass operation	3	8.8%
Struma diffusa	1	2.9%
Physical outpatient clinic visit changed to a telemedicine consultation	3	8.8%

Table 3: Overview of the different reasons or diagnosis of why the participants received the telemedicine service.

Table 3 generates an overview of the diagnoses or reasons why the participants received a telemedicine service. Half of the participants (50%) received the telemedicine service in relation to their type 1 diabetes, while 11.8% received the telemedicine service due to type 2 diabetes. The rest of the reasons why the participants received the telemedicine service were either related to comorbidities or related to organizational changes implemented to meet the COVID-19 restrictions.

3.1 The Distribution of Answers

Overall, the Danish TUQ consisted of 21 items, as did the original TUQ [19]. Table 4 gives an overview of the accumulated distribution of answers related to each item (Item 1 to Item 21) given by the 34 participants included in the study. Response category 7 was most used (48.2%), followed by response category 6 (18.8%). Response category 2 was the least used (1.6%), followed by response category 3 (2.0%). Regarding item 20, 30 out of 34 participants answered using response category 7. Item nos. 16 and 17 were considered not relevant by 29 of the 34 participants.

No.	Response category							
	1	2	3	4	5	6	7	N/A
1	2	0	1	4	8	11	7	1
2	1	0	0	0	2	5	26	0
3	3	2	2	7	9	2	9	0
4	1	0	1	2	1	6	22	1
5	1	0	0	4	1	8	20	0

6	0	1	0	2	1	11	15	4
7	1	1	0	0	3	7	22	0
8	1	1	0	4	7	7	14	0
9	2	0	0	1	1	10	19	1
10	2	0	3	3	4	6	15	1
11	2	0	1	1	0	7	22	1
12	0	0	1	1	4	4	21	2
13	2	0	0	1	3	7	20	0
14	3	2	2	0	9	5	11	2
15	4	1	3	4	8	6	6	2
16	0	1	0	0	1	1	2	29
17	1	2	0	0	0	1	1	29
18	1	0	0	0	6	4	22	1
19	1	0	0	3	2	12	16	0
20	1	0	0	1	0	2	30	0
21	1	0	0	1	0	11	21	0
Total	30	11	14	39	68	133	341	72

Table 4: The accumulated distribution of responses related to each item, including the response categories: 1,2,3,4,5,6,7 or not relevant (N/A) given by the 34 participants. The number of times a response category was used is visualized by colour ranging from green to red. Green indicates the lowest number of times a value was given, while red indicates the highest number of times a value was given.

3.2 Internal Consistency

The overall internal consistency of the Danish TUQ was measured as a Cronbach's alpha of 0.857. Table 5 gives an overview of the internal consistency for each of the five sub-groups that the original TUQ was divided into [19]. The *Reliability* group (items 15–17) showed the lowest Cronbach's alpha value (0.241), while the *Satisfaction and future use* group (items 18–21) showed the highest Cronbach's alpha value (0.857).

No.	Name	Items	Cronbach's alpha value
1	Usefulness	1–3	0.579
2	Ease of use & learnability	4–9	0.759
3	Effectiveness	10–14	0.795
4	Reliability	15–17	0.241
5	Satisfaction and future use	18–21	0.857

Table 5: Overview of the internal consistency for the five sub-groups within the Danish TUQ.

4 DISCUSSION

The aim of this study was to translate the original Telehealth Usability Questionnaire by Parmanto et al. [19] into Danish and cross-culturally validate the translation by pre-testing it and using a sample of Danish patients with diabetes who had received a telemedicine service within the last two months. The Danish TUQ was pretested on 34 participants with diabetes to ensure reliability, which is in accordance with Beaton et al. [20]. The overall internal

consistency of the Danish TUQ was found to be satisfactory (Cronbach's alpha: 0.857) [26].

The present study mirrors similar translation studies. A very similar study to this is a translational validation study by Vidal-Alaball et al., which included 33 participants. That translation showed a similar Cronbach's alpha of 0.84 in a Catalan version of the Health Optimum Telemedicine Acceptance Questionnaire [1]. In addition, Micoulaud-Franchi et al. aimed to translate and validate a French version of a six-item self-reported questionnaire that evaluates the extent to which patients find e-health systems acceptable [27]. According to the authors, their validation process revealed a satisfactory level of acceptance (i.e., a Cronbach's alpha of 0.7) [27]. When Parmanto et al. developed and tested the TUQ, they found that the Cronbach's alpha value was 0.8 [19], which is in line with the findings of the present study. Finally, in a systematic review by Weaver et al., a list of available telehealth survey instruments was examined with appertaining calculated Cronbach's alphas [16]. From their systematic assessment, Weaver et al. identified twelve telehealth communication assessment instruments and their corresponding Cronbach's alpha values, which ranged from 0.7 to 0.93. This systematic review, including the studies by Vidal-Alaball et al. and Micoulaud-Franchi et al., confirms the trend that a Cronbach's alpha value between 0.7 to 0.9 is generally considered acceptable [22]. Thus, the measured Cronbach's alpha in this study is in line with the trend.

4.1 The internal consistency

When comparing the Cronbach's alpha for each of the five sub-groups within the Danish TUQ with the same five sub-groups within the original TUQ, the values are closely related, with only a few deviations. Three out of the five sub-groups were almost identical (e.g., sub-group two: original = 0.92, Danish = 0.759; sub-group three: original = 0.86, Danish = 0.795; and sub-group five: original = 0.91, Danish = 0.857). The remaining two sub-groups deviated much from each other (sub-group one: original = 0.83, Danish = 0.579; sub-group four: original = 0.79, Danish = 0.241). There could be several explanations for the deviations seen within the two sub-groups. However, the most reasonable explanation is probably that the number of participants included in the studies has an important influence on Cronbach's alpha. In the study by Parmanto et al., 53 participants were included [19], whereas only 34 participants were included in the present study. The low Cronbach's alpha in sub-group four could be explained by the fact that 29 of the participants responded Not relevant to two of three items in this specific sub-group. These items concern potential issues regarding the specific telehealth system, and the responses indicate that the participants did not experience any issues when using the telehealth system.

4.2 Reduction in the number of questions

When looking at the accumulated distribution of answers related to each item (Table 4), the participants used response category 7 most often. This might indicate that the participants were very satisfied with the telemedicine service they received and/or that they understood the item without problems. When it comes to items 16 and 17, the participants used the response category Not relevant the most. This may indicate that there were no problems with

the system; however, it may also indicate that the translation was insufficient. When going through each of the items and the appertaining given response categories with the participants during the interviews, it appears that the participants clearly understood these items. Thus, there might be an indication of a need for a reduction in the number of questions.

To our knowledge, this study is the first initiative conducted where the original TUQ has been translated into Danish or any other Scandinavian language. A limitation of the study is that only 34 participants were included. Even though Beaton et al. argue that 30 to 40 participants is an ideal number to test on, it may have had an impact on the calculated Cronbach's alpha in this study compared to the original study where 53 participants were included [19] or other studies with several participants included [28]. Thus, it could be relevant to test the Danish TUQ on a larger scale. Moreover, telemedicine includes various technologies, users, and organizational setups [6–10]. A highly relevant next step would be to validate the Danish TUQ on different telemedicine solutions with other types of users to ensure the applicability of the Danish TUQ to all available telemedicine solutions.

4.3 Developing in a Danish context

Another limitation of the study is related to the cross-cultural differences in the healthcare sector (i.e., the context where the original TUQ was developed differed from the context where the Danish TUQ was implemented). However, this study followed the cross-cultural adaptation process prescribed by Beaton et al. [20], and these concerns should, therefore, be minimal. Finally, the study was limited by the fact that the included participants were not a representative sample of people with diabetes compared to the future users of the Danish TUQ. However, when observing the basic demographics in Table 2 and Table 3, it is evident that they represent the typical patient with diabetes in accordance with the characteristics given by the World Health Organization and several other studies describing the typical characteristics of people with diabetes [29–33].

4.4 Conclusion

A translation and cross-cultural adaptation of the original TUQ was performed into Danish. The study demonstrates an overall accepted statistical internal consistency compared with the original TUQ, which makes it a reliable tool to assess patients' perceptions after using a telemedicine service. Future work could include testing the TUQ on other groups of patients since the Danish TUQ was only pre-tested on patients with diabetes. Furthermore, it has become evident that telemedicine services will be a more integrated part of receiving healthcare services in the aftermath of the COVID-19 pandemic.

5 SUMMARY

The original TUQ was translated and cross-culturally adapted into Danish. The study demonstrates an overall accepted statistical internal consistency compared with the original TUQ, which makes it a reliable tool to assess patients' perception of telemedicine after receiving a telemedicine service.

6 REFERENCES

- [1] Vidal-Alaball J, Mateo GF, Domingo JLG, Gomez XM, Valmaña GS, Ruiz-Comellas A, et al. Validation of a short questionnaire to assess healthcare professionals' perceptions of asynchronous telemedicine services: The Catalan version of the health optimum telemedicine acceptance questionnaire. *Int J Environ Res Public Health* 2020;17:2–11. <https://doi.org/10.3390/ijerph17072202>
- [2] Beard M, Orlando JF, Kumar S. Overcoming the tyranny of distance: An audit of process and outcomes from a pilot telehealth spinal assessment clinic. *J Telemed Telecare* 2017;23:733–9. <https://doi.org/10.1177/1357633X16664851>
- [3] Westra I, Niessen FB. Implementing real-time video consultation in plastic surgery. *Aesthetic Plast Surg* 2015;39:783–90. <https://doi.org/10.1007/s00266-015-0526-4>
- [4] Nakagawa K, Umazume T, Mayama M, Chiba K, Saito Y, Kawaguchi S, et al. Feasibility and safety of urgently initiated maternal telemedicine in response to the spread of COVID-19: A 1-month report. *J Obstet Gynaecol Res* 2020;46:1967–71. <https://doi.org/10.1111/jog.14378>
- [5] Morland LA, Poizner JM, Williams KE, Masino TT, Thorp SR. Home-based clinical video teleconferencing care: Clinical considerations and future directions. *Int Rev Psychiatry* 2015;27:504–12. <https://doi.org/10.3109/09540261.2015.1082986>
- [6] Hanlon P, Daines L, Campbell C, Mckinstry B, Weller D, Pinnock H. Telehealth interventions to support self-management of long-term conditions: A systematic metareview of diabetes, heart failure, asthma, chronic obstructive pulmonary disease, and cancer. *J Med Internet Res* 2017;19. <https://doi.org/10.2196/jmir.6688>
- [7] Wootton R. Twenty years of telemedicine in chronic disease management - an evidence synthesis. *J Telemed Telecare* 2012;18:211–20. <https://doi.org/10.1258/jtt.2012.120219>
- [8] Faruque LI, Wiebe N, Ehteshami-Afshar A, Liu Y, Dianati-Maleki N, Hemmelgarn BR, et al. Effect of telemedicine on glycated hemoglobin in diabetes: A systematic review and meta-analysis of randomized trials. *Cmaj* 2017;189:E341–64. <https://doi.org/10.1503/cmaj.150885>
- [9] Mclean S, Nurmatov U, Jly L, Pagliari C, Car J, Sheikh A, et al. Telehealthcare for chronic obstructive pulmonary disease: Cochrane review and meta-analysis. *Br J Gen Pract* 2012;62:e739–49. <https://doi.org/10.3399/bjgp12X658269>
- [10] Sirintrapun SJ, Lopez AM. Telemedicine in cancer care. *Am Soc Clin Oncol Educ B* 2018;540–5. https://doi.org/10.1200/edbk_200141
- [11] Ekeland AG, Bowes A, Flottorp S. Methodologies for assessing telemedicine: A systematic review of reviews. *Int J Med Inform* 2012;81:1–11. <https://doi.org/10.1016/j.ijmedinf.2011.10.009>
- [12] Langbecker D, Caffery LJ, Gillespie N, Smith AC. Using survey methods in telehealth research: A practical guide. *J Telemed Telecare* 2017;23:770–9. <https://doi.org/10.1177/1357633X17721814>
- [13] Mann DM, Chen J, Chunara R, Testa PA, Nov O. COVID-19 transforms health care through telemedicine: Evidence from the field. *J Am Med Inform Assoc* 2020;27:1132–5. <https://doi.org/10.1093/jamia/ocaa072>
- [14] Valentino LA, Skinner MW, Pipe SW. The role of telemedicine in the delivery of health care in the COVID-19 pandemic. *Haemophilia* 2020;26:e230–1. <https://doi.org/10.1111/hae.14044>
- [15] Layfield E, Triantafyllou V, Prasad A, Deng J, Shanti RM, Newman JG, et al. Telemedicine for head and neck ambulatory visits during COVID-19: Evaluating usability and patient satisfaction. *Head Neck* 2020;42:1681–9. <https://doi.org/10.1002/hed.26285>
- [16] Weaver MS, Lukowski J, Wichman B, Navaneethan H, Fisher AL, Neumann ML. Human connection and technology connectivity: A systematic review of available telehealth survey instruments. *J Pain Symptom Manage* 2020. <https://doi.org/10.1016/j.jpainsymman.2020.10.010>
- [17] Ohannessian R, Duong TA, Odone A. Global telemedicine implementation and integration within health systems to fight the COVID-19 pandemic: A call to action. *JMIR Public Heal Surveill* 2020;6. <https://doi.org/10.2196/18810>
- [18] Gohari SH, Bahaadinbeigy K. The most used questionnaires for evaluating telemedicine services 2021:1–11
- [19] Parmanto B, Lewis, Jr. AN, Graham KM, Bertolet MH. Development of the Telehealth Usability Questionnaire (TUQ). *Int J Telerehabilitation* 2016;8:3–10. <https://doi.org/10.5195/ijt.2016.6196>
- [20] Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine (Phila Pa 1976)* 2000;25:3186–91. <https://doi.org/10.1097/00007632-200012150-00014>
- [21] Mogensen CE, Standl E. Prevention and Treatment of Diabetes Late Complications. De Gruyter, Inc.; 2010
- [22] de Vet HCW, Terwee CB, Mokkink LB, Knol DL. Measurement in Medicine. Cambridge University Press; 2011. <https://doi.org/10.1017/CBO9780511996214>
- [23] IBM. IBM Corp. Released 2020. IBM Statistics for Windows, version 27.0. Armonk 2020
- [24] Medical Association W. WMA DECLARATION OF HELSINKI – ETHICAL PRINCIPLES FOR MEDICAL RESEARCH INVOLVING HUMAN SUBJECTS 2020. <https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/>
- [25] Ministry of Health SÆ. Bekendtgørelse af lov om videnskabetisk behandling af sundhedsvidenskabelige forskningsprojekter 2017. <https://www.retsinformation.dk/eli/lt/2017/1083>

- [26] Bland JM, Altman DG, Bland JM. Statistics Notes : The Odds Ratio linked references are available on JSTOR for this article : Statistics Notes The odds ratio. *Br Med J* 1997;314:572
- [27] Micoulaud-Franchi JA, Sauteraud A, Olive J, Sagaspe P, Bioulac S, Philip P. Validation of the French version of the Acceptability E-scale (AES) for mental E-health systems. *Psychiatry Res* 2016;237:196–200. <https://doi.org/10.1016/j.psychres.2016.01.043>
- [28] Argimon-Palls JM, Flores-Mateo G, Jiménez-Villa J, Pujol-Ribera E. Psychometric properties of a test in evidence based practice: The Spanish version of the Fresno test. *BMC Med Educ* 2010;10. <https://doi.org/10.1186/1472-6920-10-45>
- [29] Okwechime IO, Roberson S, Odoi A. Prevalence and predictors of pre-diabetes and diabetes among adults 18 years or older in Florida: A multinomial logistic modeling approach. *PLoS One* 2015;10:1–17. <https://doi.org/10.1371/journal.pone.0145781>
- [30] Gucciardi E, Wang SCT, DeMelo M, Amaral L, Stewart DE. Characteristics of men and women with diabetes: Observations during patients’ initial visit to a diabetes education centre. *Can Fam Physician* 2008;54:219–27
- [31] Turi KN, Buchner DM, Grigsby-Toussaint DS. Predicting risk of type 2 diabetes by using data on easy-to-measure risk factors. *Prev Chronic Dis* 2017;14:E23–E23. <https://doi.org/10.5888/pcd14.160244>
- [32] World Health Organization T. Diabetes 2020. https://www.who.int/health-topics/diabetes#tab=tab_1
- [33] Khunkaew S, Fernandez R, Sim J. Demographic and clinical predictors of health-related quality of life among people with type 2 diabetes mellitus living in northern Thailand: A cross-sectional study. *Health Qual Life Outcomes* 2019;17:1–9. <https://doi.org/10.1186/s12955-019-1246-2>

7 ACKNOWLEDGEMENT

The authors would like to acknowledge all translators, expert committee members, and participants.

Towards Accurate Computer Vision-Based Marker Less Human Joint Localization for Rehabilitation Purposes

Thomas Hellstén¹, Jonny Karlsson², Christian Häggblom² and Jyrki Kettunen³

¹ Arcada University of Applied Sciences, School of Engineering, Culture and Wellbeing, Finland, thomas.hellsten@arcada.fi

² Arcada University of Applied Sciences, School of Engineering, Culture and Wellbeing, Finland

³ Arcada University of Applied Sciences, Graduate School and Research, Finland

Abstract

In this paper we present a computer vision (CV) based prototype application for knee range of motion analysis. The prototype is built on top of an existing CV pose estimation technique, requiring only one web camera. The aim was to investigate whether it can provide adequate measurement accuracy for rehabilitation purposes. Pilot testing were used to compare the accuracy of the prototype with universal goniometer when measuring range of motion of the knee joint. Our research indicates that sufficient accuracy for range of motion analysis of the knee can potentially be achieved in standing and lying positions by extending the underlying training dataset.

Keywords: computer vision, DensePose, marker less, telerehabilitation

1 INTRODUCTION

The COVID-19 pandemic has forced health care organizations to implement telerehabilitation (TR) as a part of health professionals' daily practice [1]. TR provides the possibility for clients to receive therapy without physically visiting a clinic or hospital. TR has also supported social isolation politics to reduce the spread of COVID-19 [1][2]. Providing easily and equally achieved TR services is a challenge due to the aging population and the concentration of healthcare services [3]. TR can be a way of improving availability of rehabilitation, and is defined as rehabilitation services that is delivered to clients through information and communication technologies (ICT) [4].

TR can involve direct online communication with a health professional, so the client and the health professional are physically at different locations, but it can also mean a technology used in health care that provides automatic feedback and support for the client [5]. Technology that can be used in TR include e.g. telephone, smartphone, computer, tablet, activity trackers, computer vision (CV), artificial intelligence (AI), virtual reality (VR) or robotics [6].

A promising and new way of implementing automatic real-time telerehabilitation services is through CV as the only technical equipment needed is one or more cameras and a computing device, such as laptop, tablet or smartphone. Tracking and analysis of human motions using CV has been an intensive research topic already for decades [7].

Traditional CV based motion analysis uses marker-based approaches, involving installation of dots or other

reflective material on key points of the body, such as knee, ankle or shoulder joints. This limitation makes routine use of motion analysis systems impractical, as they require significant technical preparations prior to rehabilitation performance. Three-dimensional (3D) CV systems, such as Vicon, have been used as golden standard in the field of CV [8], however, these include advanced and precisely calibrated equipment and are thus too expensive for home use.

The potential for providing cost-effective and easy-to-use solutions for home environments, marker-less CV solutions for rehabilitations applications have been of interest in the field of TR [9]. Recently, a comparison of marker-less vs. marker-based solutions for Gait analysis through a proof of concept study has been presented in [10]. The authors propose a multi RGB camera neural network based system for detecting and localizing key points of the human body.

In [11] a "semi-marker-less" system is proposed for knee angle measurement during lower limb rehabilitation. The solution is designed for home environments and requires only a single camera. Though, it requires the placement of three physical markers to be able to accurately localize the key joints. The system has shown promising performance results on a robotic arm.

A mobile application, based on computer vision, enabling the automatic identification of anatomical landmarks for recognition of body alignment angles is proposed in [12]. This application, known as NLMeasurer, automatically detects 17 anatomical landmarks from an image consisting of a frontal view of human body. The landmarks can be detected both with and without physical markers attached to the body. Based on these anatomical landmarks

The 18th Scandinavian Conference on Health informatics, Tromsø, Norway, August 22-24, 2022. Organized by UiT The Arctic University of Norway. Conference Proceedings published by Linköping University Electronic Press at <https://doi.org/10.3384/ecp187>. © The Author(s). This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>

NLMeasurer can assess the posture. Results from experiments indicate that NLMeasurer provides a valid solution for postural analysis from a frontal view when markers are used. However, when physical markers were not used, the measurements were not fully reliable.

In this paper, we propose a novel CV-based knee angle measurement prototype application for rehabilitation purposes. The prototype only requires a single off-the-shelf web camera and no physical markers. The CV prototype has been developed and critically evaluated within an interdisciplinary research team, including experts from the field of physiotherapy and information technology. As a guideline for the development process the Centre for eHealth Research Roadmap (CeHRes) was used [13].

The rest of the paper is structured as follows: An overview of the technical features of the CV prototype is presented in Section 2. The knee angle measurement performance is critically evaluated in Section 3. Future steps in the development process, with the purpose of extending and improving the accuracy of the measurement capabilities of the CV prototype, are discussed in Section 4. Finally, some concluding remarks are presented in Section 5.

2 COMPUTER VISION BASED KNEE ANGLE MEASUREMENT PROTOTYPE

A prototype application was developed for answering the following research question: Can existing CV-based marker-less human pose estimation techniques, based on a single camera, provide adequate joint localization accuracy for rehabilitation purposes? The technical choices and decisions made for the development are, thus, supported by a systematic review of existing 2D marker-less pose estimation systems [9].

Dense human pose estimation in the wild (DensePose) [15] is a promising technique, in terms of joint localization accuracy. DensePose uses a Region-based Convolutional Neural Network (R-CNN), based on the Mask R-CNN framework proposed in [14], for mapping all pixels of an RGB image, associated with a human, to the 3D surface of the human body. Based on these 2D to 3D mappings, known as dense correspondences, it then estimates the pose of that person. DensePose is trained on a large-scale ground-truth dataset, called DensePose-COCO [15], with 2D image to 3D surface correspondences manually annotated on 50 000 images. DensePose takes a single image as input and produces, in addition to the dense correspondences, an output image marked with the 2D coordinates of key points of the human body, including ankle, knee, hip, wrist, elbow and shoulder joints (figure 1).

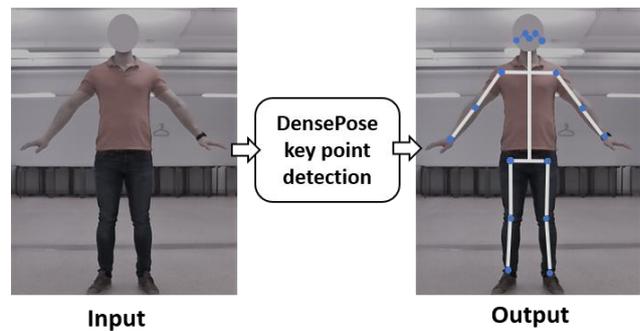


Figure 1. The key point detection feature of DensePose

The key point detection feature of DensePose has been reused in the CV prototype where 2D coordinates of the hip, knee and ankle joints are captured for each frame of the input video stream produced by a standard web camera. The knee angle is then calculated applying the law of cosines as shown in figure 2.

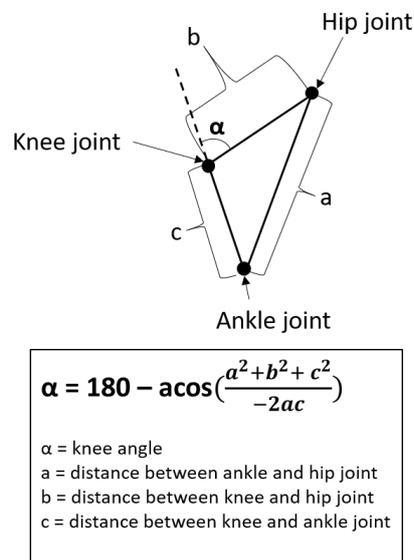


Figure 2. The knee angle measurement procedure of the CV prototype

The knee angle measurement procedure is performed for each frame and the CV prototype includes a save button allowing the user to save the knee angle to a log file at any given time. A screen shot of the output of the CV prototype when measuring the knee angle in a stand-up position is shown in figure 3.



Figure 3. The CV prototype measuring the knee angle of a client in a stand-up position

The CV prototype can be executed on a PC computer equipped with a standard web camera. For the time being, however, it requires a CUDA-enabled NVIDIA GPU. A workaround for this limitation is a distributed approach where the key point detection functionality is executed on a cloud instance and only the user interface is executed locally on the PC, tablet, or smartphone. This will, however, be prioritized in a later stage of our research project.

3 PERFORMANCE EVALUATION

For evaluating the accuracy of the CV prototype, pilot testing was applied to compare the CV prototype with universal goniometer (UG) when measuring subjects' range of motion in the knee joint. In clinical work, physiotherapists typically use UG to measure their clients' joint angles for clinical decision-making and to follow up the rehabilitation process. Goniometric joint angle measurement values can vary up to 5° from the actual angle [16]. This typically happens if the physiotherapist has improper placement of its fulcrum over the center of rotation of the joint or wrong anatomic structures [17].

In our pilot tests, healthy working-age female and male subjects (N = 30) were selected from among the Arcada University of Applied Sciences staff and students. Subjects who suffered from pain or other symptoms in the lower limbs during the preceding 3 months were excluded. Before the pilot test, subjects were provided with written informed consent and a standard written protocol was used when the joint angle measurements were performed. The pilot tests were approved research permission from Arcada University of Applied Sciences, in April 2021.

Knee angle measurement tests were performed on subjects in two different positions, i.e. standing up and lying down. When the subject was in standing position the knees were bent to maximum (deep squat) and when the subject was lying down the knee closest to the camera was bent in a randomized angle. If there was a technical issue with the CV prototype, the result was excluded. Technical issues emerged occasionally when subjects were in lying position with one leg straight and the other leg bent. In this case, the CV prototype sometimes confused the right leg with the left leg and thus produced erroneous values. An example of this problem is shown in figure 4, where the intention of the CV prototype is to measure the knee angle of the left leg, but as the knee joint of the left leg is confused with the right leg, the measurement result is incorrect.

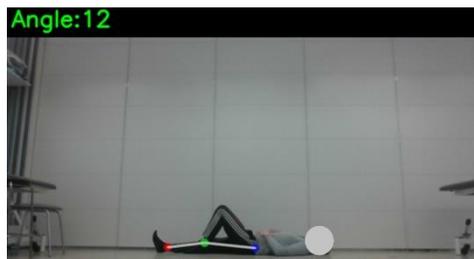


Figure 4. Example where the CV prototype confuses the knee joint of the left leg with the right leg and hence provides an erroneous measurement result

3.1 Performance results

The pilot test included 15 women (mean age 22.6 y) and 15 men (mean age 25.7 y). There was no difference (p=0.2) between genders in mean body-mass index (kg/m²; male mean 23.8, standard deviation (SD) 2.5 vs female mean 22.6, SD 2.3).

The measurement accuracy of the CV prototype was validated by calculating the mean difference between all CV prototype and UG measurement values. When the subjects was in standing position (N=30) and in lying (N=25) the mean difference between UG and CV based knee angle measurements was 3.4°. The variation of the UG and CV measurements values in standing position lied between -6.9 and 13.7 (95% CI) and in lying position between -17.4 and 24.2 (95% CI) from the mean. More detailed results are presented in figure 5 and 6.

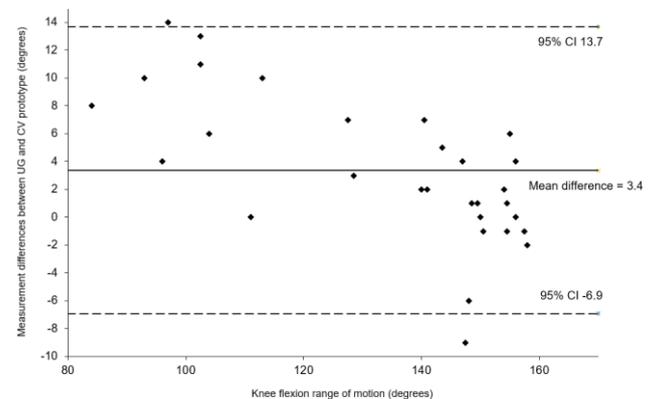


Figure 5. Bland-Altman plots showing the individual measurement differences and the mean difference between CV prototype and UG knee angle measurements in standing position with maximum knee bending. The X-axis denotes the knee bending angle. The Y-axis denotes the measurement differences in degrees between the two methods.

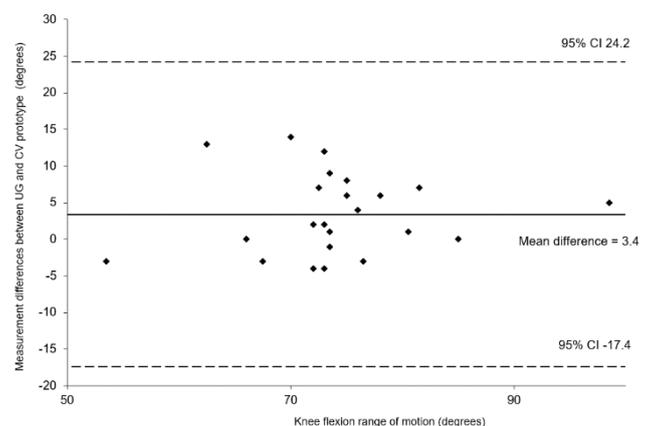


Figure 6. Bland-Altman plots showing the individual measurement differences and the mean difference between CV prototype and UG knee angle measurements for random knee angles when the subject is in lying position. The X-axis denotes the knee bending angle. The Y-axis denotes the measurement differences in degrees between the two methods.

4 DISCUSSION AND FUTURE RESEARCH

A mean difference of 3.4° in the measurement values between the CV prototype and UG can be considered as an acceptable result given UG joint angle measurements as such can vary up to $\pm 5^\circ$ in clinical use [16]. However, despite an acceptable mean difference, the variation was still occasionally high.

A possible reason for the variation between the UG and CV measurements is that UG measurements are not exact and CV prototype measurements, on the other hand, were occasionally inconsistent. The main reason for the inconsistency of measurements with the CV prototype is the lack of training data representing people in different positions in the DensePose-COCO dataset. Many of the positions during knee angle measurement are unique, e.g. one leg bent and the other extended. The DensePose-COCO dataset does not include training images with these types of positions and hence the ankle and knee joint of the leg to be measured tend to be confused with the other leg in many occasions leading to erroneous measurement results. The problem is exacerbated by the fact that most of these positions are not of as much interest in typical use, and thus these positions might not be evaluated as rigorously when the original model is trained.

Before the CV prototype is implemented in rehabilitation the accuracy demands has to be resolved, as an incorrect joint angle measurement can affect the clinical decision in rehabilitation. Therefore, as a part of future research, we will investigate how the DensePose-COCO data set could be extended to provide more accurate joint angle measurements and to prevent the mix-up of right and left leg. We will also focus on how our CV prototype application could be extended to also analyze more functional movements used in rehabilitation, such as therapeutic exercises like walking or balance tasks.

5 CONCLUSION

Computer-vision (CV) based marker-less human pose estimation is an attractive technique for telerehabilitation (TR) as it can provide clients with instant feedback on rehabilitation exercises, without the direct involvement of e.g. physiotherapist, and it does not require other hardware than a computing device with a standard web camera. CV based marker-less techniques, however, have not originally been designed for rehabilitation purposes and therefore providing an adequate level of joint localization accuracy, meeting the requirements for rehabilitation purposes, is challenging.

In this paper we have presented a novel CV-based marker-less prototype, based on DensePose, for measuring the angle of the knee joint. Results from pilot testing indicated that an acceptable measurement accuracy is achieved, although there were some errors. Sufficient accuracy for knee range of motion analysis can though potentially be achieved in both standing and lying positions by extending the underlying training dataset.

6 REFERENCES

- [1] Withers, H., Glinsky, J., Chu, J., et al. Face-to-face physiotherapy compared with a supported home exercise programme for the management of musculoskeletal conditions: protocol of a multicentre, randomised controlled trial—the REFORM trial. *BMJ open* Vol 11, Issue 5. 2021.
- [2] Cottrell, M., Russell, T. Telehealth for musculoskeletal physiotherapy. *Musculoskeletal Science and Practice* Vol 48, pp. 102-193. 2020.
- [3] Truter, P., Russell, T., Fary, R. The validity of physical therapy assessment of low back pain via telerehabilitation in a clinical setting. *Telemedicine and e-Health* Vol 20, Issue 2, pp. 161-167. 2014.
- [4] Richmond, T., Peterson, C., Cason J., et al. American Telemedicine Association's Principles for Delivering Telerehabilitation Services. *International journal of telerehabilitation* Vol 9, Issue 2, pp. 63-68. 2017.
- [5] Capecci, M., Ceravolo, MG., Ferracuti, F., et al. A Hidden Semi-Markov Model based approach for rehabilitation exercise assessment. *J Biomed Inform* Vol 78, pp. 1-11. 2018.
- [6] Salminen, A., Hiekkala, S., Stenberg, J. Etäkuntoutus. Kansaneläkelaitoksen julkaisuja. Tampere, Juvenes Print. 2016.
- [7] Debnath, B., O'brien, M., Yamaguchi, M., Behera, A. A review of computer vision-based approaches for physical rehabilitation and assessment. *Multimedia Systems*, pp. 1-31. 2021.
- [8] Aughey, R., Ball, K., Robertson, S., Duthie, G., Serpiello, F., Evans, N., et al. Comparison of a computer vision system against three-dimensional motion capture for tracking football movements in a stadium environment. *Sports Engineering* Vol 25, Issue 1, pp. 1-7. 2022.
- [9] Hellsten, T., Karlsson, J., Shamsuzzaman, M., Pulkkis, G. The Potential of Computer Vision-Based Marker-Less Human Motion Analysis for Rehabilitation. *Rehabilitation Process and Outcome*, 10. 2021.
- [10] Moro, M., Marchesi, G., Hesse, F., Odone, F., Casadio, M. Markerless vs. Marker-Based Gait Analysis: A Proof of Concept Study. *Sensors* Vol 22, Issue 5. 2022.
- [11] Yang, Y., Wang, L., Yao, Y., He, C., Yin, H. Lower Limb Rehabilitation Motion Angle Measurement Based On Deep Learning YOLOv3 Model. In *ICMLCA 2021; 2nd International Conference on Machine Learning and Computer Application*, 1-6. 2021.
- [12] Moreira, R., Fialho, R., Teles, A., Bordalo, V., Vasconcelos, S., de Moraes Gouveia, G., Bastos, V., Teixeira, S. A computer vision-based mobile tool for assessing human posture: A validation study. *Computer Methods and Programs in Biomedicine* Vol 214, 106565. 2022.
- [13] Kelders, S., Lisette van Gemert-Pijnen, Sanderman, R., Kip, H. *eHealth Research, Theory and Development: A Multi-Disciplinary Approach*, Taylor and Francis, Milton. 2018.

- [14] He, K., Gkioxari, G., Dollár, P., Girshick, R. Mask r-cnn. In Proceedings of the IEEE international conference on computer vision, pp. 2961-2969. 2017
- [15] Güler, R., Neverova, N., Kokkinos, I. Densepose: Dense human pose estimation in the wild. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 7297-7306. 2018
- [16] Aalto, T., Airaksinen, O., Härkönen, M., Arokoski, J. Effect of passive stretch on reproducibility of hip range of motion measurements. Archives of physical medicine and rehabilitation Vol 86, Issue 3, pp. 549-557. 2005
- [17] Norkin, C., White, D. Measurement of joint motion: a guide to goniometry. Fifth edition. Philadelphia: F. A. Davis Company. 2016.

7 ACKNOWLEDGEMENT

This work was supported by ‘Fonden för teknisk undervisning & forskning’ (TUF), a fund for supporting education and technical research at Arcada University of Applied Science

From paper-based to electronic prescribing of multidose drug dispensing — effects on pharmacy workload

Anette Vik Jøsendal¹ and Trine Strand Bergmo^{1,2}

¹Norwegian Centre for E-health Research, University Hospital of North Norway, 9038 Tromsø, Norway

² Department of Pharmacy, UiT, the Arctic University of Norway, 9037 Tromsø, Norway

Anette.vik.josendal@ehealthresearch.no

Abstract

Since 2014, an electronic prescribing system has been piloted for patients receiving medications as multidose drug dispensing; a system commonly used in home care services. In this longitudinal study of 499 patients, we investigate workload at the pharmacies, measured as the number of times pharmacists assess prescriptions. In the 26-week period before the implementation, 17% of the patients got their prescriptions assessed by a pharmacist every 2 weeks, in the 42 weeks after, this increased to 47%. This considerably increases the pharmacy workload, with an estimated additional 602.000 pharmacist assessments every year if all eligible patients get the new prescribing system

Keywords

Multidose drug dispensing, e-prescribing, shared medication list

1 INTRODUCTION

Multidose drug dispensing (MDD) is an adherence aid commonly used for patients in home care services. MDD is machine dispensed solid medicines (tablets and capsules) in disposable plastic bags that replaces manually filled dosettes. MDD is common in hospitals across the world but is also used in primary care in the Scandinavian countries, Finland and the Netherlands [1]. In Norway, there are about 100.000 users of MDD, most of whom get it via home care services [2]. MDD is believed to reduce administration errors, improve medication adherence, save medication costs, save working time for nurses and reduce waste of unused medications [3-11]. However, the system is also associated with more inappropriate prescribing, increased risk of errors at care transitions and patients having fewer changes in their medication treatment, compared to patients with ordinary prescribing [12-16]. There are, however, very few studies on the effects of MDD systems in primary care in general [17, 18].

In Norway, the prescriptions used to dispense MDD differs from ordinary prescriptions. Firstly, the MDD prescriptions consist of a complete list of the medication use, which includes all regular medications (both those dispensed as MDD and those who are dispensed in their original packaging), *when needed* medications, medical devices and dietary supplements. Ordinary prescriptions only contain one medicine at a time and will not contain dietary supplements. Secondly, the MDD prescriptions are paper-based, usually printouts from the GPs medication journal, faxed or sent by fax to the pharmacy. Ordinary prescriptions are electronically transferred via a national database accessible to all pharmacies and prescribers in the country. Since 2014 an electronic prescribing system for MDD patients has been piloted, the implementation is,

however, slow. At the time of writing, about 2300 patients are getting MDD based on electronic prescriptions [19].

1.1 The e-prescribing system

The electronic MDD system uses the same e-prescriptions as ordinary prescribing. However, the system also requires the GP to create a digitally shared medication list (SML) transferred via the same database. As for the paper-based MDD prescription, this SML contains a total overview of the patient's medications: regular medications, *when needed* medications and dietary supplements. Unlike the MDD prescription, however, the SML is not legally a prescription meaning that it cannot be used to dispense medications by itself. It is thus necessary with e-prescriptions for each medication on the SML to dispense medications. These accompanying e-prescriptions are identical to ordinary prescriptions and are therefore available for any pharmacy or physician in the country.

The electronic prescribing of MDD is not a new system, but rather a function in the existing electronic health record (EHR) systems the GPs are using. This function can be turned on at each installation. Once this is turned on, a GP can define a patient as using MDD. This means that the next time the GP prescribe medications for this patient, the system will automatically generate an SML for the patient along with the e-prescriptions.

We know from previous research that e-prescribing can introduce workarounds and change work practices for the personnel involved [20-22]. The physicians using the e-prescribing system for MDD have so far reported that the system is both less time-consuming and safer for the patients [23, 24]. Interviews with pharmacists and nurses have revealed that they experience the system to be more time-consuming, specifically the pharmacists described having to assess prescriptions more frequently [20].

The 18th Scandinavian Conference on Health informatics, Tromsø, Norway, August 22-24, 2022. Organized by UiT The Arctic University of Norway. Conference Proceedings published by Linköping University Electronic Press at <https://doi.org/10.3384/ecp187>. © The Author(s). This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>

Description of step	Paper-based	Electronic
Transmit Prescription sent from prescriber to the pharmacy and linked to a patient in the pharmacy dispensing program.	Prescription transmitted via fax or mail. A pharmacy employee manually link prescription and patient	Prescription transmitted via the national prescription database. Prescription automatically linked to a patient via social security number
Transcribe Prescription entered into the pharmacy dispensing program	Manual transcribing of the entire prescription: medication, dosing schedule, prescriber, reimbursement.	Semi-automatic transcribing. A pharmacy employee chose what prescription to transcribe, and most information is automatically entered except the dosing schedule
Pharmacist check A pharmacist assesses the clinical appropriateness and validity of the prescription. Contact the prescriber or home care services if there are problems.	Includes double-checking the transcription from the previous step. Should be done by another pharmacist than the one doing the transcribing.	Same as for the paper-based, but less need to check for correct transcribing.
Order The pharmacist sends an electronic order to the MDD supplier.	Sent per patient group/home care service, not per patient	Same as for the paper-based

Table 1: Steps in the prescription handling at the pharmacy

The aim of this study is to investigate whether the electronic system for MDD affects the pharmacist workload by analyzing the number of times they needed to assess the MDD prescription.

2 METHODS

We conducted a longitudinal study using the MDD prescriptions from the main MDD supplier in Norway.

2.1 The work process at the pharmacy

To dispense a medication, a pharmacist needs to check the prescription. This check includes a clinical evaluation of the appropriateness of the treatment in relation to the patient's age, gender and indication for the medication, as well as checking the validity of the prescription and ensuring that the patients get appropriate information about how to use the medications [25, 26]. For ordinary prescriptions, this assessment is done every time a medication is dispensed (usually every 3 months for regularly used medications), and only for the prescriptions that are dispensed at a given time. For MDD, however, this is done the first time a prescription is included on the MDD prescription, and then only when there are changes to the treatment [26]. The pharmacist also assesses the entire MDD prescription, regardless of whether the medicines are actually dispensed at the time of the check. Before the prescription can be checked by a pharmacist, the prescription needs to be transmitted to the pharmacy and transcribed into the MDD dispensing system. Table 1 describes these different steps, and how they differ between the paper-based and electronic MDD systems.

2.2 Data collection

We contacted the MDD supplier who started pilot testing electronic MDD in 2014 and asked for the prescriptions used in the period from June 2012 to August 2020. The supplier provided us anonymous data for all patients in the municipalities where the electronic system had been piloted.

The data contained information about when the prescription was checked by a pharmacist and for what period MDD was dispensed for each patient. Regarding the patients, the data contained age and gender. Regarding medications the prescriptions contained details of medication names,

strength and formulation, in addition to dosing schedule and dispensing type (MDD medications, regular medications not dispensed as MDD and *when needed* medications).

2.3 Analyses

For each individual, we defined the index date as the first time MDD had been dispensed on an electronic prescription. Because MDD is usually dispensed for 2 weeks at a time, we divided each patient into 2 weeks intervals, from 26 weeks before the index date to 42 weeks after.

We excluded patients if they did not have data for all intervals, i.e. patients with more than a 2-week stop in their MDD dispensing. If the patient started directly with electronic prescribing or went back to the paper-based system before 42 weeks, they were excluded. So were patients who were transferred to nursing homes during the study period.

We used Stata Stata/MP 15 for the analyses. The main outcome measure was the number of patients where a pharmacist had checked the prescription from one period to the next.

2.4 Ethics

This study was approved by the Data Protection Officer at Apotek 1 AS. The data were anonymised by the pharmacy before being given to the researchers, and the study did not require approval from the Regional Ethics Committee.

3 RESULTS

The original dataset consisted of 2102 patients who had received MDD based on electronic prescriptions at least once during the 8-year period. We excluded 1603 patients who did not have enough data before or after the interventions, and the final dataset consisted of 499 patients.

	Female N (%)	Male N (%)	Total N (%)
Sex	283 (57%)	216 (43%)	499
Age	78 (18)	66 (20)	73 (20)
Number of drugs	10.8 (5.7)	8.9 (4.7)	10.0 (5.3)

Table 2: Study population characteristics

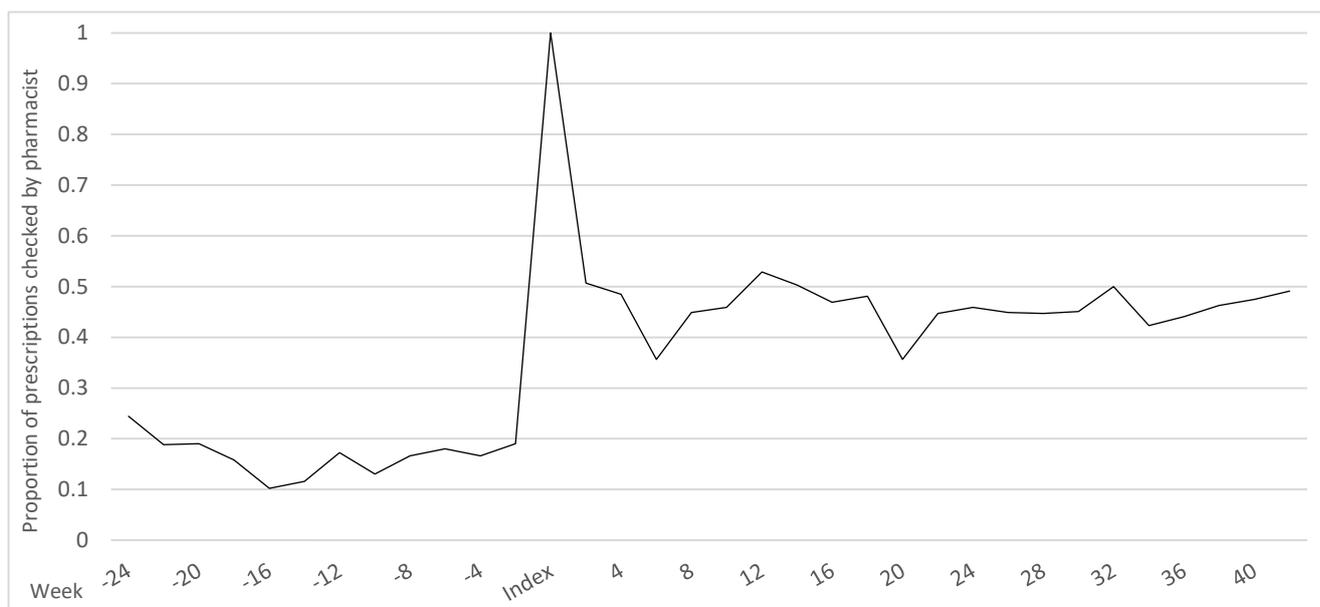


Figure 1: Number of prescription changes

From Table 2 we see that the patients were on average 73 years old, the majority were women and they had on average 10 medications on their prescriptions.

The proportion of patients who got their prescriptions assessed by a pharmacist between each MDD dispensing is shown in Figure 1. At the index date, all patients went from having paper prescriptions to electronic prescriptions, and thus all prescriptions needed to be checked by a pharmacist. From -24 to -2 weeks we find that an average of 17 % of the patients had their prescription assessed by a pharmacist between each MDD dispensing; from week 2 to 42, the average was 46%.

Given this increase, this results in an estimated 602.000 additional pharmacist checks per year if all 80.000 users of the paper-based MDD system in home care services [2] get the electronic prescribing system. In order for the electronic system to be equally efficient for the pharmacists, the process of transcribing and checking prescriptions needs to be $16/46 = 0.36$ times as efficient in the electronic system compared to the paper-based.

4 DISCUSSION

In this longitudinal study, we find a temporal association between the start of the electronic prescribing of MDD and increased number of pharmacist checks. The degree of checks is maintained throughout the 42 weeks follow-up. In terms of pharmacist checks on the prescriptions, this increased from 17 every 2 weeks per 100 patients, to 46, resulting in an estimated 602.000 additional pharmacist checks per year if all eligible MDD patients get the electronic prescribing system.

The number of pharmacist checks on the prescriptions is a reflection of the number of times the medication treatment is changed by a physician. There are several reasons why the electronic prescribing system may increase the number the prescription changes: 1) failures with the transmission are less likely 2) the GP has to send all prescriptions 3) the pharmacy is notified of all changes on a patient 4) increased renewing of prescriptions and 5) changes in prescribing patterns. Each will be discussed in more detail below.

4.1 Fewer errors in the transmission of prescriptions

In the paper-based system, the medication lists are usually faxed or sent by post to the pharmacy. These are manual processes that might fail or be forgotten. Parts of the increase in prescription checks we see in this study might thus be that the electronic system reduces errors in the transmission step. Insufficient communication regarding a patient's medication treatment and manual routines in updating the medication lists are major causes of discrepancies in medication lists [27-33], and reducing errors in transmission probably increases medication safety for MDD users.

4.2 The GP has to send all prescriptions

In the paper-based system, the GP might also choose not to send prescriptions to the MDD pharmacy. This might happen if they know the medicines will not be dispensed in MDD, either because the medicines should be taken when needed, because the formulations are not dispensable as MDD (e.g. creams, liquids and eye drops) or because the medicine should not be dispensed as MDD of other reasons (e.g. antibiotics). With the electronic system, the GP no longer has this option and all medications must be sent.

In interviews with the pharmacists piloting the electronic system, they described how they in the paper-based system typically were not notified if the patient was prescribed e.g. a course of antibiotic, while in the electronic system they would be notified about all changes regardless of whether it would be dispensed in MDD or not [20]. The current MDD dispensing system is designed so that the pharmacist at the MDD pharmacy checks the entire MDD prescription, including those medicines that are not dispensed in MDD. This enables the pharmacist to assess the entire treatment and check for drug-drug interactions of all prescribed medicines. This has previously been described as one of the benefits of the MDD system that helps improve safety [4, 34-37]. However, this is not a legal requirement [38]. Legally, only the prescriptions that are dispensed (i.e. the MDD medications) need to be assessed by a pharmacist at this stage, as the other prescriptions will have to be checked again by a pharmacist when they are actually dispensed in their original packaging at a later time. Since the pharmacy

gets notifications about all changes on a patient, including those that do not directly affect the medicines dispensed as MDD, this might also contribute to the increased number of pharmacist checks we find in this study.

4.3 Notifications about all changes

Similarly, the pharmacy is notified about all changes to the patient's medication treatment, also those done by doctors other than the GP. That the MDD pharmacy does not get automatically updated about prescriptions from e.g. the hospital is a known weakness of today's MDD system. Previous studies have shown many challenges with MDD patients during care transitions [27, 34, 39-42], and it has been shown that MDD patients have between 3 and 18 times increased risk of errors in this process [13, 14, 43]. When the MDD pharmacy is not notified about these prescriptions the home care services might have to manually correct the MDD bags, which is a time-consuming process that is also prone to errors [44-46]. Having direct access to the prescriptions from other prescribers than the GP has been described as one of the benefits of the electronic MDD system in Norway in terms of increasing medication safety [20]. However, the pharmacists also described it as being a major cause of increased workload because it results in a lot more clarifications and checks [20].

4.4 Renewing prescriptions

A fourth cause of the increased number of prescription changes we see in this study is the design of the electronic prescribing system when it comes to the validity of prescriptions. In the paper-based system, there is one MDD prescription, valid for one year supply for all the medications on the prescription, and with one expiry date. In the electronic system, there are individual e-prescriptions for each medicine, all with potentially different expiry dates. In addition, e-prescriptions contain a quantity that can be dispensed, meaning that the prescription can be emptied out before a year has passed. In fact, prescribing the wrong quantity has been shown to be among the most common errors on e-prescriptions [47-49]. When a prescription is renewed, despite the medicine, dosing schedule and prescriber being identical to the expired prescription, this requires a new check by the pharmacist as it is formally a new prescription. Expired prescriptions were described as one of the major causes why both nurses and pharmacists found the electronic MDD system more time-consuming than the paper-based system [20].

It has previously been suggested that the decrease in prescribing quality that is seen in MDD patients over time, might be due to too automatic renewing processes for these patients [50, 51]. A more frequent renewing of prescriptions might thus improve prescribing quality as the GP more often review the treatment of the patient. However, renewing prescriptions can also be a technical task and does not have to include reviewing the treatment as a whole [52]. In a cross-sectional study of 336 patients testing the electronic MDD system, it was found that 23 % were missing prescriptions on regularly used medications after the transition to the electronic system [53]. Another study interviewing pharmacists and nurses involved in the pilot also described how the patients were more frequently

missing medicines in the MDD bags after the transition to e-prescribing [20]. Considering the risk of the patients not getting prescriptions renewed in time for ordering MDD, and the time pharmacists, nurses and GPs use on the task of renewing prescriptions, it is uncertain whether the increased need for renewing prescriptions would increase medication safety, even if it results in the GP reviewing the medication treatment more frequently. A better approach to increase medication safety would probably be to shift focus away from single e-prescriptions that need individual renewals, and rather focus on the medication treatment and the SML as a whole, and do a medication review of this complete list at set intervals.

4.5 Changes in prescribing patterns.

Lastly, the electronic prescribing system might change the prescribing patterns, which results in an increased number of prescription changes. The SML system seems to improve the overview of the patient's medication use, compared to the paper-based: there are fewer discrepancies between the medication lists at different care providers, and the list is more up-to-date, including prescriptions from hospital physicians as well as dietary supplements [53, 54]. This increased overview might affect the GP's prescribing. Because the e-prescriptions have quantities this also might make the GP more aware of the amount of medications they are prescribing. This might also affect the prescribing, especially for medications with the potential for abuse. Lastly, MDD prescribing has been described as more time-consuming and complex than ordinary prescribing [5, 34, 55], and studies have shown that MDD patients have fewer changes in their medication regimens (starts, discontinuations and dose changes) than patients with ordinary prescribing [50, 51]. Having the same prescribing procedures for patients with ordinary prescribing and MDD might thus also result in the GPs more frequently making changes to the prescriptions.

4.6 Workload

In this study, we have looked at the pharmacist's workload in terms of the number of times they assessed new or changed prescriptions. In order for a pharmacist to do this task, the prescription first has to be transmitted to the pharmacy and transcribed into the MDD dispensing system. As described in Table 1, these steps are more automatic in the electronic system, and are likely less time-consuming compared to the paper-based system. This is also consistent with the pharmacists' descriptions of the new system [20]. The step of checking and correcting the prescriptions, however, might be more time consuming because there seems to be an increased need to do clarifications and contact the prescriber compared to the paper-based system [20, 56]. The finding that e-prescriptions require more frequent contact with GPs than paper-based prescriptions, is in line with previous research [57, 58].

Given the increased number of pharmacist checks performed in the electronic system as shown in this study, the process of prescription management at the pharmacy needs to be almost 3 times as efficient for the pharmacists to use an equal amount of time in the two systems. Future studies should address how this increased workload affects the financial situation of the pharmacies.

4.7 Strengths and limitations

The main strength of this study is that we have used a complete data set for almost 500 patients spanning over 18 months. This makes us able to show a relatively stable trend, increasing the validity of the results. The patients included were older adults and had on average 10 mediations on their prescriptions reflecting the challenges with polypharmacy. One weakness of this study is the lack of a control group. However, we believe that the longitudinal data is solid and capture the trend in prescription changes before and after the implementation of electronic prescribing for MDD users.

5 CONCLUSION

Going from paper to electronic prescription of MDD, increased the number of prescription changes which considerably increased the workload for the pharmacy. This study does not investigate the nature of the prescription changes, but we propose several explanations for the increase. By automating or redesigning the e-prescribing system, especially regarding renewing prescriptions, the workload for pharmacists could likely be lowered. If the extra workload persists when the system is implemented at scale, we need to analyse the economic consequences for the pharmacies in more detail. The consequences for the home care services and GPs should also be investigated. Future studies are also needed to look into whether these prescription changes are clinically relevant.

6 REFERENCES

- [1] Rechel, B., Hub-and-spoke dispensing models for community pharmacies in Europe. *Eurohealth*, 2018. 24(4): p. 3-6.
- [2] Norwegian Pharmacy Association. Apotekstatistikk (in Norwegian). 2022; Available from: <http://www.apotek.no/statistikk/apotekstatistikk/kunden>.
- [3] Wekre, L.J., et al., Multidose drug dispensing and discrepancies between medication records. *Qual Saf Health Care*, 2010. 19(5): p. e42.
- [4] Wekre, L.J., et al., GPs' prescription routines and cooperation with other healthcare personnel before and after implementation of multidose drug dispensing. *Scand J Public Health*, 2012. 40(6): p. 523-30.
- [5] Wekre, L.J., L. Melby, and A. Grimsmo, Early experiences with the multidose drug dispensing system--a matter of trust? *Scand J Prim Health Care*, 2011. 29(1): p. 45-50.
- [6] Søndergaard, B., et al., Litteraturstudie af dosisdispensering som medicinsk teknologi. Arbejdsrapport (in Danish). 2005, The Danish University of Pharmaceutical Sciences, Danish College of Pharmacy Practice, University of Southern Denmark.
- [7] Mertens, B.J., et al., Effect of multidose drug dispensing on the time in therapeutic range in patients using vitamin-K antagonists: A randomized controlled trial. *J Thromb Haemost*, 2020. 18(1): p. 70-78.
- [8] van Rein, N., et al., Multi-dose drug dispensing as a tool to improve medication adherence: a study in patients using vitamin K antagonists. *Pharmacoepidemiol and Drug Saf*, 2018. 27(1): p. 46-51.
- [9] Price Water Coopers, [Multidose drug dispensing An economic assessment of measures]. 2007, Social - og Helsedirektoratet: Oslo.
- [10] Riksförsäkringsverket, Medicin på kredit och i påse. Apotekets delbetalningssystem och dosdispenseringsverksamhet. 2001, Riksförsäkringsverket: Stockholm.
- [11] Association of Finnish Pharmacies, Annual Review: Medicines in prepacked doses. 2003.
- [12] Belfrage, B., et al., Prevalence of suboptimal drug treatment in patients with and without multidose drug dispensing—a cross-sectional study. *Eur J Clin Pharmacol*, 2014. 70(7): p. 867-872.
- [13] Caleres, G., et al., Medication Discrepancies in Discharge Summaries and Associated Risk Factors for Elderly Patients with Many Drugs. *Drugs Real World Outcomes*, 2020. 7(1): p. 53-62.
- [14] Midlöv, P., et al., Medication errors when transferring elderly patients between primary health care and hospital care. *Pharm World Sci*, 2005. 27(2): p. 116-120.
- [15] Johnell, K. and J. Fastbom, Multi-dose drug dispensing and inappropriate drug use: A nationwide register-based study of over 700,000 elderly. *Scand J Prim Health Care*, 2008. 26(2): p. 86-91.
- [16] Lea, M., et al., Prevalence and risk factors of drug-related hospitalizations in multimorbid patients admitted to an internal medicine ward. *PloS one*, 2019. 14(7): p. e0220071.
- [17] Sinnemäki, J., et al., Automated dose dispensing service for primary healthcare patients: a systematic review. *Syst rev*, 2013. 2(1): p. 1.
- [18] Halvorsen, K.H. and A.G. Granas, [Multi-dose dispensed drugs in Scandinavia - A systematic review of possibilities and limitations] in Norwegian. *Nor Farm Tidsskr*, 2012. 4: p. 22-28.
- [19] The Directorate of e-health. Multidose i e-resept (in Norwegian). 2021; Available from: <https://ehelse.no/prosjekt/multidose-i-e-resept>
- [20] Josendal, A.V. and T.S. Bergmo, From Paper to E-Prescribing of Multidose Drug Dispensing: A Qualitative Study of Workflow in a Community Care Setting. *Pharmacy*, 2021. 9(1): p. 41.
- [21] Grossman, J.M., et al., Transmitting and processing electronic prescriptions: experiences of physician practices and pharmacies. *Journal of the American Medical Informatics Association*, 2012. 19(3): p. 353-359.
- [22] Villaseñor, S. and R.J. Piscotty Jr, The current state of e-prescribing: Implications for advanced practice registered nurses. *Journal of the American Association of Nurse Practitioners*, 2016. 28(1): p. 54-61.
- [23] Bergmo, T.S., A.V. Jøsendal, and E. Johnsen, Factors easing the transition from paper to electronic prescribing of multidose dispensed drugs (MDD), in SHI2019: 17th Scandinavian Conference on Health

- Informatics, November 12-13, 2019. 2019, Linköping University Electronic Press: Oslo, Norway. p. 1-6.
- [24] Gullslett, M.K. and T.S. Bergmo, Implementation of E-prescription for Multidose Dispensed Drugs: Qualitative Study of General Practitioners' Experiences. *JMIR Hum Factors*, 2022. 9(1): p. e27431.
- [25] Norwegian Pharmacy Association, Bransjestandard for Apotek - Farmasøytikkontroll i Reseptekspedisjonen (in Norwegian). 2017.
- [26] Rule on Requisition and Distribution of Medicines from Pharmacies, Forskrift om rekvirering og utlevering av legemidler fra apotek (in Norwegian). 2019, Helse- og omsorgsdepartementet.
- [27] Reuther, L.O., et al., Multi-dose drug dispensing is a challenge across the primary-secondary care interface. *Dan Med Bull*, 2011. 58(12): p. A4341.
- [28] Glintborg, B., S.E. Andersen, and K. Dalhoff, Insufficient communication about medication use at the interface between hospital and primary care. *BMJ Quality & Safety*, 2007. 16(1): p. 34-39.
- [29] Frydenberg, K. and M. Brekke, Poor communication on patients' medication across health care levels leads to potentially harmful medication errors. *Scand J Prim Health Care*, 2012. 30(4): p. 234-240.
- [30] Frydenberg, K. and M. Brekke, Kommunikasjon om medikamentbruk i henvisninger, innleggesskriv og epikriser. *Tidsskrift for den Norske Lægeforening*, 2011. 131(9): p. 942.
- [31] Rognstad, S. and J. Straand, Vet fastlegen hvilke medisiner hjemmesykepleien gir pasientene? 2004.
- [32] Bakken, K., et al., Insufficient communication and information regarding patient medication in the primary healthcare. *Tidsskrift for den Norske lægeforening: tidsskrift for praktisk medicin, ny række*, 2007. 127(13): p. 1766-1769.
- [33] Mandt, I., A.M. Horn, and A.G. Granås, Samhandling mellom allmennleger og apotek ved korreksjon av resepter. *Tidsskr Nor Lægeforen*, 2009.
- [34] Bardage, C., A. Ekedahl, and L. Ring, Health care professionals' perspectives on automated multi-dose drug dispensing. *Pharm Pract (Granada)*, 2014. 12(4): p. 470.
- [35] Bell, H.T., A. Steinsbekk, and A.G. Granas, Factors influencing prescribing of fall-risk-increasing drugs to the elderly: a qualitative study. *Scand J Prim Health Care*, 2015. 33(2): p. 107-114.
- [36] Frøyland, H., Legers synspunkter på multidosepakkede legemidler (in Norwegian). *Nor Farm Tidsskr*, 2012. 4: p. 19-21.
- [37] Josendal, A.V., T.S. Bergmo, and A.G. Granas, The Practice Guidelines for Multidose Drug Dispensing Need Revision—An Investigation of Prescription Problems and Interventions. *Pharmacy*, 2021. 9(1): p. 13.
- [38] [Rule on Requisition and Distribution of Medicines from Pharmacies (FOR-1998-04-27-455)]. 1998, Norwegian Ministry of Health and Care Services,.
- [39] Caleres, G., et al., Elderly at risk in care transitions When discharge summaries are poorly transferred and used—a descriptive study. *BMC Health Serv Res*, 2018. 18(1): p. 770.
- [40] Lysen, C., et al., [Medication errors after hospitalisation due to multi-dose drug dispensing in the primary sector] in Danish. *Ugeskr Laeger*, 2011. 173(33): p. 1944-5.
- [41] Alasaad, A., et al., Prescription and transcription errors in multidose-dispensed medications on discharge from hospital: an observational and interventional study. *J Eval Clin Pract*, 2013. 19(1): p. 185-91.
- [42] Caleres, G., et al., Drugs, distrust and dialogue—a focus group study with Swedish GPs on discharge summary use in primary care. *BMC Fam Pract*, 2018. 19(1): p. 127.
- [43] Bergkvist, A., et al., Improved quality in the hospital discharge summary reduces medication errors--LIMM: Landskrona Integrated Medicines Management. *Eur J Clin Pharmacol*, 2009. 65(10): p. 1037-46.
- [44] Nyen, B., et al., Gjennomgang og forbedring av arbeidet med multidose i Larvik kommune. 2009.
- [45] Mertens, B.J., et al., Immediate or deferred adjustment of drug regimens in multidose drug dispensing systems. *Res Social Adm Pharm*, 2018.
- [46] Cheung, K.-C., et al., Medication incidents related to automated dose dispensing in community pharmacies and hospitals—a reporting system study. *PLoS One*, 2014. 9(7): p. e101686.
- [47] Gilligan, A.M., et al., Analysis of pharmacists' interventions on electronic versus traditional prescriptions in 2 community pharmacies. *Res Social Adm Pharm*, 2012. 8(6): p. 523-532.
- [48] Timonen, J., et al., Electronic prescription anomalies: a study of frequencies, clarification and effects in Finnish community pharmacies. *Journal of Pharmaceutical Health Services Research*, 2018. 9(3): p. 183-189.
- [49] Odukoya, O.K., J.A. Stone, and M.A. Chui, E-prescribing errors in community pharmacies: Exploring consequences and contributing factors. *Int J Med Inform*, 2014. 83(6): p. 427-437.
- [50] Wallerstedt, S.M., et al., Drug treatment in older people before and after the transition to a multi-dose drug dispensing system—a longitudinal analysis. *PLoS One*, 2013. 8(6): p. e67088.
- [51] Sjöberg, C., H. Ohlsson, and S.M. Wallerstedt, Association between multi-dose drug dispensing and drug treatment changes. *Eur J Clin Pharmacol*, 2012. 68(7): p. 1095-1101.
- [52] Oravainen, T., et al., How Physicians Renew Electronic Prescriptions in Primary Care: Therapeutic Decision or Technical Task? *Int J Environ Res Public Health*, 2021. 18(20): p. 10937.
- [53] Josendal, A.V. and T.S. Bergmo, [Better agreement between medication records with electronic multidose drug dispensing] in Norwegian. *Nor Farm Tidsskr*, 2018(4): p. 21-23.
- [54] Josendal, A.V., T.S. Bergmo, and A.G. Granas, Implementation of a shared medication list in primary care - a controlled pre-post study of medication discrepancies. *BMC Health Serv Res*, 2021. 21(1): p. 1335.
- [55] Heier, K., et al., [Healthcare providers' experience with multi-dose packaged medicines] in Norwegian.

Tidsskr Nor Laegeforen, 2007. 127(18): p. 2382-2385.

- [56] Johnsen, E., A. Jøsendal, and T. Bergmo, The e-multidose is better for patients' safety than dosette boxes and faxes. Sykepleien Forskning, 2018. 13(e-69983).
- [57] Vik, S., et al., Pharmaceutical interventions on prescriptions in Norwegian community and hospital pharmacies. Int J Clin Pharm, 2020.
- [58] Smith, M. and B. Sprecher, Pharmacy communications with physician offices to clarify prescriptions. J Am Pharm Assoc, 2017. 57(2): p. 178-182.

7 ACKNOWLEDGEMENT

We thank Apotek 1 Gruppen AS for providing data for this study.

Digital Monitoring of Antibiotic Resistance (ABR) in Low- and Middle-Income Countries: A Narrative Literature Review

Yogita Thakral^{1,2}

¹HISP Centre, Department of Informatics, University of Oslo, Norway yogitat@ifi.uio.no

²Health Information Systems Programme, India

Abstract

The objective of this narrative review is to provide an overview of the knowledge and gaps in the existing research on digital Antibiotic resistance (ABR) monitoring in Low- and middle-income countries (LMICs). ABR presents a complex threat to global health. One of the key global priorities is to address this challenge through effective monitoring. An analysis of the literature revealed the missing role of IS (Information systems) research in digital ABR monitoring. A thematic analysis of the identified literature on digital interventions for ABR revealed several gaps. This research contributes by providing potential research directions and identifying the role of IS research in ABR.

Keywords

ABR, AMR, surveillance, monitoring, LMICs, digital platforms

Abbreviations

ABR – Antibiotic Resistance; AMR – Antimicrobial Resistance; NAP – National Action Plan; GAP – Global Action Plan; WHO - World Health Organization; IS – Information System; LMICs – Low- and middle-income countries; TrACSS - Tripartite Antimicrobial Resistance Country Self-Assessment Survey; SDGs - Sustainable Development Goals

1 INTRODUCTION

Antibiotic resistance (ABR) presents a widespread, complex threat to global health and universal health coverage [1]. ABR occurs when microorganisms like bacteria, viruses, parasites, fungi, and other pathogens develop resistance to the drugs used to fight them. The term ABR is used especially for antibiotic resistance in bacteria and is a subset of Antimicrobial Resistance (AMR). Globally, an estimated 700 000 deaths are attributed to ABR annually, with a projected economic impact of US\$100 trillion by 2050 [2]. ABR threatens the effectiveness of treatment of infectious diseases and consequently the sustainability of health systems globally [3]. The adverse consequences of ABR extend to the environment, food production, poverty, health security, and attainment of the Sustainable Development Goals (SDGs) underscoring the need for both global and local research and practical action to address this huge challenge[4].

The World Health Organization (WHO) report on global surveillance of ABR in 2014 highlighted the immediate need for global action to identify actionable information on pathogens and monitor trends of resistance [5,6]. The WHO released a global action plan (GAP) in 2015 and recommended for countries to develop national action plans (NAPs), with a focus on strengthening the knowledge and evidence base through digital-based surveillance and monitoring to strengthen policy and practice[7]. At the policy level, surveillance and monitoring can help in making better estimates of geographical trends and patterns of resistance which can guide decisions related to resource allocation and the building of regulatory frameworks. At

the clinical or practice level, effective monitoring can help develop an evidence base for targeted treatment, build infection control practices, and guidelines for antibiotic prescription practices.

Despite the development of these global and national frameworks, LMICs lag far behind in their effective implementation. While 163 countries have developed NAPs to combat ABR, very few have materialized them in practice [8,9] and, ABR continues to expand mortality and morbidity rates [10]. LMICs are the worst hit with the least resources, including for diagnostic, poor regulation, ad hoc prescription practices, and limited data on the epidemiology of resistance [6,7,11,12].

Surveillance data at local, national, and international levels is needed to guide patients' treatment, inform health policies, trigger responses to health emergencies, and provide early warnings for outbreaks [1]. Current data on ABR surveillance in LMICs are fragmented and lack representativeness [13]. The major sources of ABR data in LMICs are mainly tertiary hospitals, some pharmaceutical companies, private labs, and limited academic literature on the patterns of use of antibiotics [14,15]. Identifying existing research gaps is of crucial importance and a central focus of this paper.

This study aims to provide a narrative literature review and analysis of the existing research related to the applications of digital solutions for monitoring ABR from an LMIC perspective. This paper discusses the existing literature, identifies key gaps, and makes some suggestions for strengthening these identified gaps. While ABR refers to the health of humans, animals, and the environment, referred to as One Health, this paper focuses only on human health in the context of LMICs. We particularly examine

The 18th Scandinavian Conference on Health informatics, Tromsø, Norway, August 22-24, 2022. Organized by UiT The Arctic University of Norway. Conference Proceedings published by Linköping University Electronic Press at <https://doi.org/10.3384/ecp187>. © The Author(s). This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>

what has been the contribution of Information Systems (IS) research to this domain and how can this be enhanced in the future.

2 METHODS

In this study, a narrative literature review was done for data collection to gather and summarise existing research literature on this topic, and to identify dominant themes addressed and directions for future research.

2.1 Search strategy

The existing research on ABR was searched in the AIS eLibrary which is a repository of all the major IS research, including the basket of seven articles. The initial focus of the search was to identify the information systems research related to ABR. However, the search yielded no results, and the search was then extended to Scopus to identify literature on digital monitoring of ABR as it is a repository of major life sciences, social sciences, and health sciences research. The search was broadened to use generic keywords to identify papers even outside the IS domain.

The keywords for database search were identified based on the research focus. An initial search was performed in the Scopus database using the keywords “Antimicrobial resistance”, “Antibiotic resistance”, “Surveillance” and “Monitoring”. surveillance” to obtain a better understanding of the breadth of studies and their focus. Based on the result of the initial search, the scope of the search was defined to include a focus on only literature related to digital ABR monitoring in LMICs in the human domain.

Search terms used included “antimicrobial resistance”, “antibiotic resistance”, “digital surveillance”, “digital platform”, “information system”, “digital monitoring”, LMICs, low- and middle-income countries, and developing countries. The title, abstract, and keywords were searched in May 2022, and no time filter was applied to the search.

2.2 Selection of studies and data retrieval

The metadata query with the selected keywords was used and the search was limited to scientific papers in the English language, papers published from 2011 until 2022 (as of this article’s submission date), and full author information available. These articles were manually screened to identify relevant articles while applying the following exclusion criteria:

- Duplicate articles.
- Articles (the reading of the abstract, introduction, discussion, and conclusion) that were irrelevant to the focus of the study.

The inclusion criteria applied:

- All articles published from 2011 until 2022 (at the time of submission).
- Cited and uncited articles.
- Abstracts (abstract, introduction, discussion, and conclusion) and titles relevant to the theme of study.

2.3 Data analysis

A thematic analysis was then conducted on the identified articles. The technique of thematic analysis was chosen because it is a suitable interpretive method that helps to

uncover key concepts and patterns in a data set [16]. It is a dynamic way to understand and generate explanations from data or to explore an a priori theoretical understanding of a phenomenon under study [17,18].

3 RESULTS

The initial database search in the AIS eLibrary yielded no studies on antimicrobial/antibiotic resistance in the IS domain. An ABR monitoring system is the lifeline of a surveillance and monitoring program to tackle the grand societal problem at all levels including global, national, regional, and facility-specific initiatives. Given the grand nature of the problem, a multidisciplinary approach and collaboration to act at the practice and policy levels are needed but the problem is largely invisible in IS research which could play a guiding role in the realization of the potential of the digital.

The database search in Scopus yielded 870 records which were filtered to 77 after the use of relevant keywords and after removing duplicate papers. The titles, keywords, abstracts, discussion, and conclusions of these papers were further screened, and 37 papers were removed which were found irrelevant. Further, 40 records were considered for detailed assessment of full text and excluded 28 papers not meeting one or several of the inclusion criteria. A total of 12 relevant studies were included in the detailed review. The PRISMA flow chart shows the number of records/studies at each stage (Figure 1).

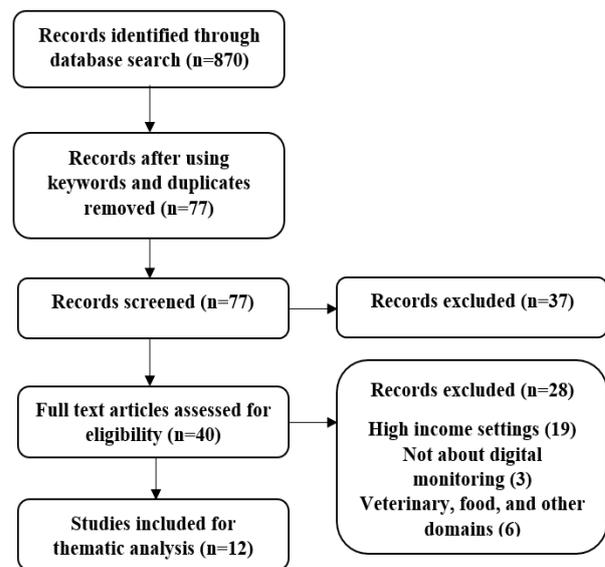


Figure 1: PRISMA flow diagram

3.1 Characteristics of the included studies

Articles from the following LMICs were identified: India, Cambodia, Uganda, Laos, Vietnam, Thailand, Iran, Nepal, India, Bangladesh, Indonesia, Maldives, and East Timor (see map below). The digital technologies discussed

included the widely used WHONET¹ for data capture, Global Antimicrobial Resistance and Use Surveillance System (GLASS²), DHIS2³ (District Health Information System), and other proprietary in-house developed applications. The red lines in figure 2 demonstrate the list of countries from the research articles included in this review.

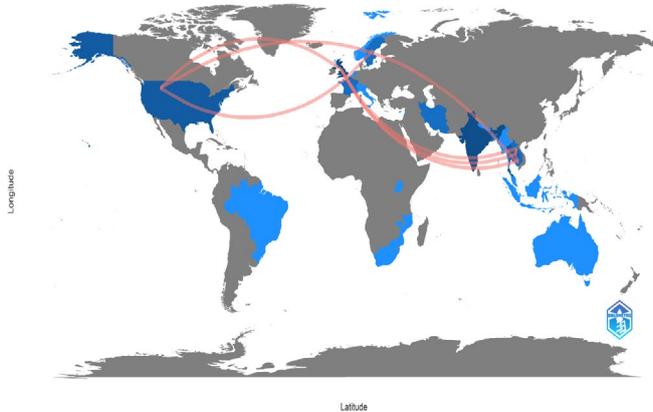


Figure 2: List of countries represented in the review articles

A summary of the key characteristics of the identified papers is summarized in the table below and is then briefly discussed.

Description	Results
Timespan	2011:2022
Sources (Journals, Books, etc)	11
Documents	12
Average years from publication	3.17
Article types	
Journal papers	9
Conference paper	1
Review	2
Sources	
Wellcome Open Research	2
Antimicrobial Resistance and Infection Control	1
BMJ (Online)	1
BMJ Global Health	1
Drug Resistance Updates	1
Frontiers In Public Health	1
IFIP Advances in Information and Communication Technology	1
International Journal of Medical Informatics	1

¹ WHONET is a desktop windows application for the management and analysis of microbiology laboratory data with a particular focus on antimicrobial resistance surveillance developed and supported by the WHO Collaborating Centre for Surveillance of Antimicrobial Resistance.

² Global Antimicrobial Resistance and Use Surveillance System (GLASS) is a global collaborative effort to standardize WHONET is used at the

JMIR Public Health and Surveillance	1
Journal Of Global Antimicrobial Resistance	1
Journal Of Medical Internet Research	1

Table 1: Key characteristics of the identified articles

The articles specifically discussing the digital monitoring of ABR were considered for the study. 12 articles from the period 2011 to 2022 were selected. The selected list included 9 research papers published in journals, 1 conference paper, and 2 review articles. These selected articles were from 11 different outlets. The articles appeared in global public health journals like BMJ global health, BMJ online, Frontiers in public health, and reviews from Wellcome open research.

However, the presence of the importance of the digital in ABR is missing from journals setting the global health agenda. There is an absence of articles in disciplines other than those clinically relevant. Most articles are from global public health journals discussing the importance of ABR monitoring from a global and national perspective. Only one conference paper from the list discussed the relevance of digital systems relevant for multiple contexts at global, national, regional, and facility levels for ABR.

The details of these papers including the main author, country of study, year of study, digital technology used, implementation context and level, and the main findings are presented in Annexure 1. The themes identified from a deeper analysis of the research articles are presented in the next section.

3.2 Analysis: Identifying themes

A content analysis based on a detailed reading of the identified papers was done. Related themes were classified and coded in groups. These codes were reassessed based on further reading and final relevant themes were identified: i) Marginal role of context; ii) Inadequate consideration of scale; iii) Relevance of open-source platforms not considered

3.1.1 Marginal role of context

Context can be defined as “situational opportunities and constraints that affect the occurrence and meaning of organizational behavior as well as functional relationships between variables”[19]. This specifies the role and importance of the development of policies, frameworks, guidelines, and technology based on the *context* where they are implemented. Context-specific development is relevant in the case of ABR since the nature of the problem varies in different contexts and especially in the case of LMICs which are burdened by multiple structural and societal issues in addition to the burden of infectious diseases.

facility/lab level to capture data which is then aggregated and imported to GLASS in a specific format annually.

³ DHIS2 is an open source, web-based platform most commonly used as a health management information system (HMIS) and for case-based data capture and analysis

Six out of twelve studies included the analysis discussing the framework for ABR monitoring at global levels using WHONET and GLASS. All these initiatives rely on good quality data from the micro or the hospital levels to enable monitoring at national and global levels [20]. One of the studies indicates that among the 136 countries reporting to the Global Database for the Tripartite Antimicrobial Resistance Country Self-Assessment Survey (TrACSS) in 2019–2020, only 32 (24%) countries include integrated multisectoral ABR surveillance and monitoring in their NAPs [20]. However, the frameworks developed for use of these applications in LMICs have limited discussions about the context-specific challenges [23,24].

Studies are done at the hospital level, or the department level to identify the need for a patient-based application that could guide them at the practice level and provide information about the local and geographical resistance profiles. Turner et. al. [25] identified the need for a clinically oriented digital tool that could guide at the hospital level as GLASS lacks clinical metadata on antibiotics prescription and use at the local level and the duration of hospitalization. Similarly, Vong et. al. [26] in their study on the use of digital applications for monitoring ABR in seven Asian countries including hospitals in Thailand, Nepal, India, Bangladesh, Indonesia, Maldives, and East Timor identified the need for patient-specific information to act at the local level.

Guidelines for technology and monitoring developed at global and sometimes national levels for countries like India with diverse health profiles in different areas that lack contextual information often fail at the implementation stage. One such example is the poor implementation of NAPs in the countries where specific challenges of implementation are not considered in the plans developed at global and national levels. For example, the guideline to develop a monitoring system at the national level without considering the local challenges of capacity and resources like poor internet, lack of manpower, etc at the contextual level.

3.1.2 Inadequate consideration of scale

Designing for scale means building relevance both for the local facility level and the multiplicity of contexts, within the framework. Such a focus continues to enhance the local value of the processes while also enabling them to be expanded easily to new contexts [27]. ABR represents a unique challenge of scale and scope both geographically and functionally, as it is a global problem without geographical constraints. Functionally, ABR data is not only needed from the microbiology lab at a hospital but also in other departments of the hospital like the antibiotics prescription patterns from the clinical prescription data, etc. to strengthen hospital-wide activities of managing hospital-acquired infections and infection prevention and control activities.

Vong et al. [26] identified challenges with the implementation and use of WHONET in the LMIC context and discussed constraints like configuration of WHONET and BacLink, system interoperability, lack of data standards, and lack of a well-trained local and national IT workforce. Another study in an LMIC context in the Republic of Laos, Vietnam, Myanmar, Thailand, and Vietnam used a locally developed offline application to

generate reports for use at the hospital level[26]. This allowed the hospital under study to generate standardized reports that allowed easy comparison of resistance among facilities. However, challenges were presented with analyzing data and generating a report as lengthy and time-consuming processes a sit required intensive manual work and trained personnel which is an existing challenge with LMICs. One of the selected studies to study the strengthening of surveillance and monitoring in India discusses the features of an in-house developed application [28] that captures and analyses the data collected from 25 tertiary hospitals from the human domain in the country. The limited data submitted to GLASS [13] by India presents the grave challenge of surveillance and monitoring as the data from a total of 71 facilities is sent to GLASS annually from a country with a population of 1.37 billion and more than 200,000 public health facilities across the country [28]. However, most digital applications in the documented articles are being implemented and used at tertiary facilities with limited discussion to scale to public and community facilities. Another study evaluated the use of WHONET and GLASS in a research project to monitor ABR from 2015 to 2020 at a few hospitals in Uganda [29]. The data collected and analyzed during the project duration is planned to be used to guide ABR policies in the country. However, the plan to scale and routinize the use of technology was not discussed in the study.

Among the articles included in the review, 5 studies [24–26,28,30] on monitoring and surveillance of ABR at regional or hospital levels identified the need for systems to collect hospital-specific information but because of the lack of standards in data collection and analysis, the information sharing becomes impossible⁴. The systems developed at the local level thus have limited considerations to scale to different contexts, both geographically and functionally. There are limited studies discussing the challenge of scale in ABR monitoring in LMICs. Only one study discussed the scaling of digital technology for monitoring ABR at multiple levels [31].

3.1.3 Relevance of open-source platforms not considered

Open-source platforms are not only cost-effective by allowing free usage of the platform without having to pay the licensing and maintenance fee, but they are also flexible and scalable. They allow the use of global standards while providing the flexibility to configure the local and user-specific requirements. The use of free and open-source software platforms for the collection, management, analysis, and use of ABR monitoring data is imperative for LMICs struggling with existing challenges of capacity and resources.

One of the main barriers to adopting digital technologies in LMICs is the cost of its purchase and maintenance, which highlights the open-source approach as a good solution for resource-constrained areas [32]. In-house development using proprietary platforms limits the scaling of the application to other contexts and is expensive to maintain. The monitoring platforms to capture and analyze data for ABR developed using proprietary sources in the reviewed articles have presented challenges like lack of system interoperability and lack of data standards. This limits the scope of the applications and limits the standardization of

data analysis [26,28]. Vong et.al. [26] in their study based on high-level discussions between SEARO countries about challenges in ABR monitoring and surveillance, collate the requirements for ABR monitoring in the participating countries. They state the need for an open-source application that is easier to maintain and enables standardized data collection, analysis, and reporting at hospital levels, and allows sharing of data in a standardized format to a central level to guide policy and necessary action. Sahay et. al. [31] discuss the geographic and functional scaling of an open-source platform to capture, analyze and use data to guide both practice and policies at multiple levels.

The studies included in the review (Appendix 1) discuss the challenges with digital platforms developed locally. Four articles included developed the technology locally for facility-specific requirements, but experienced challenges as stated above. This represents an urgent demand for both advanced knowledge and technology which is open-source, reliable, and flexible for ABR monitoring systems, especially in low-resource settings.

4 DISCUSSION

The narrative review provides an overview of the current knowledge and existing gaps in digital ABR monitoring and surveillance in LMICs. The studies presented discussed the development of frameworks and plans for ABR and the use of digital applications at global and national, regional, and facility levels. However, at the facility level, several challenges are encountered to bring the guidelines to practice during the implementation of digital technologies with limited scalability to other contexts. Based on the results, directions for future research on digital monitoring of ABR in LMICs are now discussed that could potentially guide in solving the complex and interconnected pieces of the puzzle.

4.1 Future Research Directions

Building upon our thematic analysis, we provide some suggestions on how future research in this domain of ABR monitoring in LMICs can be further strengthened.

4.1.1 Interdisciplinary research efforts

An interdisciplinary approach entails interaction, collaboration, and cooperation among scientific, academic, and non-academic disciplines, researchers, and stakeholders, to integrate scientific, technical, and non-technical knowledge as bases for policymaking at the higher level and context-specific implementation at the practice level [33]. The need for an interdisciplinary approach to tackle ABR is well documented because of the interconnected domains like human, veterinary, food, environment, etc., and the involvement of multiple stakeholders[34].

A lack of focus on ABR in the existing IS literature indicates a significant scientific and practical vacuum. This vacuum is particularly striking when we consider the magnitude of the ABR domain. In the context of increasing calls for building one-health approaches to ABR research [35], where digital monitoring is pivotal, IS research needs to become more relevant in guiding the realization of the potential of the digital. Building digital monitoring systems in LMIC settings is not limited to one hospital or nation, it

is a global interconnected, and complex issue, making it a wicked problem that demands interdisciplinary and collaborative approaches. However, 10 out of 12 of the studies identified in the literature are from public health journals written either by medical or clinical and public health professionals.

Supplementation of ABR research with a social systems approach to IS research can help in the development of monitoring systems guided by the problem context with the expertise from both clinical/medical and IS researchers and help to facilitate contribution towards antimicrobial stewardship (AMS) interventions. A social systems approach to IS discusses the problems of design and implementation of digital technology as an interplay of human, organizational, social, and technical factors[36]. It is particularly relevant for ABR and the LMICs perspective as the context-specific design and implementation of monitoring systems must involve an understanding of these factors and in which the digital technology is to be implemented and used for its adoption by the end users[34,37]. It can potentially provide insights into the specific challenges like a better understanding of the structural issues aggravating the problem to make decisions at policy and practice levels. For example: At the facility level, an ABR monitoring system could potentially make the issues visible like prescription practices of antibiotics and data quality issues at the practice level and the use of this data to make an antibiotic policy at the policy level.

4.1.2 Advocating systems thinking approaches

Systems thinking is an approach widely used to address and solve complex problems, including those relating to information systems [38]. It is the consideration of systems in their totality, as their constituent parts and their interactions, as well as their interaction with the wider environment [37]. ABR is considered to be one of the most complex problems and a global threat that cannot be solved by focusing on individual processes [39] and will benefit through the application of multiple research lenses.

It requires a focus on understanding the problems as a whole from multiple perspectives like medical/clinical, IS, public health, etc to identify different underlying components, and challenges, and predict behaviors. This could be done by a system thinking approach to examine and analyze the underlying problems and plan interventions accordingly. The participation of stakeholders and experts from different domains while using a systems approach can potentially increase stakeholder engagement and ownership of the new knowledge generated through the process by allowing ideas to the incorporation from different perspectives and encouraging a participatory approach to solving a problem [40].

The systems thinking approach has been applied to a variety of societal issues of global impact like environmental challenges and policy, climate change, and disease eradication programs [41,42]. However, the problem of ABR has remained untouched by the systems thinking approach. Considering the complexity and seriousness of the issue, a system thinking approach must be used to evaluate the problem, existing interventions, and their impacts and to plan the future interventions accordingly by considering the problem as a whole consisting of clinical, social, ecological, and cultural,

economic constituents. For example, the problem of the irresponsible use of antibiotics is the major reason for the occurrence of ABR. Antibiotics use is a complex issue resulting from a chain of events in an ecosystem with multiple subsystems and involves the actions of multiple stakeholders. E.g.: Prescription practices of physicians, dispensing practices of pharmacists, patterns of use of antibiotics by patients, etc. The practices of these stakeholders are affected by the underlying social and cultural factors and to address the issue an evidence base is needed to act. Social sciences research combined with an IS approach could potentially guide at the practice level by providing an evidence base for the physicians to prescribe antibiotics responsibly and guide the development of infection control and antibiotics use policies etc.

4.1.3 Research influencing practice

The research-practice gap occurs when knowledge acquired through research in an academic environment is not integrated with real-world clinical practice [43]. As standards of care continue to evolve, there can often seem to be a disconnect between what is considered best practice and actual practice. Several contributing factors result in the research and practice gap. For example, communication gap between researchers and practitioners, service delivery issues including lack of awareness and knowledge, lack of political and economic support, etc [44]. Several other factors have been documented like the interventions being narrowly or too broadly focused, complex, difficult, and costly, or may not engage or meet the perceived needs of the community at the practice level. ABR interventions are a classic example of the research-practice gap as there are several policies and frameworks defined at the global and national levels, GAPs and NAPs but these are poorly implemented at the practice levels [22]. Local practitioners identify challenges with the implementation of global platforms like WHONET and face challenges in configuring and interoperability etc [26] and GLASS does not provide patient-based information at the hospital level and lacks clinical metadata on antimicrobial use and duration of hospitalization [25]. The challenges in implementation are also cultural, lack of experience, and require context-specific solutions to meet global standards and to meet the needs at the practice level. IS research integrated with clinical research on ABR could help in the development of a context-specific evidence-based to taking local actions at the practice level that could potentially be scaled to other contexts.

5 CONCLUDING REMARKS

Arguably, this paper is a first step in arguing the potential role that IS research can have in strengthening ABR research and practice, and some suggestions on future areas of focus. While acknowledging this is indeed only touching the tip of the iceberg, it is required and urgent. A key role for IS research is in guiding the design, development, and implementation of context specific ABR digital interventions supplemented with expertise from other disciplines. This study proposes three future research directions which can help guide efforts and interventions for implementing digital ABR monitoring efforts in varying LMIC contexts, which would need to be applied in practice and further evolved with experiences.

6 REFERENCES

1. The antibiotic alarm. *Nature* 2013;**495**:141–141.
2. O'Neill J. *Tackling Drug-Resistant Infections Globally: Final Report and Recommendations*. Government of the United Kingdom, 2016.
3. Vos T, Barber RM, Bell B *et al*. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *The Lancet* 2015;**386**:743–800.
4. Rochford C, Sridhar D, Woods N *et al*. Global governance of antimicrobial resistance. *The Lancet* 2018;**391**:1976–8.
5. Organization WH. *Antimicrobial Resistance: Global Report on Surveillance*. World Health Organization, 2014.
6. Adeyi OO, Baris E, Jonas OB *et al*. Drug-resistant infections: a threat to our economic future. *World Bank Group Wash DC* 2017.
7. Mendelson M, Matsoso MP. The World Health Organization global action plan for antimicrobial resistance. *SAMJ South Afr Med J* 2015;**105**:325–325.
8. Okeke IN, Feasey N, Parkhill J *et al*. Leapfrogging laboratories: the promise and pitfalls of high-tech solutions for antimicrobial resistance surveillance in low-income settings. *BMJ Glob Health* 2020;**5**:e003622.
9. Ayukekbong JA, Ntemgwa M, Atabe AN. The threat of antimicrobial resistance in developing countries: causes and control strategies. *Antimicrob Resist Infect Control* 2017;**6**:1–8.
10. Laxminarayan R, Chaudhury RR. Antibiotic Resistance in India: Drivers and Opportunities for Action. *PLOS Med* 2016;**13**:e1001974.
11. Laxminarayan R, Van Boeckel T, Frost I *et al*. The Lancet Infectious Diseases Commission on antimicrobial resistance: 6 years later. *Lancet Infect Dis* 2020;**20**:e51–60.
12. de Kraker ME, Stewardson AJ, Harbarth S. Will 10 million people die a year due to antimicrobial resistance by 2050? *PLoS Med* 2016;**13**:e1002184.
13. Organization WH. Global antimicrobial resistance surveillance system (GLASS) report: early implementation 2020. 2020.
14. Ashley EA, Recht J, Chua A *et al*. An inventory of supranational antimicrobial resistance surveillance networks involving low-and middle-income countries since 2000. *J Antimicrob Chemother* 2018;**73**:1737–49.
15. Gandra S, Alvarez-Uria G, Turner P *et al*. Antimicrobial Resistance Surveillance in Low- and Middle-Income Countries: Progress and Challenges in Eight South Asian and Southeast Asian Countries. *Clin Microbiol Rev* 2020;**33**:e00048-19.
16. Boyatzis RE. *Transforming Qualitative Information: Thematic Analysis and Code Development*. sage, 1998.
17. Bryman A. *Social Research Methods*. Oxford university press, 2016.
18. Miles MB, Huberman AM. *Qualitative Data Analysis: An Expanded Sourcebook*. sage, 1994.

19. Johns G. The essential impact of context on organizational behavior. *Acad Manage Rev* 2006;**31**:386–408.
20. Oberin M, Badger S, Faverjon C *et al*. Electronic information systems for One Health surveillance of antimicrobial resistance: a systematic scoping review. *BMJ Glob Health* 2022;**7**:e007388.
21. Iskandar K, Molinier L, Hallit S *et al*. Surveillance of antimicrobial resistance in low- and middle-income countries: a scattered picture. *Antimicrob Resist Infect Control* 2021;**10**:1–19.
22. Chua AQ, Verma M, Hsu LY *et al*. An analysis of national action plans on antimicrobial resistance in Southeast Asia using a governance framework approach. *Lancet Reg Health West Pac* 2021;**7**:100084.
23. Grundmann H, Klugman KP, Walsh T *et al*. A framework for global surveillance of antibiotic resistance. *Drug Resist Updat* 2011;**14**:79–87.
24. Seale AC, Hutchison C, Fernandes S *et al*. Supporting surveillance capacity for antimicrobial resistance: Laboratory capacity strengthening for drug resistant infections in low and middle income countries. *Wellcome Open Res* 2017;**2**:91.
25. Turner P, Ashley EA, Celhay OJ *et al*. ACORN (A Clinically-Oriented Antimicrobial Resistance Surveillance Network): a pilot protocol for case based antimicrobial resistance surveillance. *Wellcome Open Res* 2020;**5**:13.
26. Vong S, Anciaux A, Hulth A *et al*. Using information technology to improve surveillance of antimicrobial resistance in South East Asia. *BMJ* 2017;**358**, DOI: 10.1136/bmj.j3781.
27. Seebregts C, Dane P, Parsons AN *et al*. Designing for scale: optimising the health information system architecture for mobile maternal health messaging in South Africa (MomConnect). *BMJ Glob Health* 2018;**3**:e000563.
28. Kaur J, Kaur J, Dhama AS *et al*. Strengthening the Surveillance of Antimicrobial Resistance in India Using Integrative Technologies. *Front Public Health* 2022;**10**.
29. Nabadda S, Kakooza F, Kiggundu R *et al*. Implementation of the World Health Organization Global Antimicrobial Resistance Surveillance System in Uganda, 2015-2020: Mixed-Methods Study Using National Surveillance Data. *JMIR Public Health Surveill* 2021;**7**:e29954.
30. Rezaei-Hachesu P, Samad-Soltani T, Yaghoubi S *et al*. The design and evaluation of an antimicrobial resistance surveillance system for neonatal intensive care units in Iran. *Int J Med Inf* 2018;**115**:24–34.
31. Sahay S, Arora G, Thakral Y *et al*. Designing for Scale: Strengthening Surveillance of Antimicrobial Resistance in Low Resource Settings. In: Bandi RK, C. R. R, Klein S, *et al*. (eds.). *The Future of Digital Work: The Challenge of Inequality*. Cham: Springer International Publishing, 2020, 251–64.
32. Labrique AB, Wadhvani C, Williams KA *et al*. Best practices in scaling digital health in low and middle income countries. *Glob Health* 2018;**14**:1–8.
33. Orderud GI, Vogt RD, Tan H *et al*. Interdisciplinary research and transdisciplinary processes for environmental management under different socio-natural conditions. *Int J Environ Stud* 2018;**75**:827–46.
34. Larson EL, Saiman L, Haas J *et al*. Perspectives on antimicrobial resistance: Establishing an interdisciplinary research approach. *Am J Infect Control* 2005;**33**:410–8.
35. Amuasi JH, Lucas T, Horton R *et al*. Reconnecting for our future: The Lancet One Health Commission. *Lancet Lond Engl* 2020;**395**:1469–71.
36. Walsham G, Symons V, Waema T. Information systems as social systems: Implications for developing countries. *Inf Technol Dev* 1988;**3**:189–204.
37. Rosenkranz C. Information systems development as a social process: a structural model. 2011.
38. Peters DH. The application of systems thinking in health: why use systems thinking? *Health Res Policy Syst* 2014;**12**:1–6.
39. Hinchliffe S, Butcher A, Rahman MM. The AMR problem: demanding economies, biological margins, and co-producing alternative strategies. *Palgrave Commun* 2018;**4**:1–12.
40. Siokou C, Morgan R, Shiell A. Group model building: a participatory approach to understanding and acting on systems. *Public Health Res Pr* 2014;**25**:e2511404.
41. Hinman A. Eradication of vaccine-preventable diseases. *Annu Rev Public Health* 1999;**20**:211–29.
42. Woodruff S, BenDor TK, Strong AL. Fighting the inevitable: infrastructure investment and coastal community adaptation to sea level rise. *Syst Dyn Rev* 2018;**34**:48–77.
43. Bansal P, Bertels S, Ewart T *et al*. *Bridging the Research–Practice Gap*. Academy of Management Briarcliff Manor, NY, 2012.
44. Mallonee S, Fowler C, Istre GR. Bridging the gap between research and practice: a continuing challenge. *Inj Prev* 2006;**12**:357–9.

Appendix 1: Details of the articles included in the review

First author, year (country)	Digital technology used	Setting and level of implementation	Main results
Kaur J, 2022 (India)	i-AMRSS (Web-based digital AMR surveillance system)	Used for data collection from 30 tertiary hospitals	Locally developed application for monitoring. The study discussed features of the tool and the possible analysis and the possibility to extend to veterinary and other domains possible
Nabadda S, 2021 (Uganda)	WHONET/GLASS	At specific surveillance sites in the country from 2015 to 2020	Data collected during the project duration to be used to guide policies. However, no plan for country-wide surveillance is described.
Iskandar K, 2021 (Review of data sources for LMICs)	Review of available data sources for LMICs Requirement assessment for LMICs. Experience from implementation in Georgia		The barriers and limitations of conducting effective antimicrobial resistance surveillance in LMICs and highlight multiple incremental approaches that may offer opportunities to strengthen population-based surveillance if tailored to the context of each country.
Sahay S, 2020 (India)	DHIS2(District health information system)	Facility/hospital level	Design and implementation of an open-source application for AMR monitoring at a facility with the possibility to scale both functionally and geographically.
Turner P, 2020 (Laos, Vietnam & Cambodia)	WHONET/GLASS	Plan to pilot in one facility each in the three countries	Digital surveillance to build on GLASS as it does not provide patient-based information at the hospital level and lacks clinical metadata on antimicrobial use and duration of hospitalization
Rezaei-hachesu P, 2018 (Iran)	Requirements analysis for a surveillance system	Neonatal Intensive care units (NICUs) at 2 tertiary hospitals in Iran	Framework for the design of an AMR/ABR surveillance system for use in the NICUs in north-western Iranian hospitals to cover information gaps and proposes three modules for monitoring: the data registry, dashboard, and decision support
Safdari R, 2017 (Iran)	GLASS	Review of literature on existing digital surveillance systems	The study developed a framework for the design and implementation of a national ABR monitoring system building on GLASS
Seale A.C, 2017 (WHO GLASS countries)	GLASS	Review of literature on existing digital platforms	A roadmap for participation in the Global Antimicrobial Surveillance System (GLASS)
Oberin M, 2022 (Review of existing digital platforms)	Review of existing digital platforms- to identify solutions for monitoring in all domains	Review of existing digital platforms	No EIS for AMR surveillance was identified that was designed to integrate a broad range of AMR data from humans, animals, and the environment, representing a major gap in global efforts to implement One Health approaches to address AMR.
Lim C, 2020 (Thailand)	Antimicrobial resistance Surveillance System (AMASS)	One hospital in Thailand	An offline application to generate standardized AMR surveillance reports in the R programming language. The challenges presented with analyzing data and generating a report as lengthy and time-consuming processes that require trained personnel.
Vong S, 2017 (Seven Asian countries)	WHONET	Global and National	Constraints of Information technology surveillance like configuration of WHONET and BacLink, system interoperability, lack of data standards, etc
Grundmann H, 2011	WHONET/GLASS	Global	Framework for AMR/ABR surveillance at global/national/regional levels

Expectations of users and non-users of wearable sensors and mobile health applications

Andre Henriksen^{1*}, Gerit Pfuhl^{1,2*}, Ashenafi Zebene Woldaregay³, David-Zacharie Issom⁴, Eirik Årsand¹, Keiichi Sato^{1,5}, Gunnar Hartvigsen¹

¹UiT The Arctic University of Norway, Tromsø, Norway, andre.henriksen@uit.no

²Norwegian University of Science and Technology, Department of Psychology, Trondheim, Norway

³University Hospital of North Norway, Norwegian centre for clinical artificial intelligence, Tromsø, Norway

⁴Geneva University Hospitals, Division of Medical Information Sciences, Switzerland

⁵Illinois Institute of Technology, Illinois, USA

Abstract

Patient self-management is vital to improved health outcomes for patients with chronic diseases. The objective of this study was to understand the role of wearable sensors in patients' self-management. A survey encompassing factors related to motivation in mHealth was conducted. Ease of use and sensory accuracy was found most important when choosing a wearable. Manual registration of most health-related information is unpopular, although some exceptions exist. Respondents valued sensor accuracy and easiness in manual registration and usage of mHealth systems. Further research is needed to pinpoint what ease of use exactly is, and how ease of use can be improved.

Keywords

mHealth, eHealth, self-management, chronic disease, persuasive design

INTRODUCTION

Persons with chronic diseases could benefit from using mobile health (mHealth) tools for self-management. Various devices, ranging from wearable sensors and apps integrated in smartphones, to health specific devices (e.g., glucometers) exist. The range of output from these devices, and the possibility for long-term unobtrusive monitoring, makes these devices uniquely supportive for continuous chronic disease self-management [1]. Fan et al. [2] recently published a review assessing the usability and effectiveness of mHealth apps in chronic disease self-management and concluded that mHealth technologies are as good as traditional care.

However, there seems to be a lack of motivation from most users to keep using these health apps over a long period of time [3]. Attig et al. [4] assessed reasons for physical activity tracker attrition and found that lack of motivation was one of the main reasons for no longer wearing such trackers. For persons with chronic disease, continuous use of physical activity trackers are reported to improve their health management [5].

Therefore, as mitigation measure and to further support the effort in self-management of chronic diseases, this study aimed to identify what features and factors motivates people with and without chronic disease to use mHealth apps and sensors.

METHOD

An anonymous online survey was distributed physically at a Swiss conference and on multiple social media fora

*Authors contributed equally

The 18th Scandinavian Conference on Health informatics, Tromsø, Norway, August 22-24, 2022. Organized by UiT The Arctic University of Norway. Conference Proceedings published by Linköping University Electronic Press at <https://doi.org/10.3384/ecp187>. © The Author(s). This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>

related to diabetes and sickle cell disease, as well as a more general site not related to chronic disease. A total of nine online platforms were used. The survey was constructed based on 16 in-person interviews, out of which 12 persons had a chronic disease [3]. Announcements were either in English, Norwegian, or French, depending on distribution site. Respondents answered questions about what motivates them to self-manage their own health or disease. The survey had seven themes with the following headings: 1) background and health goal questions, 2) use of wearables and sensors, 3) use of mobile apps, 4) data-logging, 5) data sharing and data integration, 6) social media and entertainment factors, and 7) demographic questions including age, gender, and chronic disease diagnosis.

Here we report about the use of wearables, apps and sensors, and data logging and registration. Table 1 gives the sub-set of questions used in this study, with answer options. Question using Likert scales (question 7, 9, 10, 14, 15) gave answer options from 1-4 (where 1 was the lowest score). "Don't know" was also an option.

The online survey was open for data collection between November 2018 and March 2020. Primary results comparing those with and without chronic diseases [6, 7] and investigating the role of caretakers [8] have been previously published.

A request for ethical approval was reviewed at the Regional Ethics Committee (REK) and found to be exempt from their purview (ref. 2017/562).

Data was analysed using the software RStudio [9]. We report descriptive results, and for group comparisons we used Welch's t-tests and chi-square tests, were appropriate.

No.	Question
Background	
5	Have you ever used a wearable device for collecting activity or other health data? (Yes, No)
28	Do you have a chronic disease? (Diabetes, Sickle-cell, No, Don't want to answer)
Wearables and sensors	
6	Which of these technologies for health tracking do you regularly use? (Multiple choice)
7	How important are these features for you when choosing a wearable device? (Likert 1-4, Don't know)
8	Which features would motivate you most to use a wearable sensor longer? (Single choice)
9	How important are these specific health related features for you when choosing a wearable device? (Likert 1-4, Don't know)
Mobile apps	
10	How important are these features when choosing a mobile health app? (Likert 1-4, Don't know)
11	How do you decide if a mobile app is trustworthy? (Multiple choice)
Logging/registration	
14	If you are required to do manual logging (registration) in a health mobile app, how important are these criteria for you? (Likert 1-4, Don't know)
15	How willing would you be to manually log or register the following types of data? (Likert 1-4, Don't know)

Table 1. Selected questions from questionnaire.

RESULTS

Participant characteristics

Among the 814 who responded to the survey, 300 (37%) indicated to have a chronic disease and 490 (60%) indicated not to have a chronic disease. Twenty-four (3%) respondents left this question unanswered. 272 (33%) respondents used sensors in their smartphone, 285 (35%) respondents used physical activity trackers, 255 (31%) respondents used mobile health apps, and 185 (23%) respondents used health specific measurement devices, e.g., glucometers. Multiple responses were possible. 281 (35%) respondents stated not using any wearable sensor, activity tracker, or mobile health app. Of those 85 (30%) indicated to have a chronic disease and 184 (65%) stated no chronic disease.

Motivation for prolonged use of apps and sensors

We first looked at what motivates respondents to wear and use a sensor for longer periods (Question 8), stratified into those with previous experience with wearable devices for

physical activity tracking or other health tracking, and those with no previous experience (Question 6).

Relevant personalized feedback was a main motivator for 42.5% of respondents; whereof 38% with a chronic disease and 49% without a chronic disease. This group difference was significant; $\chi^2 = 40.953$, $p < 0.001$. The second rated motivator was *ease of use* with 36.4% of respondents having this as their main motivation; whereof 43% with a chronic disease and 34% without a chronic disease. This group difference was significant; $\chi^2 = 4.268$, $p = 0.039$.

Access to aggregated data and *social media integration* were not main motivators. Table 2 shows, when ignoring those answering, "don't know" or "other", that the ranking of the motivations was similar among those already using mHealth technology and those not yet using it (Question 6). The difference was statistically significant with a small effect size; $\chi^2 = 37.323$, $p < 0.001$, Cramer's $V = 0.219$; mainly due to those not yet using any device answering, "don't know/other".

	Not using any device (n=281)	Using at least one device (n=533)
Relevant personalized feedback	113 (40%)	233 (44%)
Ease of use/non-disruptive	95 (34%)	201 (38%)
Access to aggregated summarize data on the population level	10 (4%)	25 (5%)
Social integration (e.g., Facebook)	0 (0%)	14 (3%)
Don't know/Other	50 (19%)	40 (8%)

Table 2. Question 8, Motivating factors grouped by those who have ever used a wearable device (Question 6).

Next, we looked at which features were most important to the respondents (Question 7). Answers were given on a 4-point Likert scale. Across the six features 11-20% responded with "Don't know" or left the field blank.

Figure 1 shows the rating of the six features, with *sensory accuracy* and *easy to use* being rated as very important by most respondents, and irrespective of whether they already use or not use wearable sensors. *Access to data* and *ergonomic design* were also rated as important features whereas *known brand* was least important. *Notification* on the mobile phone was rated as somewhat important. Respondents with and without a chronic disease rated these six features similarly.

All figures give boxplots where the black line is the median response, the black dot the mean, grey triangles are outliers, and the boxes represent the range between the 25th and 75th percentile

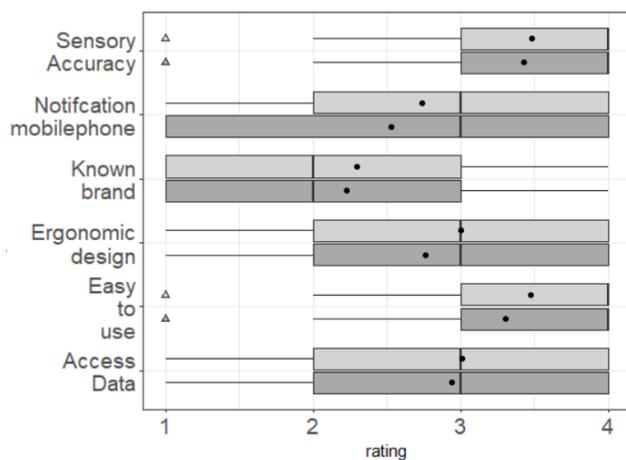


Figure 1. Question 7. Rating of importance for six mHealth features.

Regarding health specific measurements/features (Question 9) rated on a 4-point Likert scale, *physical activity tracking* was rated as most important, followed by *predicting/preventing deterioration of health, alerts, and managing the disease*. Notably, even within the chronic group, *physical activity tracking* was rated highest, followed by *managing the disease, alerts, and predicting/preventing deterioration of health*. Mean values and standard deviations, for all participants and only those with a chronic disease, are given in Table 3. The chronic group (M = 3.12) rated those four features higher than the non-chronic group (M = 2.92), Welch's t-test: $t(2259.6) = 4.8106, p < .001, 95\% \text{ CI } [.119, .284]$.

	All participants	Chronic group
Physical activity tracking	3.29 (0.91)	3.23 (0.95)
Predicting/preventing deterioration of health	2.92 (1.07)	3.03 (1.04)
Alerts	2.92 (1.05)	3.08 (1.05)
Managing the disease	2.85 (1.14)	3.14 (1.08)

Table 3. Question 9 answer option scores. Mean score (SD)

Specifically for choosing a mobile health app, we asked how important; simplicity/usability; functionality/features; price; trust/security/privacy; and personalization (tailored features) were (question 10). As Table 4 shows, all features were rated as important (mean score over 3, maximum possible score was 4).

	Chronic	Non-chronic
Simplicity	3.50 (0.775)	3.55 (0.694)
Trust, Security, Privacy	3.39 (0.896)	3.47 (0.847)
Functionality	3.33 (0.961)	3.39 (0.840)
Price	3.15 (0.979)	3.23 (0.904)
Personalization	3.15 (0.979)	3.02 (0.916)

Table 4. Question 10 answer option scores. Mean score (SD)

Following up on *trust, security, and privacy* we asked on which feature they would decide about the trustworthiness of a mobile app (Question 11). 457 respondents (56%) would use personal *experience* or other people's experience for judging trustworthiness, 430 respondents (53%) would base it on *certificates*, 359 respondents (44%) would decide

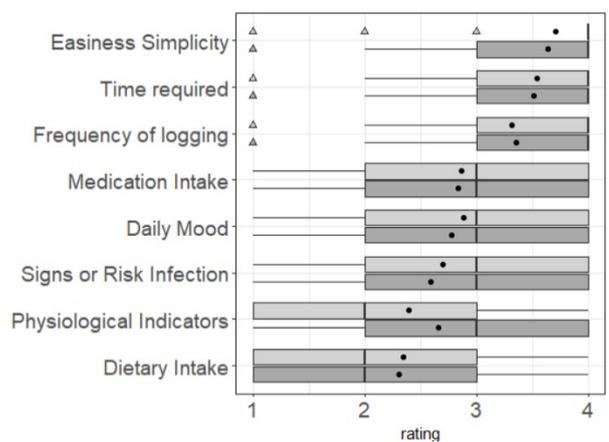


Figure 2. Question 14. Important criteria for motivating manual registration.

it based on *product specifications* and/or *provider reputation*, and 260 respondents (32%) would use Google Play or AppStore ratings.

Next, we looked at manual registration and logging. We asked about the importance of three features if required to do manual registration (Question 14); *easiness / simplicity, time required, and frequency of logging*. We also asked about how willing they are to manually register (Question 15) *daily mood, dietary intake, medication intake, physiological indicators and signs or risk of infection*. Both were answered on a 4-point Likert scale.

Figure 2 shows that *easiness/simplicity, time required, and frequency of logging* are rated as very important for manual registration. Nearly all respondents with a chronic disease strongly agreed that manual registration must be easy and simple (M = 3.64, SD = 0.68). Respondents were most willing to register their *medication intake and daily mood*, followed by *signs or risk of infection and physiological indicators*. They were least willing to register their *dietary intake*. Participants with a chronic disease (M = 2.66) were more willing to register *physiological indicators* than respondents without a chronic disease (M = 2.4), Welch's t-test: $t(576.45) = 3.143, p = 0.0018$, else the groups did not differ.

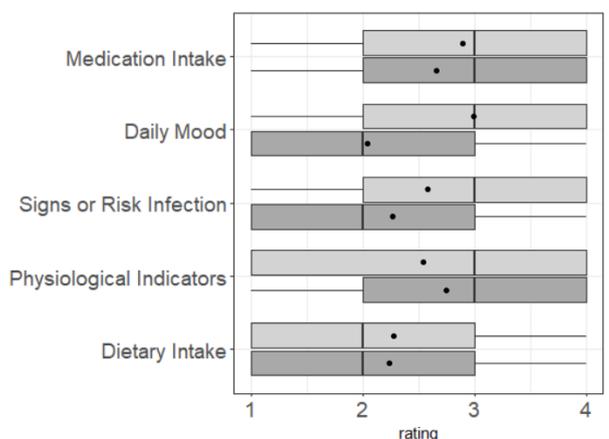


Figure 3. Question 15. Willingness for manual registration among respondents with Diabetes and respondents with other chronic diseases

Finally, we looked closer at respondents with Diabetes (n=71) and their willingness to register health-related measurements compared to respondents with other chronic

diseases (n=185). Diabetes management benefits from registering *dietary intake*.

Figure 3 shows that respondents with Diabetes were similarly unwilling to register dietary intake than respondents with other chronic diseases. Notably, respondents with Diabetes were less willing to register their *daily mood* than respondents with other chronic diseases.

DISCUSSION

As expected, *easy to use* is a highly desired feature for wearable sensors and is independent of previous experience with mHealth devices. Still, *sensory accuracy* and *personalized feedback* were rated as very important too. This is encouraging as wearable sensors shall support self-management. Notably, *personalized feedback* was more important among those not having a chronic disease. Possible explanations could be age differences or experience with feedback from wearable sensors. In the case of Diabetes, *sensory accuracy*, e.g., measuring blood glucose levels, might be regarded as one part of personalized feedback. For physical activity tracking, *personalized feedback* might be performance scores. This shows the ambiguity in what respondents might understand by *personalized feedback*.

Our survey used lay terms like *easy to use* without asking deeper what the respondents understand by this term. Not everybody may perceive small displays, colours or touch screens as easy to use, e.g., visually impaired, persons with colour deficiency, persons with Parkinson disease (e.g., [10]). Voice assistance might alleviate this, but has also its limits, e.g., persons with Amyotrophic lateral sclerosis or Huntington disease do not benefit from such features.

Over 50% of the respondents used more than one mHealth device, i.e., sensors on smartphones and physical activity trackers, or sensors on smartphones and mobile health apps. This suggests that devices are not solely used and bought for health self-management. This challenges developing tailored health apps with high sensory accuracy for automatic registration as devices vary in hardware. On the other hand, using specialized devices means to learn and use another device. Given the pace with which user interfaces of mobile devices are updated and changed, this is an enduring challenge for mHealth.

Relevant *personalized feedback* was not rated as most important among respondents with a chronic disease. This suggests that the burden of a chronic disease emphasises *simplicity* and *sensory accuracy* over *personalized feedback* in mHealth devices. *Personalized feedback* might be expected from their general practitioner.

This may touch on another important factor, namely *trustworthiness*. Do we trust the output from a digital device or rather a medical doctor? Despite the statistics favouring the digital device over the fallible human being, the opacity of how the digital device derives the personalized feedback hampers trust in technology [11, 12]. Doing manual registration and logging is part of life for many with a chronic disease. This is often the most time-consuming aspect of self-management and can feel like a burden. Technological developments have reduced the time but not eliminated it entirely [13]. Notably, diet registration is the most hindering/demotivating feature among those

respondents who would gain most from it, namely respondents with diabetes. This might indicate a conflict between knowing enough about coping with the disease and freedom to live a life without constantly thinking about the disease [14]. Indeed, food is a strong reward, and removing the joy from eating by having to register one's diet, may adversely affect well-being [15].

The study has also limitations. The quantitative survey was based on qualitative interviews [3] but the factors and motivational reasons were not exhaustive as some respondents indicated *other*. However, the main factors and reasons were covered. In this report we did not control for demographic factors. Age and gender might influence adaptation of mHealth apps but this might matter less for continuing using wearable sensors and mHealth apps, hence we did not statistically control for demographic factors but future studies should.

Wearable sensors and mHealth apps should take the cost-benefits for users into account, not least from a mental health perspective.

SUMMARY

In this study, responses from 814 participants of a survey about motivation in mHealth, was used to identify factors and features that motivates people with and without chronic disease to use mHealth apps and sensors.

Users and non-users of wearable sensors expects that mHealth apps and devices have accurate sensors, to be easy to use, and providing them with personalized feedback. The latter two can be addressed by software developers, the first often also requires appropriate hardware.

REFERENCES

- [1] Guo, Y., Liu, X., Peng, S., Jiang, X., Xu, K., Chen, C., Wang, Z., Dai, C., and Chen, W., "A review of wearable and unobtrusive sensing technologies for chronic disease management", *Comput Biol Med*, 129, 104163, 2021
- [2] Fan, K. and Zhao, Y., "Mobile health technology: a novel tool in chronic disease management", *Intelligent Medicine*, 2 (1), 41-47, 2022
- [3] Woldaregay, A.Z., Issom, D.Z., Henriksen, A., Marttila, H., Mikalsen, M., Pfuhl, G., Sato, K., Lovis, C., and Hartvigsen, G., "Motivational Factors for User Engagement with mHealth Apps", *Stud Health Technol Inform*, 249, 151-57, 2018
- [4] Attig, C. and Franke, T., "Abandonment of personal quantification: A review and empirical study investigating reasons for wearable activity tracking attrition", *Computers in Human Behavior*, 102, 223-37, 2020
- [5] Gabarron E., Dorrnzoro, E., Bradway M., Rivera-Romero O., Wynn R., and Arsand E., "Preferences and interests of diabetes social media users regarding a health-promotion intervention", *Patient Prefer Adherence*, 12, 2499-506, 2018
- [6] Henriksen A., Woldaregay A. Z., Issom DZ., Pfuhl G., Richard A., Årsand E., Sato K., Hartvigsen G., and Rochat J. (2019), 'Replication data for: User expectations and willingness to share self-collected

- health', in T. The Arctic University of Norway Ui (ed.), (V2 edn.: DataverseNO).
- [7] Woldaregay A. Z., Henriksen A., Issom D. Z., Pfuhl G., Sato K., Richard A., Lovis C., Arsand E., Rochat J., and Hartvigsen G., "User Expectations and Willingness to Share Self-Collected Health Data", *Stud Health Technol Inform*, 270, 894-98, 2020
 - [8] Bradway M., Woldaregay A. Z., Issom D. Z., Pfuhl G., Hartvigsen G., Årsand E., and Henriksen A., "mHealth: Where is the potential for aiding informal caregivers?", *Stud Health Technol Inform*, 2021
 - [9] RStudio Team (2020) 'RStudio: Integrated Development for R', RStudio, PBC, Boston, MA. <http://www.rstudio.com/>
 - [10] Engeset R. V., Pfuhl G., Orten C., Hendriks J., and Hetland A., "Colours and maps for communicating natural hazards to users with and without colour vision deficiency", *International Journal of Disaster Risk Reduction*, 76, 2022
 - [11] Materia, F.T., Faasse, K., and Smyth, J.M., "Understanding and Preventing Health Concerns About Emerging Mobile Health Technologies", *JMIR Mhealth Uhealth*, 8 (5), e14375, 2020
 - [12] Vo, V., Auoy, L., and Sarradon-Eck, A., "Patients' Perceptions of mHealth Apps: Meta-Ethnographic Review of Qualitative Studies", *JMIR Mhealth Uhealth*, 7 (7), e13817, 2019
 - [13] Choe, E.K., Klanja, P., and Pratt, W. (2021), 'mHealth and Applications', *Biomedical Informatics* (Fifth edn., Biomedical Informatics).
 - [14] Leung, L. and Chen, C., "E-health/m-health adoption and lifestyle improvements: Exploring the roles of technology readiness, the expectation-confirmation model, and health-related information activities", *Telecommunications Policy*, 43 (6), 563-75, 2019
 - [15] Tylka, T.L., Annunziato, R.A., Burgard, D., Danielsdottir, S., Shuman, E., Davis, C., and Calogero, R.M., "The weight-inclusive versus weight-normative approach to health: evaluating the evidence for prioritizing well-being over weight loss", *J Obes*, 2014, 983495, 2014

ACKNOWLEDGEMENT

We thank all respondents who answered the survey

ARMAGNI: Augmented Reality Enhanced Surgical Magnifying Glasses

Situational Awareness during surgery with AR in the Loupe

Artur Liebrecht¹, Roman Bibo¹, Bastian Dewitz¹, Sebastian Kalkhoff¹, Sobhan Moazemi¹, Markus Rollinger²,
Jean-Michel Asfour², Klaus-Jürgen Janik³, Artur Lichtenberg¹, Hug Aubin^{1*}, Falko Schmid^{1*+}

¹Digital Health Lab Düsseldorf, Department of Cardiac Surgery, Medical Faculty and University Hospital Düsseldorf, Heinrich-Heine-University Düsseldorf, Düsseldorf, Germany

²DIOPTIC GmbH, Weinheim, Germany

³JADENT GmbH, Aalen, Germany

*These authors contributed equally to this work as senior authors

+Corresponding author

Abstract

During cardiac surgery, in addition to their manual work, surgeons need to perceive a large amount of procedural intraoperative data, including information which is not directly visible to them. Therefore, we developed an augmented reality demonstrator that displays alphanumeric data into the loupe of surgical magnifying glasses. Eight cardiac surgeons tested the demonstrator in a skill task that simulates the critical part of a typical surgical procedure while being confronted to vital intraoperative parameters getting critical. The results showed a decrease of missed critical phases and improved response times when using the demonstrator instead of a customary monitor for tracking intraoperative parameters.

Keywords

Augmented Reality, Cardiac Surgery, Intraoperative Assistance

1 INTRODUCTION

Situational awareness, “the perception and understanding of the surrounding environment” [13], is important in humans’ performance fulfilling a complex task [6]. This also applies to cardiac surgeries where safety and outcome are dependent on information flow, concerning not only preoperative data but also intraoperative procedural data such as vital parameters. Although operation rooms are equipped with multiple monitors displaying standard parameters like vital signs, other information such as heartlung machine procedural data and respiratory parameters are not readily visible for the surgeon. When needed, it requires intense communication efforts which can lead to distraction and interruption of the procedure [4]. Augmented reality (AR) can be a solution to bridge the gap between information needs, communication, and display possibilities.

Head-mounted displays (HMD), including smart glasses such as Google Glass, have been playing an increasing role in the health care industry [12]. In a systematic review from 2019, Rahman et al. [12] determined 120 HMD applications in surgery, of which most were used for image guidance and AR. For instance, Liebert et al. [7] compared traditional vital signs monitor to an HMD for monitoring patient’s vitals, showing potential for increased situational awareness and improved patient safety. The overall feedback from users was positive. In a qualitative descriptive study, Enlöf et al. [5] also showed that health

care professionals have a generally positive view of using smart glasses in the medical field. In video assisted surgery AR glasses have potential to reduce bad body posture during procedure [8]. Arpaia et al. [1] presented a system for retrieving and displaying patient’s vitals and evaluated its effectiveness, transmission error rate, refresh rate and latency with confirming technical feasibility of such a system. The target group consisted of assistant surgeons and anesthetists.

The related work mentioned in the previous lines indicate benefits from using AR devices in surgery and high user acceptance, but these applications did not consider cases where users require surgery magnifying glasses. For cardiac surgeons, wearing magnifying glasses and HMDs simultaneously can be cumbersome due to interference. As one solution, smart glasses can be mounted to surgical loupes. Yoon et al. [14] successfully used such a setup for image guidance during shunt placement. The AR images were projected above the loupe and required quick eye movement to see. Qian et al. [10], on the other hand, combined a Magic Leap One with binocular magnifying loupes. They evaluated a calibration method to align virtual content to the real world in the magnified or minified field-of-view of the user. But since we are projecting simple alphanumeric procedural data that does not have to be aligned to real objects, there is no need for such a calibration method. Hence, we can use a simpler setup with lightweight hardware.

The 18th Scandinavian Conference on Health informatics, Tromsø, Norway, August 22-24, 2022. Organized by UiT The Arctic University of Norway. Conference Proceedings published by Linköping University Electronic Press at <https://doi.org/10.3384/ecp187>. © The Author(s). This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>



Figure 1. AR demonstrator consisting of an AR module and Galilean loupes mounted on a custom-made glasses frame mimicking surgical magnifying glasses.

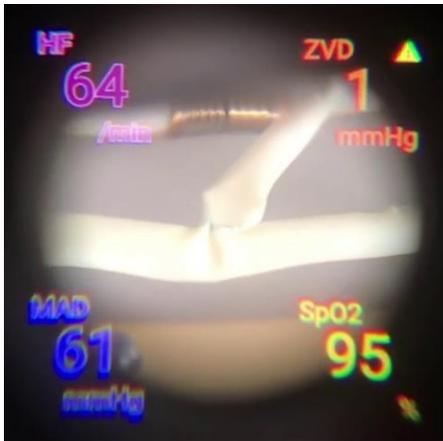


Figure 2. Image of the field of view of the AR loupe.

Our approach for displaying procedural data in the field of view of surgeons is different to the ones mentioned above. Instead of extending existing HMDs which have limitations due to weight, size, or short battery life [12], we develop typical surgical magnifying glasses with an integrated AR display into the loupe. For the proof of concept, we built an AR demonstrator that can be adjusted to the user. The prototype is shown in Figure 1 and the view of the loupe with AR visualizations in Figure 2.

In a feasibility study, experienced cardiac surgeons tested the usefulness of AR in the loupe with our demonstrator. We propose that AR magnifying glasses can increase situation awareness of surgeons. Beyond that, they will have potential to reduce surgeons' workload during cardiac surgeries and encounter high user acceptance, as they mimic the typical surgical magnifying glasses used during cardiac surgery.

2 METHODS

For this study, we asked cardiac surgeons to test our prototype referred as AR demonstrator, which we will describe in detail below along with the corresponding software, we will further outline the experimental setup.

2.1 AR Demonstrator

The prototype consists of two customary Galilean loupes with 2.5x magnification and 300 mm working distance. They were attached to a custom-built metallic mechanism, which allows different users to adjust pupil distance, inter-eye asymmetry or height of the loupes. The metallic mechanism is mounted on a plastic glasses frame made with a laser cutter.

The 2.5 cm x 2 cm x 1.5 cm sized AR module (Figure 3) was built with a combination of a micro-OLED display, a lens, and a beam splitter. For displaying AR visualizations, we used a Sony ECX336AF-6 OLED micro display. The 0.23-inch diagonal sized RGB display has a 640 x 400 pixels resolution, 800 cd/m² maximal luminance and 10,000:1 contrast ratio. We installed the micro display on the side of a 3D printed component. There, the displayed images are projected through a lens magnifying them to fully cover the view through the loupes. A beam splitter centered between the display, the loupe, and the eye, reflects the images into the eye of the user. The AR module can be attached to one of the Galilean loupes.



Figure 3. AR module attachable to loupes.

The micro display is connected to a Raspberry Pi Zero installed in the glasses frame. To be able to power the micro display with the required 1.8 volts and send images to it, the Raspberry Pi Zero was extended by a custom circuit board. A 3.6 volts rechargeable battery is powering the Raspberry Pi Zero through a 200 cm long cable. This battery is bundled in a 3D printed case with integrated electronics and a button, which allows us to turn the Raspberry Pi on and off.

2.2 Base station

To keep the AR magnifying glasses lightweight and conformable, we follow a client-server approach. One endpoint is the mentioned Raspberry Pi Zero built in the glasses frame. As the other endpoint, we are using a base station on Raspberry Pi 4B basis and the official Raspberry Pi 7-inch touchscreen (see Figure 4). The base station serves as a mobile control center for fetching patient data from simulated medical devices as well as generating visualizations and streaming them to the AR display. Furthermore, the base station is running a graphical user interface frontend application for configuring the AR visualizations using the touchscreen.

AR visualizations are generated in the frontend application and sent as PNG images to a REST server on the Raspberry Pi Zero. Using a Raspberry Pi 4B with the 64-Bit version of the Raspberry Pi OS as our base station, we can achieve approximately 5 frames per second with a delay of 500 milliseconds.



Figure 4. Base station running the software.

Our frontend application developed with Vue.js enables users to customize the AR visualizations. They can pick different parameters from connected medical device simulators and choose the desired position in the field of view. We decided the visualized medical parameters to be automatically arranged in a ring formation, so they do not overlay the working surface of the user. Though, the user is free to choose the size of the ring and by that the size of parameters as well as the number of displayed parameters. Another feature of the application allows users to finetune the whole AR visualization by moving, resizing, or rotating it with touch gestures while wearing the glasses and seeing the changes near real-time. Each user can have multiple settings profiles, which are saved in a PostgreSQL database.

Communication with the database and medical devices is performed by an Apollo GraphQL server, which exposes a GraphQL API for the frontend application. Backend was built in accordance with IEEE 11073 Service-oriented Device Connectivity (SDC) for being capable of communicating with real medical devices that also support SDC. Using SDCLib, an open-source SDC implementation by SurgiTAIX, we developed a SDC consumer micro service, which serves as a bridge between medical devices and our GraphQL server.

For the experiment, we developed a medical device simulator, that generates random but reasonable values for heart rate (labeled as HF in Figure 2), mean arterial pressure (MAD), oxygen saturation (SPO2) and central venous pressure (ZVD). The simulated values can continuously increase and decrease every 3 seconds, and every 60 – 90 seconds one of the parameters becomes critical for 15 seconds. Parameters in a critical state are displayed with a small warning sign and blink with a frequency approximately 600 ms. By means of this simulation, we evaluated the effect of the AR module on the user during a skill exercise, the setup of which we will describe next.

2.3 Experimental setup

The experiment took place in the clinic for cardiac surgery at the University Hospital Düsseldorf. We invited 8 cardiac surgeons (2 female, 6 males, between 25 and 54 years old) to perform a skill exercise while wearing our AR magnifying glasses and reacting to critical parameters. Directly afterwards, we asked them to take a survey, which included the System Usability Scale (SUS), the NASA Task Load Index (NASA-TLX) and further questions regarding the usefulness and usability of the AR.

In the beginning of the experiment, the investigator taught participants how they can adjust the AR magnifying glasses via the metallic mechanism and finetuned the AR display using the frontend application. During the experiment, frontend and backend applications were running on a laptop computer and used by investigator rather than the participant.

We followed a within-subject design with randomized order of conditions. In both cases, participants performed an anastomosis on the Arroyo's Anastomosis Simulator [11] while wearing our prototype. In the test condition, the AR module was turned on and the participants had to keep track of the displayed parameters in the loupe. In the control condition on the other hand, parameters were displayed on a customary 27-inch display only. We positioned the display on the right side of and 2 meters away from the participants, so they had to turn their head to the right by approximately 20-degree angle to see the values.

During performing the skill exercise, participants had to react to critical values as soon as possible by telling the investigator which parameter is critical and whether it is too high or too low. We recorded their response time using a timer implemented in the simulation software. Participants had a time limit of 15 seconds to notice and react to a critical phase and the investigator registered the answer as “right” or “wrong” by pushing the corresponding button on the keyboard. If the pre-set 15-second timeout passed before the participant reacted, the software automatically registered the answer as “missed” with a response time of 15 seconds.

After each trial, our simulation software generated a log file containing participant's response times to critical phases and whether the response was right, wrong, or missed.

3 RESULTS

We performed t-tests to compare our metrics between test condition (using AR) and control condition (using monitor). All metrics were tested for homogeneity of variance and normality before comparing them with a two-tailed paired t-test for dependent means. If requirements for homogeneity or normality were not met, Wilcoxon Signed-Rank test was used.

3.1 Response performance

We used a Wilcoxon Signed-Rank test to compare response performances between test and control conditions. Response performance is the count of correct responses divided by count of critical phases for each participant. The calculated data for each participant can be seen in Figure 5. It should be noted that no participant gave wrong responses.

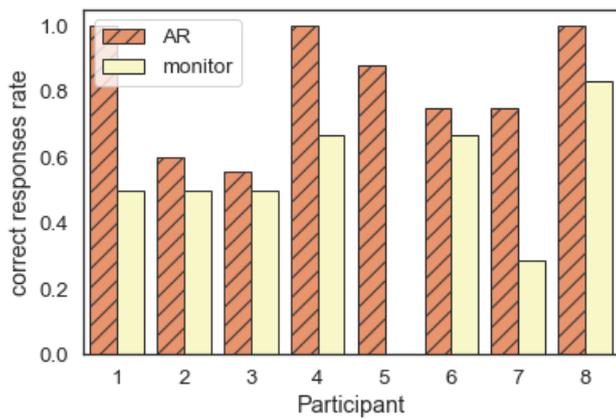


Figure 5. Rate for correct responses of each participant.

Participants showed an increase in response performance when using the AR display ($M=.8$, $SD=.2$) compared to customary display ($M=.5$, $SD=.3$), $W=0$, $p < .05$.

3.2 Response time

Data for response time is visualized in Figure 6. Since missed critical phases mean that there was no reaction for 15 seconds, data is strictly speaking not normally distributed. One participant even missed all critical phases in control condition; hence median was 15 seconds in this case.

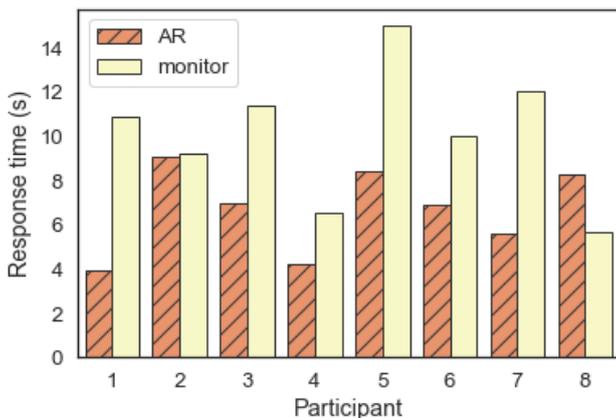


Figure 6. Median response times in seconds.

The Levene's test for response times was not significant ($F(1, 7) = 1.313$, $p = .340$) so variance homogeneity was assumed. Despite data being limited between 0 and 15 seconds, Shapiro-Wilk test for the differences between the pairs did not show a significant departure from the normality ($W(8) = .915$, $p = .433$).

Participants responded faster to critical values in the AR condition ($M = 6.7$, $SD = 1.9$) compared to monitor condition ($M = 10.1$, $SD = 3$), ($t(7) = 2.9$, $p = .025$).

3.3 Total time for trials

Participants required different amount of time to complete the anastomosis on the Arroyo's simulator. We used a t-test to compare the required time for test and control conditions.

The Levene's test for required times was not significant ($F(1, 7) = .079$, $p = .783$) so variance homogeneity was

assumed. Shapiro-Wilk test for the differences between the pairs did not show a significant departure from the normality ($W(8) = .984$, $p = .999$).

The t-test did not show any significant difference in required time for trials between AR ($M = 12.4$, $SD = 5.2$) and monitor conditions ($M = 11.4$, $SD = 5.5$), ($t(7) = 1.4$, $p = .203$).

3.4 NASA-TLX

The Levene's test for global NASA-TLX, namely the sum of individual NASA-TLX scores, was not significant ($F(1, 7) = .064$, $p = .804$) and Shapiro-Wilk test did not show significant departure from normality, ($W(8) = .92$, $p = .471$). The t-test for global NASA-TLX showed an improved workload for AR condition ($M=52.9$, $SD=22.1$) compared to monitor condition ($M = 65.4$, $SD = 24.1$), ($t(7) = 2.8$, $p = .027$).

Boxplots and significance in differences determined by ttests for NASA-TLX subscales can be seen in Figure 7. Variance homogeneity and normality criteria using Levene's test and Shapiro-Wilk test, respectively, were met. Outliers are classified as being outside 1.5 times the interquartile range.

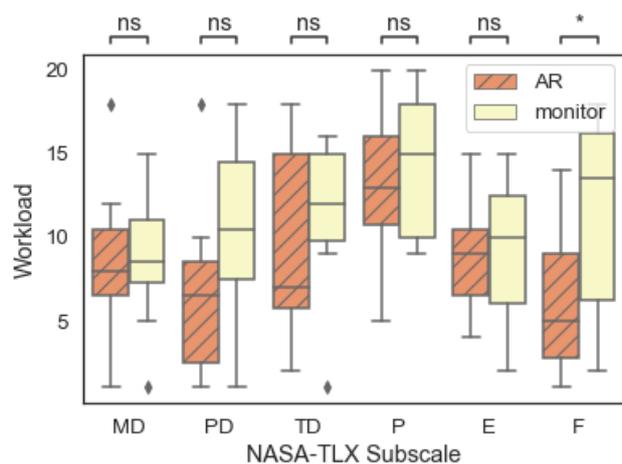


Figure 7. Mean scores for mental demand (MD), physical demand (PD), temporal demand, (TD), performance (P), effort (E) and frustration (F). ns means that a paired t-test result was not significant.

3.5 SUS Score

Mean SUS score was 66.875 (SD 12.02). Responses for individual statements are visualized in Figure 8.

3.6 Further questions

Responses to further questions about AR in the demonstrator and about AR magnifying glasses in general are summarized in Figure 9. Participants were asked to rate whether they agree or disagree with the statements on a scale of 0 – 100.

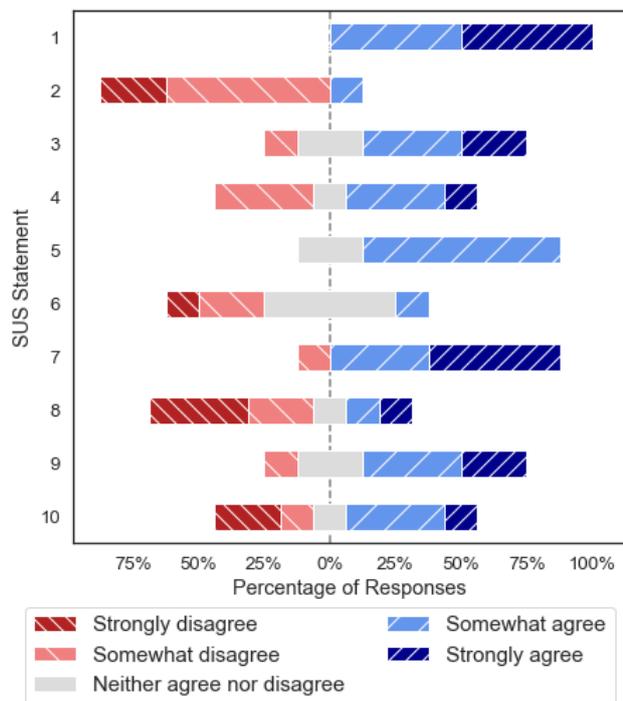
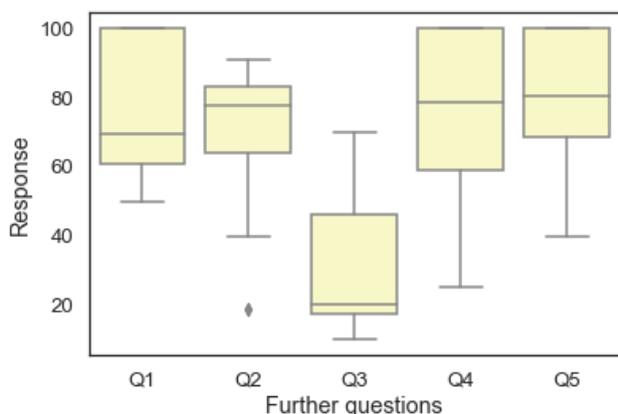


Figure 8. Responses to SUS statements.

Figure 9. Please tell us whether you agree or disagree with the following statements.

- Q1. I preferred using the AR instead of looking at the monitor to detect critical parameters.
- Q2. The medical parameters in the AR were clearly visible.
- Q3. I found that the AR visualization obscured the view over the working surface.
- Q4. I can imagine myself using surgical magnifying glasses with AR in daily OR work.
- Q5. I think magnifying glasses with AR can improve the outcome of surgeries.



4 DISCUSSION

4.1 Objective evaluation

The experiment's results showed an improvement in participant's reactions to critical phases when using the AR display compared to a common diagnostic monitor. In the test condition, not only reaction times were 51 % faster, but

also the performance for detecting critical parameters was increased by 60 %. On the other hand, given our sample size of 8 participants being small, statistical conclusions should be treated with care. Nevertheless, our initial proposition that augmented reality can improve surgeons' situational awareness could be confirmed, at least, in our experimental setup.

Circumstances during real operation are different. Firstly, OP light is much brighter than in our setup, which negatively affects the visibility of the AR visualizations. Technical suitability should be evaluated in further studies. Secondly, the question arises whether an AR display in magnifying glasses is necessary in the presence of auditory alerts and assistant doctors or anesthetists. Our experimental setup demonstrates that an AR display can help surgeons keep track of selected parameters omitting some communication efforts that can possibly interrupt the operation procedure.

4.2 Subjective assessments

There was a significant difference in global NASA-TLX between test and control condition, indicating that AR magnifying glasses can also help to improve workload in surgeons. Especially, frustration was reduced by 45 % and physical demand by 36 % (borderline not significant though) when using the AR display. Higher physical demand in monitor condition could result from the fact that participants had to turn their head to see the parameters. Also, in the control condition, participants missed more critical phases (response performance dropped by 38 %), which explains higher frustration. It might be interesting to further investigate these subscales under real circumstances, where cardiac surgeons rarely look to monitors but communicate with attendants about them.

4.3 Usability and opinions

Participant's feedback to the AR demonstrator presented were mixed and an average SUS score of 66.875 can be considered as "marginally acceptable" according to determined ranges by Bangor et al [2].

On the one hand, participants liked the AR functionality and showed interest in using such a device in the future.

On the other hand, everyone had difficulties with the prototype itself, mostly independent of the AR module. Adjustment of the magnifying glasses could take over 10 minutes but even then, the alignment of the loupes was not ideal. Sometimes participants had to re-adjust the loupes during the skill task. Another common problem was that, due to their weight, the glasses slightly slid down during skill task so that the top two parameters disappeared. These problems primarily occurred because of the mechanics for adapting the glasses to different users. Normally, surgical loupes are custom-made for each user, to fit pupil distance, height, and preferred working distances. However, we wanted one single device that could be tried out by multiple people, which would not be possible with a tailor-made device.

In the future, we are building another prototype with a different micro display that uses Bluetooth Low Energy for data transmission and a lighter battery. This omits a Raspberry Pi Zero in the glass frame reducing the weight

of the device, which should resolve most issues of the current AR demonstrator as discussed above.

5 CONCLUSION

This first feasibility study showed that ARMAGNI has the potential to increase surgeons' situation awareness during operations requiring magnification glasses while decreasing the workload for surgeons. Participants liked the concept of the product but had difficulties with the current prototype. In an upcoming improved prototype, participants feedback will be taken into account to enhance usability. In addition, a different micro display is intended to improve overall ergonomics of the device.

6 REFERENCES

- [1] Arpaia, P., Benedetto, E. D., Dodaro, C. A., Duraccio, L. & Servillo, G. Metrology-Based Design of a Wearable Augmented Reality System for Monitoring Patient's Vitals in Real Time. *IEEE Sensors Journal* **21**, 11176–11183 (2021).
- [2] Bangor, A., Kortum, P. T. & Miller, J. T. An Empirical Evaluation of the System Usability Scale. *International Journal of Human-Computer Interaction* **24**, 574–594 (2008).
- [3] Cercenelli, L. *et al.* The Wearable VOSTARS System for Augmented Reality-Guided Surgery: Preclinical Phantom Evaluation for High-Precision Maxillofacial Tasks. *Journal of Clinical Medicine* **9**, 3562 (2020).
- [4] ElBardissi, A. W., Wiegmann, D. A., Henrickson, S., Wadhera, R. & Sundt, T. M. Identifying methods to improve heart surgery: an operative approach and strategy for implementation on an organizational level. *European Journal of Cardio-Thoracic Surgery* **34**, 1027–1033 (2008).
- [5] Enlöf, P., Romare, C., Jildenstål, P., Ringdal, M. & Skär, L. Smart Glasses for Anesthesia Care: Initial Focus Group Interviews with Specialized Health Care Professionals. *Journal of PeriAnesthesia Nursing* **36**, 47–53 (2021).
- [6] Hazlehurst, B., McMullen, C. K. & Gorman, P. N. Distributed cognition in the heart room: How situation awareness arises from coordinated communications during cardiac surgery. *Journal of Biomedical Informatics* **40**, 539–551 (2007).
- [7] Liebert, C. A., Zayed, M. A., Aalami, O., Tran, J. & Lau, J. N. Novel Use of Google Glass for Procedural Wireless Vital Sign Monitoring. *Surgical Innovation* **23**, 366–373 (2016).
- [8] Lim, A. K., Ryu, J., Yoon, H. M., Yang, H. C. & Kim, S. Ergonomic effects of medical augmented reality glasses in video-assisted surgery. *Surgical Endoscopy* **36**, 988–998 (2021).
- [9] Maluf, M. A. *et al.* Cardiovascular Surgery Residency Program: Training Coronary Anastomosis Using the Arroyo Simulator and UNIFESP Models. *Revista Brasileira de Cirurgia Cardiovascular* (2015).
- [10] Qian, L., Song, T., Unberath, M. & Kazanzides, P. AR-Loupe: Magnified Augmented Reality by Combining an Optical See-Through Head-Mounted Display and a Loupe. *IEEE Transactions on Visualization and Computer Graphics* 1–1 (2020).
- [11] Raet, J. M. D. *et al.* How to build your own coronary anastomosis simulator from scratch†. *Interactive CardioVascular and Thoracic Surgery* **16**, 772–776 (2013).
- [12] Rahman, R. *et al.* Head-Mounted Display Use in Surgery: A Systematic Review. *Surgical Innovation* **27**, 88–100 (2019).
- [13] Woodward, J. & Ruiz, J. Analytic Review of Using Augmented Reality for Situational Awareness. *IEEE Transactions on Visualization and Computer Graphics* 1–1 (2022).
- [14] Yoon, J. W. *et al.* Technical feasibility and safety of image-guided parieto-occipital ventricular catheter placement with the assistance of a wearable head-up display. *The International Journal of Medical Robotics and Computer Assisted Surgery* **13**, e1836 (2017).

7 ACKNOWLEDGEMENT

This Project is supported by the Federal Ministry for Economic Affairs and Climate Action (BMWK) under grant 16KN069341 on the basis of a decision by the German Bundestag. Authors state no conflict of interest

Ubiquitous digital health-related data: clarification of concepts

Erlend Johannessen¹, André Henriksen¹, Gunnar Hartvigsen^{1,2}, Alexander Horsch¹, Eirik Årsand¹, Jonas Johansson³

¹Department of Computer Science, UiT, The Arctic University of Norway, Tromsø, Norway, erlend.johannessen@uit.no

²Department of Health and Nursing Science, University of Agder, Grimstad, Norway

³Department of Community Medicine, UiT, The Arctic University of Norway, Tromsø, Norway

Abstract

The increased development and use of ubiquitous digital services reinforce the trend where health-related data is generated everywhere. Data usage in different areas introduces different terms for the same or similar concepts. This adds to the confusion of what these terms represent. We aim to provide an overview of concepts and terms used in connection with digital twins and in a healthcare context.

Keywords

Digital Twin, Digital thread, Digital shadow, Small data, Terms and concepts, Healthcare

1 INTRODUCTION

Traditionally the healthcare field has been subject to strict privacy regimes. Patient data has only been available to health personnel with privileges to view or use this data. Details describing health conditions and health history has been and is still regarded as a matter between doctor and patient.

The development and use of ubiquitous (ever-present) digital services, online or offline, is based on the increasing trend that health-related data is generated “everywhere”. These types of data would not specifically describe the person that is using the digital services collecting these data. But when assembling various types of data for an individual, this may create a “customer” profile that can form identifying characteristics of the person. When assembling all available data for a person, we are approaching the concept of a digital twin.

A digital twin is generally defined as a digital replica of a living or non-living physical entity. The term digital twin was first presented by Michael Grieves in 2002-2003, and published in a white-paper in 2014 [1]. The term has been used in the manufacturing industry for over a decade with the aim of minimising costs, improving quality and product life, and increase efficiency in manufacturing. The concept of virtualising physical entities has since spread into other areas such as business, education, healthcare, transport, and construction.

Roberto Saracco’s article *Digital Twins: Bridging Physical Space and Cyberspace* [2] discusses the impact of digital twins and some future challenges. He defines a digital twin as consisting of a digital model, shadow, and thread. He predicts that digital twins will be applied to people in a not-too-distant future.

One problem with the concept of digital twins is that many of the terms in the digital twin field are used as synonyms. In her 2020 post [3], Lindsey Andrew argues that the term digital twin has become a buzzword, through the hype of

what can be provided versus the reality of the actual solutions. She also suggests that the term is overused and creates unrealistic expectations for the use of digital twinning. The fact that a search for “digital twin” today will result in millions of results [4] supports the hypothesis that we are witnessing a hype around the “digital twin” concept.

There are several other terms that are used alongside the term digital twin. All these terms are used under similar circumstances. To some degree they are used as synonyms, adding further to the confusion around the precise semantics of the underlying concepts. In order to be able to utilise these terms and concepts in a healthcare context, we need to clearly define them.

The aim of this paper is to provide an overview of concepts and terms used in connection with digital twins and define them in a healthcare context.

2 METHOD

The database Web of Science was searched with the queries "digital dust", "digital footprint", "digital phenotype", "digital shadow", "digital thread", "digital traces", and "small data". The goal with the search was finding explanatory articles, literary reviews, and scoping reviews where these and related terms were defined. The search query "digital twin" was expanded to "*digital twin*" AND ("*literature review*" OR "*scoping review*") to limit the results to more relevant articles.

We summarised number of papers for each search terms in a table in order to try to identify the most used concepts. The uncovered documents were skimmed for relevant articles, focussing on studies that could help explain digital twin related concepts and terms. Non-English and duplicate studies were excluded from the results.

3 RESULTS

An overview of the results from the term search can be seen in Table 1. A total number of 3727 studies were found, after

The 18th Scandinavian Conference on Health informatics, Tromsø, Norway, August 22-24, 2022. Organized by UiT The Arctic University of Norway. Conference Proceedings published by Linköping University Electronic Press at <https://doi.org/10.3384/ecp187>. © The Author(s). This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>

removing duplicates (n=223) and studies in other languages than English (n=50).

Some of these terms were already found in Roberto Saracco’s digital twin article [2] mentioned in the introduction. Especially the Semeraro et al. article *Digital Twin Paradigm: A Systematic Literature Review* [5] was useful, in that it listed 30 different definitions of what a digital twin is.

Term searched for	Results
Behavlets	1
Digital dust	4
Digital footprint	259
Digital phenotype	56
Digital shadow	73
Digital thread	103
Digital traces	427
Digital twin scoping/literature review	99
Small data	2705

Table 1. Results from literature search

Four articles in particular is worth mentioning with regard to definition and discussion of concepts.

The 2020 article *Digital Twins and the Emerging Science of Self: Implications for Digital Health Experience Design and “Small” Data* by Schwartz et al. [6] divides publicly available health data into four categories, 1) clinically generated data, 2) commercial real-world health data, 3) consumer digital health device-generated data, and 4) health-suggestive data. They also discuss the renaissance of N-of-1 or individual science. N-of-1 evaluation creates the opportunity to evaluate each individual uniquely.

Jones et al. [7], in their study *Characterising the Digital Twin: A systematic literature review*, defined 13 characteristics of digital twins, Physical Entity/Twin; Virtual Entity/Twin; Physical Environment; Virtual Environment; State; Realisation; Metrology; Twinning; Twinning Rate; Physical-to-Virtual Connection/Twinning; Virtual-to-Physical Connection/Twinning; Physical Processes; and Virtual Processes, and a framework with regard to digital twin operation. They also identified topics for future research.

Househ et al. [8], in their writeup *Big Data, Big Problems: A Healthcare Perspective*, argues that small data can be more accurate and can bring about more improved healthcare outcomes than big data systems, and that big data may cause more problems than solutions for healthcare.

In their article *Digital Twin in manufacturing: A categorical literature review and classification*, Kritzinger et al. [9] define digital twin, digital model, and digital shadow in more detail, to clear up any confusion around these terms.

4 TERMS AND CONCEPTS

This chapter outlines the terms searched for, and in addition some other terms that are interesting in a health-related

context. The terms in this chapter are divided into terms which originates in the industry and is not necessarily health-related, and those terms that are directly connected to human use of digital devices and services.

4.1 Terms used in an industrial context

The term digital twin is used in many different contexts and has been used to describe dissimilar scenarios. One problem is that the terms digital twin, digital model, and digital shadow often are used interchangeably. Other synonyms used for these include computerized counterpart, digital avatar, digital copy, digital replica, digital representation, dynamic virtual model, living model, virtual model, virtual prototype, and virtual replica. This usage may or may not include data transfer between the physical and digital entity. A summary of these terms can be found in Table 2.

Digital model

A digital model could be described as a representation of a physical entity. Digital models could be e.g., simulation models of planned entities or mathematical models of already existing entities. There is no automated data exchange between the physical entity and the digital model [9]. The dotted line in Figure 1 signifies this initial and/or manual updates.

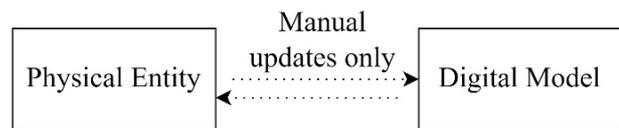


Figure 1. Digital model with initial and/or manual updates only

Digital shadow

If there exists an unidirectional flow of data from the physical entity to the digital model, then this could be defined as a digital shadow [9], visualised in Figure 2. The dataflow originates in changes of the physical entity as e.g., measured by sensors in the entity, and then creates changes to the digital model. Changes to the physical entity would happen through manual updates only.

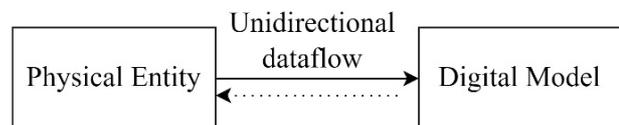


Figure 2. Digital shadow with unidirectional dataflow to update the digital model

Digital twin

If there exists a two-way data flow between the digital model and the modelled physical entity, this can be referenced to as a digital twin. In this case the data from the digital twin may control or update the physical entity, and vice versa, see Figure 3. Other related digital or physical systems or entities may control or update the paired entities, digital and physical [9].

Term	Categories	Description / Explanation	Example
Digital model	Data, Infrastructure	A digital representation of a physical entity	Having a computer-aided design (CAD) model of a building.
Digital shadow	Data, Process, Infrastructure	A digital model, where changes to the physical entity updates the digital model continuously (in real-time)	Updating the digital model with data from the actual physical construction in real time.
Digital thread	Data, Process, Infrastructure, Time	A data-driven architecture that links all information generated and stored within the digital twin, enabling it to flow seamlessly through the entire lifetime of the physical entity from invention to disposal	Logging events in the mechanisms and state of the building in order to analyse its history, to make it more efficient in the future, or to predict when to do maintenance.
Digital twin	Data, Process, Infrastructure	A digital model, where changes to the physical entity updates the digital model continuously (in real-time), and vice versa	Measuring changes to the building, updating the digital model, and automatically implement changes to the physical building based on the current state.

Table 2. Terms originating in the manufacturing industry

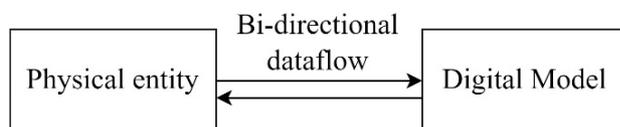


Figure 3. Digital twin with bidirectional dataflow to update both digital model and physical entity

The concept of creating or using a digital twin in order to optimise and maintain the underlying physical entity or process is also referred to as *digital twinning* [10].

Digital thread

A digital thread is recording or logging of a digital entity's lifetime, from creation to termination. Digital threads in smart manufacturing aim to show the physical entity's changes throughout its lifespan by following a product's design, performance data, product data, supply chain data, and software [11].

4.2 Terms used in a human context

There are several terms that arise from a human using digital services. This section describes most of them. A summary can be found in Table 3.

Digital traces

Digital traces (or digital trace data) are data we share when using digital devices, actively or passively [12]. As opposed to physical traces, like footprints in the sand, digital traces are the digital "footprints" we leave behind when using technical systems such as websites, social media platforms, smartphone apps, and sensors. Some of these traces are intentional, like emails, texts, blog posts, tweets, comments or likes on social media sites like Twitter and Facebook. However, many traces are invisible and unintentional, like records of our website visits and searches, or global-positioning systems (GPS), logs of movements, or phone calls.

Digital dust

Another term used as a synonym for digital traces may be digital dust [13]. This term is used in connection with online use and behavioural digital traces, see also *The Internet of Behaviour*, described below.

This term, however, has not yet been confirmed as a scientific term. It is used in several different ways in research, sometimes casually; what's left online when people die [14], digital investigation (forensics, defence) [15], criminal investigation [16], or grey literature (literature like presentations, reports, blogs, papers, produced outside traditional publishing channels) [17].

Small data

Small data is derived from our individual digital traces. Consider a new kind of cloud-based app that would create a picture of your health over time by continuously, securely, and privately analysing the digital traces you generate as you work, shop, sleep, eat, exercise, and communicate.

While there are personal devices and Internet services specifically designed for self-tracking, digital traces include a much richer corpus of data that we generate every day [18].

Digital phenotype

A digital phenotype, as defined by Jain et al. [19], can be regarded as a term for the trail of relevant health data that are left behind in people's use of the internet, social media, and digital technologies in general. Jain argues that this data largely is an untapped potential for early detection of various health conditions.

Digital phenotyping is a term introduced by Tourus et al. [20], and is described as "moment-by-moment quantification of the individual-level human phenotype in-situ using data from smartphones and other personal digital devices". Tourus et al. defines digital phenotyping as distinct from the term digital phenotype. Their goal was creating a platform to collect research-quality data from raw smartphone sensors and smartphone usage.

Term	Categories	Description / Explanation	Example
Crowdsensing	Data collection	Data collection technique where information is extracted from mobile devices such as smartphones, tablet computers, or wearables	Collecting data from mobile use on social networks, search engines, mobile operators, online games, and e-commerce sites
Digital dust	Data	Data we share when using digital devices accessing online services, actively or passively. Used as a synonym for digital traces	Data from social networks, search engines, mobile operators, online games, and e-commerce sites
Digital footprint	Data	Sum of all Digital dust, Digital traces, and/or Small data	All data from social networks, search engines, mobile operators, online games, and e-commerce sites
Digital phenotype	Data, Metadata	A human trait described from a trail of health-related data left behind through (user) interaction with technology	Data from social networks, search engines, mobile operators, online games, and e-commerce sites
Digital traces	Data	Data we share when using digital devices accessing online services, actively or passively.	Data from social networks, search engines, mobile operators, online games, and e-commerce sites
Small data	Data	Leave behind a 'trail of breadcrumbs' with our digital service providers	Data from social networks, search engines, mobile operators, online games, and e-commerce sites
The Internet of Behaviour (IoB)	Data, Psychology	Change in human behaviours based on collected and used digital dust	Recommending new user experiences, product suggestions, and company services based on the collected digital dust

Table 3. Terms originating from a human-digital context

Digital footprint

A digital footprint is the sum of all the data that we leave behind with or without our consent when we use digital services [21]. This has historically also been referred to as a digital shadow, but given Kritzinger's [9] definition of a digital shadow, we will not use this definition to mean the same as a digital footprint. Figure 4 illustrates adding up all small data (or digital traces) for a person.

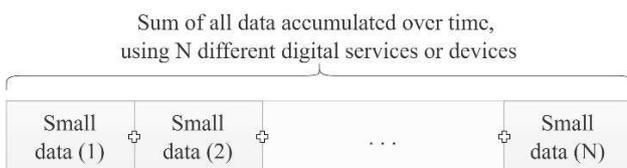


Figure 4. Digital footprint visualised

The Internet of Behaviour

The Internet of Behaviour (IoB) is about collecting and using digital dust from a variety of sources, to change behaviours using feedback loops [22-23]. User data is analysed with regards to behavioural psychology, recommending new user experiences, product suggestions, and company services based on the collected digital dust.

Behavlets

Behavlets are used in gaming (also referred to as videogaming) to extrapolate features from actions or patterns in gameplay which expose player behaviour [24]. Behavlets may be used to improve application interaction

and predict behaviour, approaching research areas such as psychology and temperament theory.

Crowdsensing

Crowdsensing, sometimes called mobile crowdsensing, is a data collection technique where a group of individuals with mobile devices such as smartphones, tablet computers, or wearables, collectively share and extract information with the intention to measure, map, analyse, estimate, or predict processes of common interest. This comes mainly in two flavours, participatory and opportunistic crowdsensing. Participatory crowdsensing is when users voluntarily participate in contributing information, opportunistic crowdsensing is when data is sensed, collected, and shared automatically without user intervention, in some cases without the user's knowledge [25-26].

The Internet of Medical Things

Internet of Things (IoT) is the technology that enables devices with communication capabilities to share data with other devices and systems over the internet [27].

The Internet of Medical Things (IoMT) is a subset of IoT for medical and health-related purposes, data collection and analysis for research, and monitoring [28]. A IoMT device can be almost anything, from personal devices like specialised implants, pacemakers, wristbands, or hearing aids, to "smart beds" or sensors within living spaces or kitchen equipment.

5 DISCUSSION

When used in different contexts, these terms may differ because the particular term synonym fits a particular usage scenario. Problems arise when the exact same term is used as a description of different phenomena. This creates communication difficulties, especially when used across different disciplines or professions, and could hold back the progress of the research field.

There might be several reasons for a term having different meanings. There might be a development over time, where a term definition has been changed due to usage in new areas or introducing new technology. Or there has been a lack of international consensus meetings to operationalise common definitions.

What is noteworthy is that there are different expressions that are applied to the same context. An example of this is the terms *digital dust*, *digital phenotype*, *digital traces*, or *small data*, which all describe the trail of data left behind when using digital services or devices. These all stem from human activity where digital technologies with data storage are used.

The definition of a digital thread and a digital footprint may on the surface look similar. The difference is that a digital thread logs all changes over time, while a digital footprint is a snapshot of all the existing data, not factoring in updates and history.

Digital twin terms, from manufacturing to healthcare

One question to ask is if terms in manufacturing belong in a human or a healthcare context. Creating a human digital twin as previously defined in manufacturing does not seem fully attainable, with a bi-directional dataflow between physical entity (the human) and the digital model. This would mean “updating the human” based on data in the digital model, although this might be partly realizable through changes in medication (e.g., for advanced insulin pumps in diabetes self-management) or by updating physical implants digitally, but not in all respects. The human body is still not fully physiologically understood, so creating a perfect digital copy with bi-directional updates seem unattainable.

The one-directional flow of data from the physical entity (human) to the digital model seems more plausible. In a human context, the *digital shadow* can be viewed as the digital equivalent of living a life, monitoring all processes

in the body as life happens, with the Human Digital Thread as a history or log of human data/events/updates.

The *digital thread* could be viewed as the log or history of all dataflows to update the digital shadow or digital twin. In a human context this can be defined as all the small data that is generated by human activity in different contexts, using phones (smart or not), general mobile app usage, computer usage, or online services. This is illustrated in Figure 5. The digital thread consists of multiple data events, here referred to as small data instances (slices), over a period of time. This describes a continuous history of the digital shadow or digital twin. The digital shadow or twin is defined by the latest iteration or data update in the lifetime of the entity. This is always the “now”, the current time, given that there are continuous updates to the digital representation. This would be the case for human digital twins.

One way of imagining a digital twin of a person, is to avoid creating an exact digital copy of the individual. Accepting that there are areas we cannot – yet – recreate digitally, opens the possibility for creating a semi-automatic feedback loop where monitoring and registering data for an individual may give direct or indirect feedback to the individual through software or technology. This is in many ways what we have today, though not as a complete system. Wearing smartwatches or using smart scales sends health-related data to servers, and feedback are given through e.g., mobile applications. These kind of data sources are narrow in scope, in that they register parameters like activity and weight, but does not consider all sorts of other sources of information like weather conditions, road dust, mental state, illnesses, season variations, and similar. Creating this wider definition of a human digital twin could in the end be valuable for public health, health care providers, policy makers, and others, as well the person him/herself. This type of system could also be valuable for population studies similar to The Tromsø Study [29].

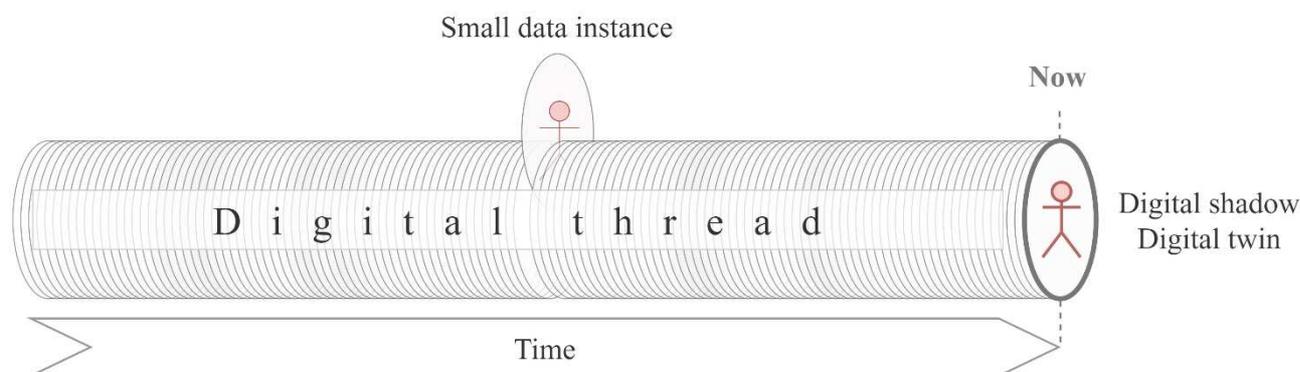


Figure 5. Digital thread, small data, and digital shadow (or digital twin) illustrated

6 SUMMARY

We have examined digital twin and related terms and concepts and how they fit in the healthcare area. A human digital twin is currently not fully realisable, but there are untapped possibilities for creating a more complete human digital twin today, by joining different types of data from a heterogenous set of services and sources. If we can exploit these possibilities, we can approach an approximation of a digital twin for use in healthcare.

7 REFERENCES

- [1] M. Grieves, 'Digital twin: manufacturing excellence through virtual factory replication', *White paper*, vol. 1, pp. 1–7, 2014.
- [2] R. Saracco, 'Digital Twins: Bridging Physical Space and Cyberspace', *Computer*, vol. 52, no. 12, pp. 58–64, Dec. 2019, doi: 10.1109/mc.2019.2942803.
- [3] Lindsey Andrews, 'The Term Digital Twin is a Buzzword', *SensrTrx | Manufacturing Analytics*, Feb. 18, 2020. <https://www.sensrtrx.com/digital-twin-buzzword/> (accessed Oct. 03, 2021).
- [4] "digital twin*" - Google Search'. https://www.google.com/search?q=%22digital+twin*%22 (accessed Feb. 24, 2022).
- [5] C. Semeraro, M. Lezoche, H. Panetto, and M. Dassisti, 'Digital twin paradigm: A systematic literature review', *Computers in Industry*, vol. 130, p. 103469, Sep. 2021, doi: 10.1016/j.compind.2021.103469.
- [6] S. M. Schwartz, K. Wildenhaus, A. Bucher, and B. Byrd, 'Digital Twins and the Emerging Science of Self: Implications for Digital Health Experience Design and "Small" Data', *Frontiers in Computer Science*, vol. 2, 2020, Accessed: May 28, 2022. [Online]. Available: <https://www.frontiersin.org/article/10.3389/fcomp.2020.00031>
- [7] D. Jones, C. Snider, A. Nassehi, J. Yon, and B. Hicks, 'Characterising the Digital Twin: A systematic literature review', *CIRP Journal of Manufacturing Science and Technology*, vol. 29, pp. 36–52, May 2020, doi: 10.1016/j.cirpj.2020.02.002.
- [8] M. S. Househ, B. Aldosari, A. Alanazi, A. W. Kushniruk, and E. M. Borycki, 'Big Data, Big Problems: A Healthcare Perspective', *Stud Health Technol Inform*, vol. 238, pp. 36–39, 2017.
- [9] W. Kritzinger, M. Karner, G. Traar, J. Henjes, and W. Sihn, 'Digital Twin in manufacturing: A categorical literature review and classification', *IFAC-PapersOnLine*, vol. 51, no. 11, pp. 1016–1022, Jan. 2018, doi: 10.1016/j.ifacol.2018.08.474.
- [10] M. M. Rathore, S. A. Shah, D. Shukla, E. Bentafat, and S. Bakiras, 'The Role of AI, Machine Learning, and Big Data in Digital Twinning: A Systematic Literature Review, Challenges, and Opportunities', *IEEE Access*, vol. 9, pp. 32030–32052, 2021, doi: 10.1109/ACCESS.2021.3060863.
- [11] robert.lipman@nist.gov, 'Enabling the Digital Thread for Smart Manufacturing', *NIST*, Dec. 04, 2014. <https://www.nist.gov/el/systems-integration-division-73400/enabling-digital-thread-smart-manufacturing> (accessed Mar. 07, 2022).
- [12] J. Howison, A. Wiggins, and K. Crowston, 'Validity Issues in the Use of Social Network Analysis with Digital Trace Data', *Journal of the Association for Information Systems*, vol. 12, no. 12, Dec. 2011, doi: 10.17705/1jais.00282.
- [13] 'What Is "Digital Dust" and Should We Worry About Ours?' <https://www.linkedin.com/pulse/what-digital-dust-should-we-worry-ours-ema-linaker> (accessed Mar. 07, 2022).
- [14] N. Wright, 'Death and the Internet: The implications of the digital afterlife', *First Monday*, May 2014, doi: 10.5210/fm.v19i6.4998.
- [15] E. Casey, 'Digital dust: Evidence in every nook and cranny', *Digital Investigation*, vol. 6, no. 3, pp. 93–94, May 2010, doi: 10.1016/j.diin.2010.02.002.
- [16] C. Bernardi, 'Digital Dust and Visual Narratives of Femicidios', *Visual Culture & Gender*, vol. 13, pp. 6–16, Sep. 2018.
- [17] 'Grey Literature: Digital scholarship or digital dust', *ABC Radio National*, Feb. 25, 2013. <https://www.abc.net.au/radionational/programs/bigideas/grey-literature/4529562> (accessed Jan. 12, 2022).
- [18] D. Estrin, 'Small data, where n = me', *Commun. ACM*, vol. 57, no. 4, pp. 32–34, Apr. 2014, doi: 10.1145/2580944.
- [19] S. H. Jain, B. W. Powers, J. B. Hawkins, and J. S. Brownstein, 'The digital phenotype', *Nat Biotechnol*, vol. 33, no. 5, Art. no. 5, May 2015, doi: 10.1038/nbt.3223.
- [20] J. Torous, M. V. Kiang, J. Lorme, and J.-P. Onnela, 'New Tools for New Research in Psychiatry: A Scalable and Customizable Platform to Empower Data Driven Smartphone Research', *JMIR Mental Health*, vol. 3, no. 2, p. e5165, May 2016, doi: 10.2196/mental.5165.
- [21] M. Madden, S. Fox, A. Smith, and J. Vitak, 'Digital Footprints', *Pew Research Center: Internet, Science & Tech*, Dec. 16, 2007. <https://www.pewresearch.org/internet/2007/12/16/digital-footprints/> (accessed Mar. 08, 2022).
- [22] 'Gartner Top Strategic Technology Trends for 2021', *Gartner*. <https://www.gartner.com/smarterwithgartner/gartner-top-strategic-technology-trends-for-2021> (accessed Mar. 07, 2022).
- [23] 'What is the Internet of Behavior and Why is it Important for Business?' <https://gbksoft.com/blog/internet-of-behaviors/> (accessed Mar. 07, 2022).
- [24] B. Cowley and D. Charles, 'Behavlets: a method for practical player modelling using psychology-based player traits and domain specific features', *User Model User-Adap Inter*, vol. 26, no. 2, pp. 257–306, Jun. 2016, doi: 10.1007/s11257-016-9170-1.
- [25] B. Guo *et al.*, 'Mobile Crowd Sensing and Computing: The Review of an Emerging Human-Powered Sensing Paradigm', *ACM Comput. Surv.*,

- vol. 48, no. 1, p. 7:1-7:31, Aug. 2015, doi: 10.1145/2794400.
- [26] R. K. Ganti, F. Ye, and H. Lei, 'Mobile crowdsensing: current state and future challenges', *IEEE Communications Magazine*, vol. 49, no. 11, pp. 32–39, Nov. 2011, doi: 10.1109/MCOM.2011.6069707.
- [27] 'Internet of Things Global Standards Initiative'. <https://www.itu.int/en/ITU-T/gsi/iot/Pages/default.aspx> (accessed Jan. 11, 2021).
- [28] A. Gatouillat, Y. Badr, B. Massot, and E. Sejdić, 'Internet of Medical Things: A Review of Recent Contributions Dealing With Cyber-Physical Systems in Medicine', *IEEE Internet of Things Journal*, vol. 5, no. 5, pp. 3810–3822, Oct. 2018, doi: 10.1109/JIOT.2018.2849014.
- [29] L. A. Hopstock, S. Grimsgaard, H. Johansen, K. Kanstad, T. Wilsgaard, and A. E. Eggen, 'The seventh survey of the Tromsø Study (Tromsø7) 2015–2016: study design, data collection, attendance, and prevalence of risk factors and disease in a multipurpose population-based health survey', *Scand J Public Health*, p. 14034948221092294, May 2022, doi: 10.1177/14034948221092294.

8 ACKNOWLEDGEMENTS

The study was funded by UiT the Arctic University of Norway.

The Perceived Usefulness of a Contactless Sleep Monitor for Adults with Type 1 Diabetes: A Qualitative Study

Julie Egmosen^{1*}, Katrine Holmstrup Sørensen¹, Tanja Fredensborg Holm¹,

Ole Hejlesen¹, Morten Hasselstrøm Jensen^{1,2}, and Stine Hangaard^{1,2}

¹Aalborg University, Department of Health Science and Technology, Aalborg, Denmark

²Steno Diabetes Center North Denmark, Aalborg University Hospital, Denmark

* Presenting author: juliee@hst.aau.dk

Abstract

When evaluating the potential of an alternative method to detect physiological changes that occur during hypoglycemia it is important to be aware of the perceived usefulness among potential users [17]. The aim of this study was to explore the perceived usefulness of a contactless monitor for detection of hypoglycemia among people with type 1-diabetes (T1D). Semi-structured interviews (n =8) were conducted after people with T1D used the contactless monitor at home. The interviews were analyzed using a thematic approach. Five overall themes were identified. Overall, the participants reported that the contactless monitor was an acceptable method to detect nocturnal hypoglycemia.

Keywords

Type 1 diabetes - Nocturnal Hypoglycemia - Contactless Monitor - Perceived Usefulness - Qualitative Interview

1 INTRODUCTION

Diabetes mellitus (DM) is one of the fastest growing health challenges of the 21st century [1]. DM increases the risk of developing several complications, depending on the degree and duration of hyperglycemia [2]. People with type 1 diabetes (T1D) are dependent on continuous exogenous insulin administration due to β -Cell loss or destruction, affecting the insulin secretion [2]. Exogenous insulin increases the risk of developing hypoglycemia. [3]. Hypoglycemia can be severe and potentially life-threatening [4], as the body depends on a continuous supply of glucose to maintain vital functions [3]. In hypoglycemia, several physiological changes occur, including increased heart rate (HR) and respiration rate (RR) [5,6].

Close to 50% of all severe hypoglycemic episodes occur during sleep. Nocturnal hypoglycemia can negatively impact the quality of life as it results in emotions such as fear, anxiety, and worry [7]. Fear of hypoglycemia may prevent insulin therapy from equaling good glycemic control [8]. To avoid hypoglycemia, people with T1D often manipulate their blood glucose levels higher, by taking less insulin than needed to achieve good glycemic control [9]. However, this increases the risks of serious late diabetic complications due to hyperglycemia [10].

Currently, a continuous glucose monitor (CGM) is used as the first and main tool for detection of hypoglycemia [11]. However, the use of CGM is associated with several disadvantages. Among other things, a deviation between measured glucose values with CGM and venous blood glucose has been observed. It remains unknown why this deviation occurs, but it may be due to people sleeping on the body site at which the CGM is placed [12]. Furthermore, skin irritation or allergic reactions may occur when using CGM [13]. People with T1D who don't have access to a CGM will have to

perform self-monitoring of blood glucose (SMBG). However, reducing complications in DM requires frequent use of SMBG, which can lead to scarring, pain, and reduced sensitivity [14,15].

There is an increasing interest in alternative detection methods related to the physiological changes that occur during hypoglycemia [11]. Detection of hypoglycemia based on physiological changes requires measuring vital parameters. In conventional clinical practice, contact-based sensors are used [16]. Contact-based sensors are inappropriate for prolonged and repeated measurements as detection of nocturnal hypoglycemia would require [16]. To the best of our knowledge, no studies have investigated the possibility of contactless detection of nocturnal hypoglycemia.

Besides the potential of contactless monitoring to detect nocturnal hypoglycemia, an alternative detection method must be aware of the acceptance among potential users, as it's an essential parameter for the successful implementation of new technology [17]. Perceived usefulness is essential for technology acceptance [17]. In this paper, potential users are people with T1D as people with T1D are the ones who are most likely to have nocturnal hypoglycemia. The aim of the study was to explore the experience of perceived usefulness of a contactless monitor for detection of nocturnal hypoglycemia among people with T1D.

2 METHODS

2.1 Participants

People with T1D aged ≥ 18 years were recruited through social media, see figure 1. The inclusion criteria were people with T1D using a Flash Glucose Monitor (isCGM) Freestyle Libre 1 or 2 and the ability to understand and speak in Danish or

English. Exclusion criteria were sleep apnoea, hypoglycemia unawareness, pregnancy, electronic implants, or children sleeping in bed. The study was performed between February 2022 and June 2022 at Aalborg University, Denmark. According to the Declaration of Helsinki, each participant signed a written informed consent.

The study wasn't found notifiable by the local ethics committee or Medical Research Ethics Committees (MREC).

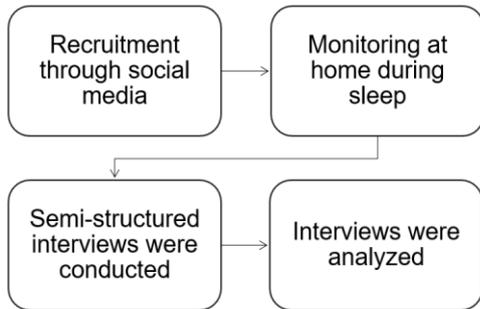


Figure 1. The flowchart illustrates the process from recruitment to analysis of the interviews.

2.2 Data and Procedures

The present study was originally a part of a mixed methods study and was therefore presented as a sub-study. The original study consisted of the present study and a quantitative part where it was investigated if contactless monitoring of heart rate and respiration rate could detect nocturnal hypoglycemia. The quantitative part is expected to be published elsewhere.

A qualitative interview study was performed to explore the perceived usefulness of a contactless monitor among people with T1D. To provide a realistic experience of using the contactless monitor, the participants were instructed to monitor HR and RR during sleep throughout a period of 5-8 weeks at home. The contactless monitoring system consisted of a monitor (Sleepiz One, Sleepiz AG, Switzerland) and a wifi hotspot (Huawei 4G Mobile WiFi). The monitor used radar technology and collected HR and RR. Each participant was instructed to place the monitor at the bedside. The hotspot was placed in the same room as the monitor. All participants were instructed to remain on their usual insulin therapy and continue their normal lifestyle throughout the study. The participants turned on the monitor before they went to bed and turned off the monitor after waking up in the morning. If the participants woke up during the night, they could get out of bed leaving the monitor turned on.

At the end of the study period, a 3-8 minute semi-structured interview was conducted with all participants. A semi-structured interview is characterized by the possibility of making the interview relevant to the desired focus by preparing an interview guide with research and interview questions in advance [18]. The interview guide was based on, perceived usefulness [17], which formed the basis for research and interview questions. Perceived usefulness was the overall keyword that formed the framework for the four interview questions, see figure 2. The interviews were conducted by one

of the researchers of the team using telephone or Microsoft Teams. All interviews were audio-recorded, transcribed verbatim, and fully anonymized.

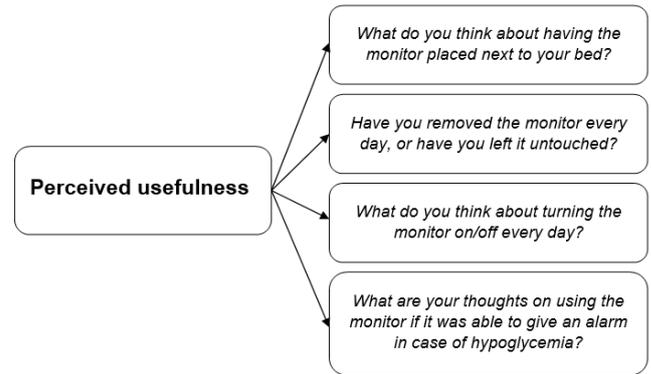


Figure 2. The keyword, perceived usefulness, and related interview questions.

2.3 Data Analysis

The interview data were analyzed using Kvale and Brinkman's meaning coding, meaning condensation, and meaning interpretation [19].

Three authors (KHS, TFH, and JE) independently did the meaning coding by reading through the transcripts, while relevant themes were identified and coded [18,20]. Afterward, the identified themes were unified. In case of overlapping themes, these were compiled. The themes were arranged in a matrix with accompanying quotations from the transcripts. The quotations for each theme were compressed into short and precise formulations by meaning condensation [18]. The results were based on self-understanding as the context of interpretation [18].

3 RESULTS

3.1 Participant Characteristics

Eight participants aged 24-68 years (48.5 ± 14.7) with T1D were included in the study. Diabetes duration ranged from 0.67-53 years. For an overview of baseline characteristics see Table 1.

Parameter	Mean	± SD
$N = 8$		
Female ($n = 4$)		
Age (years)	48.5	14.7
Diabetes duration (years)	21.7	16.7

Table 1. Baseline characteristics of participants in this study

3.2 Overall Themes

All participants completed a semi-structured interview. It was assessed that data saturation was achieved, as no new themes emerged in the final two interviews. Five overall themes were identified, see figure 3.

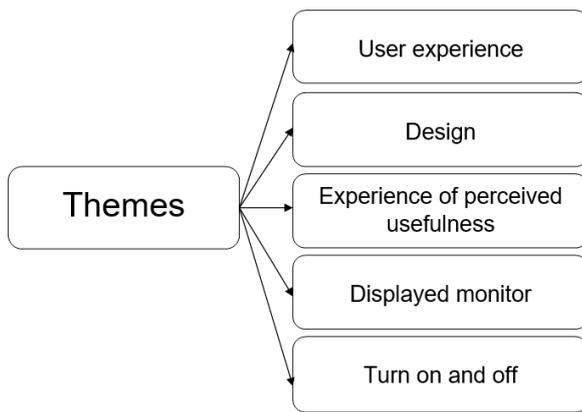


Figure 3. Overview of themes based on semi-structured interviews

3.3 User Experience

The overall experience of using the contactless monitor was that it was easy and clear to use. It quickly became a habit to use the monitor.

'It means nothing for me to turn it on, it's like turning off my night light. It's that easy and clear. (...) So for me, it just quickly became a habit, also because it just stands right there. So, it's super easy. I wish everything with diabetes was so easy. That would be great.' (P1)

A few participants described a feeling of surveillance but found this insignificant in terms of use.

'Well, something is monitoring me at night. It's strange, but there is nothing in it, I can't keep up with anything, so I don't know.' (P7)

One participant experienced some frustrations using the monitor due to repeated defects consisting of problems with transferring data. Except for problems due to defects, the participant found the actual use of the monitor unproblematic.

'Besides the fact that it sometimes doesn't work, it has not really been that bad (...) But I can just say that if it had caused any more problems, then I probably would not bother anymore.' (P3)

3.4 Design

Two participants suggested changing the design of the monitor, to improve the perceived usefulness. It was suggested to extend the range of the monitor so it could be placed on e.g. a shelf or a bedside table. Furthermore, it was suggested to change the location of the indicator light on the monitor, to avoid nuisance from the light during sleep.

'(...) You could redesign it on the back, then it would be smart. Or on the other side, but if you then sleep on the other side of the bed, it will still light right in your face. So, if it's in the back where you put the power plugin, then that's fine.' (P1)

One participant expressed concerns about using the monitor, due to lack of aesthetic expression.

'(...) It's not that pretty [contactless monitor], and I would say that in the long run, it would be annoying to have it there.' (P6)

3.5 Experience of Perceived Usefulness

All participants evaluated that if the contactless monitor could give an alarm in case of hypoglycemia, it was an acceptable supplement to existing methods of detection. Thus, they found that the monitor has the potential to provide greater security during the night, leading to better sleep quality.

'Well, I think that for me it's an indispensable tool [isCGM]. So, you can say that I experience low blood sugar occasionally, and it gives me enormous peace of mind that something is waking me up. Then I think my sleep quality is better because there is something that wakes me up. Whether it's the alarm [contactless monitor] or the device [isCGM], it wouldn't matter to me. The security about sleeping is just nice.' (P6)

Two participants did not fear nocturnal hypoglycemia, and therefore found alarms for hypoglycemia irrelevant.

'(...) I have turned off my alarms because sometimes I am too high and other times, I am too low. So, but it still runs with me, it's not like I get completely down [hypoglycemia].' (P3)

3.6 Displayed Monitor

All participants had the monitor displayed during the entire study period. The majority stated that it made it easy to use the monitor daily. Some participants indicated that it would probably cause frustration to remove it every day.

'I left it untouched [contactless monitor]. Because then it's so easy. Then it's just to press a button twice a day. If I had to set it up and remove it, I think it would be a much more difficult thing to do and a much more tiring experience.' (P2)

3.7 Turn On and Off

All participants turned the contactless monitor on and off without any problems. However, all participants experienced that it could be difficult to remember turning it on and off. Several participants described that it quickly became a routine to turn the monitor on and off during use.

'Well, it meant nothing to me. In the evening I turn it on [contactless monitor] and then I go out and brush my teeth. In the morning when I have measured blood glucose and stuff like that I turn it off. So, it's just a rhythm.' (P4)

4 DISCUSSION

4.1 Perceived Usefulness

The aim of the study was to explore the experience of perceived usefulness of a contactless monitor for detection of nocturnal hypoglycemia among people with T1D. Overall, the participants reported that the contactless monitor was an acceptable alternative for existing nocturnal hypoglycemia detection methods if the monitor was able to give an alarm in case of hypoglycemia. Some participants emphasized that their use of isCGM gave a feeling of safety during sleep and mentioned that the monitor had the potential to provide equivalent safety. These findings support previous research in which CGM was able to provide a feeling of safety during sleep [21]. Furthermore, the use of CGM gave fewer disturbances during sleep and made people with T1D fall asleep more easily [21]. Though most of the participants found the opportunity of an alarm in case of hypoglycemia essential, a few participants experienced no fear of nocturnal hypoglycemia and found the alarm less relevant. Among some participants, there were divided opinions on whether alarms from their isCGM disturbed sleep. Previous research found that alarms from CGM were essential but could disturb sleep [21]. Moreover, too many alarms were annoying and made some people with T1D turn off the alarms [21].

The participants' acceptance of the monitor as an alternative method might increase the opportunity for a successful implementation since perceived usefulness promotes use and acceptance [22]. Perceived usefulness may be affected by usability, whereby it indirectly has an impact on the acceptance of a new technology [22]. The participants stated that the contactless monitor was easy and clear to use. The use of the monitor only required the participants to turn the monitor on and off every day, which was not associated with difficulties. The usability is thus assessed to have a positive influence on perceived usefulness and therefore increase a successful implementation [22].

4.2 Strengths and Limitations

One of the main strengths of the present study is that the results address an important gap of knowledge according to perceived usefulness of an unexplored alternative detection method. Another strength is that the participants included in the study are diagnosed with T1D, who are the potential end users of the contactless monitoring system. Their perspectives according to perceived usefulness will gain important insight to promote future successful implementation [17].

A limitation is that the interview was conducted using a telephone as it can influence the interaction between interviewer and participant [18]. Another limitation is that the recruitment of participants through social media might have influenced the results. Participants using social media might have a greater acceptance and understanding of technology. One way to reduce this potential confounder could be recruitment through e.g., general practitioners or diabetes clinics. Finally, a limitation is the short duration of the interviews.

5 CONCLUSION

The participants reported that the contactless monitor was an acceptable alternative to detect hypoglycemia compared to existing detection methods. Future research would be needed to further evaluate the potential of contactless detection of nocturnal hypoglycemia.

6 REFERENCES

- [1] International Diabetes Federation. IDF diabetes atlas [Online]. 2019. Available from: https://www.diabetesatlas.org/upload/resources/material/20200302_133351_IDFATLAS9e-final-web.pdf
- [2] Ramachandran A, Snehalatha C, Nanditha A. Classification and Diagnosis of Diabetes. I: Textbook of Diabetes, pp. 23–29. United Kingdom, 2017.
- [3] Cryer PE, Arbeláez AM. Hypoglycemia in Diabetes. I: Textbook of Diabetes, pp. 513-529. United Kingdom, 2017.
- [4] Senthilkumaran M, Zhou XF, Bobrovskaya L. Challenges in Modelling Hypoglycaemia-Associated Autonomic Failure: A Review of Human and Animal Studies. *International Journal of Endocrinology*. Vol. 2016. 2016.
- [5] Thompson EL, Ray CJ, Holmes AP, Pye RL, Wyatt CN, Coney AM, m.fl. Adrenaline release evokes hyperpnoea and an increase in ventilatory CO₂ sensitivity during hypoglycaemia: a role for the carotid body. *J Physiol*. Vol. 594, pp. 4439-4452. 2016.
- [6] Ward DS, Voter WA, Karan S. The effects of hypo- and hyperglycaemia on the hypoxic ventilatory response in humans. *J Physiol*. Vol. 582, pp. 859-869. 2007.
- [7] Edelman SV, Blose JS. The Impact of Nocturnal Hypoglycemia on Clinical and Cost-Related Issues in Patients With Type 1 and Type 2 Diabetes. *Diabetes Educ*. Vol. 40, pp. 269-279. 2014.
- [8] Schechter A, Eyal O, Zuckerman-Levin N, Amihai-Ben-Yaacov V, Weintrob N, Shehadeh N. A prototype of a new noninvasive device to detect nocturnal hypoglycemia in adolescents with type 1 diabetes--a pilot study. *Diabetes Technol Ther*. Vol. 14, pp. 683-689. 2012.
- [9] Wild D, von Maltzahn R, Brohan E, Christensen T, Clauson P, Gonder-Frederick L. A critical review of the literature on fear of hypoglycemia in diabetes: Implications for diabetes management and patient education. *Patient Education and Counseling*. Vol. 68, pp. 10-15. 2007.
- [10] Flyvbjerg A, Cockram C, Flyvbjerg A, Goldstein BJ. Pathogenesis of Microvascular Complications. I: Textbook of Diabetes. pp. 543-553. United Kingdom, 2017.
- [11] Diouri O, Cigler M, Vettoretti M, Mader JK, Choudhary P, Renard E, m.fl. Hypoglycaemia detection and prediction techniques: A systematic review on the latest developments. *Diabetes/Metabolism Research and Reviews*. Vol. 37, pp. 1-19. 2021.

- [12] Lodwig V, Kulzer B, Schnell O, Heinemann L. Current Trends in Continuous Glucose Monitoring. *J Diabetes Sci Technol*. Vol. 8, pp. 390-406. 2014.
- [13] Englert K, Ruedy K, Coffey J, Caswell K, Steffen A, Levandoski L. Skin and Adhesive Issues With Continuous Glucose Monitors: A Sticky Situation. *J Diabetes Sci Technol*. Vol. 8, pp. 745-751. 2014.
- [14] Dicembrini I, Cosentino C, Monami M, Mannucci E, Pala L. Effects of real-time continuous glucose monitoring in type 1 diabetes: a meta-analysis of randomized controlled trials. *Acta Diabetol*. Vol. 58, pp. 401-410. 2021.
- [15] Heinemann L. Finger Pricking and Pain: A Never Ending Story. *J Diabetes Sci Technol*. Vol. 1, pp. 919-921. 2008.
- [16] Kebe M, Gadhafi R, Mohammad B, Sanduleanu M, Saleh H, Al-Qutayri M. Human Vital Signs Detection Methods and Potential Using Radars: A Review. *Multidisciplinary Digital Publishing Institute*. Vol. 20, pp. 1454. 2020.
- [17] Ammenwerth E. Technology Acceptance Models in Health Informatics: TAM and UTAUT. *Applied Interdisciplinary Theory in Health Informatics*. pp. 64-71. 2019.
- [18] Kvale S, Brinkmann S. *InterViews: learning the craft of qualitative research interviewing*. Third edition. pp. 405. Los Angeles, 2015.
- [19] Brinkmann S, Kvale S, Flick U. *Doing interviews*. pp. 186. Los Angeles ; London, 2018.
- [20] Leedy PD, Ormrod JE, Johnson LR. *Practical research: planning and design*. 2021.
- [21] Je P, M FH, K S. Real-time continuous glucose monitoring in type 1 diabetes: a qualitative framework analysis of patient narratives. *Diabetes care*. Vol 38. 2015.
- [22] Holden RJ, Karsh BT. The Technology Acceptance Model: Its past and its future in health care. *Journal of Biomedical Informatics*. Vol. 43, pp. 159-172. 2010.

7 ACKNOWLEDGEMENT

The authors would like to thank all the participants for kind and patient involvement in the study and for their collaboration. Furthermore, the authors thank Steno Diabetes Center Nordjylland, Denmark for providing economic funding. Moreover, the authors would like to thank Dominik Hollinger at Sleepiz AG, Switzerland, for preparation and support of the contactless monitors.

Self-imperative Care of Pregnancy using IoT Solutions

Sarala Ghimire¹, Santiago Martinez², Martin Gerdes¹

¹Department for Information and Communication Technologies, Centre for e-Health, University of Agder, Grimstad, Norway, ghimires@uia.no

²Department of Health and Nursing Sciences, Centre for e-Health, University of Agder, Grimstad, Norway

Abstract

Typically, routine prenatal care includes several in-person visits with healthcare professionals by pregnant women, where fetal and maternal assessments are performed. This paper proposes an architectural framework for prenatal care using non-invasive, simple, and low-cost internet of things (IoT) monitoring system. The aim is to design an IoT-based architecture that serves as a fundamental system for self-imperative care in regular pregnancy check-ups in the comfort of the home that offers routine prenatal screening tests. The system provides easy access to care regardless of the location and internet availability. We implemented preliminary architecture with simulated sensor data for blood pressure monitoring.

Keywords

care; consultation; internet of things; monitoring; prenatal; pregnant; virtual

1 INTRODUCTION

Maternal health is women's health condition during pregnancy, childbirth, and the postpartum period. Nearly 295,000 women died of pregnancy-related causes in 2017; 84% occurred in developing countries due to the lack of care and awareness [1]. The major causes of maternal mortality and morbidity are infection, hemorrhage, high blood pressure, and obstructed labor [2]. Moreover, hypertension is common in women during pregnancy, leading to multiple deaths worldwide. Chronic hypertension is another disorder, known as preeclampsia (that in severe cases may cause eclampsia), that occurs in approximately 3% to 5% of pregnant women in the developed world, causing severe injury or death [3]. Preeclampsia, in particular, has been found to lead to maternal and fetal mortality, intrauterine growth restriction, and preterm birth [4]. Prenatal care provided to pregnant women has reduced the rate of stillbirth, maternal mortality, and neonatal death [5-10]. Moreover, prenatal care includes several visits to HCPs by pregnant women for fetal and maternal assessment. A minimum of eight prenatal visits to HCPs is recommended by the World Health Organization (WHO) [11] for a healthy pregnant woman, and more visits for those having risks and complications [11]. Typically, physical examinations such as fetal heart rate monitoring, blood pressure monitoring, weight checks, fundal height measurements, and urine tests are performed during each routine prenatal visit, including nutrition and exercise planning [12]. Further, based on the stage of the pregnancy, some visits include additional physical examinations such as blood tests and ultrasound imaging to diagnose the risks that may affect the mother and the fetus (such as gestational diabetes and hypertensive disorders) [12]. As the increase in gestational age increases the risk of obstetric complications, the visits to HCPs are more frequent in higher gestational age [13].

Since early detection of pregnancy complications with regular checkups and monitoring can lead to a healthier birth, the development of information technology that is readily accessible and is able to provide up-to-date health status during the pregnancy is vital. Moreover, the advancements in technology have transformed healthcare delivery. Notably, the IoT exploited various sensing and communication infrastructures that offer data collection, transmission to the remote servers, and analysis and providing feedback, which is well-accepted in the healthcare sector [14]. Increased digitalization in the health care sector and the COVID-19 pandemic have underlined the need to implement accessible care for pregnant women, regarded as a sensible population [15].

Meanwhile, several techniques based on IoT have been developed to remotely monitor pregnant women's medical conditions to address these concerns. A secure IoT-based pregnancy monitoring system is proposed in [16], where three sensors - temperature sensor, memes sensor, and heartbeat sensor - were used to monitor pregnant women's health status. The system operates with a Wi-Fi module interface. The method in [17] remotely monitors the health of pregnant women and provides data visualization for their health care provider. The use of the internet allowed the visualization of sensor data by the healthcare professionals and permitted remote consultation with the pregnant women. A smartphone application is used in [18] that monitors and transmits heart rate sensor data, eating habits, and activity levels of pregnant women. In addition, it provides a medium to communicate remotely with healthcare professionals regarding pregnancy and health status. The approach in [19] presents a mobile monitoring system to identify hypertension using wearable sensors on a pregnant woman's body. It monitors changes in blood pressure and classifies the severity of hypertension using Naïve Bayes, which helps in the decision-making of health

The 18th Scandinavian Conference on Health informatics, Tromsø, Norway, August 22-24, 2022. Organized by UiT The Arctic University of Norway. Conference Proceedings published by Linköping University Electronic Press at <https://doi.org/10.3384/ecp187>. © The Author(s). This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>

care providers. Wireless Body Area Networks (WBANs) for pregnant women, based on the 802.15.4e TSCH standard, are presented in [20], which provides real-time remote monitoring of pregnant women centered on the biosensors' values. The monitoring of pregnant women's stress level, sleep, and physical activities during pregnancy based on IoT is shown in [21]. IoT-based system in [22] consists of different sensors such as sensors for non-invasive anemia and glucose rate detection, sudden fall detection, heart rate, and body temperature. The system is focused on the extensive health monitoring of pregnant women using Wi-Fi and GSM viable for telemonitoring. The system used in [23], monitors the uterine contraction to determine the possible risk of preterm delivery. The uterine contraction is detected using a wireless body sensor network, which triggers an alert message via a smartphone for values outside normal thresholds. A novel wearable system for fetal movement monitoring is presented in [24], aiming to save hospital resources by implementing monitoring at home. The movement data acquired using four accelerometers are processed and classified with a fuzzy classifier that is categorized into six different movement by which the fetal movement is inspected. However, the classifier was trained with data collected and labeled by pregnant women themselves, which was considered imprecise and highly subjective, degrading its performance.

Moreover, remote maternal monitoring has shown a comprehensive outcome and reflected numerous aspects of pregnancy care. However, conventional maternal monitoring systems are either too specific [19, 23, 24] or too general [16, 22]. They are especially focused on particular risk groups such as hypertension, preterm etc [19, 23, 24]. In addition, some are focused only on monitoring physical activities, behaviors, and habits [18, 21]. Typically, the internet is exploited for data transmission and communication, which may not be relevant for the pregnant population without internet access [16-24]. Thus, to address the health system and the limitations highlighted above, and considering the need underlined by the COVID-19 pandemic, we propose an architectural framework for home-based self-imperative care of pregnant women that facilitates IoT solutions regardless of location and internet access. The proposed system consists of two robust IoT solutions: internet-based and non-internet based. Either solution uses the IoT-based smart sensor network to monitor several physiological signals of pregnant women commonly done in routine prenatal visits and screening tests. The primary focus is to perform a general physical assessment (such as fetal heart rate monitoring, blood pressure monitoring, urine tests, and weight check), thereby reducing the frequency of in-person visits and providing easy access to care in the comfort of the home.

2 PROPOSED METHOD

The schematic diagram of the proposed framework is shown in Figure.1. The proposed system has two major adaptive aspects regardless of the location or internet access: internet based IoT solution and non-internet IoT solution. Internet-based IoT is operated by employing an

IoT gateway, while a long-range wide area networks (LoRaWAN) gateway is used for the non-internet IoT solution. These two approaches complement each other. To reduce the connectivity downtime, an IoT gateway could be used when there is internet connectivity while if there is no internet connectivity LoRaWAN could be utilized. Moreover, the overall architecture consists of four layers: sensor data acquisition, middleware gateway, data storage, and application layer. The sensor data acquisition layer consists of the sensor devices that acquire fetal heart tone, blood pressure, and weight values. Similarly, a middleware gateway is chosen based on internet availability; the LoRaWAN gateway is employed where there is no internet access at the endpoint; however, the IoT gateway is utilized in internet-accessible areas. The layer allows the transmission of data acquired from the sensor devices to a central database where the data are stored. The software solution to retrieve data from sensor devices and transmit it to the central database is included in this layer. The acquired sensor data are packaged and sent to the central database using the message queuing telemetry transport (MQTT) protocol. The transmitted data are then stored in the central database, which is later analyzed and displayed on both pregnant woman and healthcare professionals' end. Finally, the application layer consists of the end-user application for pregnant women and healthcare professionals that allows visualization of the sensor data acquired from the first layer.

2.1 Sensor Data Acquisition

The proposed architecture focuses on the self-imperative care of pregnant women for routine prenatal monitoring in the comfort of the home. Typically, the regular prenatal visits include blood pressure measurements, fetal heart tone assessments, weight checks, fundal height measurements, and urine tests; thus, their corresponding sensor devices are utilized for routine checkups to determine the potential risks and complications. The non-invasive blood pressure sensor device is used to measure pregnant women's blood pressure that detects the risk associated with hypertension and preeclampsia. The systolic and diastolic values of the blood pressure are measured by utilizing the oscillometric method, which can also measure pulse rate. Similarly, the most used Doppler device for fetal heart rate monitoring is used to assess the heart rate, rhythm, and regularity of the fetal heart tones. The pregnant woman's weight and fundal height are then self-measured, noted, and updated in the application where the healthcare professional is able to review the data. Along with blood pressure measurement, urinalysis for protein, which is specially done to diagnose preeclampsia, is performed, an examination commonly part of each prenatal screening test [25]. For this, commercially available urinalysis reagent strips that gives a rapid result are used in the proposed architecture [25]. Such results are examined with visual inspection or automated readers, and further testing is undertaken physically during in-person visits if it gives a positive outcome.

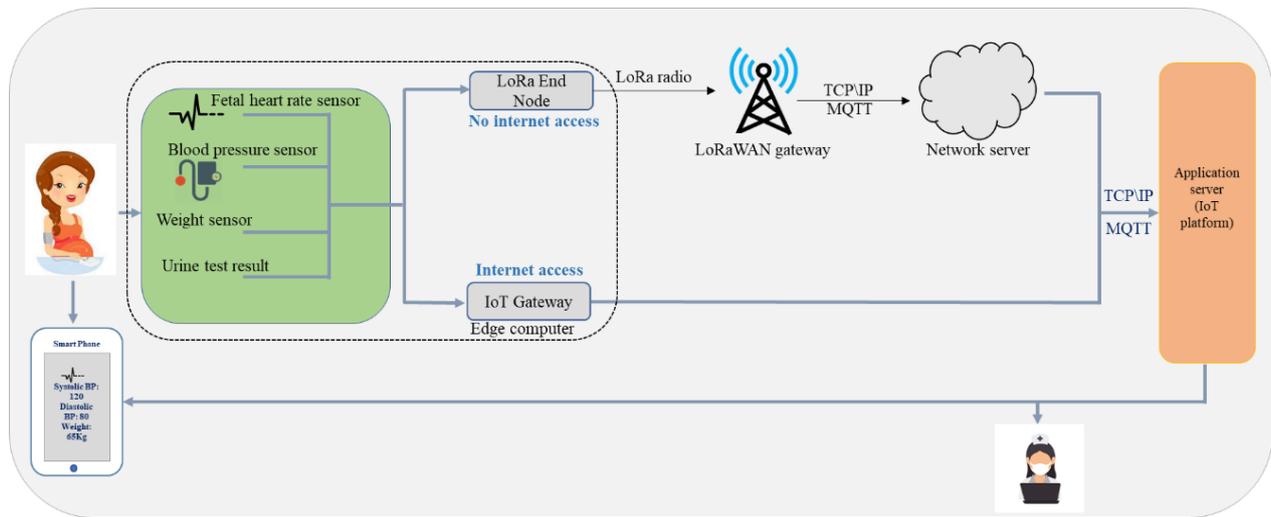


Figure 1. Schematic diagram of the proposed architecture.

2.2 Middleware Gateway

Middleware gateway layer comprises IoT gateways. The gateway on the internet-based approach consists of an edge computer wherein an IoT gateway software is installed. The IoT gateway software is responsible for converting the sensor data into a MQTT format and forwarding it to the remote server. In contrast, in non-internet-based approach, the LoRaWAN gateway receives sensor data employing LoRa radio and forwards it to the network server [26]. The gateway on the internet-based solution is limited to connecting a single pregnant woman to the application server. In contrast, star topology is followed in LoRaWAN (multiple pregnant women can be connected to a single gateway). In addition, the long-range property of LoRa radio helps connect the end nodes (pregnant women) that are deployed in remote areas (inaccessible areas that do not have internet service and are several kilometers from the gateway).

2.3 Data Storage

This layer is a remote database server that stores sensor data. The database server is accessible from the application server for further data analysis.

2.4 Application Layer

This layer is an IoT platform that provides feedback on potential risk factors or notification of risks to health care professionals and pregnant women based on the analysis of the monitored data. For example, an email or SMS can be sent to a concerned party as an alarm or alert using the IoT platform (e.g., Thingsboard [27]). Thingsboard is an open-source IoT platform. In addition, the end-user application that shows the result on the user's end (pregnant women and healthcare professionals) is also included in this layer. The end-user application provides a summary of their health status to the pregnant women and assistance to healthcare professionals to examine this status remotely.

3 RESULTS

To illustrate preliminary results, we realized the internet-based approach using simulated sensor data, as shown in the Figure 2.

A program was written to simulate Photoplethysmography (PPG) signals using different heart rates. PPG is a simple optical technique used to detect volumetric blood changes in peripheral circulation. As the latest research suggests that the PPG signal can be utilized to estimate blood pressure [28], we simulated PPG signals to realize blood pressure. The simulation software was installed on an edge computer (Raspberry Pi 3) that published MQTT packets (simulated PPG signals) every 10 seconds. Moreover, an IoT gateway software from the Thingsboard was installed on the same device, which subscribed to the simulated PPG signal and forwarded them to the Thingsboard cloud. A dashboard was created in the Thingsboard cloud to illustrate the simulated PPG signals as time series telemetry.

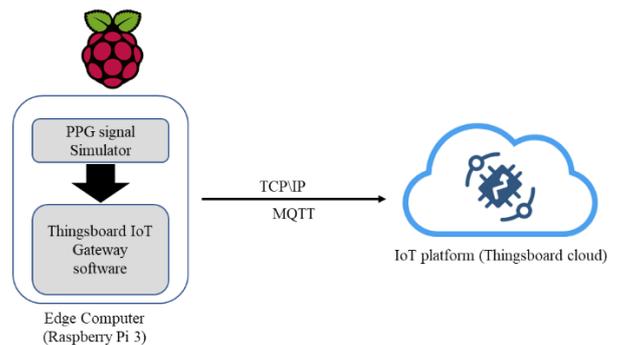


Figure 2. Architecture for the internet-based approach using simulated sensor data.

The result of the simulated PPG signal is shown in Figure 3. As shown in Figure 3., orange and green color lines depict the PPG signals using heart rates of 100 and 70, respectively. Based on the signal value, an alert message (email or SMS) can be sent from the Thingsboard cloud to the pregnant woman and healthcare professionals when the PPG signal is outside the threshold range.

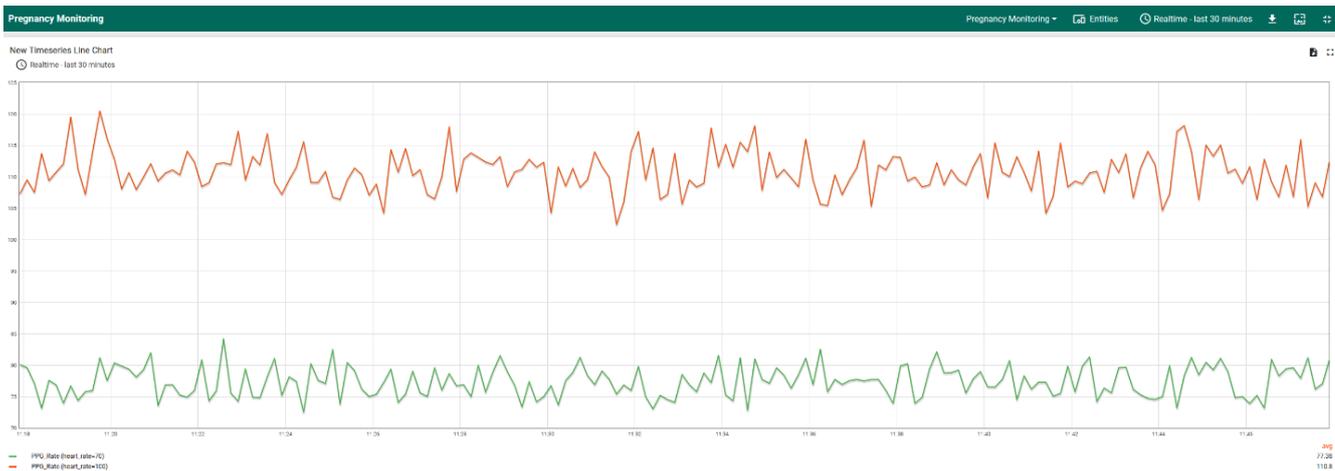


Figure 3. Simulation result of internet-based architecture for PPG signal.

4 CONCLUSION

We propose an IOT-based architectural framework for self-imperative care and monitoring of pregnant women in the comfort of the home. The proposed system provides easy access to care regardless of the location and the internet availability based on two different IoT solutions: internet-based and non-internet based. The IoT solution is based on a smart sensor network that performs physical assessments of pregnant women that are commonly carried out during routine prenatal visits and screening tests. In-home monitoring, analogous to the regular checkup in the proposed system, reduces the frequency of in-person visits to the hospital, requiring in-person visits only for further diagnosis and physical examination that require the HCP's supervision. Moreover, the system helps in the early

detection of risks, which enables adequate care and on time treatment for pregnant women. Besides, the proposed architecture prototype could reduce the gap between expectant mothers and medical specialists/resources and has a potential use in various e-Health home care applications. Also, the implementation of the proposed research work would help society move towards healthier lives and well-being among pregnant women and reduce their mortality rate. In future work, we will consider the real-time implementation of the complete architecture of the proposed framework using a real sensor, including evaluation and analysis of its performance and results. In addition, performance-wise comparison between two approaches will be performed.

REFERENCES

- [1] "Maternal mortality." <https://www.who.int/news-room/fact-sheets/detail/maternal-mortality> (accessed April 8, 2022, 2022).
- [2] "WHO | Maternal and reproductive health." [Online]. Available: <https://www.who.int/data/gho/data/themes/maternal-and-reproductive-health>.
- [3] L. Benschop, J. J. Duvekot, J. Versmissen, V. van Broekhoven, E. A. Steegers, and J. E. Roeters van Lenep, "Blood pressure profile 1 year after severe preeclampsia," *Hypertension*, vol. 71, no. 3, pp. 491-498, 2018.
- [4] F. M. Musyoka, M. M. Thiga, and G. M. Muketha, "A 24-hour ambulatory blood pressure monitoring system for preeclampsia management in antenatal care," *Informatics in Medicine Unlocked*, vol. 16, p. 100199, 2019.
- [5] N. Regassa, "Antenatal and postnatal care service utilization in southern Ethiopia: a population-based study," (in English), *Afr Health Sci*, vol. 11, no. 3, pp. 390-397, 2011. [Online]. Available: <Go to ISI>://WOS:000306171400015.
- [6] X. J. K. KD, M. SL, and T.-V. B, "Deaths: final data for 2007," 2010.
- [7] S. Partridge, J. Balayla, C. A. Holcroft, and H. A. Abenheim, "Inadequate Prenatal Care Utilization and Risks of Infant Mortality and Poor Birth Outcome: A Retrospective Analysis of 28,729,765 U.S. Deliveries over 8 Years," (in English), *American Journal of Perinatology*, vol. 29, no. 10, pp. 787-793, Nov 2012, doi: 10.1055/s-0032-1316439.
- [8] J. A. Ozimek and S. J. Kilpatrick, "Maternal Mortality in the Twenty-First Century," (in English), *Obstetrics and Gynecology Clinics of North America*, vol. 45, no. 2, pp. 175-+, Jun 2018, doi: 10.1016/j.ogc.2018.01.004.
- [9] C. J. Berg, W. M. Callaghan, C. Syverson, and Z. Henderson, "Pregnancy-Related Mortality in the United States, 1998 to 2005," (in English), *Obstetrics and Gynecology*, vol. 116, no. 6, pp. 1302-1309, Dec 2010, doi: 10.1097/AOG.0b013e3181fdff11.
- [10] M. Mrisho *et al.*, "The use of antenatal and postnatal care: perspectives and experiences of women and health care providers in rural southern Tanzania," (in English), *Bmc Pregnancy and Childbirth*, vol. 9, 2009, doi: Artn 1010.1186/1471-2393-9-10.
- [11] WHO. "ANC Guideline" https://www.who.int/reproductivehealth/news/ANC_guideline-presentation.pptx (accessed).
- [12] N. I. o. C. H. a. H. Development. "What is prenatal care and why is it important? ."

- <https://www.nichd.nih.gov/health/topics/pregnancy/conditioninfo/prenatal-care> (accessed September 15, 2020).
- [13] B. M. M. Pflugeisen, MEd, C. M. McCarren, RN, S. M. Poore, M. A. Carlile, and R. M. Schroeder, "Virtual Visits: Managing prenatal care with modern technology," *MCN: The American Journal of Maternal/Child Nursing*, pp. 24-30, 2016.
- [14] L. Ungurean and A. Brezilianu, "An internet of things framework for remote monitoring of the healthcare parameters," *Advances in Electrical and Computer Engineering*, vol. 17, no. 2, pp. 11-16, 2017.
- [15] G. Weigel, "Novel coronavirus "COVID-19": Special considerations for pregnant women," *Retrieved May*, vol. 10, p. 2020, 2020.
- [16] S. Gayathri, T. Bharathi, A. DevleenaJerusha, and A. A. Kumar, "Pregnant Women Health Care Monitoring System Based on IoT," *International Journal of Engineering Technology and Management Science, ijetms. in Issue: 1*, no. 2, 2018.
- [17] S. Veena and D. J. Aravindhar, "Remote Monitoring System for the Detection of Prenatal Risk in a Pregnant Woman," *Wireless Personal Communications*, vol. 119, no. 2, pp. 1051-1064, 2021.
- [18] G. K. Endo, I. Oluwayomi, V. Alexandra, Y. Athavale, and S. Krishnan, "Technology for continuous long-term monitoring of pregnant women for safe childbirth," in *2017 IEEE Canada International Humanitarian Technology Conference (IHTC)*, 2017: IEEE, pp. 6-10.
- [19] M. W. Moreira, J. J. Rodrigues, A. M. Oliveira, and K. Saleem, "Smart mobile system for pregnancy care using body sensors," in *2016 International Conference on Selected Topics in Mobile & Wireless Networking (MoWNeT)*, 2016: IEEE, pp. 1-4.
- [20] A. Sene, I. Niang, A. Diop, and A. Gueye, "T4PW: TSCH in WBAN for Real-Time Monitoring of Pregnant Women," in *International Conference on Industrial Networks and Intelligent Systems*, 2021: Springer, pp. 3-15.
- [21] F. Sarhaddi *et al.*, "Long-Term IoT-Based Maternal Monitoring: System Design and Evaluation," *Sensors*, vol. 21, no. 7, p. 2281, 2021.
- [22] T. G. Troyee, M. K. Raihan, and M. S. Arefin, "Health Monitoring of Expecting Mothers using Multiple Sensor Approach:"Preg Care"," in *2020 2nd International Conference on Advanced Information and Communication Technology (ICAICT)*, 2020: IEEE, pp. 77-82.
- [23] H. Allahem and S. Sampalli, "Framework to monitor pregnant women with a high risk of premature labour using sensor networks," in *2017 IFIP/IEEE Symposium on Integrated Network and Service Management (IM)*, 2017: IEEE, pp. 1178-1181.
- [24] X. Zhao, X. Zeng, L. Koehl, G. Tartare, and J. De Jonckheere, "A wearable system for in-home and long-term assessment of fetal movement," *IRBM*, vol. 41, no. 4, pp. 205-211, 2020.
- [25] K. L. Tucker *et al.*, "The feasibility and acceptability of self-testing for proteinuria during pregnancy: a mixed methods approach," *Pregnancy Hypertension*, vol. 12, pp. 161-168, 2018.
- [26] A. Farhad, D.-H. Kim, S. Subedi, and J.-Y. Pyun, "Enhanced lorawan adaptive data rate for mobile internet of things devices," *Sensors*, vol. 20, no. 22, p. 6466, 2020.
- [27] "Thingsboard IoT platform." <https://thingsboard.io/> (accessed 05-04-2022, 2022).
- [28] G. Slapničar, M. Luštrek, and M. Marinko, "Continuous blood pressure estimation from PPG signal," *Informatica*, vol. 42, no. 1, 2018.

Towards Mapping of Information Technology-Induced Alterations in Online Physicians' Professional Identities

A Conceptual Framework and Empirical Illustrations from Sweden

Michal Dolezel¹ and Åsa Cajander²

¹Prague University of Economics and Business, Department of Information Technologies, Czech Republic
michal.dolezel@vse.cz

²Uppsala University, Department of Information Technology, Sweden

Abstract

Digital Patient Contact Technologies (DPCT), including telemedicine solutions and digital tools for text-based communication between patients and physicians, play a significant role in today's healthcare. Professional identity defines norms, principles, and logic that guide one's professional actions. Presently, little research is available regarding professional identity changes in the context of DPCT implementations. This work theoretically and empirically illustrates the nature of the possible DPCT' impact on physicians' professional identities. To this end, a conceptual framework was constructed, and the interviews with eight physicians operating an asynchronous healthcare-advice chat service (1177 Vårdguiden) in Uppsala, Sweden, were examined.

Keywords

Telemedicine; eHealth; doctor-patient relationship; professional identity

1 INTRODUCTION

Due to the new capabilities of information technology (IT), one can witness a significant change in how healthcare services are delivered today [1]. The technological means for enabling digital contact with patients have been of foremost interest to practitioners and researchers in recent years [2], [3]. The pivotal role of these solutions was further strengthened by the COVID-19 pandemic crisis [4].

Examples of digital contact technologies include full-fledged telemedicine solutions transmitting audio, video, and biomedical data [2], [5]; controlled use of public video calling platforms (such as Skype, Facetime, or Zoom) [6]; specialized mHealth applications [5]; asynchronous advice-giving using specialized web services and discussion fora [7]. More recently, synchronous online chat services employing embodied conversational agents [8] have been introduced as an additional communication tool to enable digital contact between a patient and healthcare professionals, typically in the context of healthcare guidance [3], [9]. In this paper, we introduce the term Digital Patient Contact Technology (DPCT) by putting all such technologies under one conceptual umbrella. Being technologically agnostic, the concept of DPCT is to label an *aggregate of information and communication technology components enabling the delivery of health-related care over distance (distance healthcare)*.

Naturally, implementing DPCT may bring positives for some stakeholders and perceived negatives for others. This

may result in prospective DPCT users' resistance [10]. Speculatively, a number of aspects of the existing medical work will get significantly impacted, given the role that DPCT is to play in medicine in the near future [1], [11]. Overall, it becomes clear that "[t]raditional concepts in medical ethics, confidentiality, empathy, empowerment/power, efficiency and mutual responsibilities are reframed in the context of digital consulting" [4] or DPCT. One common point of criticism is the negative impact of DPCT on the patient-physician relationship [12], [13]. Another essential factor to consider is the possible change in job content and work routines [4], responsibilities and decision-making strategies [3], and the identities of healthcare professionals [14]. The latter aspect is the primary concern of the present work.

Identity research is broad and dispersed. Essential contributions to understanding professional identities in the medical domain have been made by researchers in the health sciences (e.g. [15]) as well as other fields (e.g. [16]). This paper builds upon the understanding of identity established within medical education [17]. In simple words, *identity* (and more specifically, professional identity) defines *who we are (as professionals)* (e.g. [18]), *what underlying norms, principles, and logic guide the courses of our (professional) actions* (e.g. [19]), and *what (professional) values we subscribe to* (e.g. [20]).

Presently, little research evidence is available specifically on the topic of relationships between patients and

The 18th Scandinavian Conference on Health informatics, Tromsø, Norway, August 22-24, 2022. Organized by UiT The Arctic University of Norway. Conference Proceedings published by Linköping University Electronic Press at <https://doi.org/10.3384/ecp187>. © The Author(s). This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>

physicians and of professional identity changes *when facing DPCT implementations*. Importantly, many in the medical community seem to hold strong opinion about these issues [12], [21]. Also, while certain evidence is available for some classes of the DPCT (e.g. telemedicine solutions) or related areas [22], the area of text-based communication and advice-giving through chat or on-line forums remains largely uncovered.

This research, carried out within the social strand of medical informatics [23], [24], fills the theoretical and empirical knowledge gap. This paper aims to theoretically and empirically illustrate the nature of the possible DPCT impact on the physicians' identities. To this end, a conceptual framework was constructed. In addition, qualitative data from a research project concerned with the influence of DPCT on the work routines of physicians [3] were analysed, providing some clues to direct further research in the domain of DPCT's impact on professional identity.

2 THEORETICAL BACKGROUND: SOCIAL AND PROFESSIONAL IDENTITY

People rely on their identities, serving as anchors in their work and private lives. As humans are social creatures, who need to cooperate to survive and thrive, an important part of their identity is rooted in the shared norms and values of their social groups.

This means that their self-view is not determined by personality characteristics or by their interpersonal relationships alone; how people see themselves is also determined by the groups to which they belong (that is, their social identity; [25]). [14]

Social identity is a broad concept that has attracted the significant attention of scholars from several different fields. The long-existing interest has resulted in an extensive body of knowledge which includes psychology (e.g. [26]), sociology (e.g. [27]), and the organizational sciences (e.g. [16]). The seminal foundations for this work were laid out during 1970s by Tajfel [25], a social psychologist, who defined social identity as "the individual's knowledge that he belongs to a social group together with some emotional and value significance to him of this group membership."

More recently, social identity theories found their way into the health sciences domain, where they have been employed in medical education research [28], [29]. This orientation reflects the fact that social identities are formed and developed during one's professional education, and the extensiveness of medical education provides a significant space to do so [14].

Medical education research introduced the concept of professional identity as a specific instance of social identity [28]. Professional identity "describes how we perceive ourselves within our occupational context and how we communicate this to others" [18]. More intriguingly, professional identity can be viewed as a set of "professional values, actions, and aspirations" [14] one builds upon when conducting professional duties as a member of certain

profession. However, it is important to remind that "[p]rofessional identity is not a stable entity; it is complex, personal, and shaped by contextual factors".

The pioneering study of Becker et al. [30] emphasized the process of forming professional norms and values during the medical education. Considering today's world of medical specialization, however, it is important to note that different specializations may incline to distinct sets of values. This fact was exemplified in the study of Intensive Care Unit organizational arrangements, which rendered the differences in values and beliefs between intensivists and surgeons [31]. Some medical specializations, such as surgery, attribute a great importance in identity formation processes to the factors such as perfectionism and technical skills [32]. By contrast, an important identity aspect of the healthcare provided by general practitioners (called also family physicians in some countries [33]) is "seeing the patient as a whole" [16] or being a "holistic physician" [33].

While providing important cues, the literature reviewed so far allow for a broad interpretation of what constitutes an essence of social/professional identity in the field of medicine and how to conceptualize its key dimensions. A more tangible conceptualization is needed with respect to the purpose of the present work. Introducing a pragmatic analytic framework, Hendriks [34] argued that physicians' professional identity can thought of in terms of three key components: excellence (e.g. providing state-of-art medical care), ethics (e.g. doing no harm to patients) and engagement (i.e. commitment to patients and the profession).

Stemming from a theory-focused research, another three dimensions have been popular with social psychology scholars interested in social identity work. Originating in the 1980s, these include: awareness of group membership; group interdependence; and emotional aspects of belonging to the group [35]. More recently, Cameron [36] argued for extending the three key components derived from the original work of Tajfel [25] to characterize the essence of professional identity better. Proposing and validating a new three-factor's model [36], Cameron posits identity dimensions of similar conceptual significance:

- *Cognitive centrality* is the "subjective importance of the group to self-definition", manifesting in the "frequency with which the group comes to mind" [36]. Put differently, cognitive centrality be used to quantify how intense is one's awareness of group identity.
- *In-group affect* captures an emotional component of the group identity, providing a conceptual tool to analyse subjective feelings about one's membership in the given group. These can be both positive and negative [37].
- *In-group ties* is a dimension used to describe a sense of connectedness or belonging with the group [36]. It may therefore concern both verbal and non-verbal communication acts between the members, but also perceptions "that one 'fits in,' 'has strong ties,' or 'shares a common bond' with the group and its members" [37].

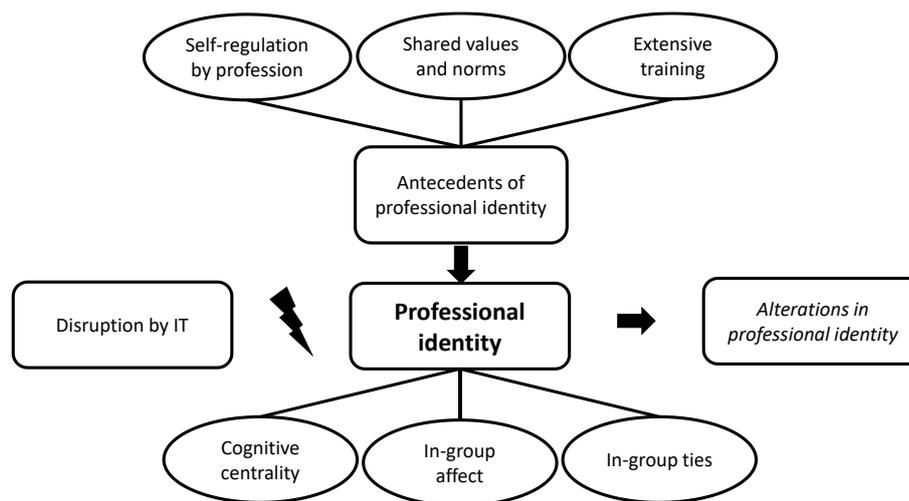


Figure 1. The sensitising framework: Disruption of professional identity caused by information technology (synthesised from [14] and [36]).

In this preliminary work on DPCT's influence on physicians' professional identity, we constructed a conceptual framework to shed better light on these problems. As our point of departure for framework construction, we took the theory development work of Molleman and Rink [14] elaborated within the domain of social science applied to health. The work was specifically focused on professional identity alterations in physicians. In addition, we employed Cameron's three factor model of social identification [36] from the domain of social psychology, due to being a competent conceptual tool suitable for qualitative identity transition work related to medical professionals [37]. It is also compatible with Tajfel's ideas, well-established in medical education research [38].

Blending both these sources of theoretical notions, we present our resulting theoretical framework in Figure 1. The framework has two principal parts. The top part illustrates the major antecedents of professional identity, in-line with the line of thought put forward by Molleman and Rink [14]. The bottom part decomposes professional identity into three principal dimensions adopted from Cameron [36] and described above.

One additional point remains to be clarified. The concept of professional identity is different from that of *professionalism* [38]. In short, the latter describes a set of desired or exemplified behaviours, which can be directly observed. Inversely, professional identity is much less pronounced, hidden from direct observations. "Professional identity is how an individual conceives of him- or herself as a doctor ..." [38]. In that sense, some authors relate it to the concept of professional culture, from which it stems [39]. Importantly, significant role in identity formation processes, notably in the medical domain, is ascribed to role modelling [40], i.e. following and internalizing another individuals' *modi-operandi* in a given area. In that sense, it appears quite problematic to think of professional identity as of a homogenous cultural entity, as

individual role models and mentors are of great significance in one's professional identity formation when becoming doctor [37].

3 METHOD

The present paper presents an illustrative theory-driven analysis, serving as a sensitising framework for our subsequent research in the area of DPCT. "Unlike a rigorous theoretical framework that is established in advance, ... [the] sensitising concepts are elastic and open to revision because of their general, empirically not contentful nature (e.g. 'culture', 'institutions', 'structure', 'roles')" [41].

The data used in this paper were collected within a research initiative mapping the impact of a DPCT in the context of primary care. The goal of the research was defined broadly: to understand the changes in the work regimes of healthcare professionals caused by DPCT [3].

3.1 Context

In Sweden, the healthcare system is decentralized, so that many nuances exist across the 21 counties [42]. Stated broadly, primary care is provided by general practice physicians working in primary care centres [43]. Patients have access to a first contact service, and a triage service titled "1177 Vårdguiden". As explained on the service website of the Uppsala region,

"1177 is manned by licensed nurses who will respond to calls to give advice, consult on the need for potential further care and give guidance to the appropriate healthcare clinic when needed" [44].

Starting in 2019, a pilot project implementing asynchronous chat service was carried out in Region Uppsala, as described in detail in [3].

3.2 Data collection

We carried out eight semi-structured interviews with the

physicians. The interviews took place at 1177's premises, i.e. in offices or a meeting room. These interviews, each of ca. 1-hour duration, were conducted by the second author in Swedish, audio-recorder and transcribed verbatim by a professional transcription service. Four participants were males and four females. Their age range was 26-49 years. Two specialists and six residents participated in the interviews. The interview guide contained broad questions such as "What do you think are the advantages of working with the chat/on the telephone?", which primarily targeted the professional work routines or the material world rather than subjective feelings about one's professional identity per se. The interview guide is available as an online appendix of the original study [3].

3.3 Data analysis

Following the principles of deductive qualitative analysis informed by Cameron's three-factor model of social identification, the first author examined the qualitative data. As the main aim of this exercise was to explore the feasibility of applying the mentioned theoretical framework in the context of DPCT initiatives, a pragmatic analytical approach was used. After employing constant comparison while iterating between the theory and data, corresponding vignettes were chosen for the present empirical illustration. The fit between the selected data and the theoretical framework was then discussed with the second author, who had designed the original study, until a consensus regarding interpretation of the data in light of the theory was reached.

4 RESULTS: ILLUSTRATIVE VIGNETTES

When presenting the results, we follow the logic of the Cameron's three-factor model of social identification. In that sense, we consider the professional identity of online doctors being an altered form of professional identity of medical generalists working in a front-line setting with patients face-to-face as this was the professional background of many of our participants, as illustrated below.

... I am a trained physician, who specialises in general practice at this health centre. (Physician 4)

4.1 Cognitive Centrality: Awareness of Necessity Going Online

For some of our participants, the reasoning behind moving to the on-line word was associated with a perceived necessity. The concrete motives included both quality-of-care concerns (Vignette 1) and political decisions (Vignette 2).

Vignette 1: I think the big concern for healthcare is that we cannot have the healthcare system we have today in 30 years. ... I mean society is different today than it was in the 90's. It will be different in 10 years, it will be different in 5 years. And then we have to change certain things. ... And need to work more efficiently without being drained of quality. And then you cannot have the attitude ... [of being] extremely sceptical. Eh... and I'm trying to be somewhere in between there I think, eh... in my attitude. (Physician 4)

Vignette 2: I see more digital care will come due to strong political will, and then I think it is better to sit in the cab and steer the bus a little closer than seeing it drive straight into the wall. (Physician 5)

4.2 In-group Affect: Feelings about Online Work

There was a significant emotional component contained in the analysed qualitative data. This fact is illustrated below by two vignettes, focusing on the physicians' concerns about practicing digital medicine. Illustratively, we thematise these vignettes as the *Fear of losing one's professional self* (Vignette 3) and the *Fear of losing the doctor's "sixth sense"* (Vignette 4).

*Vignette 3: Uh ... but at the same time if you just sat and chatted, it would ... you don't feel quite as much like a doctor maybe. You know, have a stethoscope around your neck and listen to a heart. ... Maybe just a little bit in the attributes. I'm a bit superficial so I think if you have white scrubs **then you are a doctor**. ... And we do not do that here. Both the nurses and the doctors are civilian-dressed. (Physician 6)*

Vignette 4: ... they could make incorrect assessments because they have not been able to perceive that difference, which could determine whether they decide to send a patient to the emergency room ... That it's such small things that you can actually react to. How someone expresses themselves, how someone behaves in the room. Yes, and [you can do] this just with the physical examination. (Physician 4)

4.3 In-group Ties: Diagnosing Online

The extent of the shift in the "sense of belonging" component of professional identity can be illustrated by the following vignette.

*Vignette 5: ... the online medical companies are saying that 60-70% of health centre cases could be dealt with via chat. Or not chat, but via video anyway. So that and that's probably what **divides us doctors** a little bit, how much you can actually assess [through DPCT]... (Physician 4)*

5 DISCUSSION

Our work provides some preliminary insights into the problem of professional identity alterations in reaction to DPCT introduction. We begin our discussion by highlighting the aspects of *cognitive centrality* of the identity-in-transition. Recognizing the fact that there is a need for change in the existing healthcare provision schemes, our participants internalized the notion that their work (and professional identity) will be different than it used to be. Arguably, they have accepted that this is a long-term effort, as one can speculate about the metaphorical language exemplified by Vignette 2. Still, our participants opted to actively participate in the change instead of passively standing aside. That said, the motives for participation in this particular DPCT initiative might have highly varied. For example, as our data indicate, a financial aspect might have played a role here too.

In Section 2, we highlighted the centrality of role modelling

for the processes of professional identity formation in medical education [37]. Based on that, we speculate that engaging the pilot project participants (or “early adopters” [45] of the altered professional identity) in expertise sharing activities will be of foremost importance for the future success of the DPCT initiative when implemented nation-wide. Opinion leaders acting as change agents play an essential role in adopting new technology in traditional, well-established communities [46].

We continue by discussing the importance of *in-group affect* components of professional identity. Notably, we point to the fact that professional identity is associated with values, norms and behaviours, and with material symbols, especially with the symbols serving to assert one’s status [19]. These symbols may provoke emotional reactions. This particular notion was illustrated by the point brought in Vignette 3, referring to not feeling as a doctor due to not wearing a stethoscope and scrubs [47]. Previous research in the social science domain demonstrates that white coats and scrubs are examples of material artefacts that may shape one’s sense of professional identity of being a physician [48].

Although it is impossible to draw any firm conclusions based on our illustrative data, we note that the wide spread of telemedicine and other DPCT has fundamentally changed some of the medical field’s material artifacts and behavioural norms. Not all physicians practice medicine dressed up uniformly [48], which is a significant change introduced by DPCT. In general, patients make sense of the medical world by following well-known clues of medical status, prestige and expertise [49], being guided by a customary image of the archetypical physician [47], [50]. Perhaps as a result of this, telemedicine physicians are nowadays encouraged to “dress as they would when they see a patient in person, such as a white coat, business casual clothes, or scrubs” [51]. Naturally, this makes no pragmatic sense in the domain of asynchronous chat services not implementing any video communication, as was in our case.

Finally, the remaining aspect conceptualized by the theoretical framework is *in-group ties*, i.e., sense of belonging to the group of online physicians. Illustratively, Vignette 5 highlights a role of digital divide between online physicians and proponents of medicine practiced in the “traditional way”. Evidence suggest that many concerns from the provider perspective persist even in highly-digitalized countries such as Sweden [52], where the present pilot DPCT initiative was carried out. The physicians’ concerns include doubts about DPCT safety for patients and claims about “increased risk of misunderstandings” [52]. In addition, DPCT may be seen as “stressful time-thief[s]” and representing threats to the medical profession and physicians’ own integrity [52].

Prior concluding, the limitations of this work need to be highlighted. First and foremost, this paper does not present a full-fledged empirical study. Instead, it offered a theory-building exercise supplemented with qualitative evidence of illustrative nature. Another limitation of the present paper is the early stage of development of our theoretical

concepts in terms of possible professional identity collisions, overlaps and multiple membership. By this we mean that physicians can take part in multiple social groups, including various professional groups formed by work settings and specializations. For example, if a physician worked in a front-line setting in addition to his/her DPCT engagement, his/her identity could not be considered entirely “online”.

Also, medical professionals coming from different specializations than general medicine might bring their unique professional identities, which may differ from each other quite significantly [14]. We plan to mitigate these limitations when designing our subsequent qualitative enquiry probing into the implementation of DPCT providing health guidance and consultations. Within the scope of that initiative, a similar chat service as implemented in this pilot project in the Uppsala Region is planned to be implemented Sweden-wide. That offers a possibility of executing a larger scale study exploring the professional identity component as one of its central concepts. We hope that our study will contribute to the global debate on prospects of general practice (family medicine) in today’s, increasingly digitalized world [11]. Perhaps, DPCT can help with solving the present “identity crisis” dilemma perceived by some in the general practice discipline [33].

6 CONCLUSION

The seeds of qualitative evidence presented in this paper illustrate that DPCT might have a profound impact on the existing identity of healthcare professionals. The professional identity of physicians significantly influence how they carry out their daily work. Deciphering the underlying social and cognitive processes stemming from introducing innovative DPCT into healthcare organizations is pivotal in designing and carrying out more effective, people-friendly IT change efforts in various medical organizations.

7 REFERENCES

- [1] S. Kraus, F. Schiavone, A. Pluzhnikova, and A. C. Invernizzi, “Digital transformation in healthcare: Analyzing the current state-of-research,” *J. Bus. Res.*, vol. 123, pp. 557–567, 2021, doi: 10.1016/j.jbusres.2020.10.030.
- [2] M. T. Svendsen, S. N. Tiedemann, and K. E. Andersen, “Pros and cons of eHealth: A systematic review of the literature and observations in Denmark,” *SAGE Open Med.*, vol. 9, pp. 1–10, 2021, doi: 10.1177/20503121211016179.
- [3] Å. Cajander, G. Hedström, S. Leijon, and M. Larusdottir, “Professional decision making with digitalisation of patient contacts in a medical advice setting: A qualitative study of a pilot project with a chat programme in Sweden,” *BMJ Open*, vol. 11, no. 12, pp. 1–6, 2021, doi: 10.1136/bmjopen-2021-054103.
- [4] J. Sturt *et al.*, “How does the use of digital consulting change the meaning of being a patient and/or a health professional? Lessons from the

- Long-term Conditions Young People Networked Communication study,” *Digit. Heal.*, vol. 6, pp. 4–13, 2020, doi: 10.1177/2055207620942359.
- [5] S. Khairat, S. Liu, T. Zaman, B. Edson, and R. Gianforcaro, “Factors determining patients’ choice between mobile health and telemedicine: Predictive analytics assessment,” *JMIR mHealth uHealth*, vol. 7, no. 6, p. e13772, 2019, doi: 10.2196/13772.
- [6] E. Richardson, D. Aissat, G. Williams, and N. Fahy, “Keeping what works: Remote consultations during the COVID-19 pandemic,” *Eurohealth (Lond.)*, vol. 26, no. 2, pp. 73–76, 2020.
- [7] J. Jiang, M. Yang, M. Kiang, and A. F. Cameron, “Exploring the freemium business model for online medical consultation services in China,” *Inf. Process. Manag.*, vol. 58, no. 3, pp. 1–24, 2021, doi: 10.1016/j.ipm.2021.102515.
- [8] S. ter Stal, L. L. Kramer, M. Tabak, H. op den Akker, and H. Hermens, “Design Features of Embodied Conversational Agents in eHealth: a Literature Review,” *Int. J. Hum. Comput. Stud.*, vol. 138, pp. 1–22, 2020, doi: 10.1016/j.ijhcs.2020.102409.
- [9] F. Gabrielsson-Järhult, S. Kjellström, and K. A. Josefsson, “Telemedicine consultations with physicians in Swedish primary care: a mixed methods study of users’ experiences and care patterns,” *Scand. J. Prim. Health Care*, vol. 39, no. 2, pp. 204–213, 2021, doi: 10.1080/02813432.2021.1913904.
- [10] P. Klöcker, R. Bernnat, and D. Veit, “Implementation through force or measure? How institutional pressures shape national eHealth implementation programs,” in *European Conference on Information Systems*, 2014, pp. 1–16.
- [11] R. Hays and T. Sen Gupta, “Developing a general practice workforce for the future,” *Aust. J. Gen. Pract.*, vol. 47, no. 8, pp. 502–505, 2018, doi: 10.31128/AJGP-02-18-4488.
- [12] C. Stokel-Walker, “Why telemedicine diminishes the doctor-patient relationship,” *The BMJ*, vol. 371. BMJ Publishing Group, p. m4348, 2020, doi: 10.1136/bmj.m3603.
- [13] C. Grönloh, G. Myreteg, A. Cajander, and H. Rexhepi, ““Why Do They Need to Check Me?” Patient Participation Through eHealth and the Doctor-Patient Relationship: Qualitative Study,” *J. Med. Internet Res.*, vol. 20, no. 1, p. e11, 2018, doi: 10.2196/jmir.8444.
- [14] E. Molleman and F. Rink, “The antecedents and consequences of a strong professional identity among medical specialists,” *Soc. Theory Heal.*, vol. 13, no. 1, pp. 46–61, 2015, doi: 10.1057/sth.2014.16.
- [15] A. P. Sawatsky *et al.*, “Autonomy and professional identity formation in residency training: A qualitative study,” *Med. Educ.*, vol. 54, no. 7, pp. 616–627, 2020, doi: 10.1111/medu.14073.
- [16] Y. Kyratsis, R. Atun, N. Phillips, P. Tracey, and G. George, “Health systems in transition: Professional identity work in the context of shifting institutional logics,” *Acad. Manag. J.*, vol. 60, no. 2, pp. 610–641, 2017, doi: 10.5465/amj.2013.0684.
- [17] B. Burford, “Group processes in medical education: Learning from social identity theory,” *Med. Educ.*, vol. 46, no. 2, pp. 143–152, 2012, doi: 10.1111/j.1365-2923.2011.04099.x.
- [18] S. Mangione, B. F. Mandell, and S. G. Post, “The Language Game: We Are Physicians, Not Providers,” *American Journal of Medicine*, vol. 134, no. 12. Elsevier Inc., pp. 1444–1446, 2021, doi: 10.1016/j.amjmed.2021.06.031.
- [19] Y. Kyratsis, R. Atun, N. Phillips, P. Tracey, and G. George, “Health systems in transition: Professional identity work in the context of shifting institutional logics,” *Acad. Manag. J.*, 2017, doi: 10.5465/amj.2013.0684.
- [20] H. M. Swick, “Academic Medicine Must Deal with the Clash of Business and Professional Values,” *Acad. Med.*, vol. 73, no. 7, pp. 751–755, 1998.
- [21] P. L. da Luz, “Telemedicine and the doctor/patient relationship,” *Arquivos Brasileiros de Cardiologia*, vol. 113, no. 1. Arquivos Brasileiros de Cardiologia, pp. 100–102, 2019, doi: 10.5935/abc.20190117.
- [22] A. Boonstra, J. Vos, and L. Rosenberg, “The effect of Electronic Health Records on the medical professional identity of physicians: A systematic literature review,” in *International Conference on ENTERprise Information Systems / ProjMAN - International Conference on Project MANagement / HCist - International Conference on Health and Social Care Information Systems and Technologies*, 2021, pp. 272–279, doi: 10.1016/j.procs.2021.12.014.
- [23] K. Cresswell and A. Sheikh, “Organizational issues in the implementation and adoption of health information technology innovations: An interpretative review,” *International Journal of Medical Informatics*, vol. 82, no. 5. 2013, doi: 10.1016/j.ijmedinf.2012.10.007.
- [24] L. Crystal Jiang, Z. Z. Wang, T. Q. Peng, and J. J. H. Zhu, “The divided communities of shared concerns: Mapping the intellectual structure of e-Health research in social science journals,” *Int. J. Med. Inform.*, vol. 84, no. 1, pp. 24–35, 2015, doi: 10.1016/j.ijmedinf.2014.09.003.
- [25] H. Tajfel and J. C. Turner, “The Social Identity Theory of Intergroup Behavior,” in *Psychology of intergroup relations*, Monterey, CA: Brooks-Cole, 1979, pp. 33–47.
- [26] S. A. Haslam, N. Ellemers, S. D. Reicher, K. J.

- Reynolds, and M. T. Schmitt, "The social identity perspective today: An overview of its defining ideas," in *Rediscovering Social Identity*, New York: Psychology Press, 2010.
- [27] J. Evetts, "New Professionalism and new public management: Changes, continuities and consequences," *Comp. Sociol.*, vol. 8, no. 2, pp. 247–266, 2009, doi: 10.1163/156913309X421655.
- [28] B. Burford, "Group processes in medical education: Learning from social identity theory," *Med. Educ.*, vol. 46, no. 2, pp. 143–152, 2012, doi: 10.1111/j.1365-2923.2011.04099.x.
- [29] H. D. Frost and G. Regehr, "I am a doctor: Negotiating the discourses of standardization and diversity in professional identity construction," *Acad. Med.*, vol. 88, no. 10, pp. 1570–1577, 2013, doi: 10.1097/ACM.0b013e3182a34b05.
- [30] H. Becker, B. Geer, E. Hughes, and A. Strauss, *Boys in White: Student Culture in Medical School*. New Brunswick, N.J.: Transaction Books, 1976.
- [31] J. Cassell, T. G. Buchman, S. Streat, and R. M. Stewart, "Surgeons, intensivists, and the covenant of care: administrative models and values affecting care at the end of life," *Crit. Care Med.*, vol. 31, no. 4, pp. 1263–70, 2003.
- [32] A. Cope, J. Bezemer, S. Mavroveli, and R. Kneebone, "What Attitudes and Values Are Incorporated into Self as Part of Professional Identity Construction When Becoming a Surgeon?," *Acad. Med.*, vol. 92, no. 4, pp. 544–549, 2017, doi: 10.1097/ACM.0000000000001454.
- [33] M. D. Beaulieu, M. Rioux, G. Rocher, L. Samson, and L. Boucher, "Family practice: Professional identity in transition. A case study of family medicine in Canada," *Soc. Sci. Med.*, vol. 67, no. 7, pp. 1153–1163, 2008, doi: 10.1016/j.socscimed.2008.06.019.
- [34] W. P. M. A. Hendriks, "Priced not praised: Professional identity of GPs within market-oriented healthcare reform," *J. Prof. Organ.*, vol. 5, no. 3, pp. 12–27, 2018, doi: 10.1093/jpo/jox011.
- [35] P. Obst and K. White, "Three-Dimensional Strength of Identification Across Group Memberships: A Confirmatory Factor Analysis," *Self Identity*, vol. 4, no. 1, pp. 69–80, 2005, doi: 10.1080/13576500444000182.
- [36] J. E. Cameron, "A Three-Factor Model of Social Identity," *Self Identity*, vol. 3, no. 3, pp. 239–262, 2004, doi: 10.1080/13576500444000047.
- [37] S. van den Broek, S. Querido, M. Wijnen-Meijer, M. van Dijk, and O. ten Cate, "Social Identification with the Medical Profession in the Transition from Student to Practitioner," *Teach. Learn. Med.*, vol. 32, no. 3, pp. 271–281, 2020, doi: 10.1080/10401334.2020.1723593.
- [38] I. Wilson, L. S. Cowin, M. Johnson, and H. Young, "Professional Identity in Medical Students: Pedagogical Challenges to Medical Education," *Teach. Learn. Med.*, vol. 25, no. 4, pp. 369–373, 2013, doi: 10.1080/10401334.2013.827968.
- [39] A. Niemi and L. Paasivaara, "Meaning contents of radiographers' professional identity as illustrated in a professional journal - A discourse analytical approach," *Radiography*, vol. 13, no. 4, pp. 258–264, 2007, doi: 10.1016/j.radi.2006.03.009.
- [40] S. R. Cruess, R. L. Cruess, and Y. Steinert, "Supporting the development of a professional identity: General principles," *Med. Teach.*, vol. 41, no. 6, pp. 641–649, 2019, doi: 10.1080/0142159X.2018.1536260.
- [41] A. Witzel and H. Reiter, *The Problem-Centred Interview: Principles and Practice*. London: SAGE Publications, 2013.
- [42] D. S. Kringos, W. G. W. Boerma, A. Hutchinson, and R. B. Saltman, Eds., *Building primary care in a changing Europe*. Copenhagen: World Health Organization.
- [43] A. Johansson, M. Larsson, and B. Ivarsson, "General Practitioners' Experiences of Digital Written Patient Dialogues: A Pilot Study Using a Mixed Method," *J. Prim. Care Community Heal.*, vol. 11, 2020, doi: 10.1177/2150132720909656.
- [44] 1177, "About 1177." [Online]. Available: <https://www.1177.se/en/Uppsala-lan/other-languages/other-languages/About-1177.se/about-1177-vardguiden/>.
- [45] R. Ward, "The application of technology acceptance and diffusion of innovation models in healthcare informatics," *Heal. Policy Technol.*, vol. 2, no. 4, pp. 222–228, 2013, doi: 10.1016/j.hlpt.2013.07.002.
- [46] D. Carlon, A. Downs, and F. Schiavone, "Resistance to industry technological change in communities of practice: The 'ambivalent' case of radio amateurs," *J. Organ. Chang. Manag.*, vol. 25, no. 6, pp. 784–797, 2012, doi: 10.1108/09534811211280564.
- [47] M. Zollinger *et al.*, "Understanding patient preference for physician attire in ambulatory clinics: A cross-sectional observational study," *BMJ Open*, vol. 9, no. 5, 2019, doi: 10.1136/bmjopen-2018-026009.
- [48] A. H. Vinson, "Short White Coats: Knowledge, Identity, and Status Negotiations of First-Year Medical Students," *Symb. Interact.*, vol. 42, no. 3, pp. 395–411, Aug. 2019, doi: 10.1002/symb.400.
- [49] D. Menchik, *Managing Medical Authority: How Doctors Compete for Status and Create Knowledge*. Princeton, New Jersey: Princeton University Press, 2021.
- [50] M. Chwistek, "Are You Wearing Your White Coat?: Telemedicine in the Time of Pandemic," *JAMA - J. Am. Med. Assoc.*, vol. 324, no. 2, pp.

149–150, 2020, doi: 10.1001/jama.2020.10619.

- [51] T. Elliott, E. C. Matsui, A. Cahill, L. Smith, and L. Leibner, “Conducting a Professional Telemedicine Visit Using High-Quality Webside Manner,” *Curr. Allergy Asthma Rep.*, vol. 22, no. 2, pp. 7–12, 2022, doi: 10.1007/s11882-022-01029-y.
- [52] H. Glock *et al.*, “Attitudes, barriers, and concerns regarding telemedicine among swedish primary care physicians: A qualitative study,” *Int. J. Gen. Med.*, vol. 14, pp. 9237–9246, 2021, doi: 10.2147/IJGM.S334782.

8 ACKNOWLEDGEMENT

The first author was funded by the project “International Mobility II” (CZ.02.2.69/0.0/0.0/18_053/0016932) and a fund of long-term research support (IP 400040), both administered by the Prague University of Economics and Business. This work was carried out during his research visit at the Department of Information Technology of the Uppsala University. The second author was funded by AFA grant number 180 250.

Designing an e-Health Program for Lifestyle Changes in Diabetes Care A Qualitative Pre-Study in Norway

Tina Rishaug¹, André Henriksen¹, Anne-Marie Aas^{2,3}, Gunnar Hartvigsen¹, Kåre I. Birkeland^{2,4}, Eirik Årsand¹

¹UiT The Arctic University of Norway, Department for Computer Science, Tromsø, Norway, tina.rishaug@uit.no

²Oslo University Hospital, Department of Clinical Services, Oslo, Norway

³The University of Oslo, Department of Clinical Medicine, Oslo, Norway

⁴The University of Oslo, Department of Transplant Medicine, Oslo, Norway

Abstract

Type 2 diabetes mellitus (T2D) and prediabetes prevalence rates are high. Consequences are serious, but current treatment is often not efficient for achieving remission. Remission may be achieved through lifestyle intervention. Frequent follow-up is necessary, and health care personnel (HCP) lack resources, time, and often adequate knowledge. Self-management of T2D can benefit from better use of Information and Communication Technologies (ICT), which may improve patient involvement and follow-up, and can provide new tools for lifestyle change. Virtual follow-up through ICT with HCP may reduce costs and reach more patients. An electronic health (e-health) program for T2D and prediabetes will be developed in collaboration with users, as part of a plan for a large randomized controlled trial (RCT).

Keywords

Diabetes, type 2 diabetes mellitus, e-health, technology, lifestyle intervention.

1 INTRODUCTION

The prevalence of diabetes is increasing worldwide, where type 2 diabetes mellitus (T2D) accounts for approximately 90-95% of all cases (1). Additionally, many have prediabetes which is characterized by glucose levels higher than normal, but not high enough to meet the criteria of T2D (1). T2D is a chronic, metabolic disease characterized by elevated blood glucose levels caused by relative insulin deficiency. Risk factors include age, overweight or obesity, inactivity, and genetic predisposition (2). T2D, often in combination with obesity, affects morbidity, mortality, and quality of life, and the economic costs associated with T2D in Norway in 2019 was approximately 45,7 billion NOK (3). More than 50% of people with T2D have obesity (4), which can be treated with lifestyle intervention. Nevertheless, bariatric surgery is the most effective long-term treatment because long-term weight loss maintenance through lifestyle intervention only occurs in around 20% of cases (5, 6).

A weight loss of 5-10% is sufficient to achieve important health benefits, including reduced risks of developing complications and other co-morbidities in T2D (7-9). Larger weight losses may also reverse T2D (normalization of blood glucose and symptom relief without the use of medications) (7-9), and is possible to achieve for many. However, lack of thorough and sufficient follow-up tools and strategies often leads to weight regain and return of risk factors associated with overweight and obesity. Time since diagnosis is relevant, as people diagnosed early in the disease course are more likely to succeed with lifestyle changes (10). This confirms the importance of early and efficient intervention to improve chances of long-term health benefits.

According to national and international guidelines, newly diagnosed T2D should be treated with lifestyle intervention through dietary changes and increased physical activity (PA) level (11, 12). Two studies from Finland (13) and the USA (14) showed that lifestyle intervention emphasizing weight loss through a healthy diet and increased PA, can reduce the number of new T2D cases among people at risk of developing the disease by 58%. In the DiRECT trial from the UK, a structured weight loss program with a strict energy-restricted diet and intensive follow-up, administered by the primary health care provided remission of T2D in nearly half (46%) of the participants after one year, and 36% after two years (8, 15). Nonetheless, treatment through lifestyle intervention and follow-up offered to patients today is not sufficient to reach treatment goals regarding blood glucose control and other risk factors, and the majority of T2D patients also need blood glucose lowering medication (16). Bariatric surgery is used in obesity to reach substantial weight reduction, and a T2D remission rate of 25-75% is reported depending on surgical procedure and follow-up time (17).

The burden of T2D is considerable for both the individual and society. Lifestyle changes and weight reduction have great potential in preventing T2D, improving metabolic control among those who have the disease, and may provide remission. However, no present models uses this potential and offer long-term follow-up in the Norwegian health care system today. General practitioners lack capacity to provide such follow-up and may not have the necessary knowledge and they call for better options for lifestyle follow-up to offer their patients with prediabetes or T2D.

The 18th Scandinavian Conference on Health informatics, Tromsø, Norway, August 22-24, 2022. Organized by UiT The Arctic University of Norway. Conference Proceedings published by Linköping University Electronic Press at <https://doi.org/10.3384/ecp187>. © The Author(s). This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>

Technological tools such as mobile phone-based applications with connected human sensors can contribute to, and potentially improve, self-management and follow-up of people with T2D (18, 19), and they have potential to prevent and provide support in the management of several chronic diseases (20). Such technology can provide tools for remote guidance from health care personnel (HCP) and may reduce costs and reach out to a larger group of patients. This technology also enables registration of food intake, PA, body weight, blood glucose levels, and more. In addition, video calls etc. with therapists or other HCP provides opportunity for follow-up without physical attendance. Few such technological tools, which provide lifestyle changes to prevent, treat and reverse T2D and emphasize long-term follow-up and contact with HCP, are available today.

A potential issue when designing and developing new e-health tools, is the probability that such tools are made for people in *least* need of them. This is referred to as the 'People Like Us' (PLU) problem (21) and point to the fact that people in greater need of help in their management and treatment of disease are in fact not the users in which these tools are tailored for. Furthermore, this could mean that the privileged gain even more access to health care services, and those in urgent need of such services are left out due to factors such as low health and technological literacy, motivation, and more. It is therefore necessary to recruit a diverse study sample including people with different needs and experiences in the design and development phase.

An important term in the case of disease prevention and self-treatment is *self-management*, a process which facilitates achievement of knowledge and skills necessary to manage and take control over one's disease. This is especially important in the treatment of T2D (22). People must, according to Bandura (2004), "*learn to monitor their health behavior and the circumstances under which it occurs*" (23). This is possible using technological aids, and an easy-to-use e-health program providing sufficient tools and follow-up will be of great value to patients, relatives and care takers, HCP, and society as a whole in Norway.

This project is a collaboration between Oslo University Hospital (OUS), the University of Oslo (UiO), UiT the Arctic University of Norway (UiT), and ABEL Technologies AS, a technology company based in Tromsø, Norway. ABEL Technologies has developed a tool for lifestyle follow-up which has focused mostly on exercise and PA until now. They now wish to further develop this to target the health care sector including prevention and treatment of diseases.

2 AIMS AND HYPOTHESIS

The aim of this study is to gather information about what the user group (people with prediabetes and T2D) want and need from an e-health program, including essential functionalities.

The long-term aim is to use this e-health program in a large RCT planned to take place in Norway after the present pre-study is finished, to assess its effects in prevention, self-management, and remission of T2D, compared to the effect of not using such a program.

Based on findings from the literature and the clear need for improved options for preventing, treating, and reversing

T2D, it is assumed that technology-based tools may be a more efficient way of deliver lifestyle interventions than options available today.

3 METHODS

3.1 Study design

This is a qualitative pre-study where study participants take part in physical and/or digital focus group meetings, where they will be encouraged to share information about what an e-health program should include, and how it should be designed and used. Questionnaires and audio recordings will be used to collect relevant information. Paper-prototyping (prototyping in paper format where the participants contribute to paper representations of how the functionalities should work and look like) will be used to involve the participants in the development of the program. During digital meetings, a safe communication platform (Whereby or similar) will be used, as well as a digital solution for prototyping. Whereby is currently used at the largest hospital in Norway (OUS) for video consultations with patients and is considered secure for such use.

3.2 Participants and recruitment

Approximately 32 adults (≥ 18 years) with prediabetes or T2D currently living in Norway will be recruited to this pre-study. Recruitment will mainly take place at the Norwegian Diabetes Association's web and Facebook page, in addition to flyers distributed by our collaborating partners at OUS and the University Hospital of North Norway (UNN).

Participants may sign up for the study in three ways: 1) through following a link which takes them to a secure web-based form: Nettskjema, 2) through scanning a QR code provided on the recruitment flyer which takes them to the same Nettskjema, or 3) by sending an e-mail to one of the researchers on the project, with instructions to only write the following: "I want to sign up for the study" or "I want to know more about the study". The person will then be contacted by the researcher by phone.

The web-based form (Nettskjema) contains information about the study and participation criteria. Information regarding age, gender, and education will be collected to ensure a representative group of participants. The qualified and chosen participants will then be able to read the informed consent form and take part in the study. Their e-mail addresses and phone numbers will be collected to enable further contact.

Inclusion and exclusion criteria

Inclusion criteria: HbA1c >48 mmol/mol for T2D, and elevated HbA1c above normal levels (>38 mmol/mol) or increased risk of developing T2D (e.g., first-degree relatives with T2D, elevated blood glucose levels) for prediabetes. Eligible participants must be willing to attend physical or digital meetings with other participants, and consent to audio recordings during meetings.

Exclusion criteria: health challenges which complicate ingesting a normal diet or being normally physical active (e.g., serious cardiovascular disease, lung disease etc.). Difficulties using mobile phones (e.g., reduced vision, motor ability etc.) and not able to communicate in Norwegian.

4 DETAILED PROTOCOL

4.1 Project description

In this pre-study, digital and physical focus group meetings including people with prediabetes and T2D will be held to inform the design and development of an e-health program in line with wishes and needs from a user perspective. These meetings allow the user group to contribute to the development process by sharing expectations about available functionality from an e-health program and suggest new ones. This includes factors associated with motivation and coping, what increases the likelihood of long-term use and thereby aiding in prevention, self-management, and remission of T2D.

Our aim is to develop treatment options using technological self-help tools in order to assess its long-term effect in a planned RCT in Norway, and to collect data to help answer research questions regarding the development of T2D; ranging from basic medicine related issues (importance of genetics, epigenetics, inflammation, adipose tissue changes etc.) to determining prerequisites for successful treatment with lifestyle intervention. This may provide knowledge needed for personalized treatment of prediabetes and T2D. One of the main aims of the e-health program is to tailor it to its users, building upon previous experience with tailoring of the Diabetes Diary self-management app (24). Tailoring is a process for creating individualized communications, and is used to “*determine the most appropriate information or strategies to meet the person’s unique needs*” (25). Tailoring can lead to positive outcomes and benefits for health behavior when compared to non-tailored interventions (26, 27).

4.2 Data collection

Participants will be asked to attend focus group meetings at the beginning of this study. Participants will be allocated to one of the following groups based on their disease status: T2D group or prediabetes group. The aim is to create three groups consisting of people with T2D, and one group consisting of people with prediabetes. Each group will have a maximum of eight participants. Two of the groups are planned to take place online (using Whereby or similar).

Given the short amount of time and possibilities (resources/funding etc.) available in this pre-study, it is expected that all important topics will be covered by including a variety of gender, age, and geographical spread across Norway. Data saturation is estimated to be reached at four focus groups (eight participants in each), based on previous research (28, 29). Several definitions of saturation exist, where one describes data saturation as the point at which “*new data tend to be redundant of data already collected. In interviews, when the researcher begins to hear the same comments again and again, data saturation is being reached*” (30).

Each meeting will last approximately 120 minutes (including a 15-minute break) and take place twice over a three-month period. The first half of the first meeting will be an introductory meeting where the participants get to know each other and are informed about the study, their role and what to expect from the future meetings. This part of the meeting will not be audio recorded in order for participants to share personal stories and experiences. If

any personal information is shared during later meetings, such information will be censored from the transcript.

The following meetings will cover specific topics which are presented in the first meeting to allow preparation for the participants. All participants will be asked to fill out questionnaires regarding preferences, thoughts, and opinions on present and future e-health programs (for digital meetings a secure online form will be used [Nettskjema]). Participants will be asked to identify factors that can improve the usefulness and user experience in general, and which functions are excessive in an e-health program.

During meetings (except the first part of the introductory meeting) two audio recorders (Olympus WS-852) will be used (one for backup). Audio recordings will be saved on two encrypted memory sticks (Corsair Padlock 3) locked up in a separate cabinet until transcription is completed, and content will then be deleted by reformatting the memory sticks, and the memory cards on the audio recorders. Each participant will be assigned a unique identification code used in the transcripts, to deidentify all participants. An overview table containing participant’s names and identification codes are kept locked up in a separate cabinet from the audio recordings. After deleting the audio recordings, only identification codes will be used. The table connecting participants to collected data will be deleted when relevant information has been retrieved and summarized, maximum one year after the end of the study.

Transcribing of audio recordings will be performed on an offline computer at UiT the Arctic University of Norway, by one of the researchers involved in the study, or by a certified company with education in the necessary security and privacy regulations. In the transcripts, unique identification codes will be used, and no sensitive or personal information will be transcribed. Any sensitive or personal information shared in meetings will be censored from transcripts.

Information, opinions, wishes, and needs provided by participants will be saved and used in publications, and to develop the e-health program. ABEL Technologies AS will be informed about the information that is relevant for developing the app with linked functionalities (no sensitive or other information about participants).

4.3 Dissemination

Results from this pre-study will be published as peer-reviewed article(s) and used in further research.

4.4 Ethics

This study was approved by the data protection officer at UiT the Arctic University of Norway and by the Norwegian Center for Research Data (Sikt, previously NSD). The Regional Committees for Medical and Health Research Ethics (REK) has been consulted and we received a conclusion that no application and approval was necessary for this study. Additionally, a risk assessment has been made guided by experts at UiT and has been accepted as sufficient to proceed with this pre-study.

Participation in this pre-study is informed and voluntary and all participants may withdraw from the study at any time without providing any reason(s).

5 SUMMARY

Prevalence rates of prediabetes and T2D are high and increasing worldwide. Consequences may be serious and impact morbidity, mortality, and quality of life, with economic costs exceeding 45 billion NOK in Norway in 2019. Lifestyle change is encouraged for people with prediabetes and T2D, and remission may be achieved. Moderate weight loss (5-10%) can induce important health benefits, but long-term weight maintenance is difficult and only successful in around 20% of cases. GPs and other HCP need new tools to prevent, treat and reverse T2D in Norway, and worldwide. Technological tools such as mobile phone-based applications and services can be valuable and may improve self-management and follow-up in people with T2D, encouraging and providing new and innovative ways for self-management.

The aim of this pre-study is to prepare for the design of a comprehensive e-health program for lifestyle change in this target group, through qualitative focus group meetings, including adults with prediabetes or T2D. Both digital and physical meetings will be arranged, using questionnaires, paper prototyping, and audio recordings to gather information relevant to design an e-health program tailored to the needs of people with the disease.

The long-term aim is to use these results in a large RCT, planned to take place in Norway in 2024, to collect data to help answer research questions regarding the development and treatment of prediabetes and T2D.

6 REFERENCES

1. American Diabetes Association. 2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes-2020. *Diabetes Care*. 2020;43(Suppl 1):S14-s31.
2. Cole JB, Florez JC. Genetics of diabetes mellitus and diabetes complications. *Nat Rev Nephrol*. 2020;16(7):377-90.
3. Oslo Economics. Diabetes type 2 i Norge - en analyse av forekomst, sykdomsbyrde, behandling og samfunnsvirkninger 2021 [cited 2022 July 1.]. Available from: <https://osloeconomics.no/wp-content/uploads/2021/06/Diabetes-i-Norge-En-analyse-av-forekomst-sykdomsbyrde-behandling-og-samfunnsvirkninger.pdf>.
4. Colosia AD, Palencia R, Khan S. Prevalence of hypertension and obesity in patients with type 2 diabetes mellitus in observational studies: a systematic literature review. *Diabetes Metab Syndr Obes*. 2013;6:327-38.
5. Ryder JR, Gross AC, Fox CK, Kaizer AM, Rudser KD, Jenkins TM, et al. Factors associated with long-term weight-loss maintenance following bariatric surgery in adolescents with severe obesity. *Int J Obes (Lond)*. 2018;42(1):102-7.
6. Wing RR, Phelan S. Long-term weight loss maintenance. *Am J Clin Nutr*. 2005;82(1 Suppl):222s-5s.
7. Helsedirektoratet. Nasjonal faglig retningslinje - Diabetes 2021, March 16 [Available from: <https://www.helsedirektoratet.no/retningslinjer/diabetes>].
8. Lean ME, Leslie WS, Barnes AC, Brosnahan N, Thom G, McCombie L, et al. Primary care-led weight management for remission of type 2 diabetes (DiRECT): an open-label, cluster-randomised trial. *Lancet*. 2018;391(10120):541-51.
9. American Diabetes Association Professional Practice Committee. Diabetes care - Facilitating behavior change and well-being to improve health outcomes: Standards of medical care in diabetes—2022: American Diabetes Association; 2022 [Available from: https://diabetesjournals.org/care/issue/45/Supplement_1].
10. Hansen AH, Wangberg SC, Årsand E. Lifestyle changes among people with type 2 diabetes are associated with participation in online groups and time since diagnosis. *BMC Health Services Research*. 2021;21(1):688.
11. Affinati AH, Esfandiari NH, Oral EA, Kraftson AT. Bariatric Surgery in the Treatment of Type 2 Diabetes. *Curr Diab Rep*. 2019;19(12):156.
12. Landgraf R, Aberle J, Birkenfeld AL, Gallwitz B, Kellerer M, Klein H, et al. Therapy of Type 2 Diabetes. *Exp Clin Endocrinol Diabetes*. 2019;127(S 01):S73-s92.
13. Tuomilehto J, Lindström J, Eriksson JG, Valle TT, Hämäläinen H, Ilanne-Parikka P, et al. Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. *N Engl J Med*. 2001;344(18):1343-50.
14. Knowler WC, Barrett-Connor E, Fowler SE, Hamman RF, Lachin JM, Walker EA, et al. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *N Engl J Med*. 2002;346(6):393-403.
15. Lean MEJ, Leslie WS, Barnes AC, Brosnahan N, Thom G, McCombie L, et al. Durability of a primary care-led weight-management intervention for remission of type 2 diabetes: 2-year results of the DiRECT open-label, cluster-randomised trial. *Lancet Diabetes Endocrinol*. 2019;7(5):344-55.
16. Bakke Å, Cooper JG, Thue G, Skeie S, Carlsen S, Dalen I, et al. Type 2 diabetes in general practice in Norway 2005-2014: moderate improvements in risk factor control but still major gaps in complication screening. *BMJ Open Diabetes Res Care*. 2017;5(1):e000459.
17. Chumakova-Orin M, Vanetta C, Moris DP, Guerron AD. Diabetes remission after bariatric surgery. *World J Diabetes*. 2021;12(7):1093-101.
18. Ramadas A, Quek KF, Chan CK, Oldenburg B. Web-based interventions for the management of

type 2 diabetes mellitus: a systematic review of recent evidence. *Int J Med Inform.* 2011;80(6):389-405.

19. Hansel B, Giral P, Gambotti L, Lafourcade A, Peres G, Filipecki C, et al. A Fully Automated Web-Based Program Improves Lifestyle Habits and HbA1c in Patients With Type 2 Diabetes and Abdominal Obesity: Randomized Trial of Patient E-Coaching Nutritional Support (The ANODE Study). *J Med Internet Res.* 2017;19(11):e360.
20. Kaufman N. Digital Therapeutics: Leading the Way to Improved Outcomes for People With Diabetes. *Diabetes Spectr.* 2019;32(4):301-3.
21. Showell C, Turner P. The PLU problem: are we designing personal ehealth for people like us? *Stud Health Technol Inform.* 2013;183:276-80.
22. Banerjee M, Chakraborty S, Pal R. Diabetes self-management amid COVID-19 pandemic. *Diabetes Metab Syndr.* 2020;14(4):351-4.
23. Bandura A. Health promotion by social cognitive means. *Health Educ Behav.* 2004;31(2):143-64.
24. Bradway M, Giordanengo A, Blixgård H, Wangberg SC, Årsand E. Goal-Setting and Patterns of App Use: Selected Results From the "Tailoring Type 2 Diabetes Self-Management" RCT. *Diabetes Technology & Therapeutics.* 2019;21(S1: February 20, 2019):A-162.
25. Rimer BK, Kreuter MW. Advancing Tailored Health Communication: A Persuasion and Message Effects Perspective. *Journal of Communication.* 2006;56(suppl_1):S184-S201.
26. Kroeze W, Werkman A, Brug J. A systematic review of randomized trials on the effectiveness of computer-tailored education on physical activity and dietary behaviors. *Annals of Behavioral Medicine.* 2006;31(3):205-23.
27. Noar SM, Benac CN, Harris MS. Does tailoring matter? Meta-analytic review of tailored print health behavior change interventions. *Psychological bulletin.* 2007;133(4):673.
28. Guest G, Namey E, McKenna K. How Many Focus Groups Are Enough? Building an Evidence Base for Nonprobability Sample Sizes. *Field Methods.* 2017;29(1):3-22.
29. Årsand E, Tatara N, Østengen G, Hartvigsen G. Mobile phone-based self-management tools for type 2 diabetes: the few touch application. *J Diabetes Sci Technol.* 2010;4(2):328-36.
30. Saunders B, Sim J, Kingstone T, Baker S, Waterfield J, Bartlam B, et al. Saturation in qualitative research: exploring its conceptualization and operationalization. *Qual Quant.* 2018;52(4):1893-907.

7 ACKNOWLEDGEMENTS

Novo Nordisk Norway has provided grants for the preparation of the development of the e-health program in this pre-study. Financial contributions have also been made by UiT the Arctic University of Norway, the University of Oslo/Oslo University Hospital, and ABEL Technologies AS.

Towards a New Model for Chronic Disease Consultations

Pietro Randine^{1,2}, John Graham Cooper³, Gunnar Hartvigsen^{1,4}, Eirik Årsand^{1,2}

¹UiT The Arctic University of Norway, Department of Computer Science, Tromsø, Norway pietro.randine@uit.no

²Norwegian Centre for E-health Research, University Hospital of North Norway, Tromsø, Norway

³Norwegian Quality Improvement of Laboratory Examinations, Haraldsplass Deaconess Hospital, Bergen, Norway

⁴University of Agder, Department of Health and Nursing Science, Grimstad, Norway

Abstract

Medical consultations for chronic diseases form an arena to provide information from health personnel to patients. This information is necessary for patients to understand how to deal with the possible lifelong symptoms and needed self-management activities. The amount of patient-generated health data is increasing. Today's patients gather an increasing amount of personalised health-related information. Meanwhile, the health personnel get more patients to care for and fewer resources. This paper summarises information and communication technologies possibilities for improved diabetes consultations. It aims to inform how the medical consultation for chronic diseases needs to change drastically to meet today and future's challenges.

Keywords

Health information system, Patient-generated health data, Chronic disease, Diabetes, Consultation

1 INTRODUCTION

Compared with the general population, those with chronic diseases (e.g., diabetes, cardiovascular diseases, and pulmonary diseases) require more frequent medical attention and symptom management [1]. Individuals from any age group, region and country can be affected by chronic diseases. Chronic diseases impose a significant economic burden on the global healthcare system [2], with predictions becoming more severe regarding the number of people affected and the costs to society.

Once an individual is diagnosed with a chronic disease, this often represents a life-long change in their life. First, the patient needs to understand their new condition and possibly define a health care plan together with the health personnel. This plan may include medication, exercise, therapy, and diet [3]. Then it should be explained and discussed during the medical consultations. After the first consultation, the following consultations often become periodic and are typically performed one to four times a year [4]. This scenario may be more complicated if the patient belongs to a vulnerable group such as children or the elderly. In this case, the medical consultation often involves others, such as parents, next of kin, and informal caregivers [5, 6], introducing additional challenges to the medical consultation.

1.1 Patient-generated health data

Especially in the last 5-10 years, patients have gained more access to health-related devices and information. It is now much easier to track, register and view physical activities, symptoms or treatments via smartphone applications

(apps), commercial wearable devices [7], and Internet of Things (IoT) solutions.

This new information can be used in the decision-making process during consultations and help to define an individual health care plan for the patient. The information flow also expands outside healthcare settings via patient groups on social media [8] and the possibility of sharing patient-gathered health data through apps and cloud-based solutions. Nowadays, social media groups are also used as an alternative resource to prepare before a consultation about symptoms, treatment options, related illnesses, self-management devices, and other health-related issues [9]. Social media groups may complement the consultation with information support without bypassing the health personnel [10]. However, patients seeking missing information on social media can be exposed to misinformation [11] due to difficulty ensuring information quality and accuracy [12, 13].

2 OBJECTIVE

This paper discusses a new model of medical consultation for chronic diseases. As our example, we will focus on a specific chronic condition, type 1 diabetes. First, we describe the current practice for medical consultation. Then we present the proposed model and discuss its implications.

The 18th Scandinavian Conference on Health informatics, Tromsø, Norway, August 22-24, 2022. Organized by UiT The Arctic University of Norway. Conference Proceedings published by Linköping University Electronic Press at <https://doi.org/10.3384/ecp187>. © The Author(s). This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>

3 WHAT DO WE KNOW ABOUT MEDICAL CONSULTATIONS TODAY?

Medical consultations for chronic diseases are physical or remote meetings between patients and health personnel. A medical consultation represents an opportunity to clarify the patient's understanding of their condition [14] and can provide procedures, tools and advice for managing their disease(s) and challenges. Studies have shown that if the patient is an active part of the consultation and the decision-making process, they become better informed about their treatment options [15] and self-management alternatives [16].

The success of a consultation is often determined by how well the patients and the health personnel communicate [17]. A systematic review, including studies from 67 countries, discovered that consultations usually last only a few minutes despite the importance of the health issue. Short consultations may adversely affect patients' disease management and health personnel's workload [18] and increase the risk of medical errors [19].

3.1 A Use case: type 1 diabetes consultations

Patients with type 1 diabetes may use devices from diverse vendors, such as continuous glucose monitors (CGMs), insulin pump systems, blood glucose meters or insulin pens, based on their needs and availability. These devices allow the patient to record and monitor glucose levels, medication (insulin) use, and daily food intake. The devices are connected to the vendors' technological solution, typically a smartphone app for the patient, a web interface for the health personnel, and a cloud-based infrastructure that synchronises the collected information and possibly shares it with others such as relatives, family members, and health personnel.

The goals of diabetes treatment are to prevent or delay short- and long-term complications and optimise quality of life. Treatment goals and management plans should be created together with patients based on their individual preferences, health status, and goals. People with diabetes should have at least one annual consultation. This consultation should be a comprehensive medical evaluation that includes an assessment for diabetes complications and potential comorbid conditions together with a review of previous treatment and risk factor control. Together with the patient, the health personnel should then assess the need to adjust the individual treatment targets. There may also be a need to address diabetes-related psychosocial problems. In clinical practice, the health care provider will often have to prioritise the components of the medical

assessment due to limitations in available resources and time.

Modern diabetes devices can improve diabetes care and the patients' quality of life. A downside may be that data analysis from these devices can be very time-consuming and lead to "information overload" for both health personnel and the patient.

Additionally, the health personnel need to register summary statistics about these patient-gathered data inside the electronic health record (EHR) system, often manually using vendor-specific systems in addition to the EHR system. Although other relevant information such as physical activities, sleep duration, and stress may be discussed during the consultation, this information is usually neither registered nor followed up in the next patient consultation [20].

3.2 Community-based type 1 diabetes consultations

Some type 1 diabetes patients may participate in technical advanced Do It Yourself (DIY) projects. They are often well-educated patients, or engaged relatives, who have formulated, developed, and distributed solutions that answer specific problems to their needs in managing their disease [21].

The diabetes community's effort has also been reflected in patient-started companies like Tidepool, where their technological solution can be used instead of vendor-specific solutions [22]. Their system integrates a subset of CGM devices from different vendors inside the same platform and makes the information available to the patients and health personnel.

4 PROPOSED MODEL

We argue for a new way of defining the consultation, where we propose the inclusion of three different phases: before, during and after the consultation. The motivation behind including also "before" and "after" consultations is the increased possibilities of information and communication technology (ICT) for chronic disease management. Furthermore, the consultation should be conceptualised as a continuous process over time, with a preparation (before), a physical or remote meeting (during) and a follow-up phase (after) [23]. Consequently, the proposed model aims to use various ICTs, some diseases-specific (e.g., CGMs, insulin pumps), some commercial devices (e.g., physical activity trackers, IoT devices), and introduce new practices both for patients and health personnel, supporting the increased information gathering and exchange (see Figure 1).

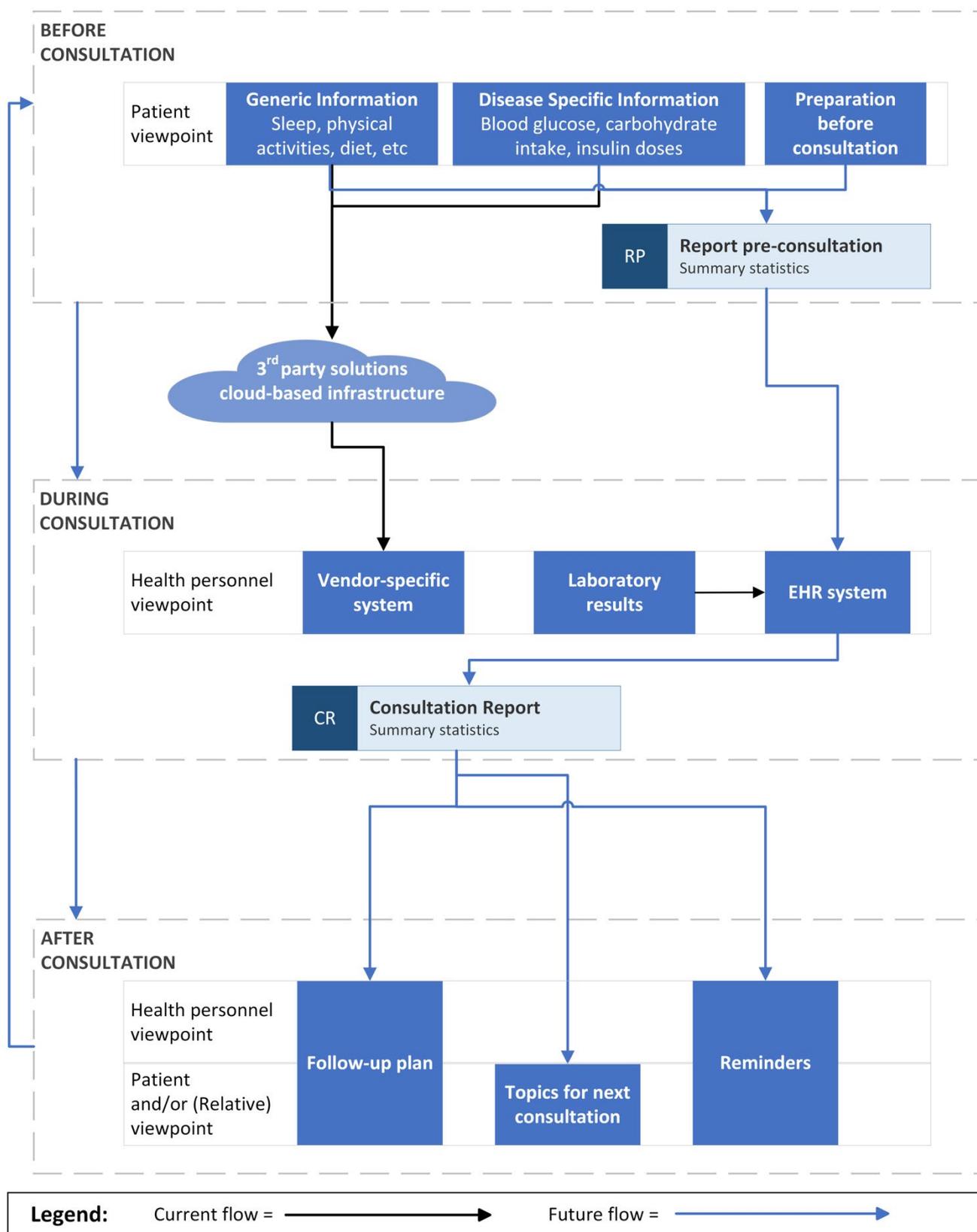


Figure 1 Proposed model, with current and future flow of information for type 1 diabetes consultation.

Before the consultation: patients could prepare themselves by looking at their self-gathered health data. Furthermore, make these available for the health personnel, e.g., physical activities, diet, and sleep, including disease-

specific data, such as blood glucose values, insulin doses, carbohydrate intake, and comprehensive summary statistics.

During the consultation: The data collected before the consultation should be reviewed and registered, preferably automatically, into the EHR system during the consultation. Meanwhile, the necessary adjustments to treatment goals and management plans could also be discussed based on this data during the medical consultation.

After the consultation: What was discussed during the consultation should be made available after the consultation. This can include understandable summaries and follow-up plans for the patients and/or their relatives. Meanwhile, the health personnel could follow up with the patients via reminders before the upcoming visits and encourage them to follow their care plans discussed in the previous consultation. Additionally, the patients can make notes about the side effects of the treatment and note down topics to discuss during their next consultation.

Overall, the presented model includes various elements of remote monitoring and envisioning the medical consultation to be extended beyond than the physical meeting between health personnel and the patient. Therefore, it could provide both parties with more information and better support in difficult situations, e.g., when the patients are not reaching their medical aims or have difficulties in their everyday life caused by their disease.

5 DISCUSSION

The adoption of commercial and medical devices in this model demands the use of third-party companies' devices and software, often located outside the European Union and European Economic Area (EEA). Due to existing regulations, such as the General Data Protection Regulation (GDPR), and their compliance, such a model may raise different critical points and challenges [23], especially from a European perspective.

5.1 Medical devices

The GDPR is not the only regulation that may impact the successful adoption of all the technologies mentioned previously. The recent European Medical Device Regulation (MDR), established in May 2021, updated and extended the definition of medical devices.

The new MDR regulation now also covers health-related smartphone apps. Partially motivated by the fact that thousands of commercial apps are publicly available, and patients with chronic diseases are one of the most prominent target groups [24]. Digital health apps are used in both developed and developing countries [25], and if the intended function of the apps is compromised, it could harm the users (aka patients) [26].

The proposed model would require trust from patients, health personnel and authorities in commercial and medical devices to be considered as a source of information for the medical consultation. In a previous study [27], health personnel ranked the main criteria for recommending medical devices such as digital health apps to patients based on information quality and usability, which employ the openness of health personnel to use these medical devices as part of the medical consultation.

An open question still remains to be answered: Will this European regulation facilitate the integration of what today

is not considered a medical device into the medical consultation? Or, on the other hand, will it slow down the integration?

5.2 Interoperability

Accepting the information gathered from commercial devices inside the medical consultation would require the information collected by the patients to be registered inside the EHR systems. Nowadays, there are technical and legal barriers to registering data generated from medical devices such as CGMs directly into EHRs.

Overall, one of the main challenges is to ensure interoperability and the possibility of data exchange using standards (e.g., FHIR, OpenEHR). Regarding the profiling of health sensor data, standardisation today is limited as well as the adoption of such standards for medical consultation for chronic diseases [23]

5.3 European health data space

The GDPR established in May 2018 has emphasised the potential value and challenges of managing e-health data, especially in terms of security and privacy issues.

Respecting patients' privacy and confidentiality are increasingly becoming more critical, and they represent two of the core values in health care [28]. A key to adopting such technologies is the security and privacy of data, considering the highly sensitive nature of medical data (confidentiality, availability, integrity).

The proposed model in Figure 1 describes an extensive data transmission with many security and privacy challenges. Using these ICTs give access to a vast amount of personally identifiable information and possible target of cyber-attacks. For the following reason, the proposed model will be further worked on in collaboration with the EU-funded HEIR project – a secured Healthcare Environment for Informatics Resilience (grant agreement No 883275).

In the coming years, new legislation, such as the Data Governance Act in 2023, may potentially impact access to more data within the EEA and open the possibility of a Health Data Space. The model presented in Figure 1 could align with such regulations and facilitate data exchange across EEA countries.

5.4 Strengths and limitations

The model presented reflects the findings from previous studies [23, 27, 29], where ICTs were used in intervention for chronic diseases [23] or specifically for diabetes self-management [29]. Many health-related ICTs of today have significant relevance for daily clinical practice, and this model empathises how ICT and interoperability standards may impact future clinical practice.

Therefore, caution should be exercised in evaluating the feasibility of such a model. The medical devices used daily by patients with diabetes are, in practice, the intellectual property of third-parties companies. Consequently, health-related information is often accessible only via proprietary systems, limiting the execution of such a model.

5.5 Future research

Since this study represents an early stage of a new model for medical consultations, future research is now required to interpret this model as a proof of concept to demonstrate

its feasibility. In addition, resource implications and limitations regarding medical device accessibility should be considered.

6 CONCLUSION

Chronic disease consultations are complex. Multiple and diverse stakeholders are often involved, such as health personnel, policymakers, vendors, relatives, and patients. Unclear definitions of the involved technologies [30] and the absence of a shared language in describing them make it harder to integrate apps and new services with health sector stakeholders [31].

This latest introduction of a vast number of medical devices, and commercial wearable devices that enable patients to collect health-related data themselves, calls for new routines and a revision of today's consultation model. Technological innovations are, to an increasing degree, being used by people with chronic conditions. However, consultations are still considered physical or remote meetings only and do not utilise all the potential that self-reported/gathered data can provide.

Regardless of the enthusiasm about these emerging technologies, we must address the adverse effects and risks these technologies can have on data security and privacy issues. Furthermore, we must facilitate the process and assume that patients will wear and adopt consumer technologies in everyday life and that health personnel will use them as part of the medical consultation.

In conclusion, such a model is technologically feasible, and its implementation in clinical practice will be dependent on the policymaking decision in the coming years.

7 SUMMARY

This paper has discussed a new model that views medical consultation as a continuous process in terms of preparation (before), a meeting (during), and a follow-up phase (after).

We are in a phase where patients have more access to health-related information such as physical activities, symptoms or treatments via diverse technologies or social media communities. This new information can be used in the decision-making process during consultations and be used in refining the individual health care plan for the patient.

Designing a system that can possibly be integrated with the clinical EHR systems used for patient treatment and follow-up is conceptually possible. Although, mainly security, privacy and interoperability issues slow down the integration of such innovation in the medical consultation and the healthcare systems.

8 REFERENCES

- [1] Bernell, S. and S.W. Howard, *Use your words carefully: what is a chronic disease?* *Frontiers in public health*, 2016. 4: p. 159.
- [2] Roglic, G., *WHO Global report on diabetes: A summary*. *International Journal of Noncommunicable Diseases*, 2016. 1(1): p. 3.
- [3] Warburton, D.E., C.W. Nicol, and S.S. Bredin, *Health benefits of physical activity: the evidence*. *Cmaj*, 2006. 174(6): p. 801-809.

- [4] Gask, L., *Role of specialists in common chronic diseases*. *Bmj*, 2005. 330(7492): p. 651-653.
- [5] Martin, C.M., *Chronic disease and illness care: adding principles of family medicine to address ongoing health system redesign*. *Canadian Family Physician*, 2007. 53(12): p. 2086-2091.
- [6] Von Korff, M., et al., *Collaborative management of chronic illness*. 1997, American College of Physicians.
- [7] Brooke, S.M., et al., *Concurrent validity of wearable activity trackers under free-living conditions*. *Journal of strength and conditioning research*, 2017. 31(4): p. 1097-1106.
- [8] Kinnersley, P., et al., *Interventions before consultations to help patients address their information needs by encouraging question asking: systematic review*. *Bmj*, 2008. 337.
- [9] Bartlett, Y.K. and N.S. Coulson, *An investigation into the empowerment effects of using online support groups and how this affects health professional/patient communication*. *Patient education and counseling*, 2011. 83(1): p. 113-119.
- [10] Smailhodzic, E., et al., *Social media use in healthcare: A systematic review of effects on patients and on their relationship with healthcare professionals*. *BMC health services research*, 2016. 16(1): p. 1-14.
- [11] Waszak, P.M., W. Kasprzycka-Waszak, and A. Kubanek, *The spread of medical fake news in social media—the pilot quantitative study*. *Health Policy and Technology*, 2018. 7(2): p. 115-118.
- [12] De Martino, I., et al., *Social media for patients: benefits and drawbacks*. *Current reviews in musculoskeletal medicine*, 2017. 10(1): p. 141-145.
- [13] Gautschi, O.P., et al., *Web-based audiovisual patient information system—a study of preoperative patient information in a neurosurgical department*. *Acta neurochirurgica*, 2010. 152(8): p. 1337-1341.
- [14] Organization, W.H., *General principles of good chronic care: integrated management of adolescent and adult illness*. 2003: World Health Organization.
- [15] Stacey, D., et al., *Decision aids for people facing health treatment or screening decisions*. *Cochrane database of systematic reviews*, 2017(4).
- [16] Mosen, D.M., et al., *Is patient activation associated with outcomes of care for adults with chronic conditions?* *The Journal of ambulatory care management*, 2007. 30(1): p. 21-29.
- [17] Gask, L. and T. Usherwood, *The consultation*. *Bmj*, 2002. 324(7353): p. 1567-1569.
- [18] Irving, G., et al., *International variations in primary care physician consultation time: a systematic review of 67 countries*. *BMJ open*, 2017. 7(10): p. e017902.
- [19] Ancker, J.S., et al., *The invisible work of personal health information management among people with multiple chronic conditions: qualitative interview study among patients and providers*. *Journal of medical Internet research*, 2015. 17(6): p. e137.

- [20] Larbi, D., et al., *Do diabetes mHealth and online interventions evaluate what is important for users?* 2019.
- [21] Omer, T., *Empowered citizen 'health hackers' who are not waiting*. BMC Med, 2016. 14(1): p. 118.
- [22] Neinstein, A., et al., *A case study in open source innovation: developing the Tidepool Platform for interoperability in type 1 diabetes management*. Journal of the American Medical Informatics Association, 2016. 23(2): p. 324-332.
- [23] Randine, P., et al., *Information and Communication Technology-based Interventions for Chronic Diseases Consultation: Scoping Review*. International Journal of Medical Informatics, 2022: p. 104784.
- [24] Research2guidance. *mHealth App Economics 2017: Current Status and Future Trends in Mobile Health*. 2017; Available from: <https://research2guidance.com/product/mhealth-economics-2017-current-status-and-future-trends-in-mobile-health/>.
- [25] Agarwal, S., et al., *Guidelines for reporting of health interventions using mobile phones: mobile health (mHealth) evidence reporting and assessment (mERA) checklist*. bmj, 2016. 352: p. i1174.
- [26] Lewis, T.L. and J.C. Wyatt, *mHealth and mobile medical apps: a framework to assess risk and promote safer use*. Journal of medical Internet research, 2014. 16(9): p. e210.
- [27] Larbi, D., et al., *Criteria for Assessing and Recommending Digital Diabetes Tools: A Delphi Study*. Studies in health technology and informatics, 2021. 281: p. 850-854.
- [28] Williams, J.R., *The Declaration of Helsinki and public health*. Bulletin of the World Health Organization, 2008. 86: p. 650-652.
- [29] Larbi, D., et al., *Methods and Evaluation Criteria for Apps and Digital Interventions for Diabetes Self-Management: Systematic Review*. Journal of Medical Internet Research, 2020. 22(7): p. e18480.
- [30] Hou, C., et al., *Do mobile phone applications improve glycemic control (HbA_{1c}) in the self-management of diabetes? A systematic review, meta-analysis, and GRADE of 14 randomized trials*. Diabetes Care, 2016. 39(11): p. 2089-2095.
- [31] Labrique, A.B., et al., *mHealth innovations as health system strengthening tools: 12 common applications and a visual framework*. Global health: science and practice, 2013. 1(2): p. 160-171.

9 ACKNOWLEDGEMENT

We would like to thank Celia Nilssen for her assistance and role as an “internal reviewer”. Furthermore, we thank Bruna Martins Catarino for her feedback on the figure and the EU-project” HEIR – a secured Healthcare Environment for Informatics Resilience (grant agreement No 883275). This work is part of the first author’s PhD project, funded by UiT The Arctic University of Norway.

Sorterius - An Augmented Reality App for Encouraging Outdoor Physical Activity for People with Intellectual Disabilities

Magnus Stellander¹, André Henriksen¹, Henriette Michalsen^{1,2}, Audny Anke^{1,2}, Daniel Ursin¹, Santiago Martinez³, Susanna Pelagatti⁴, Keiichi Sato^{1,5}, Vebjørn Haugland¹, Erlend Johannessen¹, Juan Carlos Torrado⁶, and Gunnar Hartvigsen^{1,3}

¹UiT The Arctic University of Norway, Tromsø, Norway, andre.henriksen@uit.no

²University Hospital of North Norway, Tromsø, Norway

³University of Agder, Grimstad, Norway

⁴Università di Pisa, Pisa, Italy

⁵Illinois Institute of Technology, Chicago, IL 60616, USA

⁶University of Bergen, Bergen, Norway

Abstract

Many with intellectual disabilities (ID) have difficulties adhering to current physical activity guidelines. The goal of this study was to develop a mobile app for assisting people with ID to be more physically active. We implemented a solution that combines the digital and real world using augmented reality (AR). Eight people working with people with ID (special education teachers, social workers, psychologists, and researchers) tested the app and completed a usability test. Results indicate that a mobile app focusing on everyday life scenarios can have a potential value for the targeted user group, but AR solutions can be challenging.

Keywords

Motivation, mHealth, mobile health, exergames, steps, smartphone

1 INTRODUCTION

Physical activity (PA) provides significant health benefits [1], and the World Health Organization (WHO) recommends young adults to perform at least 150 minutes of moderate PA every week [2]. However, people with intellectual disabilities (ID) are known to have difficulties achieving these recommendations [3]. Compared to the general population, they have lower PA levels and worse health [4; 3]. One of the barriers for participation in PA for individuals with ID is lack of interest and low PA related self-efficacy [5]. However, motivation for PA could be triggered by fun, use of rewards, and technology [6].

The availability of mobile applications (apps) for improving and motivating PA has greatly improved the last decades [7]. However, apps tend to generally be too complex for people with ID. It follows that people with ID need tailored apps that can motivate them and increase their PA [8]. The goal of this study was to develop an app for assisting people with ID to be more physically active. The app is part of an ongoing intervention with people with ID and is currently being tested and evaluated [9; 10].

2 METHOD

We implemented Sorterius, a Pokémon Go-inspired cross-platform app using the Unity game engine (Unity Technologies, SF, US, v2019.4).

The game uses augmented reality (AR), where players observe the real world through the smart phone camera. Digital content, in the form of 3D garbage objects, appears on the screen as the player walks around. Sorterius is based on previous work by Haugland et al. [11; 12]. A thorough description of the design and implementation of Sorterius is described by Stellander [13].

The player's goal is to help the game mascot, *Sorterius*, to clean the world. This is achieved by picking up virtual garbage as it appears on the screen (by tapping them). Players must correctly choose between several containers to throw the garbage in, depending on difficult level. The player can choose between three difficulty levels: *easy* (one garbage container), *medium* (two containers), and *hard* (four containers). An internal step counter is used to determine how often new garbage objects should appear. Players are rewarded with daily virtual rewards, in the form of *stars* and positive feedback. A caretaker menu allows customization of level difficulty and to define the daily step goal and weekly star goal. Completing the daily step goal awards three stars. In addition, to further motivate usage, physical rewards can be defined (e.g., movie tickets), which will be awarded by the caretaker upon achieving the weekly star goal. Text-to-speech options can be enabled for players with limited reading skills.

The 18th Scandinavian Conference on Health informatics, Tromsø, Norway, August 22-24, 2022. Organized by UiT The Arctic University of Norway. Conference Proceedings published by Linköping University Electronic Press at <https://doi.org/10.3384/ecp187>. © The Author(s). This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>

Due to restrictive infection control measures related to the Covid-19 pandemic, we were unable to test the app in the target group. However, eight people with relevant background, i.e., special education teachers, social workers, psychologists, and researchers working with people with ID, tested the app, completed a System Usability Scale (SUS) questionnaire, and gave general open-ended comments to the solution. A SUS is used for collecting feedback on subjective aspects of usability of a system [14]. The scale is a 10-question questionnaire, where answers are given on a 5-point liker scale, ranging from *strongly disagree* to *strongly agree*. We customized the SUS questionnaire to address the limitation of who completed the questionnaire. Table 1 gives a list of the 10 customized questions.

#	Question
Q1	I think the application would be regularly used by people with an intellectual disability
Q2	I think the application is too complicated for people with an intellectual disability
Q3	I think the application is easy to use for people with an intellectual disability
Q4	I think a person with an intellectual disability would need support to use the application
Q5	I think a person with an intellectual disability would think the different parts of the application are well connected
Q6	I think a person with an intellectual disability would think there are too many inconsistencies in the application
Q7	I think a person with an intellectual disability would be able to learn to how to use the application
Q8	I think the application is too difficult to use for a person with an intellectual disability
Q9	I think a person with an intellectual disability would be comfortable using the application alone
Q10	I think it will require extensive training before this application can be used

Table 1. Customized System Usability Scale (SUS) questions

Individual SUS scores for each question were calculated by subtracting one point (i.e., score-1) from all odd questions (positive polarity), whereas for even questions (negative polarity), participant responses were subtracted from five (i.e., 5-score) This gives 0-4 points for each question. Scores for all participants were added and multiplied by 2.5 to create a scale from 0-100 for each question. According to Sauro et al. [15], a score of 68 is an average score when analysed as a percentile rank and can be considered as a “Satisfactory” system. Alternatively, Bangor et al. [16] defined an acceptability scale, where a SUS score above 70 is considered “Acceptable”.

3 RESULTS

Figure 1 shows two game play screenshots, showing the

visualization of the garbage on screen using AR. The app achieved an overall SUS score of 61. This is somewhat lower than the average score mentioned above of 68 [15] and corresponds to a “marginally acceptable” [16] system. Participant’s individual SUS scores ranged from 42.5 to 80.0, where half of participants gave a SUS score of 65 or higher. Individual raw (i.e., before inverting negatively framed questions and before multiplying with 2.5 to create 0-100 scale) SUS scores ranged from 2.1 to 3.9. An overview of SUS (range 1-5) mean and standard deviation for each question is given in Table 2. For questions Q1, Q3,

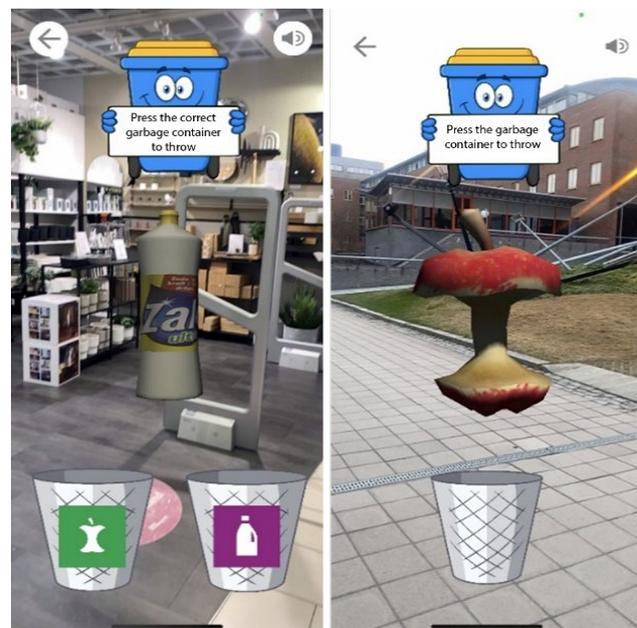


Figure 1. Example gameplay screenshots. Plastic garbage on medium difficulty (left). Food garbage on easy difficulty (right).

Q5, Q7, and Q9 (banded in Table 1), higher score is better. For questions Q2, Q4, Q6, Q8, and Q10 (non-banded in Table 1), lower score is better.

Question	Q1	Q2	Q3	Q4	Q5
Mean	3.4	2.4	3.3	3.6	3.5
SD	0.48	0.99	0.66	0.99	0.87
Question	Q6	Q7	Q8	Q9	Q10
Mean	2.1	3.9	2.3	3.1	2.6
SD	0.60	0.33	0.83	1.05	1.22

Table 2. SUS Mean score and standard deviation (SD) for each question. Score range 1 to 5.

Regarding open-ended comments, a device with a larger screen (e.g., tablets) was suggested by several participants who reported difficulties when trying to observe objects in detail. Other participants also experienced glitches during testing and emphasized the need to resolve such issues before exposing the app to the target user group, to prevent irritation. One participant suggested to log failed sorting attempts and use this data to increase learning potential when sorting garbage. The same participant also

suggested to expand Sorterius and use it as a learning tool for sorting other item categories. Overall, participants gave positive comments of the general usability of the game.

4 DISCUSSION

The mobile app Sorterius show promising potential for being used by individuals with ID, and hopefully influence PA levels. Participants in this study believed people with ID would be able to learn how to use the app (Q7) and that they would use it regularly (Q1). However, one question (Q4) affected the overall score negatively, showing a high score for thinking the target group would need support to use the app. For this user group, relying on support from caretaker is common. Because of this, scoring high in this question may therefore not necessarily substantially affect the usability. In fact, engagement of support people may increase motivation for PA [6]. In the previous work by Haugland [12], this question was scored with a similar high SUS score. Haugland also implemented a Pokémon Go-inspired AR game. One possible explanation to this score, may be that people with ID can struggle with abstract concepts, and AR games may therefore be challenging for the target group.

There was also a large score difference between the individual with the highest score (SUS=80) and the individual with the lowest score (SUS=42.5). When completing the questionnaire, several participants indicated that it was difficult to generalize, because they worked with people with different levels of ID. Although one participant thought the app was usable for people with moderate to severe ID (with support), others thought that it would be too difficult for people with a moderate severity level. Variety in SUS-questionnaires is not unusual, but the SUS range was nonetheless wide in this usability test. This SUS-questionnaire may be hard to evaluate precisely without testing the app on the target user group.

This research has some limitations. The SUS questionnaire was translated to Norwegian and modified to target people working with ID. This could potentially affect the meaning of the questions. A validation of the SUS questionnaire in the Norwegian language is thus required as further research. In addition, due to the restrictions caused by the COVID-19 pandemic, we could not invite people from the actual target user group to test the app. These issues may affect the validity of results and have been specifically addressed in the ongoing pilot and feasibility study [9; 10].

5 CONCLUSION

The benefits of PA could also be achieved by people with ID. Using technology is one way to provide incentives to a target group outside the traditional marketing campaigns of technology development. The apps available in the market tend to be too complex for people with ID. This research has presented a tailored app aiming to motivate the target user group to do PA. The main contribution of this project is a cross-platform AR app combining a

motivational tool for PA participation and learning (e.g., how to sort garbage), for a group that is often neglected in technology intervention in society. The app includes goal setting, involving the support people around the users with ID, which previously is shown to improve motivation towards PA in this user group. Future testing of the app should include users with ID to ensure the generalizability of this research and usability of the app. The evaluation of the test results shows that, although some of the test scores were low, we have created a "marginally acceptable" user interface for the targeted user group. The developed solution can be expanded to target a broader range of users and projects. Using AR to address meaningful societal aspects, such as the environment, can be a helpful tool in raising awareness while at the same time strengthening the knowledge towards serious topics in an engaging digital environment. The final product has received exciting feedback from experts and testing participants.

Sorterius is currently being used in a pilot and feasibility study for a randomized control trial (RCT) intervention where one of the goals is to investigate whether tailored mHealth support can stimulate PA for individuals with ID [9; 10]. A Norwegian version is available in for free download in Apple's AppStore [17] and in Google Play [18]. English, Portuguese, Italian, and Spanish versions are currently under development as part of the MOVE-IT project [19].

6 REFERENCES

- [1] Warburton, D. E., Nicol, C. W., and Bredin, S. S., "Health benefits of physical activity: the evidence", *CMAJ*, 174 (6), 801-9, 2006
- [2] WHO 'Physical activity', <<https://www.who.int/news-room/factsheets/detail/physical-activity>>, accessed 2022-06-20.
- [3] Wouters, M., Evenhuis, H. M., and Hilgenkamp, T. I. M., "Physical activity levels of children and adolescents with moderate-to-severe intellectual disability", *J Appl Res Intellect Disabil*, 32 (1), 131-42, 2019
- [4] Walsh, P. N., "Health indicators and intellectual disability", *Curr Opin Psychiatry*, 21 (5), 474-8, 2008
- [5] Dixon-Ibarra, A., Driver, S., Vanderbom, K., and Humphries, K., "Understanding physical activity in the group home setting: a qualitative inquiry", *Disabil Rehabil*, 39 (7), 653-62, 2017
- [6] Michalsen, H., Wangberg, S. C., Anke, A., Hartvigsen, G., Jaccheri, L., and Arntzen, C., "Family members and health care workers' perspectives on motivational factors of participation in physical activity for people with intellectual disability: A qualitative study", *J Intellect Disabil Res*, 64 (4), 259-70, 2020
- [7] Gal, R., May, A. M., van Overmeeren, E. J., Simons, M., and Monnikhof, E. M., "The Effect of Physical Activity Interventions Comprising Wearables and Smartphone Applications on Physical Activity: a

- Systematic Review and Meta-analysis", *Sports Med Open*, 4 (1), 42, 2018
- [8] Torrado, J. C., Jaccheri, L., Pelagatti, S., and Wold, I., "HikePal: A mobile exergame to motivate people with intellectual disabilities to do outdoor physical activities", *Entertainment Computing*, 42, 2022
- [9] Michalsen, H., Wangberg, S. C., Hartvigsen, G., Jaccheri, L., Muzny, M., Henriksen, A., Olsen, M. I., Thrane, G., Jahnsen, R. B., Pettersen, G., Arntzen, C., and Anke, A., "Physical Activity With Tailored mHealth Support for Individuals With Intellectual Disabilities: Protocol for a Randomized Controlled Trial", *JMIR Res Protoc*, 9 (6), e19213, 2020
- [10] Michalsen, H., Wangberg, S. C., Hartvigsen, G., Henriksen, A., Pettersen, G., Jaccheri, L., Jahnsen, R., Thrane, G., Arntzen, C., and Anke, A., "Mobile health support to stimulate physical activity in individuals with intellectual disability: Protocol for mixed methods pilot study", *JMIR Preprints*, 2022
- [11] Berg, V., Haugland, V., Wiik, M. F., Michalsen, H., Anke, A., Muzny, M., Gomez, J., Martinez, S. G., Martinez-Millana, A., Henriksen, A., Sato, K., and Hartvigsen, G. (2020), 'eHealth Approach for Motivating Physical Activities of People with Intellectual Disabilities', in Ilias O. Pappas, et al. (eds.), *Digital Transformation for a Sustainable Society in the 21st Century* (Cham: Springer International Publishing), 31-41.
- [12] Haugland, V. (2019), 'CorpOperatio: Game-inspired App for Encouraging Outdoor Physical Activity for People with Intellectual Disabilities'.
- [13] Stellander, M. (2021), 'Sorterius: Game-inspired App for Encouraging Outdoor Physical Activity for People with Intellectual Disabilities'.
- [14] Brooke, J. (1995), 'SUS - A quick and dirty usability scale', *Usability Evaluation in Industry*.
- [15] Sauro, J. (2011), *A Practical Guide to the System Usability Scale: Background, Benchmarks & Best Practices* (Denver, CO: Measuring Usability LLC).
- [16] Bangor, A., Kortum, P. T., and Miller, J. T., "An empirical evaluation of the System Usability Scale", *International Journal of Human-Computer Interaction*, 24 (6), 574-94, 2008
- [17] UiT The Arctic University of Norway (2022), 'Sorterius (AppStore)', <<https://apps.apple.com/ng/app/sorterius/id1610130479>>, accessed 04-01.
- [18] UiT The Arctic University of Norway (2022), 'Sorterius (Google Play)', <<https://play.google.com/store/apps/details?id=no.uit.ifi.sorterius>>, accessed 06-25.
- [19] moveIT project (2022), 'moveIT', <<https://moveit.webs.upv.es>>, accessed 06-10.

Student–staff Co-creation of Serious Games

- Lessons Learned

Hege Mari Johnsen¹ and Jannicke Rabben¹

¹University of Agder, Department of Health- and Nursing Science, Grimstad, Norway, hege.mari.johnsen@uia.no

Abstract

Few papers have described academic/faculty staff's experiences with co-creation, or partnering with students in cross-disciplinary collaborations. The purpose of this paper is to share challenges and outcomes from two interdisciplinary student–staff co-creations of serious games for use in a Bachelor of Nursing program in Norway. Our experiences are discussed against an evidence-informed model of student–staff co-creation in higher education. Based on the lessons learned from these two projects, we propose ten key points for planning and conducting cross-disciplinary student–staff co-creation of serious games.

Keywords

Collaboration, design, education, games, interdisciplinary

1 INTRODUCTION

Stuckless, Hogan, and Kapralos [1] define a serious game (SG) as “an interactive computer application that (1) has a challenging goal, (2) is fun to play and/or engaging, (3) incorporates some concept of scoring, and (4) impacts to the user a skill, knowledge, or attitude that can be applied to the real world” (p. 146). Serious games used in health care can provide different forms of simulated environments (virtual reality, video, picture, animation), which provide opportunities to practice clinical reasoning and decision-making skills in realistic and safe environments [2]. However, for SGs to enable active, experiential, and problem-based learning, special efforts need to be made in the design and development of SGs [3]. For example, academic/faculty staff need to design the educational content in consideration of the target users, genre/story/context, and learning objectives and ensure evidence-based content. In addition, user–computer interaction design elements must be ensured [3] [4] [5]. Hence, it is important to employ interdisciplinary collaboration which ensure competence within relevant areas and disciplines.

Cross-disciplinary student–staff co-creation of educational tools, such as SGs, may facilitate two dual-value creation dimensions: co-production and value-in-use [6]. Such collaboration and value creation are also in line with the university's vision and strategy for 2021–2024 [7], the White Paper [8], and the United Nations' sustainability goals [9]. Few papers have described academic/faculty staff's experiences with co-creation, or partnering with students in cross-disciplinary collaborations [10] [7]. Mercer-Mapstone et al. [10] call for papers sharing in more detail the challenges and possible negative outcomes of such partnerships. Hence, the purpose of this paper is to share challenges and outcomes from two SG co-creation projects. First, we present the two SGs. Second, we

describe and discuss our experiences against an evidence-informed model of student–staff co-creation in higher education [6], which provides an interdisciplinary view for conceptualizing, designing, implementing, and evaluating co-creation in education. This model considers the following three key elements of co-creation: inputs (individual and environmental considerations), processes (barriers that can arise in co-creation, mechanisms needed to support student–staff co-creation, co-production, and value-in-use), and outputs (benefits for students and staff). Finally, we propose ten key points for planning and conducting cross disciplinary student–staff co-creation of serious games.

2 THE TWO GAMES

2.1 “I-cannot-breathe”

The single-player SG “I-cannot-breathe” [Jeg-får-ikke-puste] aimed to teach nursing students clinical reasoning and decision-making skills in care for patients with COPD [11]. It was video-based and contained different quiz-based tasks that needed to be solved before continuing the game. The game was linear; users could not direct the nurse in the game or change paths based on their choices. The correct answers were demonstrated by the nurse through the video, and the answers could be viewed on the screen. Figure 1 shows a screenshot from one of four different scenarios.

This game was developed as part of the first author's PhD project connected to the Department of Health and Nursing Science. The PhD candidate collaborated with health personnel and supervisors in developing the educational and evidence-based content. Four students from the Department of Information and Communication Technology chose development software, recorded the videos, assembled video clips and quiz-based tasks, and integrated these with necessary instructions on how to use the SG. The development software chosen by the students

The 18th Scandinavian Conference on Health informatics, Tromsø, Norway, August 22-24, 2022. Organized by UiT The Arctic University of Norway. Conference Proceedings published by Linköping University Electronic Press at <https://doi.org/10.3384/ecp187>. © The Author(s). This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>

were Adobe Captivate 8, Adobe Premiere Pro CC, and Adobe Photoshop CS6. HTML5 was chosen for uploading to an internet address. Available platforms were PCs, laptops, and the newest tablets [12].



Figure 1. A screenshot from “I-cannot-breathe”

2.2 “Hans and the welfare technology”

The online, single-player SG “Hans and the welfare technology” [Hans og velferdsteknologien] was an interactive, non-linear story; the user’s choices in the quiz-based tasks determined the story. It aimed at teaching students to identify and reflect upon ethical and legal aspects in the use of welfare technology. Figure 2 shows a screenshot from the game, where the elderly man Hans rejects the offer of using a GPS tracker.



Figure 2. A screenshot from the SG “Hans and the welfare technology”.

This SG can be categorized as a visual novel, where a story is told using graphics and text and contains a low level of gameplay [13] [14]. A progress tracker based on points

given for the different choices made in the dialogue and quizzes (yellow line in Figure 2) increased the level of gamification and resulted in automated feedback to the students on their overall achievement. They also received feedback and correct answers in the form of text.

This SG was developed as part of a larger project in connection to the Department of Health and Nursing Science, led by the second author. The second author collaborated with the first author and a nurse from the municipality health care services in developing the educational and evidence-based content.

Four students from the Department of Information and Communication Technology were responsible for the development of the SG—three for programming [13] and one for graphical design [14]. The development/programming software used by the students were JetBrains Rider, JetBrains WebStorm (for JavaScript, CSS, and HTML-elements), Git, Bitbucket, and Overleaf (LaTeX-editor). The design software chosen was Pixel art.

3 LESSON LEARNED

The challenges and outcomes of the two projects will now be described and discussed in light of the three key elements of supporting student–staff co-creation: inputs, processes, and outputs [6].

3.1 Inputs

Individual and environmental factors, such as individuals’ previous history and experiences with student–staff co-creation, individuals’ motivation, authenticity, and clarity of the activity, constitute the foundations of a co-creation activity before it begins. According to Dollinger and Lodge [7], these inputs are critical to the subsequent processes and outcomes of the experience.

3.1.1 Individual Considerations

In relation to individual experiences, the managers of the two projects (authors) were both nurses with no previous technical experience in SG development. They also had limited experience with student–staff co-creation and project management. The developers of the SG were bachelor students from the Department of Information and Communication Technology with different levels of design and programming experience. The students were recruited through the university’s own website, “Kompetansetorget” [The Competence Square].

The motivations of students and staff were quite different. The students conducted this project with the intention to pass their exam and get a bachelor’s degree. The academic staff aimed to implement the SGs as part of their course, conduct related educational research, and publish the results. As suggested by Dollinger and Lodge [6], motivation and aims should have been addressed at an early stage. Then, the effort for value creation could have been more intertwined.

3.1.2 Environmental Considerations

Important environmental considerations for a co-creation project include clarity of the activities or tasks that should be conducted in the student–staff co-creation process [6]. The most serious pitfall for a project is if one chooses a level of detail in the task description that is either too coarse or too fine [15]. In the first project, a formal agreement

between the two faculties was signed. This specified roles, contributions, and the sharing of resources and costs for development. In the second project, a formal agreement was intended, but only an informal agreement was made. This was due to many involved parties and the lack of a standard procedure for co-creation. In retrospect, the agreements (formal/informal) could have benefited from more details concerning supervisors' roles and responsibilities and the handling of possible risks (3.1.2).

For SG development, it is important to make a detailed specification of the SG in collaboration with the developers [5]. These specifications include, for example, the game engine, database, software applications that fit with planned features in the SG, platform (touch-tablet, laptop, personal computer (PC), smart-phone), and, if desired, compatibility with a Learning and Management System (LMS). In addition, there are many other things that must be considered. What format and user-computer interaction design do we want (videos, graphics/photos, text) and how should users interact with the game (visual/audio, mouse/touch)? Should the SG include different types of questions (e.g., single or multiple answer, drag-and-drop questions), and should it provide the ability for users to choose wrong answers deliberately and view the consequences of their choice? How about in-game assessment? Should users answer questions or complete tasks before they can continue, and do they receive points and a get a final score? What kind of feedback should the user receive during gameplay? And do we want a single- or multiplayer SG?

Depending on whether the game is linear or the user has the option to choose different paths, a storyboard and a decision tree must be developed. One should construct a storyboard for each SG scenario [16] [17]. Each storyboard must contain a detailed description of the SG story, educational content, actions in each video clip or screen with related quiz-based tasks, and questions with answers. Finally, depending on what type of game is developed, one may need special equipment during development (e.g., video cameras and microphones).

In both projects, the description or specifications of the game could have been more detailed, as indicated by the necessity of several adjustments during development.

3.2 Processes

3.2.1 Barriers

Certain mechanisms are needed to support student-staff co-creation. Some barriers that may arise during the processes of co-creation include role confusion, need for student-staff guidance, inexperience of participants, inclusion of assessment, time, and power imbalance [6].

In relation to organization and roles, the PhD project had a quite simple organization compared to the other project. Here, the SG development team consisted mostly of the PhD candidate and the four recruited students from the bachelor program in Multimedia Technology and Design. The students received supervision from domain experts, who attended the first meetings. After that, most communication with the students' supervisors was through e-mail.

For the second project, the development team consisted of two faculty members (authors) from the Department of Health and Nursing Science, as well as four students and their supervisors from the Department of Information and Communication Technology (one from the bachelor program in Multimedia Technology and Design and three from the bachelor program in Computer Engineering). In addition, one nursing student and a consultant from the ICT department was involved in the co-creation of the SG.

In relation to barriers such as the inexperience of participants, we had no information about the level of experience of each student or how much guidance the students received from their supervisor during the SGs' development. The two teachers (authors) responsible for the SGs' development were nurses, not technicians, and were unaware of all aspects and risks in designing SGs. This caused a power imbalance between the students and academic staff. To carry out the SG development, these two projects were highly dependent on the students and supervisors from the bachelor program in Multimedia Technology and Design.

Dependence on others to be able to carry out a project may be indicative of a high-risk project [11]. Hence, recruiting students with desired skills and motivation, as suggested by Dollinger and Lodge [6], may decrease the risk of project failure. Further, it is important that the students can collaborate as a team. Hence, assistance in recruitment from teachers who knows the individual students may decrease the chances of project failure.

However, even if the right students are chosen, one has no insurance that the project will not be delayed. For example, the game "I-cannot-breathe" was not quite finished when the students submitted their bachelor thesis. The PhD candidate had to cover the expenses when one of the bachelor students finished the game. The consequences of not finishing this SG and the PhD project could have been great. As part of the environmental considerations (3.1.2), there should have been a backup plan for development software and the involvement of the ICT department from day one. Based on experiences from the PhD project, such a backup plan was made for the game "Hans and the welfare technology." Furthermore, a technical solution was chosen that made the clients less dependent on the students' work in relation to adding text.

According to Dollinger and Lodge [6], lack of time and included assessment may cause barriers in the co-creation process. The students in the first project managed to participate in usability testing of the game "I-cannot-breathe" before submitting the bachelor thesis. The testing was beneficial to both the students and the PhD candidate: the students got input from users to improve the game, and the PhD candidate included the usability test as part of her research. In the project "Hans and the welfare technology," a nursing student followed the whole design process and tested the game frequently. Usability testing was planned as part of the project. However, it could not be carried out due to trouble with recruiting nursing students. This was partly due to an overload of other course evaluations at that time and other research requests. In retrospect, we learned that one should plan the recruitment of students for educational research activities at an earlier stage and in better collaboration with leaders. It is important that such

evaluation activities are planned according to the overall education program and course evaluations. In addition, educational research activities in a study program should preferably be aligned to prevent evaluation overload and the consequences of technology from co-creation projects not being formally and scientifically evaluated. The latter may decrease the value of the final co-creation outcome.

3.2.2 Co-production

Communication and follow-up are important elements in student-staff co-creation [6], as in project management [11]. In retrospect, we recognize that meetings between the development team members and their supervisors could have been held on a more regular basis to agree on design and to discuss challenges in the design and development process. However, not all supervisors attended the scheduled meetings or contacted the project manager during the development process.

As mentioned in 3.1.2, we experienced that the specification of both games should have been more detailed. In retrospect, the two games could have benefited from earlier involvement of the students and their supervisors in planning the games. Then, some of the technical issues probably could have been avoided. For example, with “I-cannot-breathe,” we experienced that the chosen software did not quite fit with the desired functionalities/features within the SG. For example, there were too few possible options that could be chosen in the quiz-based tasks. In addition, the scoring of the tasks needed to be changed. To fit the desired design, the students had to make changes in the software (scripts). Unfortunately, these changes caused some technical issues. In the “Hans and the welfare technology” project, we tried to prevent similar problems by involving an ICT consultant in the meetings with the students. However, in the end, they still chose solutions that made it difficult to upload the editorial solution to the UiA’s servers. Thus, in retrospect, desired functionalities and options within the SG could have been better communicated to the students. Then, the software could have been chosen based on these needs.

In the “I-cannot-breathe” SG, all text and videos had to be included and assembled by the students. The first author spent much time reviewing the text for errors and the desired user interface design. In retrospect, it could have been advantageous to have used software that enabled the teacher to include and edit text and videos, rendering the teacher less dependent on the students. Hence, for the game “Hans and the welfare technology,” this kind of editorial solution was chosen.

3.2.3 Value-in-use

In a university context, value-in-use is related to how students or staff create value for themselves through the co-creation activity. This paper focus on the staff’s point of view. The game “I-cannot-breathe” was uploaded to the university’s database. When the SG was released for use, the number of students that could play the SG at the same time was underestimated. This caused technical glitches, such as sound and video lag when the game was tested by a large group of students for the first time. Fortunately, this was quickly adjusted by the ICT department. Assistance from the ICT department was also necessary when the

game suddenly became unavailable. It turned out that the ICT department had to renew the license.

The game “Hans and the welfare technology” was not integrated into the university’s database but was only made available from an online address paid for by the students. Like the other SG, it suddenly became unavailable due to an unpaid license. However, in this case, all edited text in the game also disappeared and had to be reread.

The responsibility for follow-up for the two games could have been planned and formally handed over to the ICT department earlier. Maybe the risk of unpaid licenses and technical glitches could have been decreased, and the value-in-use increased.

3.3 Outcomes

There are a lot of benefits of co-creation of SGs with students [6] [10]. In our two projects, we had never been able to develop these SGs without collaboration with students from the Department of Information and Communication Technology. This was due to the high cost of purchasing consultant services and commercial SG software. Co-creation of SGs also enables teachers to contribute to game development and develop their own pedagogical and technological competences. As for the students, they can apply for real client projects and contribute to developing solutions based on clients’ descriptions. This can facilitate fulfilment of their course learning outcomes, such as the requirement to apply current knowledge and technology to analyses and solve problems for industry and the public sector. It could also be beneficial for students’ future employability. Hence, co-creation can be a win-win-situation across education programs.

However, co-creations of SGs are often low-budget projects that may decrease the options of functionalities. In addition, co-creation with students is a time- and resource-intensive activity with many risks involved. Therefore, the value of the co-creation process and possible risks must be weighed against the possibilities of using available open-source editorial tools or purchasing SG development software. However, for the latter, we need anchoring from our own department when it comes to financial resources, but also supervision from the ICT department to choose and evaluate appropriate tools for each project. It is important to ensure that the chosen software fit with the desired SG functionalities. Then, development and implementation of SGs as pedagogical tools could be a more available option for pedagogical staff. Perhaps, in the future, SGs could become the new slideshows.

The creation or use of open-source editorial software represents the most sustainable solution [9]; these enable faculty to improve or develop new SGs that align with changes in learning objectives and ensuring evidence-based content.

To enhance the outputs and value of co-creation [6], the university may benefit from providing some guidelines, technological infrastructure, and support for similar student-staff co-creation projects.

5 CONCLUSION

Based on our experiences of challenges and outcomes from the two co-creation projects, we end this paper by

proposing ten key points for planning and conducting cross-disciplinary student–staff co-creation of SGs:

- Make a decision on what level of fidelity and gamification is feasible in your project according to students' technical skills, available time, and budget.
- If available in your organization, use a project template/handbook for student–staff co-creation projects.
- Create a strong interprofessional team (including the students) with well-defined responsibilities and agreements, preferably led by a person with competence in project management.
- The project should obtain an overview of its uncertainty/risk and how it should be handled.
- Develop comprehensive and detailed specification requirements for the SG in collaboration with the interprofessional team.
- Do not underestimate the time needed for faculty members and other professionals to develop and quality-assure a storyboard and decision tree and the pedagogical academic content in the SG's design.
- Chose a SG editorial software where the educational staff can be less dependent on students and technical consultants concerning adding and adjusting text and other content.
- There should be an established communication pattern in the follow-up; ensure that the responsibility for sustainability is formally handed over to the ICT department.
- Evaluation activities must be planned and anchored early in project development.
- Experience and knowledge from the project should be disseminated to serve as inspiration for similar student–staff co-creation projects.

6 REFERENCES

- [1] Stuckless, P., Hogan, M., & Kapralos, B. (2014). Virtual simulations and serious games in community health nursing education: A review of the literature. In M. Ma, L.C. Jain, & P. Anderson (Eds.), *Virtual, augmented reality and serious games for healthcare* (pp. 145-158). London, UK: Springer
- [2] Cant, R.P., & Cooper, S.J. (2014). Simulation in the internet age: The place of web-based simulation in nursing education. *An integrative review. Nurse Education Today*, 34, 1435- 1442.
- [3] Laamarti, F., Eid, M., & El Saddik, A. (2014). An overview of serious games. *International Journal of Computer Games Technology*, 2014, 1-15.
- [4] Annetta, L.A. (2010). The "I's" have it: A framework for serious educational game design. *Review of General Psychology*, 14, 105-112.
- [5] Wattanasoontorn, V., Boada, I., García, R., & Sbert, M. (2013). Serious games for health. *Entertainment Computing*, 4, 231-247.
- [6] Mollie Dollinger & Jason Lodge (2020) Student-staff co-creation in higher education: an evidence-informed model to support future design and implementation, *Journal of Higher Education Policy and Management*, 42:5, 532-546.
- [7] University og Agder. Strategy 2021 → 2024. <https://www.uia.no/en/about-uia/organisation/strategy-2021-2024>
- [8] The Ministry of Education & Research. (2017). White Paper No. 16 (2016-2017), Culture for quality in higher education [Kultur for kvalitet i høyere utdanning]. <https://www.regjeringen.no/no/dokumenter/meld.-st.-16-20162017/id2536007/>.
- [9] The United nations. Sustainable development goals. <https://www.un.org/sustainabledevelopment/sustainable-development-goals/>
- [10] Mercer-Mapstone, L., Dvorakova, S. L., Matthews, K. E., Abbot, S., Cheng, B., Felten, P., Knorr, K., Marquis, E., Shamma, R., & Swaim, K. (2017). A Systematic Literature Review of Students as Partners in Higher Education. *International Journal for Students As Partners*, 1(1).
- [11] Andersen, E. S., Grude, K. V., & Haug, T. (2016). *Måltrettet prosjektstyring* (7. utgave). Fagbokforlaget.
- [12] Johnsen, H.M., Fossum, M., VivekanandaSchmidt, P., Fruhling, A., & Slettebø, Å. (2016). Teaching clinical reasoning and decision-making skills to nursing students: Design, development, and usability evaluation of a serious game. *International Journal of Medical Informatics*, 94, 39-48.
- [13] Møller-Stray, D., Lien, C.T., Skogen, S.E., Birgitsdatter Haugland, E. (2015). *Nettbasert e-læringsverktøy for opplæring av spl.studenter. DAT303 Bacheloroppgave*.
- [14] Lunde, H.B., Alvir, L., Larsen, M.A.S. (2021). *Sykeplayer - En visuell roman tilnærming for e-læring i sykepleien. Bacheloroppgave DAT304*.
- [15] Jensen, Eddy. (2021). *Kunsten av å kommunisere følelser med animasjon - en interaktiv E-lærings modul. Bacheloroppgave DAT303*.
- [16] Kaczmarczyk, J., Davidson, R., Bryden, D., Haselden, S., & Vivekananda-Schmidt, P. (2015). Learning decision making through serious games. *Clin. Teach.*, 12, 1-6.
- [17] Olsen, T., Procci, K., & Bowers, C. (2011). *Serious Games Usability Testing: How to Ensure Proper Usability, Playability, and Effectiveness*. In Marcus, A (Ed.), *Design, User Experience, and Usability. Theory, Methods, Tools and Practice. DUXU 2011. Lecture Notes in Computer Science*, vol 6770. Springer, Berlin, Heidelberg

7 ACKNOWLEDGEMENTS

First, we would like to give special thanks to all the students who contributed. Second, a great thanks to the head of the involved education programs in the Department of Information and Communication Technology and all the supervisors of the two projects. We would also like to thank the ICT department for its assistance.

A Scoping Review of Diabetes Telemedicine Research in Norway

Dillys Larbi^{1,2}, Henriette Lauvhaug Nybakke¹, Karianne Lind¹, Eirin Rødseth¹ and Elia Gabarron^{1,3}

¹ Norwegian Centre for E-health Research, University Hospital of North Norway, Norway, dillys.larbi@ehealthresearch.no

² Department of Clinical Medicine, UiT The Arctic University of Norway, Norway

³ Institute for Education, ICT and Learning, Østfold University College, Norway

Abstract

The recent pandemic highlighted telemedicine's potential for continuity of remote diabetes patients' care. The study objective was to identify diabetes telemedicine services, benefits, and challenges in Norway. We searched for publications on the topic in PubMed, ScienceDirect, CINAHL, and Nora. Most of the included studies (7/15) focused on telemedicine for type 2 diabetes. Telemedicine benefits include improved self-management and cost and time effectiveness. Challenges include organizational and technical issues. To optimize the health system, telemedicine can be used for highly engaged diabetes patients. Creating clear and practical national and organizational telemedicine guidelines for diabetes management could solve the identified challenges.

Keywords

Diabetes; Telemedicine; Remote consultation; Videoconferencing; Remote sensing technology

1 INTRODUCTION

About 245,000 Norwegians (1 in 20) have been diagnosed with diabetes, and it is estimated that about 60,000 more could have undiagnosed diabetes in Norway [1]. The World Health Organization (WHO) describes diabetes as a chronic metabolic disease characterized by elevated levels of blood glucose or blood sugar [2]. There are serious and life-threatening complications associated with diabetes, including damage to vital organs such as the eyes, kidneys, nerves, and the heart [2,3].

To prevent these complications and premature death, people living with diabetes need constant, continuous, and coordinated care. This, however, becomes challenging when it involves travelling long distances to healthcare facilities or restricted access to healthcare due to disease outbreaks. The use of telemedicine could offer alternatives to standard care for people with diabetes [4]. Telemedicine basically involves healthcare delivery across geographical locations via electronic communication technologies [5].

Telemedicine solutions offer opportunities for the healthcare system to continue the care of long-distance patients living with diabetes, for example during infectious disease outbreaks [4,6]. Although Norway as a country adopted telemedicine at an early stage [7], very little is known about the benefits and challenges of using telemedicine solutions for diabetes care in Norway. Therefore, we conducted a scoping review to provide an overview of the evidence as preliminary assessment of the size and scope of available literature in this field.

The objective of this article is to identify what is known about the benefits, and challenges of telemedicine for diabetes in Norway. This includes benefits and challenges for all types of diabetes and within all areas where telemedicine for diabetes has been used in order to provide an overview of the field.

The 18th Scandinavian Conference on Health informatics, Tromsø, Norway, August 22-24, 2022. Organized by UiT The Arctic University of Norway. Conference Proceedings published by Linköping University Electronic Press at <https://doi.org/10.3384/ecp187>. © The Author(s). This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>

2 METHODS

We searched for publications with the keywords “Diabetes” AND “Telemedicine” OR “Remote consultation” OR “Videoconferencing” OR “Telemetry” OR “Remote sensing technology” in titles/abstracts AND Norway (where possible). The search was carried out in the following databases: PubMed, ScienceDirect, CINAHL, and Nora (Norwegian knowledge repository). Articles were included if they were 1) studies on use of telemedicine for diabetes care 2) conducted in Norway. No year or language limitations were used. The full search strategy is available in the Zenodo data repository [8].

All references were uploaded to EndNote 20 and X9 by Clarivate™, and duplicates removed. After that, the references were uploaded to Rayyan and DL, KFL, and EG carried out the first screening of titles and abstracts. In the second phase, the full texts of the remaining articles were downloaded. Two pairs of reviewers (HLN and KFL; DL and EG) carefully examined these full texts in order to confirm their eligibility. Eligibility incongruences were discussed with the other pair of reviewers until reaching an agreement. Articles that met the inclusion criteria were included in the qualitative synthesis. Benefits and challenges linked to technology were identified, summarized, and categorized using an inductive thematic analysis in NVivo 12 Pro for windows by QSR International.

3 RESULTS

A total of 542 records were identified and 15 met the inclusion criteria (see Figure 1). The search strategy and the list of excluded articles in the full-text screening, and the reasons for exclusion are available in the data repository [8]. Table 1 summarizes the publications included in this review.

3.1 Publication and targeted diabetes type

The 15 included studies were published between 2003 and 2020. Among these, 6 were randomized controlled trials [9-14]; 4 observational or qualitative studies [15-18]; 2 study protocols [19,20]; 1 feasibility and usability study [21]; 1 project summary [22] and 1 editorial article [6].

Of the 15 included studies, 7 were about type 2 diabetes only (7/15) [9,13,14,18,20-22]; 4 referred to both type 1 and type 2 diabetes [10-12,19]; 3 to diabetes in general [15-17]; and 1 publication was about type 1 diabetes [6].

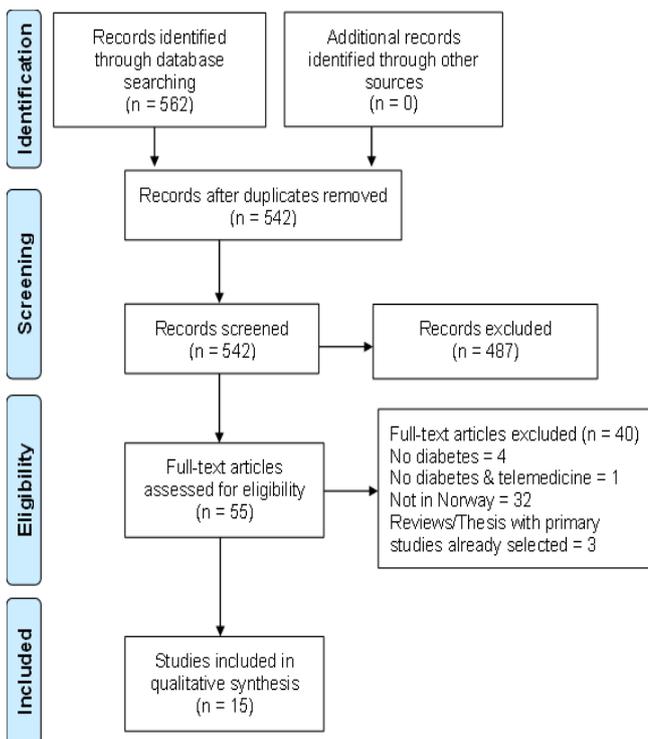


Figure 1. Flowchart of the selection procedure.

3.2 Diabetes telemedicine technologies

Six of the included articles were about telemedicine services via interactive wound platforms [10-12,16,17,19]; 5 articles were about the use of self-management systems via mobile phones [9,13,14,20,21]; 2 focused on telemedicine as a general concept [6,15]; and 2 referred to the use of image sharing technologies for diabetes management and care [18,22].

In 7 of the studies, diagnosing and counteracting diabetes complications were achieved through the use of telemedicine technologies [10-12,15-17,19]; and also ophthalmology complications [14]. Some studies (6/15) used telemedicine technology to monitor the blood glucose levels of people living with diabetes [6,9,13,14,20,21]. One of the publications did not specify the use of telemedicine technology [22].

3.3 Diabetes telemedicine benefits and challenges

The identified benefits of using diabetes telemedicine technologies include cost effectiveness [22] increased knowledge of managing diabetes complications such as wound assessment [15,17], and time effectiveness [18].

The included studies identified challenges associated with the use of telemedicine technologies for diabetes in Norway. These challenges include technical and practical issues associated with outdated equipment and inadequate technological skills among staff [12,17,21], organizational issues as a result of interorganizational and interdisciplinary collaboration to deliver telemedicine services [11,16,18], and communication and information issues due to missing elements associated with traditional delivery of care such as physical examination and non-verbal communication.

All the identified benefits and challenges are summarized in Table 2.

4 DISCUSSION

This review aimed to identify what is known about the benefits and challenges of telemedicine for diabetes in Norway. Most of the included studies focused on the use of telemedicine for type 2 diabetes, followed by type 1 diabetes.

We did not find any publication on telemedicine for gestational diabetes, which could be of relevance, for example during infectious diseases outbreaks. The results are summarized in table 2.

Although the included studies reported no significant changes in HbA1c between the intervention and the control groups [9,13], there was a decline in HbA1c levels among all participants. Similarly, mobile applications as an advanced telediabetology method have been shown to reduce HbA1c levels in people living with type 1 and 2 diabetes, as well as minimize the occurrence of hypoglycaemic events [4,23]. There is also a reported increase in quality of life among individuals with type 1 and 2 diabetes who use mobile applications for self-management. This is due to the improved design and development of these technological tools with features such as bolus calculator, carbohydrate counting, automated glucose pattern feedback, and the ability to share one's data with healthcare providers [4].

Iversen et al. [10] found that there was no significant difference between the use of telemedicine and usual care for treating foot ulcers. Cost and time effectiveness can be realised from the use of telemedicine as an alternative to usual care in remote regions in Norway since patients can visit their local health facilities while having access to specialist care. Lee et al. [24] associated cost-effectiveness with the use of telemedicine for retinal screening, telemonitoring, and telephone reminders in diabetes management. It seems that telemedicine can be offered as a low-cost alternative to people living with diabetes in Norway who are already engaged in the health system to ensure cost-effectiveness without compromising on the quality of care [25].

The use of telemedicine can increase the knowledge and skills of healthcare providers in diagnosing and treating diabetes complications [15,17]. However, this acquired knowledge will not be useful if there are identified organizational challenges associated with the use of telemedicine services and technologies. Challenges encountered during the use of telemedicine for diabetes care were; The problem of identifying who is responsible

for the care of the patient because healthcare professionals are working across different management systems and organizational structures [11,16,18], technical, and practical challenges where available equipment are outdated, and applications do not work as intended [12,17,21]. These challenges need to be properly and effectively addressed. Aberer et al. [4] suggest political and structural adjustments as a solution to diabetes telemedicine-related challenges. Adjustments made at the national and organizational level for the use of telemedicine for diabetes can contribute immensely to improved telemedicine services and quality of life of people living with diabetes in the country.

4.1 Limitations

Our review focused only on telemedicine and diabetes in the Norwegian context, and we used only few keywords in our data search. Therefore, we might have missed relevant publications on the topic carried out within the Norwegian context. Our findings cannot be generalized to other countries or to other fields where telemedicine is used. All publications we found were focusing on Type 1 and Type 2 diabetes, and none in gestational diabetes. Future research could also study the potential benefits of telemedicine for managing gestational diabetes in Norway.

5 CONCLUSIONS

There is evidence to suggest that telemedicine for diabetes management and care can be adopted in Norway to supplement the usual care. People living with diabetes who are interested in and enthusiastic about monitoring their condition can be offered telemedicine services as an alternative to usual care, especially those who live in remote areas and for follow-up purposes. Telemedicine as a supplement to usual care is beneficial to both the health system as a whole and to people living with diabetes in terms of reduced cost and efficiency of care.

Creating clear and practical national and organizational guidelines for telemedicine for diabetes care in Norway could be a way to solve the various identified challenges associated with its use.

6 AUTHORS CONTRIBUTIONS

Conceptualization: EG, ER, DL, KL, HLN; Literature searching: EG, KL; Title, abstract, and full-text screening: EG, KL, DL, HLN; Data extraction: EG, KL, DL, HLN; Data analysis and interpretation: EG, ER, DL, KL, HLN. Article Writing: EG, ER, DL, KL, HLN. All authors have read and agreed to the published version of the manuscript.

7 ACKNOWLEDGEMENT

This project is funded by the EEA and Norway Grants.

8 REFERENCES

[1] Norwegian Institute of Public Health. "Diabetes in Norway" 2021 Available from: <https://www.fhi.no/en/op/hin/health-disease/diabetes-in-norway---public-health/> (Accessed: 18th November, 2021).

[2] World Health Organization. "Diabetes" 2022 Available from: https://www.who.int/health-topics/diabetes#tab=tab_1 (Accessed: 19th May, 2022).

[3] World Health Organization. "Diabetes" 2020 Available from: <https://www.who.int/news-room/facts-in-pictures/detail/diabetes> (Accessed: 19th May, 2022).

[4] Aberer F, Hochfellner DA, Mader JK. "Application of Telemedicine in Diabetes Care: The Time is Now" in *Diabetes Ther* Vol. 12, Issue 3, pp. 629-39. 2021.

[5] Zanaboni P, Knarvik U, Wootton R. "Adoption of routine telemedicine in Norway: the current picture" in *Glob Health Action* Vol. 7, p. 22801. 2014.

[6] Birkeland KI. "Some Lessons Learned About Diabetes and COVID-19 During the Early Stage of the Epidemic in Norway" in *J Diabetes Sci Technol* Vol. 14, Issue 4, pp. 718-9. 2020.

[7] Zanaboni P, Wootton R. "Adoption of routine telemedicine in Norwegian hospitals: progress over 5 years" in *BMC health services research* Vol. 16, p. 496. 2016.

[8] Larbi D, Nybakke Lauvhaug H, Lind K, et al. "Telemedicine for diabetes in Norway: Scoping review appendices [Data Set]" *Zenodo*; 2021 Available from: <https://doi.org/10.5281/zenodo.5718818>.

[9] Holmen H, Torbjørnsen A, Wahl AK, et al. "A Mobile Health Intervention for Self-Management and Lifestyle Change for Persons With Type 2 Diabetes, Part 2: One-Year Results From the Norwegian Randomized Controlled Trial RENEWING HEALTH" in *JMIR Mhealth Uhealth* Vol. 2, Issue 4, p. e57. 2014.

[10] Iversen MM, Iglund J, Smith-Strøm H, et al. "Effect of a telemedicine intervention for diabetes-related foot ulcers on health, well-being and quality of life: secondary outcomes from a cluster randomized controlled trial (DiaFOTo)" in *BMC Endocr Disord* Vol. 20, Issue 1, p. 157. 2020.

[11] Smith-Strøm H, Iglund J, Østbye T, et al. "The Effect of Telemedicine Follow-up Care on Diabetes-Related Foot Ulcers: A Cluster-Randomized Controlled Noninferiority Trial" in *Diabetes Care* Vol. 41, Issue 1, pp. 96-103. 2018.

[12] Smith-Strøm H, Iversen MM, Graue M, et al. "An integrated wound-care pathway, supported by telemedicine, and competent wound management-Essential in follow-up care of adults with diabetic foot ulcers" in *Int J Med Inform* Vol. 94, pp. 59-66. 2016.

[13] Torbjørnsen A, Jennum AK, Småstuen MC, et al. "A Low-Intensity Mobile Health Intervention With and Without Health Counseling for Persons With Type 2 Diabetes, Part 1: Baseline and Short-Term Results From a Randomized Controlled Trial in the Norwegian Part of RENEWING HEALTH" in *JMIR Mhealth Uhealth* Vol. 2, Issue 4, p. e52. 2014.

[14] Torbjørnsen A, Småstuen MC, Jennum AK, et al. "Acceptability of an mHealth App Intervention for Persons With Type 2 Diabetes and its Associations With Initial Self-Management: Randomized Controlled Trial" in *JMIR Mhealth Uhealth* Vol. 6, Issue 5, p. e125. 2018.

- [15] Kolltveit BC, Gjengedal E, Graue M, et al. "Telemedicine in diabetes foot care delivery: health care professionals' experience" in *BMC Health Serv Res* Vol. 16, p. 134. 2016.
- [16] Kolltveit BH, Gjengedal E, Graue M, et al. "Conditions for success in introducing telemedicine in diabetes foot care: a qualitative inquiry" in *BMC Nurs* Vol. 16, p. 2. 2017.
- [17] Kolltveit BH, Thorne S, Graue M, et al. "Telemedicine follow-up facilitates more comprehensive diabetes foot ulcer care: A qualitative study in home-based and specialist health care" in *J Clin Nurs* Vol. 27, Issue 5-6, pp. e1134-e45. 2018.
- [18] Rotvold GH, Knarvik U, Johansen MA, et al. "Telemedicine screening for diabetic retinopathy: staff and patient satisfaction" in *J Telemed Telecare* Vol. 9, Issue 2, pp. 109-13. 2003.
- [19] Iversen MM, Espehaug B, Hausken MF, et al. "Telemedicine Versus Standard Follow-Up Care for Diabetes-Related Foot Ulcers: Protocol for a Cluster Randomized Controlled Noninferiority Trial (DiaFOTo)" in *JMIR Res Protoc* Vol. 5, Issue 3, p. e148. 2016.
- [20] Ribu L, Holmen H, Torbjørnsen A, et al. "Low-intensity self-management intervention for persons with type 2 diabetes using a mobile phone-based diabetes diary, with and without health counseling and motivational interviewing: protocol for a randomized controlled trial" in *JMIR Res Protoc* Vol. 2, Issue 2, p. e34. 2013.
- [21] Arsand E, Tataru N, Østengen G, et al. "Mobile phone-based self-management tools for type 2 diabetes: the few touch application" in *J Diabetes Sci Technol* Vol. 4, Issue 2, pp. 328-36. 2010.
- [22] Hernández C, Alonso A, Garcia-Aymerich J, et al. "Integrated care services: lessons learned from the deployment of the NEXES project" in *Int J Integr Care* Vol. 15, p. e006. 2015.
- [23] Doupis J, Festas G, Tsilivigos C, et al. "Smartphone-Based Technology in Diabetes Management" in *Diabetes Ther* Vol. 11, Issue 3, pp. 607-19. 2020.
- [24] Lee JY, Lee SWH. "Telemedicine Cost-Effectiveness for Diabetes Management: A Systematic Review" in *Diabetes Technol Ther* Vol. 20, Issue 7, pp. 492-500. 2018.
- [25] McDonnell ME. "Telemedicine in Complex Diabetes Management" in *Curr Diab Rep* Vol. 18, Issue 7, p. 42. 2018.

Reference	Publication type	Diabetes type
Rotvold GH, et al. (2003) [18]	Observational / Qualitative study	Type 2 diabetes
Årsand E, et al. (2010) [21]	Feasibility and usability study	Type 2 diabetes
Ribu L, et al (2013) [20]	Study protocol	Type 2 diabetes
Holmen H, et al. (2014) [9]	RCT	Type 2 diabetes
Torbjørnsen A, et al. (2014) [13]	RCT	Type 2 diabetes
Hernández C, et al. (2015) [22]	Summary of project	Type 2 diabetes
Iversen MM, et al. (2016) [19]	Study protocol	Type 1 and Type 2 Diabetes
Kolltveit BC, et al. (2016) [15]	Observational / Qualitative study	Diabetes in general
Smith-Strøm H, et al. (2016) [12]	RCT	Type 1 and Type 2 Diabetes
Kolltveit BH, et al. (2017) [16]	Observational / Qualitative study	Diabetes in general
Kolltveit BH, et al. (2018) [17]	Observational / Qualitative study	Diabetes in general
Smith-Strøm H, et al. (2018) [11]	RCT	Type 1 and Type 2 Diabetes
Torbjørnsen A, et al. (2018) [14]	RCT	Type 2 diabetes
Birkeland KI. (2020) [6]	Editorial	Type 1 diabetes
Iversen MM, et al. (2020) [10]	RCT	Type 1 and Type 2 Diabetes

Table 1. Summary of publications included in the review (n=15)

Benefits	Challenges
Improved self-management and technology used as a self-help aid [9,13,14,21]	No significant effect or impact (No differences in HbA1c-level, no significant effect on foot ulcers treatment, no significant difference in consultations, no significant impact on self-management) [9,10,13]
Alternative or supplement to usual care [10-12]	Technical and practical challenges (outdated equipment, technical problems, technological skills among the staff, app-related) [12,17,21]
Increase in wound assessment knowledge and skills in the nursing staff [15,17]	Organizational challenges (health care professionals working across different management systems and organizational structures, between the primary care sector and the specialist care sector – who is responsible for the care?) [11,16,18]
Cost-effectiveness (avoidance of costly institutional care, e.g., hospital admissions) [22]	Communication and information challenges (lack of information from physical examinations and nonverbal communication, communication among stakeholders across health care tiers) [6,22]
Time effectiveness (quicker to grade the level of retinopathy) [18]	The documentation process was time-consuming [15]
Greater reach and reduced travel distance [10,11]	The Norwegian legislation on data privacy and transfer was identified as a major limitation for the deployment of integrated care [22]

Table 2. Benefits and challenges associated with the use of telemedicine technology

Development of an Interactive Communication Model with Integrated Teach-Back

– using a web-based IT solution to create synergy between research and practice.

Lisa Korsbakke Emtækær Hæsum^{1,2} and Ole Kristian Hejlesen²

¹Department of Nursing, University College of Northern Denmark, Aalborg, Denmark. lkeh@ucn.dk

²Department of Health Science and Technology, Faculty of Medicine, Aalborg University, Aalborg, Denmark

Abstract

This paper describes the development of an easily applicable web-based IT solution that enhances interactive communication (the interactive communication model) and ensures comprehension in nursing practice. The model seeks to identify knowledge and skills allowing tailored communication and seeks to ensure comprehension and recall between nurses and patients. Results from testing the model shows that it has the potential to enhance self-care in citizens receiving community nursing and creates a basis for a more holistic nursing approach.

Keywords

Health literacy, web-based, IT, user-involvement, communication.

1 INTRODUCTION

Modern healthcare systems are developing in a way that compel patients and citizens to become more active in the management of their own disease(s), health, and life situation. This development changes the role of modern patients and the skills needed to navigate the healthcare system. There is increasing pressure on healthcare resources due to demographic changes with more elderly people and thus, an increase in the prevalence of chronic diseases [1]. A strategy for handling the increasing pressure on healthcare resources is to reduce the length of stay in hospitals and promote more healthcare in patients' and citizens' own homes. This strategy calls upon more (self) rehabilitation actions, where the goal is to strengthen self-management and self-care among citizens and patients. This requires a general adherence to treatment (during the self-management of chronic disease(s)) and an ability to be an active part in shared decision making with healthcare professionals, thus requiring that citizens and patients increase their understanding and application of health information [2–4]. Overall, the development in modern healthcare systems set some requirements to citizens' and patients' level of self-management and patient empowerment. Health literacy (HL) can be considered a prerequisite for these concepts and plays an important role in strengthening them [5]. HL is defined by the World Health Organization as *'the cognitive and social skills which determine the motivation and ability of individuals to gain access to, understand and use information in ways which promote and maintain good health'*[6]. As seen in the definition, HL is a complex and broad concept that comprises multiple personal skills and different levels (access, understand and use), which make it difficult to

access [7].

A Danish study shows that 10-20% of the Danish population experience difficulties with the ability to interact with healthcare professionals and the ability to understand health-related information well enough to act on it (two important dimensions of HL) [8]. This suggest that they do not have the necessary HL skills to handle their health sufficiently, which produces a higher risk of adverse health effects for example improper use of healthcare services and medications, poor self-management, and poor health outcomes [9]. This is problematic when considering the development in healthcare systems that emphasis a higher level of self-management and self-care. Based on this, clear communication and ensuring recall and comprehension becomes more essential in patient-provider interactions. It is particularly important in relation to patients with chronic disease(s) because self-management educational efforts and counselling are key elements in handling everyday life with chronic disease(s) and strengthening self-care. Clear communication and ensuring comprehension become even more important if patients with chronic disease experience challenges in accessing, understanding, and applying health-related information necessary to efficiently manage their health and make informed health decisions [10]. Patient-provider communication has been proposed as a potential pathway through which HL might influence health outcomes, especially in individuals with chronic disease(s). Hence, tailoring communication to the individual level of HL, might have a positive influence on health outcomes [11,12]. Nurses play an essential role in providing education and counselling to patients with a chronic disease through communication – it is considered a core competency in nursing practice [13]. However, nurses rarely use

The 18th Scandinavian Conference on Health informatics, Tromsø, Norway, August 22-24, 2022. Organized by UiT The Arctic University of Norway. Conference Proceedings published by Linköping University Electronic Press at <https://doi.org/10.3384/ecp187>. © The Author(s). This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>

communication that focus on recall and comprehension in their interaction with patients [14]. On this basis, this pilot study focuses on the development of a web-based interactive communication model (ICM) that both seeks to identify knowledge and skills (HL) allowing tailored communication and seeks to ensure comprehension and recall between nurses and patients.

2 METHODS

This study includes two phases: the development of the ICM and a pilot study of the model in community nursing.

2.1 Development of the interactive communication model

The IT-based ICM is designed based on data from 335 community nurses in Aalborg Municipality. The 335 community nurses participated in workshops (14 workshops) centred on user-involvement. The community nurses discussed and made a comprehensive answer to several questions about existing practices and challenges in (self) rehabilitation. The discussion was carried out in groups of 4-6 community nurses. The knowledge and experience gathered from practice formed the basis for the development of the ICM, which was subsequently qualified via recognized methods, to make citizens more self-reliant (regardless of HL level), identified via systematic literature search. In this way, synergy has been created between theory and practice, and a best-practice approach was mapped for the development and application of strategies that strengthen self-care and the degree of self-help in citizens with chronic disease(s). The workshops were facilitated by one researcher (LKEH, first author of the present paper) who's a registered nurse (RN). LKEH have a solid experience of work in healthcare. The content of the ICM was created based on the material from the workshops combined with the best-practice approach identified through the systematic literature search. The ICM is implemented as a web-based IT solution to enable a straight-forward integration with the software on the nurses' working tablet. The IT solution contained all the steps and information needed to use the ICM in everyday nursing practice and it is easy to make accessible on the start page of their working tablet.

2.2 Pilot study

Following an observational approach, the focus was on observing nurses' actions in an everyday community nursing context and the aim was to observe responses and challenges related to the use of the ICM as part of community nursing practice. One district of community nursing was selected as setting for the pilot study. The district comprises six community nurses, who were included in the study. The community nurses were introduced to the ICM prior to observation. Participant observations and informal interviews were used as general data collection methods.

The observations were conducted by LKEH. The presence of the researcher could, however, affect the observations and this must be considered. To minimize the effect of the researcher, she adapted to the clinical environment by wearing the same uniform as the community nurses and no nameplate. In this manner, there was access to the healthcare environment while at the same time showing not

to be seen as healthcare personnel handling citizens with care needs. Thus, the researcher could participate without being directly involved or being a distraction.

Participant observations were carried out for five weeks and guided by the aim of the pilot study. During the observations the researcher took thorough notes guided by an observational guide. Informal interviews were primarily conducted in the community nursing cars when driving between visits among citizens with care needs, but also sometimes in the coffee room. These informal interviews were mainly conducted to add further information to the observations and to create a more comprehensive understanding of what had been observed. Notes from both the observations and the informal interviews were transcribed after each observation. Six community nurses completed a brief survey (8 questions) about their use of the ICM at the end of the pilot study. The survey was constructed based on literature and focuses on evaluation of using the ICM in community nursing; for instance, they were asked to evaluate if the model enhances their awareness on knowledge and competences among their citizens on a Likert scale (see *Appendix A*). SurveyXact was used to distribute and administer the surveys.

2.2.1 Ethical considerations

Permission to carry out the pilot study was given by the head nurse in the selected district. Oral consent was obtained from the six community nurses in the district in accordance with the Declaration of Helsinki [15] after they were informed of the purpose, method and publication of the pilot study, that participation was voluntary, and they could withdraw at any time. No ethical approval was required for this type of pilot study. Citizens present in the observations were informed of the purpose of the pilot study and that the researcher was bound by professional secrecy in her role as health care professional.

2.3 Analysis

2.3.1 The interactive communication model (ICM)

Data from the 14 workshops (335 participating community nurses) were analysed using thematic analysis [16]. Thematic analysis is a basic qualitative analytical approach that seeks to identify themes or patterns in the gathered material (in this case material gathered from the 14 workshops). The analysis was characterized by an inductive data-driven approach, which resulted in themes that are closely linked to data. The coding of data can take place without relation to a specific framework or a specific analytical (pre) understanding.

The material from the 14 workshops were read and coded openly conducting descriptive coding [17]. The material was read several times and meaning units related to challenges with conducting community nursing aiming to strengthen self-care among citizens receiving community nursing were detected. These meaning units were then assigned a code describing their content. Meaning units and data text were scrutinized. On this basis, relevant categories and themes and categories were identified throughout the data material. The analysis comprised continuous discussions between authors and a lecturer in thematic analysis to reach consensus. The original data was re-examined in case of discrepancies.

2.3.2 The pilot study

The transcribed notes from observations and informal interviews were also analysed using thematic analysis [16]. The material was read several times and meaning units related to using the ICM were identified. These meaning units were then assigned a code describing their content. Meaning units and data text were scrutinised thoroughly. On this basis, relevant categories and themes, related to challenges with using the ICM to strengthen self-care in community nursing, were identified throughout the data material.

Data from the brief surveys were analysed using basic statistics [18]. Data from the observational study, and the informal interviews, was combined with data from the surveys, hence, data triangulation was carried out.

3 RESULTS

This presentation of results will comprise three sections: the content of the ICM, results from the pilot study and results from the survey.

3.1 The interactive communication model

3.1.1 Content in the ICM

The material from the 14 workshops, in conjunction with relevant research literature, has formed the basis for the development of the ICM with integrated teach-back. The ICM is inspired by Schillinger et al. [19] and comprises the following components:

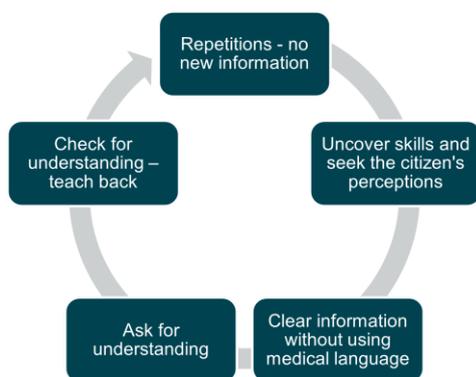


Figure 1. Illustration of the interactive communication model (ICM) inspired by Schillinger et al. [19].

Step 1 - Repetitions: The model is initiated with a repetition of key points (very few sentences) from the last visit / learning session. The community nurse simply repeats key points and provide no new information at this stage. By the first visit, the community nurse cannot perform repetitions, but instead the citizen is presented to the plan for enhancing their knowledge and self-care.

Step 2 - Uncover skills and seek the citizen's perceptions: This allows the community nurse and citizen to reach a common understanding. Moreover, it helps the community nurse assess how to tailor information and instructions for the individual citizen.

Step 3 - Clear information without using medical language: The community nurse provides information, instructions, and guidance at a very easy-to-understand level without using medical language. The information, instructions, and

guidance are provided based on predefined material made available to the community nurse from the start page of her /his working tablet. The information, instructions and guidance are supplemented by the delivery of easy-to-understand material to the citizen.

Step 4 - Ask for understanding: The community nurse asks the citizen directly if they have understood the given information and what has been discussed.

Step 5 - Check for understanding – teach back: The community nurse uses the teach-back technique to ensure that the citizen understands the information/instructions and what has been discussed. The community nurse asks the citizen to retell (in their own words) the information and instructions provided. This stage may also involve the citizen having to demonstrate the performance of a ‘task’, for example, the community nurse asks the citizen, with chronic obstructive pulmonary disease, to demonstrate their inhalation technique. This an indirect method to check for understanding.

3.1.2 Development of the web-based IT solution

The ICM must be easily accessible and an efficient tool for community nurses, therefore it is made available as a web-based IT solution. The tool is entitled ‘personalised support for self-care’ and can be accessed from the start page of the community nurses’ working tablet.



Figure 2. Screenshot of the start page of the web-based IT solution ‘Personalised support for self-care’.

The content of the IT solution is selected based on the most prominent diseases and reasons for receiving community nursing in Northern Jutland. People receiving community nursing are often very vulnerable with multiple health-related challenges and comorbidities – setting high requirements to their level of HL. The idea is that the ICM (made available in the web-based IT solution) supports community nurses in accommodating citizens' broad range of information and communication preferences. Instead of assuming what people know, the model uncovers what they know, thus, it is a more objective approach to accessing knowledge and skills.



Figure 3. Screenshot of the pull-down menu with items representing the most prominent diseases and reasons for receiving community nursing: ‘The interactive communication model’, ‘COPD’, ‘Diabetes and insulin’, ‘Diabetes and diabetic ulcers’, ‘Venous ulcers’, ‘Arterial ulcers’, ‘Compression’, ‘Heart failure’, and ‘Medicine dispensing’.

The IT solution comprises a general description of the ICM and how to perform each individual communicative step in the model (the first item in the pull-down menu – see figure 3). The description is supported with a graphical illustration of the different steps in the model. The instructions on how to use and implement the different communicative steps in the model and incorporate it in practice creates the

foundation for emphasising use of interactive communication in community nursing.

Further, the IT solution demonstrates how to use the ICM in relation to the most prominent diseases such as COPD, diabetes, cardiovascular diseases etc. It is demonstrated how the nurse should perform the individual steps related to the specific disease - all the way down to how questions should be formulated and asked to uncover citizens' knowledge and level of skills (*step 2 - ask for understanding*) and how information is provided in a concise easy-to-understand manner without medical language (*step 3 – clear information without using medical language*). Hence, the nurses do not have to decide for themselves what characterises easy-to-understand language as it is provided by the web-based system. The information and instructions formulated in the system has been developed in the following manner: information about the specific diseases is inspired by an acknowledged national patient centred webpage (sundhed.dk) and pedagogical tools with a focus on increasing and ensuring understanding with the recipient [20,21]. The material in the system is kept on a readability of 6-7th grade level and, thus, the community nurse provides information and guidance on this level when using the ICM. The material was read thoroughly by independent reviewers to ensure concise and easy-to-understand language not exceeding 6-7th grade level. It should be noted that the system also provides different links and video material (easy-to-understand that can be used by the community nurses in *step 3 – clear information without medical language*) relevant for the specific diseases.

Even though, the information and instructions in the system serves as a guide to learn the ICM technique, the goal is that the community nurses adapt the communication technique into their everyday practice. Hence, the system is intended to facilitate the use of a very easy-to-understand language, as well as to check the understanding of information by the recipient.

The web-based IT solution was informally tested and evaluated by two nurses prior to the pilot study.

3.2 Results from the pilot study

3.2.1 Proper introduction and training prior to use

The use of the web-based ICM model requires adequate training and support. It requires understanding and experience with the different steps in the model and an understanding of the purpose in each step. When the community nurses started using the model (in the beginning of the pilot study), they were a little unsure of how they could integrate it into their communication without it being ‘staged’ or ‘forced’. As a result, it took some time for the community nurses to get comfortable with using the ICM. However, with enough practice and use, the community nurses managed to integrate it over time and, thus, making their communication more interactive. On this basis, the ICM can serve as a development and training of community nurses' interactive communication skills.

The community nurses were briefly introduced to the use of the ICM prior to the pilot study. When the community nurses adapt the ICM, it can be performed in approximately five minutes – and subsequently have an overall time reducing potential. The time reducing potential of the ICM

is supported by the results from the survey – see paragraph 3.3 below.

3.2.2 A more uniform communication practice

A prominent result of the pilot study is that the web-based ICM ensures clear communication without use of medical jargon and make tailored communication easily accessible. It emphasises a focus on reducing the use of medical words and how difficult it can be to use very simple language when explaining health-related information. The community nurses became more aware of their own communication technique and use of words. Concrete formulations/explanations with regards to the most widespread diseases and community nursing services is available in the system and therefore the community nurses do not subjectively have to decide what is easy to understand health-related information. It became clear in the pilot study that the ICM contributes to a more uniform approach to communication among community nurses. It was also observed that some of the community nurses preferred to use the system (with the ICM) as preparation prior to visits with citizens needing community nursing and others used it as a guide during the visits.

3.2.3 Assessment of understanding and allowing a more holistic approach

The ICM allows identification of understanding and skills among citizens receiving community nursing. An important step in the ICM is the check for understanding via the teach-back technique – results from the pilot study shows that the community nurses often thought that their citizens had more knowledge of their disease and situation than they had. The community nurses expressed that the ICM allowed a more accurate identification of their citizens' understanding and skills.

Further, the community nurses expressed that the ICM allowed for a more holistic nursing approach as it made it easier to identify the individual challenges among their citizens and tailor communication accordingly. They felt that the ICM allowed them to access 'the whole situation' and not just was allocated for the nursing visit. These results are in accordance with the results from the survey – see paragraph 3.3 below.

3.3 Results from the survey

The following will present a short summary of the results from the brief survey at the end of the pilot study. Six community nurses were asked to rank their answers on a Likert scale: *not at all*, *a little*, *some*, *greatly* and *completely* (see appendix A).

1. 83 % of the community nurses think that the ICM greatly or completely increases citizens' knowledge of their own disease(s) and life situation, while the remaining 17% ranked it to some extent.
2. 67 % of the community nurses think that the ICM greatly or completely increases citizens' self-care, while the remaining 33 % think its increases self-care to some extent.
3. All community nurses (100%) think that the ICM completely or greatly enhances their experience of more holistic nursing.

4. All community nurses (100%) think that the ICM completely or greatly help them find out what and how much a citizen knows about their own disease(s) and life situation.
5. 50 % of the community nurses think that the ICM completely or greatly increases their focus on not using medical terms, while 33 % think it increases their focus to some extent. The remaining 17 % think it does not increase their focus on this at all.
6. All community nurses (100%) think that the ICM completely or greatly increases their focus on 'checking' citizens' understanding of the information and guidance they provide as nurses.
7. 66% of the community nurses think that the web-based system is useful, while 17 % find it useful to some extent. 17 % does not find it useful at all.
8. 33 % of the community nurses think that the ICM greatly reduces their time consumption with citizens, while 33 % think it reduces time consumption a little. 33 % does not think the ICM reduces time consumption at all.

4 DISCUSSION

The aim of this pilot study was to develop an easily applicable communication model that enhances interactive communication and ensures comprehension in nursing practice. This has been accomplished by the development of the ICM and making it available as a web-based IT solution. A strength of this pilot study is that the ICM has been developed based on a high degree of user-involvement – the end users (in this case the community nurses) were involved in the development, which create a sense of ownership. The ICM proved able to identify level of knowledge and skills among citizens receiving community nursing and thus, allowing community nurses to target their communication accordingly. It is known from the literature that nurses often overestimate the level of HL (e.g., level of knowledge and skills) [22,23]. Results from the pilot study suggest that the ICM provides a more objective approach to assessing the level of knowledge and skills allowing a more accurate estimation. The more accurate assessment of knowledge and skills (and thereby indication of HL level), further, promotes targeted communication in the citizen-nurse interaction and might enhance self-care among citizens receiving community nursing. The community nurses, further, expressed an experience of providing more holistic nursing, which is rather interesting considering the increasing scarce health resources that put restraints on the time available to perform nursing tasks. Based on this, it is reasonable to assume that the ICM can support efficient and holistic nursing at the same time. In this regard, it should, however, be mentioned that the use of the ICM requires sufficient training and practice prior to use to achieve this. From the pilot study it could be seen that the introduction should have been more in depth, preferably with focus on training the use of ICM for example the community nurses could have trained the use on each other. More in depth training of using the ICM can probably reduce the time for integrating it in practice among nurses and quicker make the nurses more comfortable using it. E.g., enough training and practice are needed to promote the time reducing potential of the ICM.

It became clear in the pilot study that the ICM is a communication technique that nurses can adapt into their everyday practice with adequate training and support. However, it takes time to incorporate the ICM (communication technique) into existing nursing practice, so an initial extra time consumption must be expected in this regard. The literature shows that nurses can be dubious of using universal precautions and simple language because they are concerned it will offend or patronise highly literate citizens and patients [24]. Even though, the ICM emphasis simple language without medical terms, it seeks to uncover skills (in step 2) that allows tailored information to the individual citizen or patient and primarily aims to meet them at their individual level.

The ICM is made available as a web-based IT solution that can easily be accessed from community nurses' working tablet, which promoted the use. It was observed in the pilot study that some of the community nurses preferred to use the system (with the ICM) as preparation prior to visits with citizens needing community nursing and others used it as a guide during the visits. In this regard, it should be noted that there is a possibility of the system 'stealing' the attention, and therefore, efforts should be made to ensure that nurses become so well versed in the ICM model that this is not the case. The ICM seemingly contributes to a development and expansion of communication techniques among community nurses. One could argue that nurses are already sufficiently trained in communication skills and educated to use communication efficiently in their practice, but on the contrary literature shows that basic communication techniques are rarely used [13,14]. Additionally, the nursing profession is undergoing a change due to the development in modern healthcare; the conducting part of nursing is increasingly reduced, while the communicative/guiding part is emphasised to support the self-care line of thought in modern healthcare systems and thus, make citizens and patients more active [25]. This changes the requirement to communication skills and techniques in nursing care, as it becomes one of the most valuable factors in supporting self-care in modern healthcare systems. If nurses overestimate HL (knowledge and skills) and assume their citizens and patients have more knowledge and skills than they actually have, then it becomes difficult for nurses to target the communication according to individual informative and communicative preferences (which is rather important to support self-care in people) [22,23]. The ICM can be considered an objective supplement to nurses' subjective estimation of knowledge and skills in citizens and patients, as it uncovers this by simple communication techniques. At the same time, the ICM also ensures recall and comprehension when providing information and guidance. When the community nurses were introduced to the ICM prior to the pilot study, they expressed concerns with regards to time consumption. The results, however, shows that this is not the case if they receive enough training and adapt the ICM into their everyday working practice (e.g., that they develop and expand their interactive communication techniques).

The web-based IT solution was informally evaluated prior to use but useability test or heuristic evaluation were not performed. However, the community nurses found it intuitive and easy to assess the material in the system

during the pilot study. The challenges were as mentioned, to learn the communication techniques and adapt them into their nursing practice. The web-based IT system serves as an important platform to learn and adapt the communication techniques that the ICM comprises.

A limitation to this pilot study is that observations were carried out by one researcher. The observations were carried out continuously for five weeks and allowed the researcher to adapt to the environment and build trust. Another limitation is that the survey results are based on six respondents; it was conducted at the end of the pilot study and contributed to the primary data material gathered through observation and informal interviews.

Hopefully, the results from this pilot study can create the foundation for testing the web-based ICM in a larger scale.

5 REFERENCES

- [1] World Health Organization. The global burden of chronic diseases [Internet]. 2016 [cited 2016 Mar 29]. Available from: http://www.who.int/nutrition/topics/2_background/en/
- [2] Jordan JE, Osborne RH. Chronic disease self-management education programs: challenges ahead. *Med J Aust.* 2007 Jan 15;186(2):84-7. doi: 10.5694/j.1326-5377.2007.tb00807.x. PMID: 17223770.
- [3] Coleman K, Austin BT, Brach C, Wagner EH. Evidence on the Chronic Care Model in the new millennium. *Health Affairs.* 2009 Jan;28(1):75–85.
- [4] Joosten EAG, DeFuentes-Merillas L, de Weert GH, Sensky T, van der Staak CPF, de Jong CAJ. Systematic review of the effects of shared decision-making on patient satisfaction, treatment adherence and health status. *Psychotherapy and Psychosomatics.* 2008 May;77(4):219–26.
- [5] Canadian Public Health Association. Low health Literacy and Chronic Disease prevention and Control- Perspectives from The Health And Public Health Sectors. 2006;(July):16. Available from: www.cpha.ca/uploads/portals/h-l/kl_summary_e.pdf
- [6] Nutbeam D. Health promotion glossary. *Health Promot.* 1998;1(1):113–27.
- [7] Duell P, Wright D, Renzaho AMN, Bhattacharya D. Optimal health literacy measurement for the clinical setting: A systematic review. Vol. 98, *Patient Education and Counseling.* Elsevier Ireland Ltd; 2015. p. 1295–307.
- [8] Bo A, Friis K, Osborne RH, Maindal HT. National indicators of health literacy: ability to understand health information and to engage actively with healthcare providers - a population-based survey among Danish adults. *BMC Public Health.* 2014 Oct 22;14:1095. doi: 10.1186/1471-2458-14-1095. PMID: 25339154; PMCID: PMC4286937.
- [9] Berkman ND, Sheridan SL, Donahue KE, Halpern DJ, Viera A, Crotty K, Holland A, Brasure M, Lohr KN, Harden E, Tant E, Wallace I, Viswanathan M. Health literacy interventions and outcomes: an updated systematic review. *Evid Rep Technol Assess (Full Rep).* 2011 Mar;(199):1-941. PMID: 23126607; PMCID: PMC4781058.

- [10] Berkman ND, Davis TC, McCormack L. Health literacy: What is it? Vol. 15, *Journal of Health Communication*. 2010. p. 9–19.
- [11] Hironaka LK, Paasche-Orlow MK. The implications of health literacy on patient-provider communication. *Arch Dis Child*. 2008 May;93(5):428-32. doi: 10.1136/adc.2007.131516. Epub 2007 Oct 4. PMID: 17916588.
- [12] Al Sayah F, Williams B. An Integrated Model of Health Literacy Using Diabetes as an Exemplar. *Canadian Journal of Diabetes*. 2012 Feb;36(1):27–31.
- [13] Kourkouta L, Papathanasiou I. Communication in Nursing Practice. *Materia Socio Medica*. 2014;26(1):65.
- [14] al Sayah F, Williams B, Pederson JL, Majumdar SR, Johnson JA. Health literacy and nurses' communication with type 2 diabetes patients in primary care settings. *Nursing Research*. 2014 Dec 10;63(6):408–17.
- [15] Medical Association W. WMA DECLARATION OF HELSINKI – ETHICAL PRINCIPLES FOR MEDICAL RESEARCH INVOLVING HUMAN SUBJECTS 2020.
- [16] Braun V, Clarke V. Using thematic analysis in psychology. *Qualitative Research in Psychology*. 2006;3(2):77–101.
- [17] Saldana J. *The coding manual for qualitative researchers*. SAGE Publications ; 2015.
- [18] Everitt B. *The Cambridge Dictionary of Statistics*. 2nd ed. Cambridge: Cambridge University Press; 2002.
- [19] Schillinger D, Piette J, Grumbach K, Wang F, Wilson C, Daher C, Leong-Grotz K, Castro C, Bindman AB. Closing the loop: physician communication with diabetic patients who have low health literacy. *Arch Intern Med*. 2003 Jan 13;163(1):83-90. doi: 10.1001/archinte.163.1.83. PMID: 12523921.
- [20] Yen PH, Fnp-Bc D, Leasure AR, Yen P. Use and Effectiveness of the Teach-Back Method in Patient Education and Health Outcomes.
- [21] Talevski J, Shee AW, Rasmussen B, Kemp G, Beauchamp A. Teach-back: A systematic review of implementation and impacts. Vol. 15, *PLoS ONE*. Public Library of Science; 2020.
- [22] MacAbasco-O'Connell A, Fry-Bowers EK. Knowledge and perceptions of health literacy among nursing professionals. In: *Journal of Health Communication*. 2011. p. 295–307.
- [23] Dickens C, Lambert BL, Cromwell T, Piano MR. Nurse overestimation of patients' health literacy. *Journal of Health Communication*. 2013 Dec 4;18(SUPPL. 1):62–9.
- [24] Brooks C, Ballinger C, Nutbeam D, Mander C, Adams J. Nursing and allied health professionals' views about using health literacy screening tools and a universal precautions approach to communication with older adults: a qualitative study. *Disability and Rehabilitation*. 2020 Jun 18;42(13):1819–25.
- [25] Kwame A, Petrucka PM. A literature-based study of patient-centered care and communication in nurse-patient interactions: barriers, facilitators, and the way forward. Vol. 20, *BMC Nursing*. BioMed Central Ltd; 2021.

Appendix A: Results from brief survey about community nurses use of the interactive communication model

	Not at all	A little	Some	Greatly	Completely
1. To what extent do you think the interactive communication model increases citizens' knowledge of their own disease(s) and life situation?	0%	0%	17%	50 %	33 %
2. To what extent do you think the interactive communication model increases citizens' self-care?	0%	0%	33%	50%	17 %
3. To what extent do you think the interactive communication model enhances your experience of a more holistic nursing?	0%	0%	0%	50%	50%
4. To what extent do you think the interactive communication model helps you find out what and how much a citizen knows about their own disease(s) and life situation?	0%	0%	0%	33%	67%
5. To what extent do you think the interactive communication model increases your focus on NOT using medical terms in communication with citizens?	17%	0%	33%	33%	17%
6. To what extent do you think the interactive communication model increases your focus on 'checking' citizens' understanding of the information / guidance you provide as a nurse?	0%	0%	0%	83%	17%
7. To what extent do you think the web-based IT solution is useful?	17%	0%	17%	33%	33%
8. To what extent do you think the interactive communication model reduces your time consumption with citizens?	33%	33%	0%	33%	0%

Social media, physical activity and autism: better or bitter together?

A scoping review

Elia Gabarron^{1,2}, André Henriksen³ and Anders Nordahl-Hansen¹

¹Department of Education, ICT and Learning, Østfold University College, Norway, elia.gabarron@hiof.no

²Norwegian Centre for E-health Research, University Hospital of North Norway, Norway

³Department of Computer Science, UiT The Arctic University of Norway, Norway

Abstract

This review provides an overview of the existing research on social media, autism, and physical activity. We searched for publications on PubMed, PsycInfo, Embase, Education source, ERIC, IEEE Xplore, and the proceedings from conferences on health informatics and autism. Eight studies were included in this review. Studies reported mixed results on the link between social media, physical activity, and autism. Technology usage is related to sedentary time. However, physical activity interventions delivered through social media provide several benefits. Further research with stronger designs is needed to increase the knowledge of the role of social media on physical activity and autism.

Keywords

Social Media; Autism Spectrum Disorder; Autistic Disorder; Physical Activity; Exercise

1 INTRODUCTION

Autism spectrum disorder (ASD) is a highly heritable neurodevelopmental condition characterized by distinctive patterns of social interaction and communication, and restricted and repetitive behaviours [1]. Limited evidence shows that autistic children and adolescents have a moderately decreased physical activity level compared with their neurotypical peers [2]. They are therefore a special risk group for health challenges, including obesity, diabetes, and depression. Autistic adults also report being less frequently physically active, showing less positive attitudes towards physical activity, having less perceived behavioural control of performing physical activity, and encountering more physical activity barriers [3].

Regular physical activity helps to prevent and manage noncommunicable diseases, to maintain healthy body weight and can also improve mental health, quality of life and well-being [4]. The World Health Organization recommends physical activity and limitation of the time spent being sedentary for both children, adolescents, adults, and also to individuals with disabilities, including autism [5].

Children and adolescents living with a disability such as ASD are recommended to spend at least 60 minutes a day in moderate-to-vigorous-intensity physical activity at least three days a week [5]. Adults living with a disability are recommended to spend at least 150 minutes in moderate or higher intensity physical activity throughout the week [5]. In addition to the benefits on health outcomes, regular physical activity might be beneficial for autistic children and adolescents to improve manipulative- and motor skills [6-9], social functioning [10, 9], communication and social skills [8, 9], and has a positively impact on sleep and mood [6].

The 18th Scandinavian Conference on Health informatics, Tromsø, Norway, August 22-24, 2022. Organized by UiT The Arctic University of Norway. Conference Proceedings published by Linköping University Electronic Press at <https://doi.org/10.3384/ecp187>. © The Author(s). This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>

Both sedentary behaviour [11, 12] and physical activity behaviour [12] have been linked to the use of social media in neurotypical individuals. However not much is known about how social media are linked to physical activity and autism in the literature.

2 OBJECTIVE

The objective of this review is to provide an overview of the existing research linking social media, autism, and physical activity.

3 METHOD

In order to provide an overview of the existing research linking the use of social media, physical activity and autism, a scoping review was carried out. This scoping review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analysis, extension for scoping reviews (PRISMA-ScR) [13].

3.1 Search strategy

This scoping review includes a secondary analysis of a broader review on social media and autism. The main search was conducted to identify publications that included in their titles and abstracts terms related to autism and terms related to social media. Figure 1 gives the search terms that were used in the query.

The search covered the following six databases: PubMed, PsycInfo, Embase, Education source; ERIC, and IEEE Xplore. No date or language limitations were used. The search was carried out on the 30th of March 2022.

Articles specifically dealing with social media, physical activity, and autism were identified in the main databases. An additional search was carried out across the latest online available abstract books or conference proceedings from three conferences in the field of health informatics: Scandinavian Conference on Health Informatics (SHI,

period 2012-2019), Medical Informatics Europe (MIE, period 2014-2021), and the World Congress in Medical and Health Informatics (MedInfo, period 2010-2019). We also searched proceedings from two conferences focused on autism: Autism Europe (period 2010-2019), and the International Society for Autism Research (INSAR, period 2010-2021).

```

("Autism Spectrum Disorder" OR
 "Autism" OR
 "Asperger Syndrome" OR
 "Pervasive Developmental Disorder")
AND
("Social media" OR "Facebook" OR
 "Twitter" OR "Instagram" OR
 "YouTube" OR "TikTok" OR
 "WhatsApp" OR "WeChat" OR "Weibo" OR
 "Telegram")

```

Figure 1. Search query

3.2 Inclusion and exclusion criteria

Publications were included in this review if they referred to: 1) social media, 2) autism, and 3) physical activity, and were primary studies or review papers reporting results. Opinion papers, study protocols, editorials, and letters to editor not reporting results, were excluded.

3.3 Eligibility, data extraction and quality of the evidence assessment

All captured references from the initial search were uploaded to EndNote 20 (Clarivate Analytics, Philadelphia, PA, US) and Rayyan (Rayyan, Cambridge, MA, US) [14]. After removing duplicates, the eligibility of the papers was assessed in two steps: first by checking titles and abstract, and then by checking full text. The selection of articles was done by one reviewer (EG) and verified by a second reviewer (ANH). A third reviewer (AH) looked for possible full-text articles related to the included conferences abstracts. One reviewer (EG) extracted the following data from the articles: country of origin, publication year, study design, participant characteristics, and main findings.

The quality of evidence of the included studies on the findings linking social media, physical activity, and autism was assessed by one reviewer (EG) by following the Grading of Recommendations Assessment, Development and Evaluation (GRADE) guidelines [15].

4 RESULTS

4.1 Sample description

A total of 1998 references were identified in the database search. After removing 188 duplicates, a total of 1810 records from the search in databases were assessed for eligibility. 1703 references were excluded during title and abstract screening. A total of 107 publications dealing with autism and social media were reviewed in full text. Five publications met the inclusion criteria [16-20]. Three additional abstracts were obtained from the manual search in conference proceedings and were included in this review [21-23]. These abstracts were replaced with their respective full-text articles [24-26]. The PRISMA flowchart of the selection procedure is reported in Figure 2.

A summary of all included studies, sorted by GRADE score, can be found in Table 1. All eight included studies were published between 2015 and 2021. Two of these studies were from Turkey [16, 20], two from USA [17, 18], two from USA and UK [24, 25], one from USA, Canada and UK [26], and one from Ireland [19].

Among the eight included publications, four were interventional studies that used specific social media channels to deliver a physical activity intervention to parents of autistic children [16-18, 20]. The other four publications were observational studies that used social media to recruit autistic adults or parents of children diagnosed with autism. These studies used interviews or surveys to get information on different aspects related to their physical activity behaviour [24-26, 19].

The GRADE assessments related to level of evidence of the findings linking social media, physical activity and autism were graded as moderate in two of the studies [26, 20], low in five studies [24, 25, 16-18], and very low in one study [19].

4.2 Participants and main findings

Three of the included studies were specifically focused on autistic adults [24-26]. One of these studies (Kim et al.) [26] used an online survey that involved 229 autistic adults plus 10 interviews (four interviews with experts on autism and six to autistic adults). That study shows that there is moderate evidence linking sedentary time in autistic adults with technology usage time [26]. The two other studies (Balgrave et al.; and Colombo-Dougovito et al.) [24, 25] interviewed 23 autistic adults. These two studies indicate that there is low evidence referring to the features or functionalities that a technology-based intervention could have to increase physical activity in adults (i.e., importance of creating non-competitive experiences, importance of considering sensory factors, importance of listening to insight of autistic adults, as well as encouraging and providing social support) [24, 25].

One further survey study (Kindregan et al.) [19] gathered information from 221 parents, 48% of which had a child with a diagnosis of ASD. Answers from these parents indicate that autistic children spend significantly less time per week being physically active compared to neurotypical kids, and significantly more time is spent being sedentary (including watching television or using a computer) [19].

Four included articles were intervention studies involving families with a child who had a diagnosis of ASD [16-18, 20]. Two of these publications are findings from the same 4-week intervention using Facebook private groups (Healy et al.; Healy et al.) [17, 18]. The intervention consisted of delivering instructional and motivational strategies to parents of 13 families and were aimed at stimulating their autistic child aged 6-16 years old [17, 18]. These two studies indicate that there is low evidence indicating that parents perceived the private *Facebook* groups intervention successful as a source of motivation, as a reminder to take action, and as a source of social support [17, 18].

Two other intervention studies analysed the feasibility and effects of a physical activity intervention delivered to parents of autistic children on a *WhatsApp* group (Esentürk et al.; Yarimkaya et al.) [16, 20]. The feasibility study included 14 parents and showed that a 4-week intervention was feasible to increase physical activity levels of their autistic child aged 9-14 years old ((Esentürk et al.)[16]. The subsequent randomized 6-week intervention involving 42 families indicate moderate evidence linking a significant increase in the physical activity level of autistic children aged five years old with the experimental intervention delivered through *WhatsApp* (Yarimkaya et al.)[20].

5 DISCUSSION

5.1 Summary of findings

There is very little evidence linking social media, physical activity, and autism. The present review shows that the usage of social media as platforms that can be used for physical activity training, but also that this topic is understudied, and the rigour of studies can be improved in future research projects.

5.2 Cons and pros of social media and physical activity in autism

Although physical activity as an alternative or complementary intervention shows promise for autistic individuals [6-8, 10, 9, 27], currently there are not any stand-out recommended interventions with clear gains within developmental domains such as social or motor development [1, 28].

Findings from two of the included studies indicate that using social media and other technologies is related to an increased sedentary time among both autistic children and adults [26, 19]. These findings agree with previous research with neurotypical population [11, 12].

However, social media could also be used to promote physical activity among their users [29]. The ubiquity and high usability of social media make these channels an environment of great potential to motivate their users to increase physical activity, including autistic children and adults. Research shows that interventions delivered through private social media channels, such as *Facebook* private groups and *WhatsApp*, can be successful sources of motivation [16-18, 20]. These channels can be used to send reminders to parents of autistic children, but also represent powerful environments of social support through interaction with others on the platform that can be helpful to instigate taking action and engage in physical activity [29]. Besides, the availability of social media for parents and autistic adults might be preferred by many as compared to more time-consuming interventions delivered by on-site interventionists.

5.3 How social media interventions should be to have a positive effect on physical activity for autistic individuals?

Interventions delivered through social media have proven to be effective to increase physical activity among neurotypical population [30] and probably could be beneficial among individuals with ASD too. Findings in our review indicate that there are a series of functionalities and features that should be considered when designing a

physical activity intervention to be delivered through social media and targeting individuals with ASD.

Although there is very little evidence, using a participatory approach, where insight and perspectives of autistic adults are included, could make interventions addressed to them more accessible and increase their use [24]. According to autistic adults these interventions should be encouraging and providing social support for physical activity participation [24]. However, participation of autistic individuals should be social but not competitive [22]. In addition, because sensory factors play a key role in the success of physical activity experiences for autistic individuals [22], considering sensory stimuli when planning for physical activity participation seems to be of relevance [24].

5.4 Activity trackers for physical activity interventions

The connection of activity trackers to social media could also be considered in order to increase physical activity among autistic individuals. Activity trackers have proven to be an effective feature in promoting a positive behaviour change [31, 32].

The use of consumer-based physical activity trackers as a motivational tool in physical activity interventions are becoming more common. Although there are few intervention studies using such trackers in neurodivergent populations, a study by Garcia et al. [33] indicate that it is feasible to use a Fitbit in this setting for youth with ASD. A similar study by Michalsen et al. [34] is currently ongoing among people with intellectual disabilities.

However, using consumer-based physical activity trackers have some privacy related challenges that have to be considered. For instance, the ownership of collected data is often unclear (i.e., is it owned by the vendor or the tracker owner?), and it is rarely possible to prevent data from being stored in vendor-owned cloud servers (often located abroad).

Integrating behaviour change techniques that have proven to be effective among neurotypical population [31], such as goal setting, feedback, and monitoring, prompts or rewards, might also help to increase physical activity among autistic individuals.

5.5 Limitations

Only one author screened the papers and assessed the quality of evidence. Although we did not use any language or year limitation on our search, we have only identified eight studies dealing with autism, physical activity, and social media. We believe the search engine and covered databases and conference proceedings allowed us to capture relevant publications in this field. However, we might have missed studies that included different keywords, were indexed in different databases, or presented in different conferences.

5.6 Conclusions

Social media could represent an effective media for promoting physical activity to autistic children and adults. However further research with stronger design is needed to understand the current role of these channels in relation to physical activity and sedentarism, and the potential of social media on physical activity and autism.

6 CONFLICT OF INTERESTS

Authors declare no conflict of interests.

7 REFERENCES

- [1] Davis, R., den Houting, J., Nordahl-Hansen, A., and Fletcher-Watson, S. (2022), 'Helping autistic children', in P.K. Smith and Hart. C. H. (eds.), *The Wiley-Blackwell Handbook of Childhood Social Development*.
- [2] Li, R., Liang, X., Zhou, Y., and Ren, Z., "A Systematic Review and Meta-Analysis of Moderate-to-Vigorous Physical Activity Levels in Children and Adolescents With and Without ASD in Inclusive Schools", *Front Pediatr*, 9, 726942, 2021, Rech, J. P., Irwin, J. M., Rosen, A. B., Baldwin, J., and Schenkelberg, M., "Comparison of Physical Activity Between Children With and Without Autism Spectrum Disorder: A Systematic Review and Meta-Analysis", *Adapt Phys Activ Q*, 1-26, 2022
- [3] Hillier, A., Buckingham, A., and Schena, D., 2nd, "Physical Activity Among Adults With Autism: Participation, Attitudes, and Barriers", *Percept Mot Skills*, 127 (5), 874-90, 2020
- [4] World Health Organization 'Physical activity', <<https://www.who.int/news-room/fact-sheets/detail/physical-activity>>, accessed 27 April 2022.
- [5] World Health Organization 'WHO Guidelines on Physical Activity and Sedentary Behaviour', <<https://apps.who.int/iris/rest/bitstreams/1315866/trieve>>, accessed 27 April 2022.
- [6] Brand, S., Jossen, S., Holsboer-Trachsler, E., Pühse, U., and Gerber, M., "Impact of aerobic exercise on sleep and motor skills in children with autism spectrum disorders - a pilot study", *Neuropsychiatr Dis Treat*, 11, 1911-20, 2015
- [7] Hassani, F., Shahrbanian, S., Shahidi, S. H., and Sheikh, M., "Playing games can improve physical performance in children with autism", *Int J Dev Disabil*, 68 (2), 219-26, 2022
- [8] Healy, S., Nacario, A., Braithwaite, R. E., and Hopper, C., "The effect of physical activity interventions on youth with autism spectrum disorder: A meta-analysis", *Autism Res*, 11 (6), 818-33, 2018
- [9] Huang, J., Du, C., Liu, J., and Tan, G., "Meta-Analysis on Intervention Effects of Physical Activities on Children and Adolescents with Autism", *Int J Environ Res Public Health*, 17 (6), 2020
- [10] Howells, K., Sivaratnam, C., May, T., Lindor, E., McGillivray, J., and Rinehart, N., "Efficacy of Group-Based Organised Physical Activity Participation for Social Outcomes in Children with Autism Spectrum Disorder: A Systematic Review and Meta-analysis", *J Autism Dev Disord*, 49 (8), 3290-308, 2019
- [11] Alley, S., Wellens, P., Schoeppe, S., de Vries, H., Rebar, A. L., Short, C. E., Duncan, M. J., and Vandelanotte, C., "Impact of increasing social media use on sitting time and body mass index", *Health Promot J Austr*, 28 (2), 91-95, 2017
- [12] Shimoga, S. V., Erlyana, E., and Rebello, V., "Associations of Social Media Use With Physical Activity and Sleep Adequacy Among Adolescents: Cross-Sectional Survey", *J Med Internet Res*, 21 (6), e14290, 2019
- [13] McGowan, J., Straus, S., Moher, D., Langlois, E. V., O'Brien, K. K., Horsley, T., Aldcroft, A., Zarin, W., Garity, C. M., Hempel, S., Lillie, E., Tunçalp, Ö, and Tricco, A. C., "Reporting scoping reviews-PRISMA ScR extension", *J Clin Epidemiol*, 123, 177-79, 2020
- [14] Ouzzani, M., Hammady, H., Fedorowicz, Z., and Elmagarmid, A., "Rayyan-a web and mobile app for systematic reviews", *Syst Rev*, 5 (1), 210, 2016
- [15] Balshem, H., Helfand, M., Schünemann, H. J., Oxman, A. D., Kunz, R., Brozek, J., Vist, G. E., Falck-Ytter, Y., Meerpohl, J., Norris, S., and Guyatt, G. H., "GRADE guidelines: 3. Rating the quality of evidence", *J Clin Epidemiol*, 64 (4), 401-6, 2011, Guyatt, G. H., Oxman, A. D., Vist, G., Kunz, R., Brozek, J., Alonso-Coello, P., Montori, V., Akl, E. A., Djulbegovic, B., Falck-Ytter, Y., Norris, S. L., Williams, J. W., Jr., Atkins, D., Meerpohl, J., and Schünemann, H. J., "GRADE guidelines: 4. Rating the quality of evidence--study limitations (risk of bias)", *J Clin Epidemiol*, 64 (4), 407-15, 2011
- [16] Esentürk, O.K. and Yarımkaaya, E., "WhatsApp-Based Physical Activity Intervention for Children With Autism Spectrum Disorder During the Novel Coronavirus (COVID-19) Pandemic: A Feasibility Trial", *Adapt Phys Activ Q*, 38 (4), 569-84, 2021
- [17] Healy, S., Marchand, G., and Williams, E., "'I'm not in this alone" the perspective of parents mediating a physical activity intervention for their children with autism spectrum disorder", *Res Dev Disabil*, 83, 160-67, 2018
- [18] Healy, S. and Marchand, G., "The feasibility of Project CHASE: A Facebook-delivered, parent-mediated physical activity intervention for children with autism", *International Journal of Disability, Development and Education*, 67 (2), 225-42, 2020
- [19] Kindregan, D., Gallagher, L., and Gormley, J. (2015), 'Physical activity in children with autism spectrum disorders', *European Child and Adolescent Psychiatry* (24; Madrid, Spain), S206.
- [20] Yarımkaaya, E., Esenturk, O.K., İlhan, E. L., and Karasu, N., "A WhatsApp-delivered intervention to promote physical activity in young children with autism spectrum disorder", *International Journal of Developmental Disabilities*, 2021
- [21] Colombo-Dougovito, A.M.; Blagrove, J; Healy, S. (2019), 'A New Grounded Theory of Physical Activity Participation in Autistic Adults: Preliminary Findings', in INSAR International Society for Autism Research (ed.), *INSAR Annual Meeting* (Montreal, Canada).
- [22] Colombo-Dougovito, A.M., Blagrove, J., and Healy, S. (2021), "'Just Invite Us": Autistic Adults' Recommendations for Developing More Accessible Physical Activity Opportunities', in INSAR

- International Society for Autism Research (ed.), *INSAR 2021* (Digital conference).
- [23] Lee, D.; Cothran, D.J.; Kennedy, J.M.; Shih, P.C.; Golzarri-Arroyo, L.; Dickinson, S.; Frey, G.C. (2019), 'Association between Physical Activity, Sedentary Time, and Technology Use in Adults with Autism Spectrum Disorder', in INSAR International Society for Autism Research (ed.), *INSAR Annual Meeting* (Montreal, Canada).
- [24] Blagrove, A. J., Colombo-Dougovito, A. M., and Healy, S. (2021), "Just Invite Us": Autistic Adults' Recommendations for Developing More Accessible Physical Activity Opportunities', *Autism in Adulthood* (3), 179-86.
- [25] Colombo-Dougovito, A. M., Blagrove, A. J., and Healy, S., "A grounded theory of adoption and maintenance of physical activity among autistic adults", *Autism*, 25 (3), 627-41, 2021
- [26] Kim, B., Lee, D., Min, A., Paik, S., Frey, G., Bellini, S., Han, K., and Shih, P. C., "PuzzleWalk: A theory-driven iterative design inquiry of a mobile game for promoting physical activity in adults with autism spectrum disorder", *PLoS One*, 15 (9), e0237966, 2020
- [27] Hynes, J. and Block, M., "Effects of Physical Activity on Social, Behavioral, and Cognitive Skills in Children and Young Adults with Autism Spectrum Disorder: a Systematic Review of the Literature", *Review Journal of Autism and Developmental Disorders*, 1-22, 2022
- [28] Fragala-Pinkham, M. A., Haley, S. M., and O'Neil, M. E., "Group swimming and aquatic exercise programme for children with autism spectrum disorders: a pilot study", *Dev Neurorehabil*, 14 (4), 230-41, 2011
- [29] Gabarron, E., Larbi, D., Årsand, E., and Wynn, R., "Engaging Social Media Users with Health Education and Physical Activity Promotion", *Stud Health Technol Inform*, 281, 789-93, 2021
- [30] Dixit, S. and Nandakumar, G., "Promoting healthy lifestyles using information technology during the COVID-19 pandemic", *Rev Cardiovasc Med*, 22 (1), 115-25, 2021, Günther, L., Schleberger, S., and Pischke, C. R., "Effectiveness of Social Media-Based Interventions for the Promotion of Physical Activity: Scoping Review", *Int J Environ Res Public Health*, 18 (24), 2021
- [31] Laranjo, L., Ding, D., Heleno, B., Kocaballi, B., Quiroz, J. C., Tong, H. L., Chahwan, B., Neves, A. L., Gabarron, E., Dao, K. P., Rodrigues, D., Neves, G. C., Antunes, M. L., Coiera, E., and Bates, D. W., "Do smartphone applications and activity trackers increase physical activity in adults? Systematic review, meta-analysis and metaregression", *Br J Sports Med*, 55 (8), 422-32, 2021
- [32] Tomaszewski, B., Savage, M. N., and Hume, K., "Examining physical activity and quality of life in adults with autism spectrum disorder and intellectual disability", *J Intellect Disabil*, 17446295211033467, 2021
- [33] Garcia, J. M., Leahy, N., Brazendale, K., Quelly, S., and Lawrence, S., "Implementation of a school-based Fitbit program for youth with Autism Spectrum Disorder: A feasibility study", *Disabil Health J*, 14 (2), 100990, 2021
- [34] Michalsen H, Wangberg SC, Hartvigsen G, Henriksen A, Pettersen G, Jaccheri L, Jahnsen R, Thrane G, Arntzen C, Anke A, "Mobile health support to stimulate physical activity in individuals with intellectual disability: Protocol for mixed methods pilot study", *JMIR Preprints*, 09/03/2022:37849, 2022

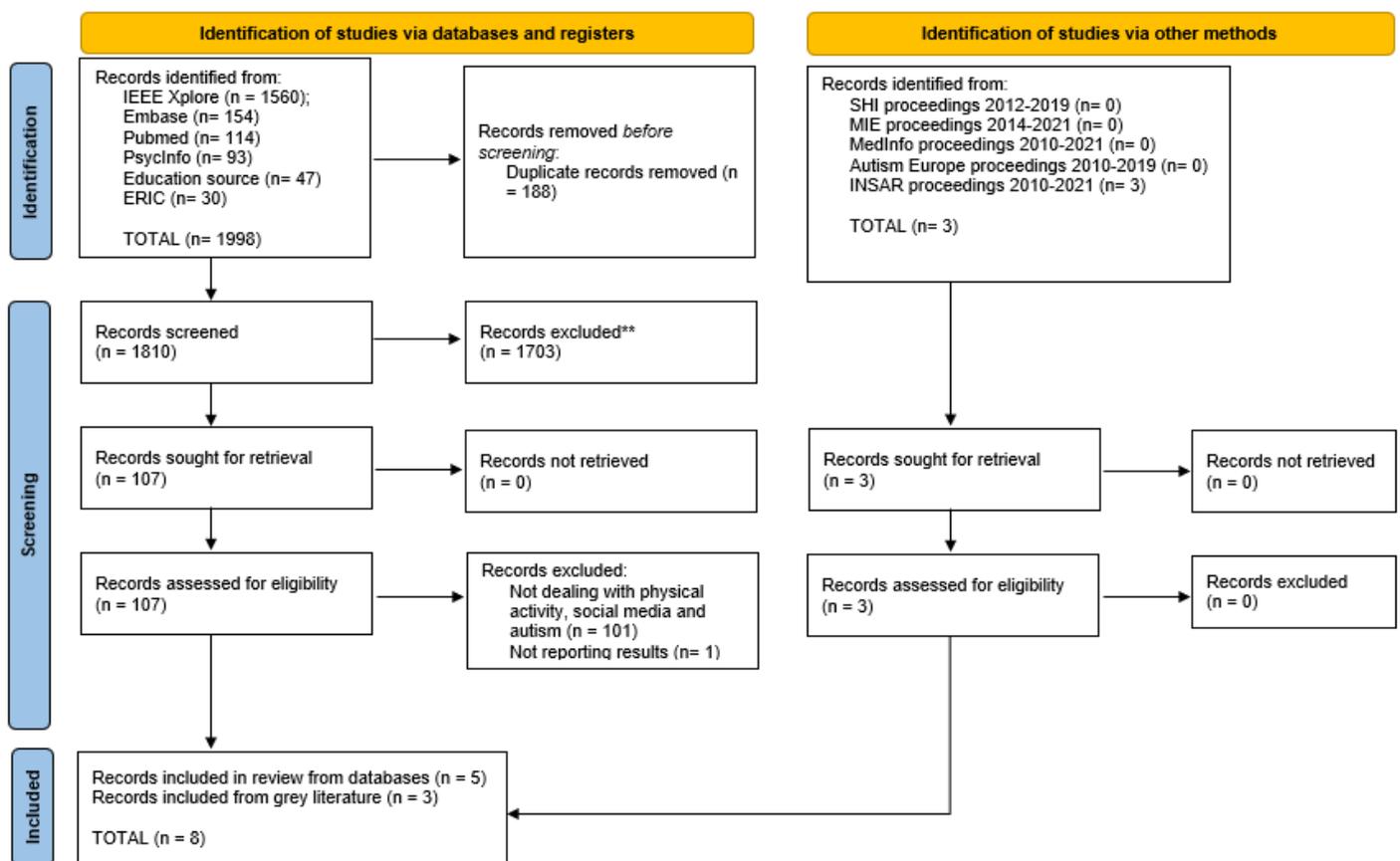


Figure 2. PRISMA flowchart of the selection process

Paper	Study design and method	Participants	Findings	GRADE level of evidence
Yarimkaya et al. [20]	Randomized study (6-weeks intervention to parents on WhatsApp group)	42 families (parent and child dyads). 21 intervention/21 control group	Significant increase in the physical activity level of children with ASD in the experimental group compared to the control group. Parents reported benefits (increased levels of physical activity; promoted family participation; improved movement skills; and reduced technological tool addiction); and usefulness of the intervention (support for physical education; useful information provided by intervention; support to create new routines)	Moderate
Kim et al. [23]	Observational study (online survey and interviews)	229 autistic adults	Adults with ASD used their smartphones more than four hours per day. Sedentary time was significantly correlated with technology usage time ($r = 0.34, p < 0.001$).	Moderate
Colombo et al. [22]	Observational study (interviews)	23 autistic adults	Importance of creating non-competitive and social experiences. Sensory factors play a key role in the success of physical activity experiences for autistic adults.	Low
Balgrave et al. [24]	Observational study (interviews)	23 autistic adults	Physical activity can be made more accessible for autistic adults by: (1) listening to the perspectives and insight of autistic adults, (2) considering sensory stimuli when planning for physical activity participation, and (3) encouraging and providing social supports for physical activity participation.	Low

Esenturk et al. [16]	Intervention study (4-weeks intervention to parents on WhatsApp group)	14 parents and their children with ASD	WhatsApp-based physical activities were a feasible intervention to increase the physical activity level of their children with ASD and stated that the contents of the physical activity shared in the WhatsApp group were useful.	Low
Healy et al. [18]	Feasibility study (4-weeks intervention to parents on private Facebook group)	13 families with a child (age 6-16 years) with ASD	All parents were satisfied or very satisfied with their overall experience of the project.	Low
Healy et al. [17]	Intervention study (4-weeks intervention to parents on private Facebook group)	13 families with a child (age 6-16 years) with ASD	Parents reported an overall positive perspective of the intervention. The parents perceived the intervention to be particularly successful as a source of motivation, a reminder for them to take action, and as a source of social support.	Low
Kindregan et al. [19]	Observational study (online survey)	221 parents responded survey. 48% of them had a kid with ASD	56 % of children with ASD spent over 6 h per week in sedentary behaviour such as watching television or on a computer, significantly more than typically developing children at 33 % ($p<0.05$).	Very low

Table 1. Summary of included studies

User preferences for a physical activity chatbot connected to an activity tracker and integrated into a social media platform

Dillys Larbi^{1,2}, Helle Sandsdalen³, Elia Gabarron^{1,4}, Eirik Årsand^{3,1}, and André Henriksen³
¹Norwegian Centre for E-health Research, Tromsø, Norway, dillys.larbi@ehealthresearch.no
²Department of Clinical Medicine, UiT The Arctic University of Norway, Tromsø, Norway
³Department of Computer Science, UiT The Arctic University of Norway, Tromsø, Norway
⁴Institute for Education, ICT and Learning, Østfold University College, Halden, Norway

Abstract

Performing regular physical activity can be challenging. Integrating chatbots with social media platforms and physical activity sensors can potentially increase physical activity. The objective of this study was to identify design preferences for integrating an activity tracker supported chatbot in a social media platform. Norwegian adults (n=120) responded to an ad-hoc online survey. User preferences included adding a step goal feature that can be renewed every week and communicating with the chatbot once per day. Preferences of all types of potential users for a social media chatbot for physical activity should be explored to produce a well-accepted intervention.

Keywords

Physical activity; social media; chatbot; sensor; behaviour change

1. INTRODUCTION

1.1 Physical activity

Efforts are being made in recent years to encourage regular physical activity among the global population. To that effect, the World Health Organization (WHO) has developed physical activity guidelines and recommendations for individuals of all ages and health status. For adults (age 18-64), these guidelines states that at least 150-300 minutes of moderate physical activity, or at least 75-150 minutes of vigorous physical activity, or a combination of these should be performed each week [1]. Although there is an increase in the promotion of regular physical activity and its benefits, such as improved cognitive function and quality of life, the prevalence of physical inactivity remains high [1, 2]. Like any healthy habit, it is challenging to start and continue doing physical activity on a regular basis [3]. Innovative and engaging solutions may increase motivation to maintain regular physical activity, for instance by using chatbots [4].

1.2 Use of internet, social media, and physical activity tracking in Norway

There is an increasing trend in internet use globally. In Norway, about 99% of the population have access to internet [5]. Similarly, there was an increase in the number of social media users, with about 83.2% of the Norwegian population using social media in 2021 [5]. In 2021, 24% of the Norwegian population had access to a smartwatch [6], while 96% had access to a smartphone in 2020 [7]. In addition to wrist-worn physical activity trackers, smartwatches and smartphones can also be used to track physical activity [8]. Dalene et al. [9] observed that Norwegian adolescents have increased their sedentary time with the availability of smartphones and internet access [9]. Laranjo et al. [10], in a meta-analysis of physical activity

interventions, concluded that the use of smartphones is effective in increasing physical activity behaviour.

1.3 Chatbots as digital health interventions

A recent introduction to the healthcare system and digital health interventions is the chatbot, also called conversational agent, an application software that employs simulated text and speech to communicate with individuals. In healthcare, chatbots have been used to promote health information, reproductive health, healthy diet, weight loss, mental health, and physical activity [3, 11, 12].

Both rule-based and artificial intelligence chatbots with or without a social media element have been developed for promoting physical activity [3, 11-14]. It has been suggested that integrating chatbots with social media and sensors might increase their utility and subsequently enhance their ability to motivate users [11].

1.4 User preferences for chatbots or digital physical activity intervention

The integration of chatbots with social media and sensors following a participatory health approach would potentially be more powerful. In a participatory approach, the user becomes the central focus [15], and therefore their opinions and interests are relevant for developing health interventions. Involving users and users' preferences when designing a social media chatbot, aimed at increasing physical activity, could have several advantages, such as increasing the interest, motivation, and engagement of future intervention end-users, and/or raising their awareness of the importance of being physically active [16].

The objective of this study was to determine the design preferences of Norwegian adults for an activity tracker-supported chatbot integrated into a social media platform.

The 18th Scandinavian Conference on Health informatics, Tromsø, Norway, August 22-24, 2022. Organized by UiT The Arctic University of Norway. Conference Proceedings published by Linköping University Electronic Press at <https://doi.org/10.3384/ecp187>. © The Author(s). This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>

2 METHOD

We designed an anonymous survey to gain insight into potential users' preferences for the chatbot. To ensure the anonymity of the participants, no direct or indirect personal questions were included in the survey.

2.1 Questionnaire design

Nettskjema [17], an online tool for creating questionnaires was used. *Nettskjema* is a tool developed by the University of Oslo where security and privacy are ensured when gathering data through online questionnaires. It is available to researchers and students at all universities in Norway through their university license.

We created an *ad-hoc* survey that included questions about physical activity behaviour and social media habits. In general, the survey was designed to get information that would help with the design- and development of a chatbot for physical activity. The link to the survey was posted on the second author's (HS) Facebook page and a Discord page for members at the Department of Computer Science at UiT The Arctic University of Norway. Discord is a social media platform originally created to enhance communication while playing games online [18]. The questionnaire was available for 1235 potential respondents. Questions were written in Norwegian and contained 16 questions. Table 1 lists an English translation of each question in the questionnaire. Data were collected in September 2021.

#	Questions and answer options
1	What is your gender? (<i>Male, Female, other/do not want to answer</i>)
2	What is your age? (<i>10-year age groups</i>)
3	Which social media platform do you spend most time on, most days? (<i>Facebook, Messenger, Snapchat, Instagram, WhatsApp, Discord, Slack, Reddit, Other</i>)
4	Do you use an activity tracker? (<i>Yes, No</i>)
5	If yes, which activity tracker brand do you own? (<i>Garmin, Apple watch, Fitbit, Samsung watch, Samsung Galaxy Fit, Polar Watch, Other</i>)
6	How often are you physically active for at least 30 minutes during a week? Everything from walking to exercising at the gym. (<i><1 time/week, 1-2 times/week, 2-4 times/week, 5-7 times/week, >7 times/week</i>)
7	Which physical activity do you do most often? (<i>Walking, Gym, Jogging, Mountaineering, Skiing, Home exercising, Swimming, Other</i>)
8	Would you consider using a social media chatbot, connected to a physical activity tracker, to motivate you to do physical activity? (<i>Yes, Maybe, No</i>)
9	Which language do you prefer to communicate with a chatbot on? (<i>Norwegian, English, Option to choose between the two, Don't know</i>)
10	How many times a day do you prefer to communicate with a chatbot about becoming more physically active? (<i><1 time, once, 1-2 times, 2-3 times</i>)

11	When do you prefer to communicate with/receive reminders from a chatbot? (<i>Morning, before noon, afternoon, night</i>)
12	Do you prefer the chatbot to initiate the conversation? (<i>Yes, No, Don't know</i>)
13	Do you want a function for setting daily step goals? (<i>Yes, No, Don't know</i>)
14	If you wanted a function for setting daily step goals, how often would you like to set new goals? (<i>Daily, weekly, monthly, anytime, do not want that function</i>)
15	Do you want to communicate with the chatbot using free-text or choose between pre-defined questions and answers? (<i>Free-text, pre-defined questions/answers, want both options</i>)
16	From which device would you like to use the chatbot? (<i>PC, Mobile phone, Tablet, Watch</i>)

Table 1. Questionnaire translated to English. Answer options in parenthesis.

2.2 Statistical analysis

For the general survey participants, descriptive statistics were used to analyse the results for gender and age groups. We used Crosstabulation to analyse the correlation between the following: how often the participants are physically active for at least 30 minutes during a week, and their use of activity trackers; and participants' willingness to use a social media chatbot connected to a physical activity tracker and their use of activity trackers. Further analyses were done on the results of participants who were willing to use a social media chatbot connected to a physical activity tracker. Descriptive statistics were used to analyse the results for these participants' preferences for whether the chatbot initiates the conversation and the mode of communication with the chatbot i.e., using free-text or pre-defined questions and answers. We used Bar charts to illustrate how often these participants are physically active for at least 30 minutes during a week, and the number of times a day these participants prefer to communicate with a chatbot about becoming more physically active. Crosstabulation was used to analyse the correlation between this group of participants' preferences for a step goal feature and how often they prefer to set new goals. All statistical analyses were done using SPSS (version 25; IBM Corp).

3 RESULTS

3.1 Participant characteristics

A total of 120 individuals answered the online survey, of which 58.3% (70/120) identified themselves as male. There were 72 participants aged between 18 and 25 years (60%). Only 3.3% (4/120) of the survey participants were aged between 36 and 45 years. Approximately 6% (7/120) of the participants were aged between 46 and 55 years, and eight participants were aged between 56 and 65 years (6.7%).

3.2 Self-reported physical activity and use of activity tracker

The majority of the survey participants (43.3%; 52/120) did at least 30 minutes of physical activity 2-4 times per week. Thirty participants (25%; 30/120) indicated doing at least 30 minutes of physical activity more than five times per

week. A total of 67 participants among the 120 respondents indicated they use activity trackers (55.8%). Of those who did at least 30 minutes of physical activity 2-4 times a week, 59.6% (31/52) use an activity tracker. Approximately 30% (20/67) of those who did at least 30 minutes of physical activity more than five times a week use an activity tracker (see Table 2).

Physical activity per week ≥ 30 mins	Activity Tracker Use		
	No	Yes	Total (%)
< once	10	5	15 (12.5)
1-2 times	12	11	23 (19.2)
2-4 times	21	31	52 (43.3)
5-7 times	7	15	22 (18.3)
> 7 times	3	5	8 (6.7)
Total (%)	53 (44)	67 (56)	120 (100)

Table 2. Participants' physical activity per week and use of activity tracker (n=120).

3.3 Willingness to use a social media chatbot integrated with an activity tracker?

A total of 29 participants answered *Yes* to being willing to use a social media chatbot integrated with their activity tracker for motivation to do physical activity (24.2%). 25% (30/120) answered *No* and 50.8% (61/120) answered *Maybe* to the same question. Among the respondents who would like to use a social media chatbot integrated with an activity tracker, 69.0% (20/29) currently used an activity tracker; versus 36.7% (11/30) of those who answered *No*, and 59.0% (36/61) of those who answered *Maybe* (see Table 3). There is a significant association between the use of an activity tracker and the willingness to use a social media chatbot integrated with an activity tracker (Chi-Square = 6.748; p value = 0.034).

Use chatbot connected with activity tracker	Activity Tracker Use		
	No	Yes	Total (%)
Yes	9	20	29 (24.2)
Maybe	25	36	61 (50.8)
No	19	11	30 (25.0)
Total (%)	53 (44)	67 (56)	120 (100)

Table 3. Participants' willingness to use a social media chatbot integrated with an activity tracker (n=120). Chi-Square= 6.748; p-value<0.05

3.4 Respondents interested in a social media chatbot connected to an activity tracker

About 34.5% (10/29) of the participants interested in a social media chatbot connected to an activity tracker indicated doing at least 30 minutes of physical activity 2-4 times per week. Similarly, another 34.5% (10/29) of the same group of participants reported doing at least 30 minutes of physical activity more than 5 times per week (see Figure 1).

More than 50% (15/29) of these participants would prefer to communicate once per day with a chatbot for physical activity. About 10.3% (3/29) of participants would prefer to communicate with a chatbot for physical activity either less than once per day or 2-3 times per day, respectively (see Figure 2).

Most participants (89.7%; 26/29) would prefer the chatbot to start the conversation about becoming more physically active. Two of the 29(6.9%) participants did not know whether or not they would prefer the chatbot to start the conversation, and one person said No.

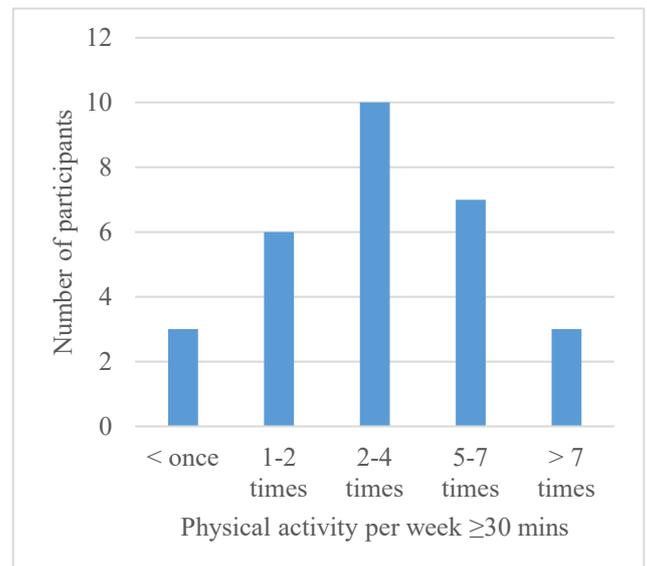


Figure 1. Physical activity per week of participants willing to use a social media chatbot (n=29).

Most participants (55.2%; 16/29) also preferred to use both predefined text and free text to communicate with the chatbot. Among those aged between 18 and 25 years, 68.8% (11/16) would prefer this option to use both predefined and free text. More participants (31.0%; 9/29) preferred to use predefined text than free text (13.8%; 4/29) to communicate with the chatbot.

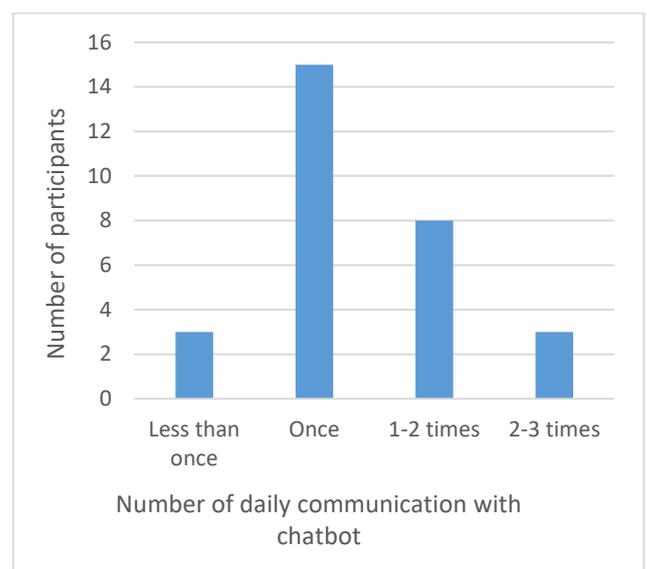


Figure 2. Participants' preferred number of daily communications with a social media chatbot (n=29).

Twenty-six (89.7%) of the 29 participants would prefer to have a step goal feature for the chatbot. Most of these participants (42.3%; 11/26) would prefer to set a step goal

on a weekly basis, followed by 34.6% (9/26) who would prefer to set a step goal on every month (see Table 4).

Frequency of setting new step goal	Step Goal Feature			Total (%)
	No	Yes	Don't know	
Anytime	0	4	0	4 (13.8)
Daily	0	2	0	2 (6.9)
Weekly	0	11	1	12 (41.4)
Monthly	0	9	1	10 (34.5)
No answer	1	0	0	1 (3.4)
Total (%)	1 (3)	26 (90)	2 (7)	29 (100)

Table 4. Participants' preferences for setting step goal (n=29).

4 DISCUSSION

In this study, we aimed to determine the design preferences of Norwegian adults for a chatbot integrated into a social media platform. The majority of the survey participants were young adults aged between 18 and 25 years. About one in three participants did at least 30 minutes of physical activity 2-4 times per week. More than half of the survey participants used an activity tracker and more than half of these do at least 30 minutes of physical activity 2-4 times per week.

Compared to previous chatbot studies addressed in this paper [3, 11-14], the present study focuses on a young Norwegian population, and it includes a bigger sample.

Twenty-nine participants, approximately 24%, and more than half of the total survey participants said *Yes* and *Maybe*, respectively, to using a social media chatbot connected to an activity tracker. The preferences of these 29 participants include 1) communicate with the chatbot once per day 2) the chatbot starts the conversation about physical activity 3) use both predefined and free text to communicate 4) a step goal feature and 5) set a step goal every week.

4.1 Using a social media chatbot connected to an activity tracker

Our findings show that most of the participants who were willing to use a social media chatbot connected to an activity tracker were active individuals who already used an activity tracker. This suggests that such a chatbot would most likely be used by individuals who are already interested in doing physical activity. Increasing physical activity among the general population, especially those with a sedentary lifestyle is the motivation for developing a social media chatbot for physical activity.

More than 40% of the survey participants did not use activity trackers, and about 80% of them said they would either not use the proposed chatbot or maybe use it. Not owning a wearable activity tracker could have been a deterrent to using a social media chatbot integrated with an activity tracker. Potential users may be unwilling to buy a wearable activity tracker. However, activity trackers are often included in phone applications, some of which are available for free. The participants' responses could have been as a result of misunderstanding the use of the word *activity tracker*.

Regardless, digital interventions, especially the ones intended for the promotion of a healthy lifestyle, should be developed in an all-inclusive manner [19]. Users should have the opportunity to customize or tailor these interventions, thereby attracting the interest of the majority of the potential users in the general population.

4.2 Participatory health: User preferences

End users or potential users should be an integral part of the design and development of services and solutions targeting them, including digital health interventions [19, 20]. Taking into account user preferences when designing the social media chatbot will help produce a chatbot that is easily accepted and adopted for long-term use.

Furthermore, it will allow users to actively participate in a decision about their current and/or future health. In our study, survey respondents indicated they preferred the chatbot to start the communication, and they preferred to communicate with it only once per day. Previous research shows that youth are interested in interacting with small talks with chatbots [21]. The human semblance of chatbots, use of emojis and use of colloquial tones that emulate human emotional expressions are also preferred user features that increase their engagement [22].

Our study results also revealed that potential users of a social media chatbot connected to an activity tracker preferred to use both predefined text and free text to communicate. Having the option to use both predefined questions and answers and free text allows the user to choose which mode of communication is most appropriate at a particular time. For example, if the user is walking while communicating with the chatbot, the user might prefer to use predefined questions and answers since it might be the most convenient. Providing users with different ways to use a feature of functionality could be one way to potentially increase their engagement with the social media chatbot.

The participants in our study preferred including a step goal feature in the social media chatbot, which they can set weekly. Since most of the participants use activity trackers, they may be familiar with the concept of goal setting, which is available on most modern activity trackers. Setting goals is a means to personalize and promote physical activity among users of digital health interventions, especially when it can be done using practical and complementary interactive tools [23, 24]. For most people, achieving a set goal may provide motivation and a sense of satisfaction. Achieving set step goals using a social media chatbot could encourage users to be more motivated to increase their physical activity in general.

4.3 Strengths and Limitations

Findings may be somewhat skewed because most respondents were young adults aged between 18 and 25 years. In addition, the survey is likely to mostly have been answered by individuals with knowledge in computer science, due to the background of the second author and the channels used for sharing the survey link. The findings may therefore not be generalizable to the general population.

To the best of our knowledge, our study is the first to explore the preferences of potential users in Norway for a

chatbot connected to an activity tracker and integrated into a social media platform.

5 CONCLUSION

The development of a social media chatbot connected to an activity tracker for increasing physical activity is aimed at individuals with little to no physical active, however, it may be more attractive to active individuals who already use an activity tracker. The results of this survey underscore the importance of exploring user preferences for physical activity digital interventions, including social media chatbots.

By doing so, important issues like exploring the preferences of individuals with sedentary lifestyles for such digital interventions will be highlighted. Preferences of all types of potential users for a social media chatbot for physical activity should be explored to produce a well-accepted intervention.

Further research should explore individuals with sedentary lifestyles' preferences for a social media chatbot connected to an activity tracker for increasing physical activity. In addition, exploring users' preferences for interacting with other users of the social media chatbot could contribute to enriching the experiences of potential users.

6 AUTHOR CONTRIBUTIONS

Conceptualization: DL, HS, EG, and EÅ. Methodology: DL, HS, EG, and EÅ. Data Analysis: DL, HS, EG, EÅ and AH. Article Writing: DL, HS, EG, EÅ and AH. All authors have read and agreed to the published version of the manuscript.

7 ACKNOWLEDGEMENT

We are grateful to all volunteers who participated.

8 REFERENCES

- [1] World Health Organization (2022), 'Physical Activity', <<https://www.who.int/news-room/fact-sheets/detail/physical-activity>>, accessed 15th May.
- [2] World Health Organization 'Physical activity', <<https://www.who.int/news-room/fact-sheets/detail/physical-activity>>, accessed 27 April 2022.
- [3] Luo, T. C., Aguilera, A., Lyles, C. R., and Figueroa, C. A., "Promoting Physical Activity Through Conversational Agents: Mixed Methods Systematic Review", *J Med Internet Res*, 23 (9), e25486, 2021
- [4] Piao, M., Ryu, H., Lee, H., and Kim, J., "Use of the Healthy Lifestyle Coaching Chatbot App to Promote Stair-Climbing Habits Among Office Workers: Exploratory Randomized Controlled Trial", *JMIR Mhealth Uhealth*, 8 (5), e15085, 2020
- [5] Kemp, S. (2022), 'Digital 2021: Norway', <<https://datareportal.com/reports/digital-2021-norway>>, accessed 15th May.
- [6] Statista.com (2022), 'Share of individuals who have access to a smartwatch in their household in 2020, by country', <<https://www.statista.com/statistics/1107874/access-to-smartwatch-in-households-worldwide/>>, accessed 2022-06-15.
- [7] Statista.com (2022), 'Share of individuals who had access to a smartphone in Norway from 2012 to 2020', <<https://www.statista.com/statistics/631747/norway-smartphone-user-penetration/>>, accessed 2022-06-27.
- [8] Henriksen, A., Haugen Mikalsen, M., Woldaregay, A. Z., Muzny, M., Hartvigsen, G., Hopstock, L. A., and Grimsgaard, S., "Using Fitness Trackers and Smartwatches to Measure Physical Activity in Research: Analysis of Consumer Wrist-Worn Wearables", *J Med Internet Res*, 20 (3), e110, 2018
- [9] Dalene, K. E., Kolle, E., Steene-Johannessen, J., Hansen, B. H., Ekelund, U., Grydeland, M., Anderssen, S. A., and Tarp, J., "Device-measured sedentary time in Norwegian children and adolescents in the era of ubiquitous internet access: secular changes between 2005, 2011 and 2018", *Int J Epidemiol*, dyac063, 2022
- [10] Laranjo, L., Ding, D., Heleno, B., Kocaballi, B., Quiroz, J. C., Tong, H. L., Chahwan, B., Neves, A. L., Gabarron, E., Dao, K. P., Rodrigues, D., Neves, G. C., Antunes, M. L., Coiera, E., and Bates, D. W., "Do smartphone applications and activity trackers increase physical activity in adults? Systematic review, meta-analysis and metaregression", *Br J Sports Med*, 55 (8), 422-32, 2021
- [11] Gabarron, E., Larbi, D., Denecke, K., and Arsand, E., "What Do We Know About the Use of Chatbots for Public Health?", *Stud Health Technol Inform*, 270, 796-800, 2020
- [12] Oh, Y. J., Zhang, J., Fang, M. L., and Fukuoka, Y., "A systematic review of artificial intelligence chatbots for promoting physical activity, healthy diet, and weight loss", *Int J Behav Nutr Phys Act*, 18 (1), 160, 2021
- [13] Larbi, D., Gabarron, E., and Denecke, K., "Social Media Chatbot for Increasing Physical Activity: Usability Study", *Stud Health Technol Inform*, 285, 227-32, 2021
- [14] Zhang, J., Oh, Y. J., Lange, P., Yu, Z., and Fukuoka, Y., "Artificial Intelligence Chatbot Behavior Change Model for Designing Artificial Intelligence Chatbots to Promote Physical Activity and a Healthy Diet: Viewpoint", *J Med Internet Res*, 22 (9), e22845, 2020
- [15] Coughlin, S., Roberts, D., O'Neill, K., and Brooks, P., "Looking to tomorrow's healthcare today: a participatory health perspective", *Intern Med J*, 48 (1), 92-96, 2018
- [16] Gabarron, E., Dorrnzoro, E., Bradway, M., Rivera-Romero, O., Wynn, R., and Arsand, E., "Preferences and interests of diabetes social media users regarding a health-promotion intervention", *Patient Prefer Adherence*, 12, 2499-506, 2018

- [17] University of Oslo (2022), 'Nettskjema (Online Form)', <<https://www.nettskjema.no>>, accessed 15th May.
- [18] Discord (2022), 'Create Space for Everyone to Find Belonging', <<https://discord.com/company>>, accessed 15th May.
- [19] Fotopoulou, A., Barratt, H., and Marandet, E., "A data-based participatory approach for health equity and digital inclusion: prioritizing stakeholders", *Health Promot Int*, 2021
- [20] Rivera-Romero, O., Konstantinidis, S., Denecke, K., Gabarron, E., Petersen, C., Househ, M., Merolli, M., and Mayer, M. A., "Ethical Considerations for Participatory Health through Social Media: Healthcare Workforce and Policy Maker Perspectives", *Yearb Med Inform*, 29 (1), 71-76, 2020
- [21] Maenhout, L., Peuters, C., Cardon, G., Compennolle, S., Crombez, G., and DeSmet, A., "Participatory Development and Pilot Testing of an Adolescent Health Promotion Chatbot", *Front Public Health*, 9, 724779, 2021
- [22] Chew, H. S. J., "The Use of Artificial Intelligence-Based Conversational Agents (Chatbots) for Weight Loss: Scoping Review and Practical Recommendations", *JMIR Med Inform*, 10 (4), e32578, 2022
- [23] Ferney, S. L. and Marshall, A. L., "Website physical activity interventions: preferences of potential users", *Health Educ Res*, 21 (4), 560-6, 2006
- [24] Ghanvatkar, S., Kankanhalli, A., and Rajan, V., "User Models for Personalized Physical Activity Interventions: Scoping Review", *JMIR Mhealth Uhealth*, 7 (1), e11098, 2019

Privacy-preserving Polygenic Risk Scoring using Homomorphic Encryption

Shaedul Islam¹, Huseyin Demirci² and Gabriele Lenzini²

¹University of Luxembourg, FSTC, Luxembourg,

²University of Luxembourg, SnT, Luxembourg huseyin.demirci@uni.lu

Abstract

The availability of direct-to-consumer genetic testing services and genome sequencing data bring novel opportunities for applications like genomic risk scoring where a polygenic disease risk score is calculated considering the statistical distribution of the disease associated SNPs. Nowadays, various websites are offering polygenic risk score estimations for various complex diseases. However, these services require the upload of the genomic data to their sites, which is a fairly sensitive personal data. Since, genome data uniquely identifies a person, anonymization is not sufficient alone and may become a threat in the long run. A potential solution is the use of cryptographic techniques along this goal. We propose to deploy homomorphic encryption, a technique which enables to do computation in encrypted data, for a web server providing polygenic risk score estimation. We implemented a proof-of-concept software to measure the performance of such a service with current technology. We also developed a GUI which facilitates the usage of homomorphic encryption for non-technical users. We conclude that recently developed homomorphic encryption libraries enable practical privacy-preserving genomic risk scoring services. Homomorphic encryption is becoming a strong alternative for practical secure privacy-preserving personalized medicine applications.

Keywords

Genomic privacy, Polygenic risk scoring, Homomorphic Encryption

1 INTRODUCTION

With the availability of direct-to-consumer genetic services, new types of web services are becoming available such as genotype imputation and polygenic disease risk prediction. A polygenic risk score (PRS) is an estimation of an individual's tendency to diseases which is calculated by considering statistical distribution of SNPs. Although, PRS cannot be used directly for diagnosis, it provides valuable information for risk stratification, prediction of the drug response or prognosis. To benefit from such services, the upload of the genomic data to the server side is required. On the other hand, genome data is strictly personal and sensitive. Uploading genomic data to a web server leads to privacy issues. There is a requirement for a new generation of genomic services which are privacy-preserving services and compatible with regulations like GDPR and CCPA.

Contribution.

In this work we demonstrate a proof-of concept implementation of a privacy-preserving web server which computes genomic disease risk scores from encrypted data. We accomplish this task by using new Homomorphic Encryption tools. In our use case scenario, the user sends her genomic variants encrypted to the insecure cloud server where all the operations are performed in the encrypted domain. The user gets back the results where she will decrypt the results on the fly. In this way, privacy

preserving versions of the existing web applications become practical. This is of critical importance, considering thousands of genomes are now being sequenced each day. This opens the way to get genomic consultancy about complex diseases without revealing sensitive variants and can be considered as a step towards the goal private personalized medicine services.

The source code of the implementation of the case study is available in the GitHub repository:

https://github.com/Shaedul/GenomeAnalysis_PySEAL

2 RELATED WORK

It is well known that genomic data is sensitive: it contains personal and confidential information such as the ancestry of an individual and of his/her kin, and their susceptibility and predisposition to specific diseases such as Alzheimer's, schizophrenia, and cancer. Therefore, the leakage of genomic data leads to irreversible ethnic and social discrimination (e.g., see [12]). Genomic data should be stored, processed, and shared in a privacy preserving manner.

Several methods have been proposed to enable genomics privacy. The most common choice, anonymization, is however provably ineffective in this case [9]. Very little piece of information, like 100 independent SNPs are enough to identify a person uniquely [13] and other sources

The 18th Scandinavian Conference on Health informatics, Tromsø, Norway, August 22-24, 2022. Organized by UiT The Arctic University of Norway. Conference Proceedings published by Linköping University Electronic Press at <https://doi.org/10.3384/ecp187>. © The Author(s). This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>

of information such as social media can be used to link with personal attributes [9]. Therefore, the natural choice is to resort to cryptographic methods. A practice which is also compliant with legal requirements in directives such as the Europe's General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA), is to encrypt the databases. This step is not enough though to ensure protection from privacy violation: if, for the purpose of processing, encrypted data sets are shared but decrypted before use, the risks of privacy violation remain.

Alternative methods are using cryptographic protocols such as Differential Privacy, Secure Multiparty Computation (SCM) and Homomorphic Encryption (HE). Each method has its own pros and cons and has different use cases in genomics.

Secure multiparty computation allows to collaboratively evaluate a function without revealing information to the parties. Jagadeesh et al. proposed to use SMC to identify diagnosis for monogenic rare diseases while preserving privacy for the remaining variants [11]. Later Akgun et al. improved the performance of the protocol [2]. Cui et al. proposed to use secure function evaluation to implement a privacy-preserving Human Leukocyte Antigen (HLA) matching [6].

Homomorphic encryption technique enables to perform computation inside encrypted data. In this way, computationally expensive operations can be privately outsourced to an insecure party like a public cloud. The existence of a fully homomorphic scheme is first proven by Gentry [8]. The first implementations of fully Homomorphic Encryption were far from being practical. It was considered as much slower than Secure Multiparty computation [11]. However, recent improvements in the implementation techniques and novel libraries enabled to perform practical applications.

Erman et al. used Paillier encryption system to implement a privacy-preserving cardiovascular disease risk analysis [3]. In [15] the authors developed a secure framework to conduct the rare variants analysis with a small sample size. Blatt et al. demonstrated that GWAS analysis of a real data set consisting of 25.000 individuals can be executed practically on encrypted data [4]. Recently, Harmanci et al. developed a secure imputation web server based on homomorphic encryption where untyped variant data is predicted from available genotype data with the help of a reference panel [10]. The applications of HE is becoming more available and practical in genomics area.

2.1. Polygenic risk scoring

Estimating the susceptibility to a disease is invaluable in medicine considering outcomes such as early diagnosis and prevention of common adult-onset conditions. Mendelian traits can point out significant outcomes such as the use of BRCA mutations, but they generally cover a small fraction of the population since they rely on rare variants. However, recent GWAS studies pointed out that, for most complex human diseases, joint consideration of common and low-frequency genetic variants that individually contribute small effects and provide much stronger predictions [14]. Generally, a polygenic risk score is calculated by computing the sum of risk alleles that are weighted by risk allele effect sizes. The effect sizes are estimated using

GWAS studies. We refer the interested reader to [5] for a comprehensive tutorial of polygenic risk scoring.

In this paper we utilize the framework of Impute.me [7] for PRS scoring. For specific diseases such as Breast cancer, Type-1 Diabetes, Alzheimer disease and Celiac disease, we have considered the effect sizes provided in this work. For the top-SNP approach, where the authors consider the most significant SNPs, we benefit from the same list of SNPs. In this model the following scores were used:

$$\begin{aligned} \text{PopulationScore}_{\text{snp}} &= \text{Frequency}_{\text{snp}} \times \beta_{\text{snp}}; \\ \text{ZeroCenteredScore} &= \\ &\beta_{\text{snp}} \times \text{Effect-allele-counts}_{\text{snp}} - \text{PopulationScore}_{\text{snp}}; \\ \text{Z-score} &= \text{ZeroCenteredScore} / \sigma_{\text{population}} \end{aligned}$$

where β is the reported effect size, $\text{Frequency}_{\text{snp}}$ is the allele frequency for the effect allele and $\text{Effect-allele-counts}_{\text{snp}}$ is the allele count of the genotype data (0,1 or 2) and $\sigma_{\text{population}}$ is the standard deviation of the population. This final Z-score is considered as the normalized metric for disease probability.

2.2. Homomorphic encryption

This cryptographic technique enables to perform computation on encrypted data with the help of a Homomorphic property, i.e. for every input plaintext pair x, y , we have:

$$\text{Enc}_k(x + y) = \text{Enc}_k(x) \oplus \text{Enc}_k(y),$$

where Enc denote the encryption function, k the encryption key and \oplus the homomorphic addition, respectively. The user encrypts the sensitive data by a public key and sends the encrypted data to an insecure party like a public cloud. Here, the insecure party performs operations in the encrypted data, without reaching the exact values of the data and sends the encrypted results to the user back. Here, the user decrypts the data with the secret key and obtains the desired result in his side. In this way, Homomorphic Encryption (HE) enables private outsourcing of computationally expensive operations to public clouds. This brings great flexibility from the security regulations (like GDPR) points of view since, even in the case of a data breach at the cloud side, no sensitive information is lost.

Although the idea of a HE system existed previously, the existence of a fully homomorphic system was only shown in 2009 by Gentry [8] using lattice algebra. There are different types of Homomorphic Encryption operations such as Partially, Somewhat and Fully Homomorphic systems. Partially Homomorphic systems preserve the operations for one single operation whereas Somewhat Homomorphic systems support homomorphism for two different types operations (i.e. addition and multiplication) for a bounded number of operations. Finally, Fully Homomorphic systems preserve the homomorphism for both types of operations unbounded number of times. We refer the reader to [1] for a comprehensive survey on the theory and implementations of Homomorphic Encryption schemes. The initial implementations of HE were not very practical. It was taking days to make simple calculations. It has been suggested that HE computation is about 5.000-10.000 times slower when compared to other privacy preserving techniques such as Secure Multiparty Computation [11]. Recently, new libraries and

implementations such as Microsoft’s SEAL and IBM’s Fully Homomorphic Encryption Toolkit For Linux have been made available for developing HE applications. These tools make HE applications more practical and available for a variety of cases including genomics analysis [15,10,4].

3 PRIVACY PRESERVING GENOMIC RISK SCORING

We briefly explain our privacy preserving scheme. When a user, Alice, who owns her genomic variants (as a .vcf file) decides to benefit from the PPPRS (Privacy- preserving Polygenic Risk Score) web server, she carries out the following tasks, respectively:

1. Alice generates a pair of Public and Private keys, denoted by PK and SK, respectively. She sends a copy of the Public Key to the cloud server and keeps the secret key to herself.
2. She requests a PRS for a specific disease from the web server.
3. The web server sends the list of related SNPs which had previously been determined with GWAS studies.
4. Alice encrypts the list of related variants and sends the encrypted list to the cloud.
5. The cloud server performs PRS calculations using encrypted values according to Algorithm and sends the encrypted PRS result to the user.
6. Alice uses her secret key to decrypt her PRS score.

The general outline of this application is depicted in Figure 1.

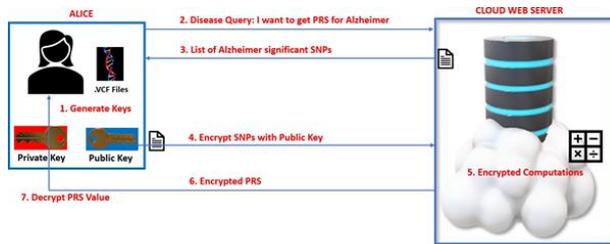


Figure 1. Privacy Preserving PRS.

4 IMPLEMENTATION

4.1. User side

To facilitate the use of the system, we have developed a Graphical User Interface (GUI). The structure of the GUI can be observed in Figure 2. The main aim of the GUI is to facilitate the use privacy preserving operations for the disease risk scoring. The operation of the client site is completed in a few steps. First, the user must register and log in with an anonymous user id and email address. Then, she generates the Public-Secret key pair, where the Public key will be shared with the cloud server to enable to encrypt the data and send to cloud. The user needs to store the secret key privately. Later, the user browses the genome variants data in VCF file format and selects the trait to interpret the disease risk score. After this step, the genomic variants data are encrypted, and the status field shows the progress of the encryption process. Finally, the encrypted data is sent to the cloud for further analysis. After the analysis is performed inside the cloud, the user receives the encrypted results and decrypts in this GUI platform using the Secret key. A the

end of the process the user could observe both the calculated normalized Z-score as well as the graphical interpretation of this Z-score.

4.2. Cloud side

We have implemented the homomorphic operations in the cloud site by using Amazon Web Service (AWS), EC2 environment. The cloud server has GWAS reference data in unencrypted form to support the PRS calculation. The reference data obtains the related Chromosome Number, SNPs ID, Effect size, Minor/Allele Frequency (MAF). When the cloud receives any encrypted genome variants data with the request of PRS for a specific disease, then reference data related to the disease will also be encrypted using the user public key. Afterwards, the cloud performs the computation for the risk score and sends the result to the client in encrypted form. We setup the Microsoft PySEAL (Python version of Simple Encrypted Arithmetic Library) library to implement the homomorphic encryption environment inside the cloud.

Computational Performances.

The time performance of the HE experiments is provided in Table 1. These experiments are carried out with a laptop with i-7 1.8- 2.3 GHz. CPU and 16 GB RAM with 64-bit Windows Enterprise operating system. We used Python version 3.7.4 on the Spyder Environment for the implementation.

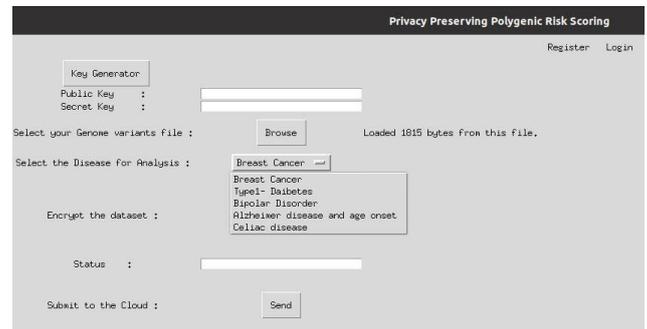


Figure 2. GUI of the Privacy-Preserving PRS Software.

Disease	# SNPs	Enc.	Dec.	Cloud
Celiac	19	1.557	0.007	1.88
Alzheimer	33	2.803	0.012	3.27
T I Diabetes	45	3.564	0.016	4.47
Breast Cancer	635	255.422	0.23	262.83
Whole Genome	3800000	1528	1.406	376.10

Table 1. Time performance of HE operations (seconds).

5 DISCUSSIONS

What we have exposed so far shows that is feasible to conduct secure genomic polygenic risk scoring using homomorphic encryption. If the number of SNPs associated with the disease is around 100, then whole operation can be executed in seconds. We note that another possibility is the encryption of the whole variant set (around 3.5-5 million variants) in the .vcf file once and then carry

out the required operations. In this case, the dominant part of the computation is the encryption part.

6 CONCLUSION

In this work we have presented a proof-of-concept secure web server for the estimation of polygenic risk scores by implementing an algorithm using Homomorphic Encryption (HE). With the recently deployed secure imputation this study fills the gap for privacy preserving genomic analysis. Our study demonstrates that recent HE libraries enable to do risk score estimations without revealing the sensitive genome information. We have conducted experiments for various diseases and number of SNPs.

7 REFERENCES

- [1] Aksu, H., Uluagac, A.S., Conti, M.: “A survey on homomorphic encryption schemes: Theory and implementation” in *ACM Computing Surveys (CSUR)* 51(4), 1–35, 2018.
- [2] Akgün, M., Ünal, A.B., Ergüner, B., Pfeifer, N., Kohlbacher, O.: “Identifying disease-causing mutations with privacy protection” in *Bioinformatics*, 2020.
- [3] Ayday, E., Raisaro, J.L., McLaren, P.J., Fellay, J., Hubaux, J.P.: “Privacy-preserving computation of disease risk by using genomic, clinical, and environmental data”, in: *2013 USENIX Workshop on Health Information Technologies (HealthTech 13)*, 2013.
- [4] Blatt, M., Gusev, A., Polyakov, Y., Goldwasser, S.: “Secure large-scale genome-wide association studies using homomorphic encryption”, in “*Proceedings of the National Academy of Sciences*, 117(21), 11608–11613, 2020.
- [5] Choi, S.W., Mak, T.S.H., O’Reilly, P.F.: “Tutorial: a guide to performing polygenic risk score analyses”, in *Nature Protocols*, 15(9), 2759–2772, 2020.
- [6] Cui, J., Li, H., Yang, M.: “Privacy-preserving computation over genetic data: Hla matching and so on”, *IACR Cryptol. ePrint Arch.* 2019, 1305, 2019.
- [7] Folkersen, L., Pain, O., Ingason, A., Werge, T., Lewis, C.M., Austin, J.: “Impute. me: an open-source, non-profit tool for using data from direct-to-consumer genetic testing to calculate and interpret polygenic risk scores” in *Frontiers in genetics* 11, 578 2020.
- [8] Gentry, C.: “Fully homomorphic encryption using ideal lattices” in: *Proceedings of the forty-first annual ACM symposium on Theory of computing*, pp. 169–178, 2009.
- [9] Gymrek, M., McGuire, A.L., Golan, D., Halperin, E., Erlich, Y.: “Identifying personal genomes by surname inference” in *Science* 339(6117), 321–324, 2013.
- [10] Harmanci, A.O., Kim, M., Wang, S., Li, W., Song, Y., Lauter, K., Jiang, X.: “Open imputation server provides secure imputation services with provable genomic privacy” *bioRxiv*, 2021.

- [11] Jagadeesh, K.A., Wu, D.J., Birgmeier, J.A., Boneh, D., Bejerano, G.: “Deriving genomic diagnoses without revealing patient genomes” in *Science* 357(6352), 692–695, 2017.
- [12] Naveed, M., Ayday, E., Clayton, E.W., Fellay, J., Gunter, C.A., Hubaux, J.P., Malin, B.A., Wang, X.: “Privacy in the genomic era” in *ACM Comput. Surv.* 48(1), Aug 2015.
- [13] Pakstis, A.J., Speed, W.C., Fang, R., Hyland, F.C., Furtado, M.R., Kidd, J.R., Kidd, K.K.: “Snps for a universal individual identification panel” in *Human genetics* 127(3), 315–324, 2010.
- [14] Torkamani, A., Wineinger, N.E., Topol, E.J.: “The personal and clinical utility of polygenic risk scores” in *Nature Reviews Genetics* 19(9), 581–590 2018.
- [15] Wang, S., Zhang, Y., Dai, W., Lauter, K., Kim, M., Tang, Y., Xiong, H., Jiang, X.: “Healer: homomorphic computation of exact logistic regression for secure rare disease variants analysis in GWAS” in *Bioinformatics* 32(2), 211–218, 2016.

8 ACKNOWLEDGEMENT

This work has been supported by the EU 956562, MSCA-ITN-2020 - Innovative Training Networks, “Legality Attentive Data Scientists” (LeADS) project. The authors wish to thank Dr. Patrick May for valuable discussions on the topic of polygenic risk calculation.

Citizens' use of Health Information Technology between 2013-2021 in Denmark: A longitudinal study

Jeppe Eriksen¹, Kristina Tornbjerg Eriksen¹ and Pernille Bertelsen¹

¹Aalborg university, Department of Planning, Denmark je@plan.aau.dk

Abstract

An increasing number of citizens with multiple chronic conditions and technological innovations enabling new types of treatments pressure the Danish healthcare sector economically.

The solution so far has been increased patient responsibility and the application of digital healthcare solutions.

This longitudinal study examines how Danish citizens between 2013-2021 interact with Health Information Technology (HIT) and digital data.

Results show that the Danes' use of HIT and digital data has increased over the period. Additionally, the numbers reveal that education, gender, age and chronic conditions influence how HIT and digital health data are used, which is relevant from a health inequity perspective.

Keywords

Citizen survey, Health Information Technology, Digital health, Digital data, eHealth

1 INTRODUCTION

Citizens' use of digital health, eHealth and Health Information Technology (HIT) are becoming increasingly important within a Danish context. This is largely due to the demographic changes resulting in an increasing number of elders, citizens with multiple chronic conditions and relatively fewer labour-active citizens [1], [2]. Simultaneously, technological innovations and cultural alterations in the healthcare sector have transformed treatment options and the role of healthcare professionals (HCPs) and patients [2]. An increased implementation and integration of digital health that affects clinical practice; requiring, adaptiveness and continuously updated digital knowledge and skillsets among healthcare professionals [1]–[4]. Digital skills that are equally important for patients to acquire and possess as the use of telemedicine and patient-reported outcomes (PROs) are spreading across the Danish healthcare system [1], [5]. The digital development has gotten momentum after the COVID-19 crisis that forced and enabled the healthcare sector to apply known technologies, such as video consultations, more extensively and introduce new solutions, for example, the corona-chatbots used to answer citizen questions related to COVID-19 [4]. The growing use of eHealth means that accessibility and security are pivotal; especially, since current health policies emphasise that patients' health data should be utilised to the greatest extent [1], [3], [4]. Hence, the current healthcare and societal challenges are countered through the use of HIT and eHealth innovations potentially enabling better use of resources while placing increased individual responsibility on the citizens [1], [3], [5]. Thus, a digitalised and functional healthcare sector requires actively participating patients who are willing - and able - to self-manage their conditions; a challenge, considering

that not all citizens have the same digital capabilities and diverse participation preferences when handling their health [1], [5]. Hence, the intentions incorporated into the political strategies are one thing, another is how citizens factually make use of HIT and digital data as part of their everyday life. Thus, questions indicating how often and in what way citizens make use of HIT and digital healthcare are relevant. Moreover, in a time where health inequities are focal in health politics, empirical data elucidating how gender, health status, chronic illness, education and age are linked to citizens' actions in a digitalised healthcare system.

It is in this context 'The citizen survey' has been conducted by researchers at the Department of Planning at Aalborg University biannually since 2013. The five surveys examine citizens':

- Use of HIT and digital data collection
- Communication with HCPs
- Attitudes towards HIT
- Use of health portals
- Perceptions of how HCPs make use of HIT in clinical practice [6]–[9]

Five relevant themes; however, to narrow the scope this paper focuses primarily on Danish citizens' use of HIT and digital data collection over time.

2 METHOD

The results are based on longitudinal data collected every second year between 2013-2021 via a cross-sectional design. To ensure the quality and representativeness of data, the surveys have been conducted by MEGAFON, who is a company specialized in collecting data through the use of quantitative methods [10]. The original conceptualisation and methodology behind the cross-

The 18th Scandinavian Conference on Health informatics, Tromsø, Norway, August 22-24, 2022. Organized by UiT The Arctic University of Norway. Conference Proceedings published by Linköping University Electronic Press at <https://doi.org/10.3384/ecp187>. © The Author(s). This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>

sectional and longitudinal design emerged through a collaboration between the authors and colleges at the Department of Planning at Aalborg University. The formal analysis, data curation, visualisation and writing of the original draft in the present study were conducted by the first author while the other authors contributed by reviewing and editing the paper.

Since the longitudinal data is based on five different cross-sectional studies, the total number of participants slightly varies, and respondents might not be identical for each iteration. However, the samples are randomly selected based on members of the MEGAFON panel, which is a group of citizens carefully and systematically selected to ensure representativeness and generalisability [10]; mandatory validation factors when conducting and assessing quantitative studies [11]. In addition to the quality of the sample, the size also affects the validity of data. Hence, Table 1, lists the number of respondents in each of the five surveys.

Year	Respondents (N)	Internet/phone interviews (n)
2013	1058	931/127
2015	1059	950/109
2017	1033	933/100
2019	1055	955/100
2021	1002	900/102

Table 1 Number of respondents and type of collection

As Table 1 illustrates, respondents vary between 1002-1059 over the years. The number of respondents completing the internet questionnaire is between 900-955 while the number of citizens included through telephone interviews ranges between 100-127. Hence, around 10% of the respondents are included through telephone interviews which is a deliberate choice, to reach some of the citizens that (for different reasons) are excluded when studies are based merely on internet questionnaires. All respondents are anonymised.

In the respective study, a statistical significance level of 95% (p-value of 0.05%) is applied. Accordingly, with 1000 respondents, which is included in all the surveys, there is a 95% probability that results are real and reflect the examined population +/- 1,35% probability that results are caused by randomness.

Hence, the results in this paper are statistically representative in a Danish context and describe citizens' use of HIT and digital healthcare over time. Due to continuous new knowledge, changing societal tendencies, technological innovations and varying health policies, questions have over time been altered, removed and replaced; a process displayed in Table 2.

Question	Year				
	2021	2019	2017	2015	2013
01	x	x	x	x	x
02	x	x	x	x	x
05	x	x	x	x	x
06	x	x	x	x	
07	x	x	x	x	

08	x	x	x	x	
09	x	x	x	x	x
09A	x	x	x	x	x
10	x	x			
11	x	x			
12A	x	x	x	x	
12C	x	x			
18A	x	x	x		
14F	x	x	x		
19	x		x	x	
19B	x	x	x		
19C	x	x			
20	x	x	x	x	
25	x	x			

Table 2 Questions included across surveys over time

Table 2 shows to what extent the questions included in this paper are included in the five surveys. The cross symbolises continuation and blank spaces indicate change - or that no such question has been included formerly. Hence, the reason several of the boxes are blank in 2013 and 2015, as new types of questions or alterations were added in later iterations of the survey. Consequently, it differs how far back data pertaining to each variable has been collected.

The latest questionnaire from 2021 consists of 34 different questions including background variables. Table 3 provides an overview of the variables included in this paper. Only the essentials of the questions are categorised and displayed in the table; however, the exact questions are in most cases included in the results section.

Question	Category	Type of variable
01	Age	Interval
02	Gender	Nominal
05	Education	Ordinal
06	Supporting/helping an elder, handicapped, or ill person in their contact with the healthcare system	Nominal
07	Types of IT used	Nominal
08	Health status	Ordinal
09	Chronic condition	Nominal
09A	Type of Chronical condition	Nominal
10	Interaction with health institutions (physical/phone)	Nominal
11	Interaction with health institutions (digital)	Nominal
12A	Type of IT used when communicating with GP digitally	Nominal
12C	Reasons for communicating with GP digitally	Nominal
18A	Citizen asked to collect data	Nominal

14F	Citizen asked to complete a questionnaire on treatment	Nominal
19	How do you follow your public health data	Nominal
19B	Individual initiated collection of health data	Nominal
19C	Information search on health and diseases	Nominal
20	How have you used the internet and mobile apps	Nominal
25	Granted relatives access to healthcare data	Nominal

Table 3 Type of content and variables

As table 3 shows most of the variables included in this paper are nominal and concern:

- a) The type of HIT citizens uses
- b) How citizens use HIT
- c) Citizens' experiences with HCPs' use of HIT
- d) Citizens' attitudes towards HIT

The nominal variables are divided into different qualitative response categories and in most cases, respondents are allowed to provide more than one answer to a question. The background variables (Age, Gender, Education, Health status and Chronic condition) differentiate by either being on a different scale and/or having different response categories.

Age (avg.)(n/%)		Gender (avg.)(n/%)		Education (avg.)(n/%)		Health status (avg.)(n/%)	
18-29	201/19,3%	Male	528/50,7%	Primary/ elementary school	65/6,3%	Very good	248/23,85%
30-39	159/15,3%	Female	513/49,2%	Secondary/ middle school	18/2,2%	good	493/47,53%
40-49	182/17,5%			High school	100/9,6%	Neither bad or good	195/18,83%
50-59	175/16,8%			Vocational education	188/18%	Bad	88/8,5%
60-69	157/15,1%			Further education (< 3 years)	132/12,7%	very bad	12/1,1%
70-	168/16,1%			Bachelor's education (3-4 years)	309/29,7%		
				Higher education (4 years+)	221/21,2%		

Table 4 Population characteristics

The variables age and gender are weighted to ensure that they resemble the Danish population. Around 50% of the Danish population have a bachelor's degree or higher education; in contrast, around 18% have finished high school or a shorter type of school as their highest completed education.

Citizens' perceptions of their health status are relatively consistent between 2013 and 2021 with an average majority of 71,38% of the citizens who believe their health status is good/very good. In the same period, the number of citizens suffering from one or more chronic conditions has increased by 14,9 percentage points, from 32,3% in 2013 to 47,2% in 2021 (Figure 1). Hence, it is interesting how the

3 RESULTS

The results section is divided into the following five areas 'Population', 'Communication', 'Collection of data', 'Data monitoring' and 'Relatives' access to data', describing different aspects of the citizens' use of HIT and digital data. In each section, the numbers comprising the graphs and tables are scrutinized and a profile of the typical citizen linked to a specific subject is constructed. These profiles are based on absolute numbers; consequently, these profiles are strongly shaped by the characteristics of the population that are listed in Table 4. In other words, the profiles do not indicate how different groups of the population relatively use HIT and digital healthcare.

Some of the tables contain the (n/%) -symbol, which is used to explain the displayed type of numbers. The n-symbol indicates the actual number of citizens and the %-symbol is the percentual size. As an example, 201/19,3% of the citizens participating in the study are between 18 and 29 years of age (Table 4).

3.1 Population

In this section general characteristics of the population are described pertaining to age, gender, education, health status and chronic conditions, primarily based on the numbers included in Table 4. These factors are considered background variables and used to analyse citizens' use of HIT and digital data in more detail. The included abbreviation avg. in Table 4. refers to the fact that the numbers are averages of the examined period.

Danes' perception of their health status remains constant while the number of citizens suffering from one or more chronic conditions increases.

In the segment of the population suffering from a chronic condition between 2013 and 2021, 90,9-95,9% state that they have a physical condition whereas 11,6-18,3 % are challenged by mental issues (Question 09A). Hence, numbers indicate that some of the citizens suffer from both physical and mental conditions.

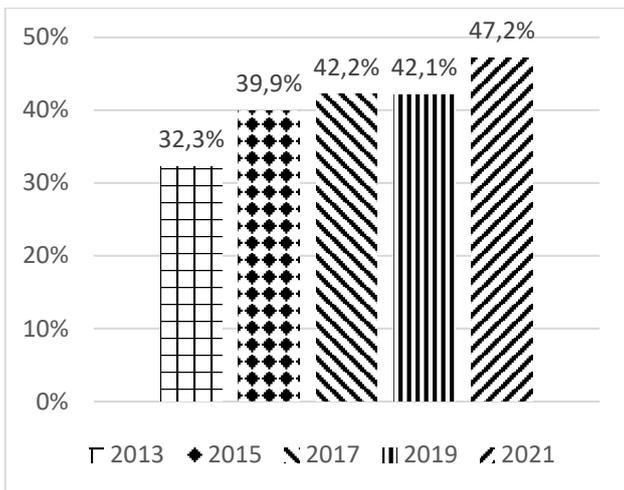


Figure 1 Citizens with one or more chronic conditions.

Considering the increasing number of elderly and citizens who are suffering from one or more chronic conditions, one might assume that the number of citizens supporting/helping an elder, handicapped, or ill person in their contact with the healthcare system has increased from 2013-2021 (Question 6).

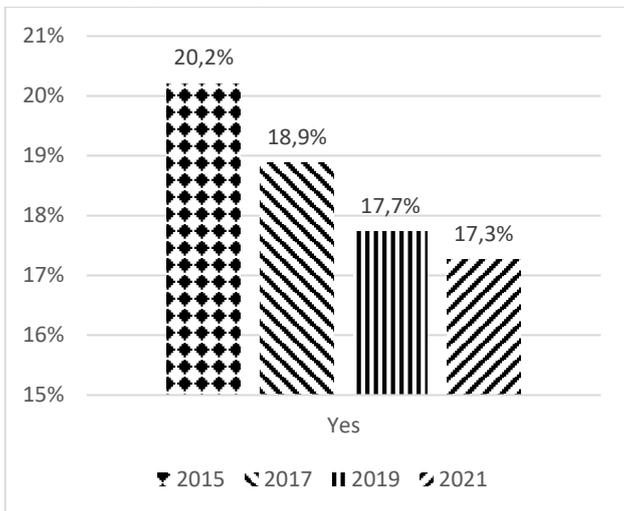


Figure 2 Supporting/helping an elder, handicapped, or ill person in their contact with the healthcare system.

However, as Figure 2 reveals, there has on the contrary been an incremental decrease in the number of citizens supporting an elder, handicapped, or ill person in their contact with the healthcare system.

The characteristics of the typical citizen who supports/helps an elder, handicapped, or ill person in their contact with the healthcare system is a male (55,5%), 40-69 years old (34%), holding a bachelor's degree or higher education (55,3%), with a good/very good health status (71,5%).

3.2 Communication

This section describes how citizens make use of HIT when communicating with their GP from at home and in these cases for what purposes HIT is used.

Specifically, citizens are asked ‘When communicating with your GP from home what equipment are you then using?’ (Question 12A) and ‘For what purposes have you used IT when communicating with your GP?’ (Question 12C).

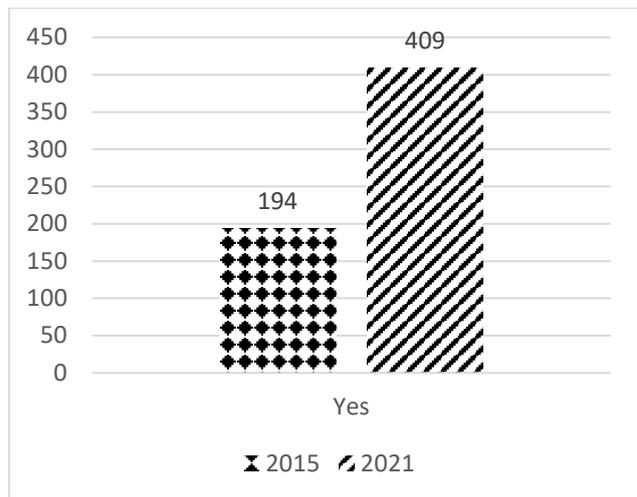


Figure 3 Citizens using smartphone at home to communicate with their GP.

In this context, smartphones are a focal technology, which citizens increasingly use to communicate with their GP (Figure 3). In 2015, 18% of the population used smartphones in their communication with their GP, which increased to 41% in 2021, an increase of 111%. In the same period, the relative number of citizens who made use of iPad/tablet when communicating with their GP shifted from 15 to 17%, whereas the use of computers rose from 43 to 52%. Thus, the use of smartphones has increased relatively most, compared to iPad/tablet and computers. Nonetheless, computers are still the preferred HIT when communicating with the GP.

The characteristics of the typical citizen who uses a smartphone at home to communicate with their GP is a male (53,6%), 18 to 49 years old (62,7%), holding a bachelor's degree in education (34,2%), and with good health status (45,5%).

For what purposes citizens then use IT when communicating with their GP is a different matter, which is displayed in Table 5.

Purposes	2019 (n/%)	2021 (n/%)
eConsultation (email)		465/46%
Video consultation		65 /6%
Uploaded readings from home monitoring		68 /7%
Booking an appointment	467/44%	402/40%
Prescription renewal	383/36%	424/42%
Referral to a specialist doctor	81/8%	81/8%
Corona test		229 (23%)
Answers test results	252 (24%)	219 (22%)
Answers blood test/lab results from specialist doctors or hospital	251 (24%)	273 (27%)

Table 5 Types of purposes when using IT in the communicating with the GP

Table 5 indicates that the purpose of using IT when communicating with ones GP is relatively stable from 2019 to 2021. Under normal circumstances, this would not be a

surprise considering the short period between the surveys; however, the surveys were conducted before and after the COVID-19 outbreak in 2020, which should be taken into account when interpreting the data. In this context, 6% using video consultation in 2021 is a relatively low number, whereas 46% having used eConsultation in the communication with their GP is less surprising, making it the most frequent purpose of use.

The characteristics of the typical citizen who uses eConsultation to communicate with their GP is a female (51,6%), citizen of all ages, holding a bachelor's or higher degree of education (58,6%), with a good health status (44,8%) and with one or more chronic conditions (55,3%).

3.3 Collection of data

In this section different types of data collection conducted by the citizens are described.

On the subject of data collection, citizens were asked: “Did your practitioner ask you to collect data meant for your treatment?” (Question 18A), referring to digital as well as a paper-based collection of data.

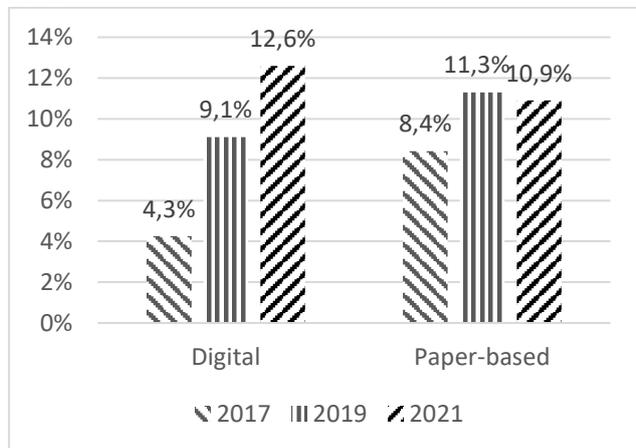


Figure 4 Citizens' collection of data for treatment requested by a practitioner.

From 2017 to 2021 citizens' digital data collection requested by a practitioner increased by 8,3 percentage points, from 4,3% to 12,6%, a relative increase of 193%. Paper-based data collection has fluctuated since 2017 but is at 10,9% in 2021. Hence, the number of citizens collecting digital and paper-based data is close to similar in 2021 but the tendency is that digital data collection requested by a practitioner is a growing phenomenon.

The characteristics of the typical citizen who is requested to collect digital data for treatment by a practitioner is a female (54%), 50 years or older (57,9%), holding a bachelor's or higher degree of education (58,9%), with a good health status (44,5%) and with one or more chronic conditions (66,7%).

In question 19B citizens are asked, “Have you on your own initiated collection of health data (e.g. from a fitness tracker or an app) and showed it to your health practitioner?” (Figure 5).

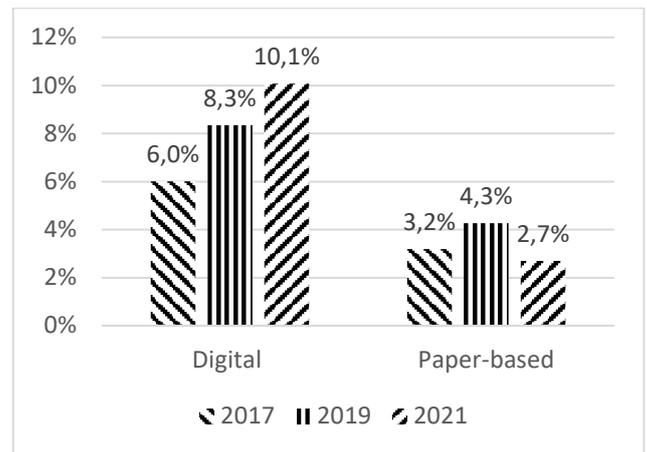


Figure 5 Individual initiated health data collection.

Even though the data type and the used HIT might be different, the same pattern is repeated. From 2017 to 2021, the number of citizens collecting digital data on their own initiative increased by 68,3% (4,1 percentage points) and the paper-based collection of data varied between 2,7 and 4,3%. Comparing the results displayed in Figures 4 and 5, what is interesting is the relative difference between data collected digitally versus paper-based in 2021. The frequency of the two approaches is close to similar when initiated by a practitioner, 12,6% and 10,9% (Figure 4), whereas 4 times as many initiated digital data collection on their own compared to the paper-based collection, 10,1% and 2,7% (Figure 5).

The characteristics of the typical citizen who initiates the collection of digital health data on their own are independent of gender, mainly between 18 and 29 years (21,6%) and 50 to 59 years of age (20,8%), holding a bachelor's or higher degree of education (55,8%), a good health status (40,2%) and one or more chronic conditions (56%).

The third question in this section concerns citizens' evaluation of treatment based on digital and/or paper-based questionnaires (Figure 6). Specifically, citizens are asked: “In your contact with the healthcare services, have you then been asked to complete a questionnaire to evaluate your treatment?” (Question 14F).

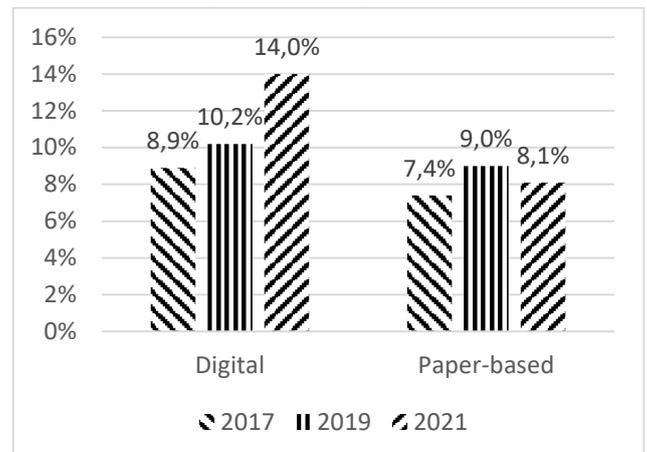


Figure 6 Citizens' completion of questionnaires to evaluate treatment.

Figure 6 shows that citizen evaluations via digital questionnaires are a growing phenomenon indicated by a

5,1 percentage point increase, which is an increase of 57,3% between 2017 and 2021. Once again, the same pattern is repeated, as the rate at which paper-based questionnaires are collected over the period fluctuates and seems to be declining.

The characteristics of the typical citizen who has evaluated their treatment via a digital questionnaire are independent of gender and age, with a bachelor's degree in education (31,5%), a good health status (47,5%) and one or more chronic conditions (59,7%).

3.4 Data monitoring

Section 3.4 concerns citizens' use of digital platforms and websites. Examining this issue, citizens were asked: 'How do you follow your public health data on the internet?' (Question 19).

Websites or apps	Computer		Smartphone or tablet	
	2015	2021	2015	2021
e-Boks	37,2%	45,2%	19,3%	47,1%
GP's homepage	22%	17,9%	9,5%	20,4%
Sundhed.dk	33,4%	50,5%	7,7%	55,1%
Borger.dk	29,6%	29,6%	8,8%	26%

Table 6 Websites and apps used by citizens to follow their public health data

The data displayed in table 6. reveals a significant increase in the use of smartphones/tablets in a Danish context, independent of the website. The most pronounced increase concerns the Sundhed.dk app, the Danish health portal where citizens can check their COVID-19 results. It was used by 7,7% of the citizens in 2015 and by 55,1% in 2021; an increase of 616% (47,4 percentage points). Another frequently used website is eBoks, visited by 45,2% (via computer) and 47,1% (via smartphone or tablet) of the population in 2021, highlighting how eBoks is a key website in mediating communication between citizens and the public sector. Based on these numbers, smartphones/tablets were in 2021 used as frequently as computers by citizens when accessing public health data.

When asking about the citizens' use of the Shared Medication Record (app), the data does not allow to distinguish between computers and smartphones/tablets why this information is not included in Table 6. However, the results are similar to the notable development in the use of smartphones and tablets; hence, 6% of the Danish population made use of the Shared Medication Record (app) in 2017, which increased to 29,7% in 2021, which is an increase of 396% (23,7 percentage point).

In answering Question 19, citizens were also able to reply that they did not follow their public health data online. Hence, 42,3% did not follow their public health data online in 2015 a number declining to 12,6% in 2021. In other words, 87,4% of the population followed their health data online in 2021 compared to 57,7% in 2015, which is an increase of 51% (29,7 percentage points).

The characteristics of the typical citizen who follows public personal health data on Sundhed.dk via smartphone or tablet are, a male (53,7%), between 18 and 29 (20,9%) or 40 and 49 years of age (20%), holding a bachelor's or

higher degree of education (55,1%) and a good health status (47,6%).

The citizens were also asked: 'Have you, within the latest year, had experience with any of the following activities by using the internet, email or an app?' (Question 20).

Activities	2015 (n/%)	2021 (n/%)
Health apps	215/20,3%	661/66%
Questions for HCPs	264/24,9%	189/18,9%
Communication with peers	87/8,2%	106/10,6%
Questions on diet and training	58/5,5%	88/8,8%
Discussions on diet and training	66/6,2%	62/6,2%
Diaries on diet and training	146/13,8%	207/20,7%
Self-help programs	212/20%	220/22%
Healthcare services in another language	213/20,1%	203/20,3%

Table 7 Citizens' activities on the internet, email and apps.

As displayed in Table 7, the most significant change in citizens' activities concerns the use of health apps as it has increased by 45,7 percentage points (225%) between 2015 and 2021. Communication with peers, the use of self-help programs as well as discussions and questions on diet and training are as frequent activities in 2015 as in 2021, while questions for HCPs have dropped by 6 percentage points (25%).

The characteristics of the typical citizen who uses health apps are male (54,9%), 18-29 years old (26,7%), with a bachelor's or higher degree of education (55,9%), a good health status (47,6%) and not suffering from one or more chronic conditions (55,3%).

3.5 Relatives' access to data

Since December 2015 patients have, through a digital power of attorney, been able to grant relatives access to their health data [12]. Thus, citizens are asked: 'Have you made use of the option to provide your relatives with access to your health data?' (Question 25)

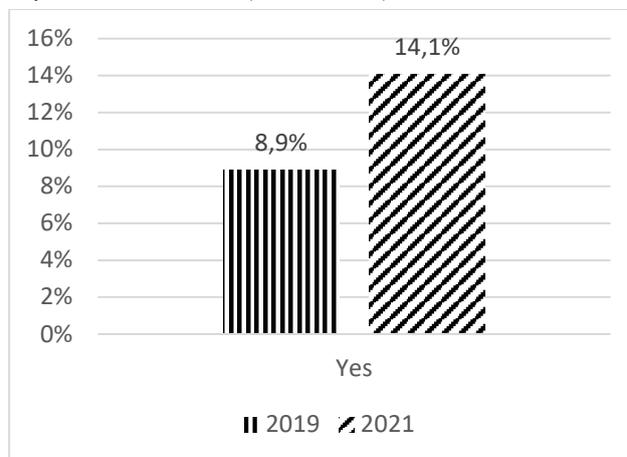


Figure 7 Citizens who granted relatives access to their health data.

As Figure 7 illustrates, citizens who allow relatives to access their health data has from 2019 to 2021, increased by 5,2 percentage points (58,1%), which might reflect the growing amount of elderly and citizens with one or more chronic conditions.

The characteristics of the typical citizen who granted relatives access to their health data are largely independent of gender, 70 years or older (23,9%), holding a bachelor's or higher degree of education (55,5%), a good health status (41,3%) and one or more chronic conditions (55,3%).

4 DISCUSSION

Based on the results section, three patterns are apparent; firstly, the Danish population are increasingly making use of HIT and digital healthcare services, secondly, the collection of digital patient data is becoming increasingly important and thirdly, the majority of citizens, and therefore most of the digital health users, have a bachelor's or higher degree of education and perceive their health status as good or very good.

4.1 Use of HIT

The widespread use of digital solutions is in the results section exemplified by citizens' extensive use of smartphones when communicating with a GP (Table 3), the Sundhed.dk app (Table 6) and health apps (Table 7), which have increased by 111%, 616% and 225%, respectively, from 2015 to 2021. Another noticeable finding is the number of eConsultations, which were used by almost half of the population in 2021 (Table 5). Results aligned with studies conducted by The Danish Organization of General Practitioners (PLO), revealing that eConsultations comprised 3,5% of all consultations in 2008 and 19,2% in 2021 [13], indicating increased use of eConsultation in healthcare. A development synergising with the results from a recently published report by Statistics Denmark on the use of IT in the Danish population in 2021. Accordingly, 92% between 16-74 years communicated with the public sector digitally, 68% made use of digital self-service solutions and 85% were active on social media [14]. Based on these numbers, the use of HIT in Denmark seems to be growing, which to some extent probably is caused by the COVID-19 crisis, similar to the experiences in other countries [15].

4.2 Patient-generated health data

Another main finding is how the collection of digital data by citizens is becoming a more normalised practice. Hence, from 2017 to 2021 the collection of data requested by a practitioner (Figure 4), digital data collection initiated by the citizen (Figure 5) and completion of digital questionnaires (Figure 6), has increased by 193%, 68,3% and 57,3%, respectively. Numbers verifying the increased importance of patient-generated and patient-reported data in a Danish context and 'the active patient', reflecting the need for increased patient participation if a healthcare system is to be based on patient-reported data [5].

4.3 The typical HIT-user

The third main finding concerns how particular segments of the population are the primary users of digital healthcare. Accordingly, in absolute numbers, the typical citizen using the digital health solutions has a bachelor's or higher degree

of education, perceives their health status as good or very good and often has one or more chronic conditions. Regarding age and gender, there seems to be no clear pattern. These findings probably mirror the fact that the majority of citizens, relatively consistently over the examined period, categorise their health status as good/very good (71,38%) and have a bachelor's or higher degree of education (50,9%). Simultaneously, the number of citizens with one or more chronic conditions has increased by 46,1% from 2015 to 2021 (Figure 1). In this light, it is noteworthy how the number of citizens who support an elder, handicapped, or ill person in their contact with the healthcare system has declined incrementally from 20,2% in 2017 to 17,3% in 2021 (Figure 2). A relatively moderate number considering that on average 40% of the Danish population were engaged in voluntary work in 2021 [16].

4.4 Consequences and solutions

On the one hand, these numbers might indicate that citizens to a higher degree are becoming digital competent and capable of self-managing their conditions. This could be perceived as a manifestation of current health policies' emphasis on citizens' individual responsibility and engagement in their health and disease handling [4], [5]. On the other hand, these results might also explain why inequity in health persists as a pivotal problem in a Danish context. Thus, this study confirms a concern raised by Professor Morten Sodemann in the article '*The healthcare system is drowning in trivialities*' [17], in which he claims that the healthcare system is shaped according to the needs of the middle class who represents the majority of the population, making this approach a good business. Consequently, the minority of the population, the resource-demanding citizens who are most in need, are neglected and excluded from vital health services [17]. Arguably, awareness of access barriers (e.g. access to the internet and digital services), and capability barriers (e.g. literacy, digital skills, language difficulties and domain knowledge) are important to consider in order to ensure an inclusive digital healthcare system [18]. To achieve this, the design and usability of digital healthcare systems are decisive [4], which is a focal point in Chris Showell's and Paul Turner's work titled '*The PLU Problem: Are We Designing Personal ehealth for People Like Us?*' [19]. In this paper, a dichotomy between People Like Us (PLUs), the designers of eHealth, and the Disempowered, Disengaged and Disconnected (DDD), the citizens most in need of the services offered by the digital healthcare services, demonstrates how potential inequities in eHealth emerge and persist due to inherent biases in the design phase favouring the preferences of the PLUs [19]. A problem confirmed by the results in this study considering the similarity between the characteristics of the typical HIT-user and the profile of the majority of the population. Therefore, to ensure that digital solutions are useable and accessible to the DDDs, the eHealth systems should be designed according to their needs [19].

4.5 Scandinavian similarities

When comparing the results to studies conducted in Sweden and Norway similar patterns occur. In the paper '*Mobile Access and Adoption of the Swedish National Patient Portal*' [20], Hågglund et al. (2020) show that the

use of Mobile phones when accessing the national health portal has increased from 38 to 77% and that the monthly number of visits has increased by 400% over 7 years, 2013-2020 [20]. In a Norwegian context, Zanaboni et al. (2022), finds that the majority (50,7%) of the respondents consider their self-reported health as good/very good and that 58,6% visit their electronic health record regularly or when needed [21]. Thus, numbers that might indicate similarities in HIT use and perceptions of health status across the Scandinavian countries.

5 CONCLUSION

This study reveals that the use of HIT and digital solutions is increasing in a Danish context. Accordingly, the use of health apps, smartphones and the national health portal (Sundhed.dk) has increased over the examined period. Moreover, the study discloses that the typical user of HIT and the eHealth system has a bachelor's or higher degree of education, a good/very good health status and often suffers from one or more chronic conditions. Based on the study findings we conclude that citizens' increased use of the digital healthcare system and the growing production of patient data are consistent with current Danish health policies and digital strategies; underscoring, the importance of patient participation, self-management and PROs. However, the minority of citizens, the 10-20% of the population with short education and a bad/very bad health status might not be able or wish to participate and self-manage their health conditions. Therefore, we encourage awareness when designing the future digital healthcare system to ensure that solutions are based on the preferences and needs of the DDDs.

Abbreviations

DDD	Disempowered, Disengaged, Disconnected
HCP	Healthcare professional
HIT	Health Information Technology
PLO	The Danish Organization of General Practitioners
PLU	People Like US
PRO	Patient-Reported Outcome

6 REFERENCES

- [1] Regeringen. *Danmarks digitaliseringsstrategi*. København K, 2022.
- [2] Højgaard, B., Kjellberg, J. *Fem megatrends der udfordrer fremtidens sundhedsvæsen*. 2017.
- [3] Regeringen. *Sundhedsreformen - Gør Danmark sundere* København K, 2022.
- [4] Danske Regioner, *Sundhed for dig – Regionernes samarbejde om digitalisering*. København Ø, 2022.
- [5] Eriksen, J. *PURPOSE, FUNCTIONALITY AND RECONCEPTUALISATION OF PATIENT-REPORTED OUTCOME A PATIENT PARTICIPATION PERSPECTIVE ON PRO IN CLINICAL PRACTICE POST ITS DIGITALISATION*. Aalborg University, 2021.
- [6] Tornbjerg, K., Bertelsen, P. *Undersøgelsen af borgernes anvendelse af sundhed-it 2013, - en udforskning af danskernes kendskab, holdninger og forhold til it, til gavn for eget helbred*. Aalborg universitet, 2014.
- [7] Petersen, L.S., et al. *Undersøgelsen af borgernes anvendelse af sundhed-it 2015, - en udforskning af danskernes kendskab, holdninger og forhold til it, til gavn for eget helbred*. Aalborg universitet, 2016.
- [8] Bertelsen, P., Tornbjerg, K. "Danish Citizens' Expectations to the Use of eHealth" in *Driving Quality in Informatics: Fulfilling the Promise*, vol. 208, pp. 78–82. Aalborg University 2015.
- [9] Bertelsen, P., et al. *Undersøgelse af borgernes anvendelse af sundhed-it i 2019 - en udforskning af danskernes kendskab, holdninger og forhold til it, til gavn for eget helbred*. Aalborg Universitet, 2020.
- [10] MEGAFON, "MEGAFON," 2022. [Online]. Available: <https://megafon.dk/>. [Accessed: 15-May-2022].
- [11] Bryman, A. *Social research methods*, 4th ed. New York: Oxford University Press, 2012.
- [12] Sundhedsdatastyrelsen. *Vejledning om informations- sikkerhed i sundhedsvæsenet*. København S, 2016.
- [13] PLO, *PLO ANALYSE - Almen praksis er blevet mere digital*. København Ø, 2019.
- [14] Tassy, A., Berg, C. E. *It-anvendelse i befolkningen*. København Ø, 2022.
- [15] Robbins, T., et al., "COVID-19: A new digital dawn?" in *Digit. Heal.*, vol. 6, pp. 1–3, 2020.
- [16] Sørensen, M.L., et al. *Det frivillige sociale engagement i Danmark - Frivilligrapport 2019-2021*. Odense, 2021.
- [17] Sodemann, M. "Sundhedsvæsenet drukner i middelklassens små problemer" in *Mandag Morgen*, 2019.
- [18] Consortium "Digital Europe," *Monitoring Digital Inclusion in the Nordic-Baltic region*. Copenhagen, 2022.
- [19] Showell, C., Turner, P. "The PLU problem: Are we designing personal Ehealth for people like us?" in *Stud. Health Technol. Inform.*, vol. 183, pp. 276–280, 2013.
- [20] Hägglund, M., et al. "Mobile Access and Adoption of the Swedish National Patient Portal" in *Integrated Citizen Centered Digital Health and Social Care*, vol. 275, *Studies in Health Technology and Informatics*, IOS Press, pp. 82-86, 2020.
- [21] Zanaboni, P., et al. "Patient Use and Experience With Online Access to Electronic Health Records in Norway: Results From an Online Survey" in *J Med Internet Res*, vol. 22(2), 2022.

Normativity assumptions in the design and application of social robots for autistic children

Henrik Skaug Sætra¹, Anders Nordahl-Hansen², Eduard Fosch-Villaronga³ and Christine Dahl⁴

¹ Østfold University College, Faculty of Computer Science, Engineering and Economics, Norway henrik.satra@hiof.no

² Østfold University College, Department of Education, ICT and Learning, Norway

³ Leiden University, eLaw Center for Law and Digital Technologies, The Netherlands

⁴ Høgskolen Kristiania, Faculty of Health Science, Norway

Abstract

Social robots interact with human beings and are used for a variety of therapeutic purposes, for example in interaction with children with neurodevelopmental disorders. A key ethical issue related to the application of social robots in these contexts is the idea of normativity, involved in both the design of social robots, i.e., the use of such robots to portray or mimic what is normal and to identify deviant behaviour or development. The article presents the beginnings of a framework for incorporating divergent opinions of normal social functioning, particularly neurodiversity, into the design and application of social robots.

Keywords

Social robots, neurodiversity, normativity, ethics, therapy, diagnostics, testing, Autism Spectrum Disorder, autistic

1 INTRODUCTION

Social robots are designed to interact with human beings and are increasingly used in social contexts. Such robots have different embodiments and can resemble, for example, toys, pets, or humans. These robots are useful for therapeutic purposes, for example in interventions with children with neurodevelopmental disorders, as the robot's behaviour is both predictive and repetitive, and less complex or intimidating than humans', which positively impacts the development of specific therapies and interventions, including those geared towards autistic children.

An important question is whether autistic children are open and willing to interact with social robots. As a group, autistic children are heterogeneous, meaning that just as in other populations the interest to interact with technology such as robots will vary from person to person. Therefore, usage of technology-based interventions for each individual must be monitored and tailored to fit the child's needs and preferences. Nevertheless, there are studies that report that autistic persons report higher use of technology and choose different computer-based learning forms more often than other groups [1]. Further, acceptability of, for example, usage of technological tools such as virtual reality head mounted displays seems to be broadly accepted by autistic persons [2], and acceptance rates are reported to be high among autistic children in interventions using social robots [3].

A critical ethical issue related to the application of social robots in these contexts is the idea of normativity [4, 5], which is relevant both in relation to the design of social robots and in the use of such robots to portray or mimic that which is *normal* and to identify deviant behaviour or development. Normativity in this article relates to the idea that a particular form of neurological functioning and development is normal. What is normal is again often associated with what is good or societally acceptable. We mainly refer to *neuronormativity*, which connects our endeavour to the terms neurotypicality and neurodiversity. Neurodiversity as a concept is usually associated with an objection to the deficit perspective in Autism spectrum disorder (autism from hereon). From the neurodiversity perspective it is often argued that autism is a result of natural genetic variation, and not a medical disorder or condition to be 'fixed' [6]. However, the notion of normativity also apply to a broader context, as seen for example by the notion of *heteronormativity* in feminist and queer scholarship [7].

Our main concern is not to determine what *is* normal, but to highlight how the very act of portraying something as normal involves an exercise of power. We argue that normativity permeates the design of social robots and their use in the context of autism. While we do not argue that normativity in itself is necessarily bad, policymakers, designers, researchers, and therapists, and others who work directly with autistic children should be aware of the potentially problematic issues involved in portraying some things as normal and others as deviant.

The 18th Scandinavian Conference on Health informatics, Tromsø, Norway, August 22-24, 2022. Organized by UiT The Arctic University of Norway. Conference Proceedings published by Linköping University Electronic Press at <https://doi.org/10.3384/ecp187>. © The Author(s). This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>

Normativity assumptions are necessarily involved in the design and application of social robots, as they manifest, demonstrate, and teach behaviour and characteristics that are perceived as both normal and desirable, either implicitly or explicitly. What constitutes ‘normal’, and to what degree interventions aim to ‘normalise’ autistic children to comply to neuronormativity, is a key focal point of heated debates within the autism intervention literature and in the autism research field as a whole [8].

The purpose of this article is to explain how and when normativity related issues become relevant for the design and application of social robots for autistic children. By highlighting these issues, we aim to make developers and designers, therapists, and policymakers better equipped to answer questions regarding when robot normativity is warranted and when it should be challenged and mitigated. The authors conducted a mapping of the potential issues related to normativity in the context of robot and autistic children based on their combined experience and expertise and have sought to highlight these challenges through a review of the literature. As one potential solution we propose a *disclosive robot ethics* approach, in line with Brey’s [9] notion of disclosive computer ethics. This article presents the major considerations involved in making normativity assumptions explicit, and this will help the various actors involved in the design and application of social robots to consider divergent opinions of normal social functioning, and in particular, neurodiversity. The framework proposed in the article involves a particular form of inclusive and value sensitive design (VSD) process [10]. However, our approach is not only about design justice [11], but relates equally to the various applications of social robots.

2 NORMATIVITY AND THE APPLICATION OF SOCIAL ROBOTS

Social robots are designed and deployed with the purpose of both mimicking human behaviour and to engender particular forms of interaction with humans [12]. However, the lack of diversity in the groups of people designing robots is garnering increasing attention [13]. This is potentially problematic, as an homogenous group of designers might be likely to design a product from a shared background and shared biases, increasing the risks that the end product might not be as useful for all, and potentially also outright harmful for some, users [13]. One example is how robots rely on gender stereotyping, and how this generates ethical issues in the application of social robots in, for example, eldercare [14]. The broader issue we focus on in this article is how robots enact and embody a problematic normativity which might cause harms to marginalized users.

In the context of autism, the express purpose of the use of social robots is often to demonstrate, teach or encourage what is perceived to be *normal behaviour*. For example, many interventions include training in eye-contact for autistic children, as this is seen as key to productive social communication. However, many autistic people experience eye-contact as highly negative both emotionally and physiologically [15]. Other examples may relate to what is discussed within the autism literature as ‘masking’, which means autistic individuals develop an ability to present as

non-autistic. This suppression of their own behaviours and/or traits is associated with anxiety, depression, and suicidality [16]. This type of camouflaging can also be found in childhood [17]. Here it is also worth noting that autism is a very heterogeneous condition and that what might be viewed as natural behaviours, traits, and thought patterns for one autistic individual can be very different for another [18].

We use the concept of normativity to represent how particular behaviour’s, actions or traits are considered *good*, and sometimes also right and just [4, 5]. Normativity relates to moral judgements, and whenever ethical standards are promoted, they represent claims on us, as ‘they command, oblige, recommend, or guide’ [19]. Regardless of the creator’s intentions, we argue that robots inevitably perform some form of normalcy and support certain ethical standards when they are used in social settings. They can become role models of sorts. Robot portrayal of purported normal behaviours might be problematic in two respects. Firstly, robots with behaviour modelled on normal behaviour will make those with divergent behaviour stand out from their peers. Secondly, these robots run the risk of making those same children aware that they are different.

Next, social robots are also used to teach and foster normal behaviour. This is distinct from the previous case, as this involves someone consciously using social robots to train or untrain particular behaviours that are perceived as desirable or undesirable. By doing so, social robots become the instruments of normalcy more directly, and robot makers and those employing them should be aware of the implications of such use of social robots. On the command of those behind the intervention, robots may demonstrate the difference between right and wrong in these interventions. This intervention might involve explaining what sort of behaviour is desirable, but it could also explain *why* this behaviour is desirable. The last sort of intervention’s effectiveness depends on the child’s age and general cognitive functioning level. It is also important to note that such interventions *could* be carried out by arguing that certain behaviours are good and correct. However, they could also be based on the approach that learning a different set of behaviours is merely useful for the child, as the child needs to learn how others perceive different situations and behaviours.

Lastly, social robots may be used to detect and monitor normalcy [20, 21]. This relates to robots that rely on gender stereotyping either in their design or in how they operate [14]. From a clinical point of view, considering the complexity and heterogeneity that is a hallmark of the autism diagnosis, it is difficult to see social robots completely substituting as opposed to supporting professionals in any near future. However, robotics has become sufficiently sophisticated to possibly tempt actors to develop software and functionalities allowing social robots to become agents of an enforcement and surveillance scheme built to identify and eliminate abnormal behaviour. Social robots have a wide range of sensors, including eye-tracking, facial recognition, and voice recognition. These capabilities can easily be coupled with software that aims to identify, for example, the emotions and the behaviour patterns of those it observes [22], and as pattern recognition

is at the core of AI, detecting normal behaviour and potential outliers is a natural part of what the software of a social robot does [23].

Imagine a scenario in which a government has mandated that all kindergartens should have a social robot equipped with new and impressive software aimed at screening and pseudo-diagnosing all children present. The robot is deployed in the kindergarten, and as children play with it – or not – it registers and monitors their behaviour and interactions. At the end of each week, the managers get a report where children with suspect behaviour are flagged and sent to a professional for proper testing.

Such use of social robots is even more problematic because it is a proactive tool for enforcing normalcy. However, it is also easy to see the pleas of the advocates of such a scheme, as there is great potential to identify more children that would potentially benefit from being diagnosed with autism to receive an intervention. Even if such benefits are both plausible and vital, normativity is deeply involved in social robots used in such a manner, and its designers and regulators will have to decide how they deal with such issues. This is particularly important because then it means that robot developers are becoming the determinant parties of what is normal and what is not, something that goes beyond democracy [24].

3 CONSIDERATIONS FOR INCORPORATING A DIVERSITY PERSPECTIVE IN THE DESIGN AND APPLICATION OF SOCIAL ROBOTS

Autism is increasingly seen as a social construct, and normativity and power theories are therefore useful for understanding the implications of social robot normativity. Approaches to issues of robot normativity and gender stereotyping are relevant to the issue at hand, and some solutions mentioned in the literature is *explanation*, *neutralization*, and *'queering'* [14]. Explanation refers to providing an explanation of why the robot is designed as it is and could be coupled with providing reasons aimed at dispelling the promotion or superiority of 'normal' features or capabilities. This approach is indeed important, and it is one to which we return below. Neutralization entails attempts at making a robot non-gendered in order to avoid stereotypes. A related approach related to autism could entail having the robot act in ways untypical of both autistic and other children, but this would often defy the purpose of using the robots in such settings, as the purpose will often be to demonstrate and encourage a particular set of behaviours. Finally, the analogy to the queering of robots would be to have the robots act in ways more similar to the autistic children. While this would indeed potentially solve some problems related to normativity, sceptics could plausibly argue that such robots might not be very effective at what they are currently being deployed to achieve. The problems are, we argue, too complex to be solved by adhering to just one of these approaches.

By adopting a post-structuralist stance, we might argue that the structures most commonly used to interpret both normalcy and desirability are themselves social structures open for debate [25]. This invites a fundamental humility with regards to values and the enforcement of norms, and we here argue that this humility is beneficial for anyone

interested in developing and using robots in a value sensitive manner.

The neurodiversity movement is relevant for examining the design and application of social robots designed for autistic children (or for robots designed for any purpose, and that can be used in interventions with autistic children). Lewin and Akhtar [26] write about neurodiversity and how stereotyped representations of autistic people, and what they call *the deficit perspective*, is potentially problematic. Pesonen, et al. [27] also analyse the framing of autism in media and argue that how autism is portrayed shapes the public's perception of the phenomenon and will, in turn, influence both autistic people and their subsequent societal acceptance. The core of the neurodiversity movement is that the neurological characteristic of autism represents a natural variation and that there is no error, fault, or something in need of fixing [26].

With such a perspective, autistic people are not deficient but slightly different from others. While these *others* may be a majority, this does not mean that the minority is an abnormality that must be 'fixed' and turned into something resembling or pleasing the majority population. This is often referred to as ableism, and entails the uncritical assertion that doing what is normal is good, and entails a depreciation of various disabilities [28]. Hehir [28] argues that 'ableist assumptions in the education of children with disabilities not only reinforces prevailing prejudices against disability', and that this also leads to a range of problems related both to educational attainment and employment. Instead, we could argue that the majority needs to learn how to interact with and understand autistic people – quite the opposite approach of trying to teach autistic people to learn and mimic the behaviour perceived as more normal. The design and use of social robots entails making decisions with consequences for the 'distribution of benefits and burdens between various groups of people' [11], and we argue that traditional distributions of the burdens could and should be continually evaluated.

Realizing such a shift may be difficult without health policies based on a thorough understanding of the societal ramifications of user-centred approaches. We understand that many children on the spectrum have difficulties to such a degree that labelling them just another kind of normal makes little sense. Many autistic children need both intervention and substantial assistance in both kindergarten, school, and life in general. In this case, an inclusive and value-sensitive approach that takes neurodiversity into account in the design and use of social robots seem to be an appropriate avenue for action. If social robots continue to be used in an interventional setting, efforts should be made to analyse and account for the normativity implications and effects these have on autistic people.

The approach we label *disclosive design* could be beneficial for ensuring that a diversity perspective is both considered in the design process, and that it is possible to evaluate products on the basis of normativity implications. This would also enable designers to provide specific explanations of why the robots are designed and operate the way they do, and such explanations are potentially of great importance for enabling therapists, parents, and some children with autism to consider and understand the why

the robots are and operate the way they do. The term disclosive design is based on Brey's [9] notion of a disclosive computer *ethics*, and disclosive *design* would entail a) systematically considering and b) disclosing in written documentation the finding from the designer's analysis and their intentions with regards to the product's embedded values and norms. This is also compatible with design justice [11], which emphasises community-involvement and participation in design, with a particular emphasis on awareness of the *matrix of domination* and the 'the equitable distribution of design's benefits and burdens'.

An initiative not entirely unlike the one here proposed is Value Sensitive Design (VSD), which promotes reflection on the consequences of technological research and development (R&D) outcomes and incorporating them into the research process [10]. While important, this article suggests that taking this a step further and actually disclosing the processes related to values explicitly *and* implicitly embedded in products are important.

In this context, the European Union pushed for the notion of responsible research and innovation (RRI) to establish an overarching process capturing essential steps researchers must take to ensure that science, research, and innovation have positive, socially acceptable, and desirable outcomes [29]. RRI is a science policy framework aimed at inclusive and sustainable technology, through which the European Research Council sought to align technological innovation with societal values, and to provide ways to deal with the uncertainty that various actors encounter in innovation processes [30].

We argue that adopting the principle of disclosive design might both improve the products and their use *and* the lives of those with perspectives that are often omitted from both research and design on social robots and the use of such robots with autistic children. The obvious consequences of such an approach might be that the robots are made available in a range of colours and that they won't necessarily display the *normal* features of human beings (with two arms, legs, etc.). Of more interest, however, are the prospects of building machines with more varied behaviors that take into account the heterogeneity of autistic persons. This might be one facet of a regular robot, and such functionality could, for example, be used to teach other children how to effectively interact with autistic children. Turning the table, so to speak, or at least working on mutual understanding instead of just trying to 'fix' the autistic child. Robots could also be designed to be more responsive to the behaviour of autistic children, and thus make their interactions with the robots more effective while not highlighting the fact they interact socially in a different manner from most.

4 CONCLUSION

In this discussion, we have focused on how social robots are not exempt from portraying and reinforcing social constructs, as they manifest a certain normativity. We have analysed them in relation to their use in the context of intervention with autistic children. Social robot-assisted interventions may indeed be effective in some instances, but it is nevertheless necessary to take a step back and

analyse the conceptions of normalcy implemented in a) the design and b) the application of social robots.

This discussion has suggested a diversity perspective in designing and deploying social robots. The realisation of this requires legislators, designers, and those who employ social robots to apply a precautionary principle with respect to diversity and the inclusivity of their creations and use of social robots. The processes described entails that autistic people or their representatives should be included in discussions regarding how social behaviour is modelled and represented in social robots, how social robots are used in interventions, and in particular if social robots should be equipped with the ability to identify and potentially perform parts of the diagnosis of autism. This aligns with the call of Bertilsdotter Rosqvist, et al. [6] to include the voice of autistic people in research endeavours in general. However, others have pointed out user centred and participatory design is no silver bullet, and that such processes might also lead to stereotyped and problematic representations of the target groups [31].

Regardless of the choice of design process, then, we argue that it should be accompanied by the approach of disclosive design, which entails that the creators and users of social robots make explicit and disclose the processes carried out in advance of production and deployment and their justifications for their choices.

5 REFERENCES

- [1] M. O. Mazurek and C. Wenstrup, "Television, video game and social media use among children with ASD and typically developing siblings," *Journal of autism and developmental disorders*, vol. 43, no. 6, pp. 1258-1271, 2013.
- [2] A. Dechsling, S. Sutterlin, and A. Nordahl-Hansen, "Acceptability and normative considerations in research on autism spectrum disorders and virtual reality," in *International Conference on Human-Computer Interaction*: Springer, 2020, pp. 161-170.
- [3] I. van den Berk-Smeekens *et al.*, "Adherence and acceptability of a robot-assisted Pivotal Response Treatment protocol for children with autism spectrum disorder," *Scientific reports*, vol. 10, no. 1, pp. 1-11, 2020.
- [4] J. Rouse, "Social practices and normativity," *Philosophy of the social sciences*, vol. 37, no. 1, pp. 46-56, 2007, doi: <https://doi.org/10.1177/0048393106296542>.
- [5] J. J. Thomson, *Normativity*. Chicago: Open Court, 2015.
- [6] H. Bertilsdotter Rosqvist *et al.*, "Doing it differently: emancipatory autism studies within a neurodiverse academic space," *Disability & Society*, vol. 34, no. 7-8, pp. 1082-1101, 2019.
- [7] S. Jackson, "Interchanges: Gender, sexuality and heterosexuality: The complexity (and limits) of heteronormativity," *Feminist theory*, vol. 7, no. 1, pp. 105-121, 2006.
- [8] R. Davis, J. den Houting, A. Nordahl-Hansen, and S. Fletcher-Watson, "Helping autistic children," *The Wiley-Blackwell Handbook of Childhood Social Development*, pp. 729-746, 2022.

- [9] P. Brey, "Disclosive computer ethics," *ACM Sigcas Computers and Society*, vol. 30, no. 4, pp. 10-16, 2000.
- [10] B. Friedman, "Value-sensitive design," *interactions*, vol. 3, no. 6, pp. 16-23, 1996.
- [11] S. Costanza-Chock, *Design justice: Community-led practices to build the worlds we need*. Cambridge: The MIT Press, 2020.
- [12] H. S. Sætra, "The Parasitic Nature of Social AI: Sharing Minds with the Mindless," *Integrative Psychological and Behavioral Science*, vol. 54, pp. 308–326, 2020, doi: <https://doi.org/10.1007/s12124-020-09523-6>.
- [13] K. Pereida and M. Greef. "Diversity in robotics: From diverse teams to diverse impact." <https://www.dynsyslab.org/wp-content/papercite-data/pdf/pereida-icra19b.pdf> (accessed June 29, 2022).
- [14] M. Weßel, N. Ellerich-Groppe, and M. Schweda, "Gender stereotyping of robotic systems in eldercare: An exploratory analysis of ethical problems and possible solutions," *International Journal of Social Robotics*, pp. 1-14, 2021.
- [15] D. A. Trevisan, N. Roberts, C. Lin, and E. Birmingham, "How do adults and teens with self-declared Autism Spectrum Disorder experience eye contact? A qualitative analysis of first-hand accounts," *PloS one*, vol. 12, no. 11, p. e0188446, 2017.
- [16] S. A. Cassidy, A. Robertson, E. Townsend, R. C. O'Connor, and J. Rodgers, "Advancing our understanding of self-harm, suicidal thoughts and behaviours in autism," vol. 50, ed: Springer, 2020, pp. 3445-3449.
- [17] H. Wood-Downie, B. Wong, H. Kovshoff, W. Mandy, L. Hull, and J. A. Hadwin, "Sex/gender differences in camouflaging in children and adolescents with autism," *Journal of Autism and Developmental Disorders*, vol. 51, no. 4, pp. 1353-1364, 2021.
- [18] C. Lord *et al.*, "The Lancet Commission on the future of care and clinical research in autism," *The Lancet*, vol. 399, no. 10321, pp. 271-334, 2022.
- [19] C. M. Korsgaard *et al.*, *The sources of normativity*. Cambridge: Cambridge University Press, 1996.
- [20] K. Chawarska and F. Shic, "Looking but not seeing: Atypical visual scanning and recognition of faces in 2 and 4-year-old children with autism spectrum disorder," *Journal of autism and developmental disorders*, vol. 39, no. 12, pp. 1663-1672, 2009, doi: <https://doi.org/10.1007/s10803-009-0803-7>.
- [21] T. Falck-Ytter, S. Bölte, and G. Gredebäck, "Eye tracking in early autism research," *Journal of neurodevelopmental disorders*, vol. 5, no. 1, pp. 1-13, 2013, doi: <https://doi.org/10.1186/1866-1955-5-28>.
- [22] B. Scassellati *et al.*, "Improving social skills in children with ASD using a long-term, in-home social robot," *Science Robotics*, vol. 3, no. 21, p. eaat7544, 2018.
- [23] E. F. Villaronga, "'I Love You,' Said the Robot: Boundaries of the Use of Emotions in Human-Robot Interactions," in *Emotional Design in Human-Robot Interaction*: Springer, 2019, pp. 93-110.
- [24] S. Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: PublicAffairs, 2019.
- [25] J. Williams, *Understanding poststructuralism*. Chesham: Acumen, 2005.
- [26] N. Lewin and N. Akhtar, "Neurodiversity and deficit perspectives in The Washington Post's coverage of autism," *Disability & Society*, vol. 36, no. 5, pp. 812-833, 2021.
- [27] H. Pesonen, T. Itkonen, M. Saha, and A. Nordahl-Hansen, "Framing autism in newspaper media: an example from Finland," *Advances in autism*, 2020.
- [28] T. Hehir, "Eliminating ableism in education," *Harvard educational review*, vol. 72, no. 1, pp. 1-33, 2002.
- [29] B. C. Stahl, N. McBride, K. Wakunuma, and C. Flick, "The empathic care robot: A prototype of responsible research and innovation," *Technological Forecasting and Social Change*, vol. 84, pp. 74-85, 2014.
- [30] RRI-Practice. "What is RRI?" <https://www.rri-practice.eu/about-rri-practice/what-is-rri/> (accessed April 4, 2022).
- [31] S. Serholt, S. Ljungblad, and N. Ní Bhroin, "Introduction: special issue—critical robotics research," ed: Springer, 2021, pp. 1-7.

Development of medical applications based on AI models and register data – regulatory considerations

Jaakko Lähteenmäki, Juha Pajula and Emmi Antikainen
VTT Technical Research Centre of Finland Ltd., Finland
jaakko.lahteenmaki@vtt.fi

Abstract

Artificial intelligence based methods, especially machine learning (ML), are increasingly used in healthcare for automatic medical image analysis and clinical decision support systems. Development and validation of ML models involve processing of large volumes of personal data. We analysed regulatory impacts on ML based application development especially from the perspective of privacy protection and usage of ML models as a basis for software under medical device regulation (MDR). We present best practices for ML application development and personal data usage in a use case of predicting elderly individuals' future need for healthcare and social welfare services.

Keywords

Artificial intelligence, machine learning, medical device regulation, MDR, privacy protection

1 INTRODUCTION

There is a growing interest towards the use artificial intelligence (AI) to improve healthcare [1]. Machine learning (ML) is extensively used for automatic medical image analysis for supporting and improving human interpretation. It is increasingly also used to support precision medicine by predicting patient outcomes, identifying patients with elevated risk and suggesting most favourable care pathways and services for the patients. Machine learning models empower decision support applications providing guidance to healthcare professionals and patients [2, 3]. Such applications potentially affect the healthcare of an individual patient, and are considered to be Medical Device Software (MDSW) falling under the Medical Device Regulation (MDR) [4].

Regulatory compliance is based on rigorous risk management and release acceptance processes. This is a challenge for ML based applications, which shall be validated against personal health data, and typically would need to be frequently updated as new data becomes available. Also, the use of agile and continuous development approaches causes the need for frequent software updates challenging the conventional “waterfall-type” development process [5]. The challenges caused by frequent software changes in the context of medical devices have been addressed in several earlier studies and reports [6–8].

Only a few earlier papers address the challenges related to the need to use sensitive healthcare register data in the development of medical device software [8]. In the present paper, we will analyse the regulatory impacts on ML based application development from the perspective of sensitive

personal data usage. We will address the relevant development phases starting from research and modelling activities and extending to medical software development and deployment. Detailed analysis of the development phases is beyond the scope of the paper. Our purpose is to provide an overview of the topic highlighting observations that we have made during the planning and data collection phase of the MAITE project, which aims at data-driven prediction of the need for health and social services.

2 RELEVANT REGULATION

2.1 Personal data protection

Access to individual level health data is a precondition for ML model development in the health domain. In most cases the data need to be acquired from one or more health data registers, such as electronic health record systems (EHRs). Real-world data (RWD) accumulated in EHRs can be made available by the respective data controller for so called *secondary use* referring to the usage of data for another purpose than the purpose for which the data was originally collected [9]. Secondary use of data may take place without the consent of the data subject based on *public interest* in the area of public health or scientific research as defined by the General Data Protection Regulation (GDPR), article 9(2) i, j. Some countries, e.g. Finland, have national legislation regulating secondary use of health data, and a related European-level legislative action is on-going [10]. Another data access option is the usage of data specifically collected for research, for example, based on the biobank consent given by the data subject [11].

Data sets available for scientific research - either based on public interest or consent - are typically pseudonymized.

The 18th Scandinavian Conference on Health informatics, Tromsø, Norway, August 22-24, 2022. Organized by UiT The Arctic University of Norway. Conference Proceedings published by Linköping University Electronic Press at <https://doi.org/10.3384/ecp187>. © The Author(s). This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>

Even though this encompasses removal of direct person identifiers of the data released for research use, the data could still be reidentified and, therefore, fall under the GDPR. Privacy of the data is in most cases protected by limiting its use to a secure processing environment (SPE) separated from the environment where the software is developed [12].

2.2 Medical device software

Legal and regulatory requirements for medical devices (including MDSW) in Europe are set out in the Medical Device Regulation 745/2017 (MDR) [4]. MDR classifies medical devices into risk groups (1, 2a, 2b or 3) with respective conformity assessment procedures. In practice, development under MDR requires the manufacturer to have a certified quality system (e.g. complying with the ISO 13485 standard) covering management processes, product requirements management, product realization, customer feedback and support. Food and Drug Administration (FDA) is responsible for the corresponding regulation in the United States, where the term Software as a Medical Device (SaMD) is used instead of MDSW [13].

ML based applications are problematic from the regulatory perspective as they may need to be updated when new data comes available [8, 14, 15]. FDA in the United States has published an action plan with concrete proposals to enable software changes to be implemented in a controlled way without a new regulatory approval. The FDA approach is based on a predetermined *change control plan* and *algorithm change protocol* which the manufacturer needs to specify upon product approval [16]. EU has chosen to provide related guidance through Artificial Intelligence Act (AIA) draft proposal, but is less explicit in defining the procedures to be adopted.

2.3 AI regulation

Specific EU-level regulation, the Artificial Intelligence Act (AIA) is currently under development [17] and will affect the development and use of AI based applications. AIA defines all AI systems under the Union harmonization legislation (including MDR) to be included in the high-risk category.

The AIA regulation complements the MDR in addressing aspects related to the quality of training, validation and testing data sets. The regulation specifically requires the manufacturer to record detailed documentation on the AI system development, including data cleaning and model training methodologies as well as usage of third-party tools. AIA also addresses several ethics-related issues, such as interpretability of results produced by an AI system and the need for human oversight in the service provision context.

2.4 Healthcare information systems

Medical device regulation, referred above, is focused to ensure the safety and performance of the product. ML based medical software typically needs to be integrated in the information system environment of a healthcare service

[18, 19]. Such integration may be subject to additional regulation besides the MDR [20]. Healthcare infrastructures are to a large extent country-specific and regulated by national laws and requirements [21]. Additionally, GDPR and AIA set considerable limits to the use of AI in automatic decision making and profiling.

3 APPLICATION DEVELOPMENT PHASES

ML application development phases and their main linkage to regulation are indicated in Table 1. The application development lifecycle is divided into three major phases: research, software development, and application deployment.

3.1 Research

The research phase starts by defining the study approach and objectives in co-operation with the relevant healthcare professionals and other domain experts. The research plan describes the target population (inclusion criteria), research methodology, and data contents to be used. The development of an ML model typically falls to the category of exploratory research, where study endpoints are not known at the time of data permit application and are not explicitly defined in the study protocol. The research plan is an important part of the data permit application and shall comply with standard scientific research practices to be aligned with the GDPR requirements concerning access to data for secondary purposes.

After positive data permit decision, the data resources are made available for the data user. Data usage is subject to several restrictions due to the sensitivity of the data. Most typically, the data is made available for research in a secure processing environment (SPE), which provides the tools, storage capacity and computing resources needed for data processing, but does not allow exporting the data out of the environment. The installation of additional data analytics tools may also be subject to approval by the permit authority, and the availability of high performance computing (HPC) resources may be limited.

These limitations may complicate ML model development. On the other hand, the SPE approach can also be seen as an enabler for using data resources, which would not be accessible otherwise. Furthermore, an external SPE may be an attractive alternative for a research organization, which can avoid to invest in its own computing resources. SPE's are still emerging and expected to be improved in terms of services and computing performance.

3.2 Software development

Depending on the results of the research phase, the software development activity in line with MDR requirements may be started after completion or during the research phase. The ML model developed in the research phase should include only anonymous information (e.g. tuned model coefficients), which enables the model to be exported from the SPE and used in the medical device software development process.

Regulation	Research	Software development	Application deployment
GDPR	Privacy and ethics of data usage for research and development	Privacy and ethics of the product	Privacy and ethics of the operational service
AIA	Privacy and ethics of data usage for research and development	Privacy, ethics, resilience and performance of the product	Privacy, ethics, resilience and performance of the operational service
MDR	ML-model documentation to enable traceability	Product safety and clinical performance	Enabling post-market surveillance
National regulation	Regulation on secondary use of data	Privacy, security, interoperability, functionality (digital health applications)	Privacy, security, interoperability, functionality

Table 1. Regulatory objectives in the development phases of ML based health applications.

The basic requirement for medical device software is that the released product is safe and provides the declared clinical benefits. When significant changes are introduced for class 2a devices or higher, they need to be approved by a Notified Body [4]. Applicable standards (in particular IEC 62304:2006 medical device software – software life cycle processes) expect the development cycle to be divided into phases such as product planning, product design, design transfer, product realization and release.

Each phase ends in a design review, where final versions of the created documents and other artefacts are reviewed and approved. As indicated above in Section 2.2., new approaches for enabling agile updating of AI based applications are being introduced both in the USA and in Europe.

3.3 Application deployment

ML based applications are typically not stand-alone applications, but need to be integrated in health and social services information system environment. For example, a decision support application needs to be integrated with an EHR system to get access to patient records. Such deployment may be subject to additional national regulation besides the MDR [21]. The purpose of such national requirements is to ensure correct exchange of information between software components, appropriate personal data protection and resilience towards cyber-attacks. Certification demonstrating compliance with national requirements may be required. Also, joint testing with other software providers may be necessary to demonstrate interoperability [20].

Efficient clinical use of ML based applications requires, besides technical interoperability, also seamless integration with the clinical process. Although process-level integration would not be directly covered by regulation, it is a prerequisite for positive impact and clinical benefits of ML based applications. Therefore, it is important that ML based applications are reliable, compatible with current care guidelines and practices, show direct benefit for healthcare professionals and customers [22]. Also, final responsibility of treatment choices should always rest with

the healthcare professional, and ML based applications should only be use as assistive tools [23].

4 BEST PRACTICES - CASE “MAITE”

In the following, we will analyse typical challenges in developing and deploying ML based applications and deploying them in health and social services. As an example, we will use the MAITE project (Data-driven identification of elderly individuals with future need for multi-sectoral services), where VTT Technical Research Centre of Finland Ltd. (VTT) is responsible for ML model development. We will identify the best practices to be adopted in the MAITE project to overcome the challenges of ML application development.

4.1 Case overview

The MAITE project emerges from the observation that health and social services expenditure is concentrated to a small fraction of the population [24]. It is expected that future heavy users of services could be identified based on their current health and social status and service usage history. The objective of the project is to develop an ML based model and proof-of-concept (PoC) application for predicting future service usage of elderly individuals. The model would support personalized and group-level preventive interventions to avoid excessive service need in the future.

VTT is responsible for the ML model development based on register data of Päijät-Häme Joint Authority for Health and Wellbeing (PHHYKY), a public health and social services provider with catchment area of 200 000 inhabitants in Southern Finland. The Finnish institute of health and welfare (THL) is responsible for coordinating the co-operation of stakeholders and ensuring continuous interaction between developers and end-users. The project is currently in the research phase with the data permit application recently accepted. The data permit covers the data resources listed in Table 2.

Data resource	Data description
Health and social services encounter data (Effica and Lifecare systems)	Basic demographic information: age, sex, municipality of residence. Information about health and social services visits: primary and specialized healthcare outpatient and inpatient visits, mental health services, substance abuse services, social services for elderly, dental care services. For each visit: (1) visited service, (2) date, (3) main diagnose and operation related to the visit, (4) medication data (ATC groups: M01, N05, N06), (5) laboratory results (HbA1c, glucose, hemoglobin, ferritin, vitamin D, PEth)
Social services decisions (Effica and Lifecare systems)	Information about social services decisions: rehabilitation, elderly housing service, caregiver support, homecare, transportation services For each decision: (1) service decision category, (2) decision outcome (positive/negative), (3) date
Service need and physical function assessment (Raisoft)	Information about service need and physical function assessments.

Table 2. Data resources covered by the data permit of the MAITE project.

4.2 Protecting privacy

ML based models typically require large data sets in terms of data subjects and variables. Extracting such large cohorts to be processed in external computing environments may involve high privacy risks. Consequently, register controllers have been reluctant to grant data permits to external research users, especially if data would need to be transferred across country borders.

Privacy risks can be reduced by limiting the data processing to a closed SPE with a remote desktop user interface for the researcher. Such approach is currently mandatory for register based research studies in Finland.

In the case of the MAITE project, we submitted the data permit application to the data controller (PHHYKY). The positive decision was received within two weeks. The data will be next transferred from the data controller to the SPE (“Kapseli”) hosted by the Finnish Social and Health Data Permit Authority (Findata) [25]. Although the closed SPE limits the researcher’s freedom in data processing, we consider the related benefits to be higher. Privacy protection in the MAITE project is of high priority as the targeted number of study participants is high (N=33 000) and a wide set of data for each individual will be used in the research and ML model development. We expect that our commitment of using the certified SPE had a positive impact on the data permit application process by convincing the register controller about sufficient privacy protection.

After the development phase, machine learning models may still be vulnerable to security threats compromising the privacy of the training data. Adversarial attacks may for instance attempt to identify and de-pseudonymize individuals included in training or reconstruct the training data [26]. Privacy concerns can be partly reduced by controlling access to the model parameters and output at deployment, but further defence mechanisms can also be employed when developing the model. Differential privacy mechanisms constitute the state-of-the-art and improves privacy by adding noise to the data or the algorithm itself,

e.g., in the objective function or in the gradients at each training iteration [26, 27]. Such mechanisms will also be considered in the MAITE project. We will identify and compare suitable privacy-preserving methods to ensure secure future integration of the ML model in commercial software.

4.3 From research to medical software

Medical device software development based on ML models originating from research projects can be a challenge for traceability. Even if the ML model development is carried out in the closed SPE, the traceability requirements of MDR and AIA should be fulfilled. This means that data pipelines, processing algorithms, data sets (tuning and testing) and respective version information shall be carefully documented. This may be a challenge for an SPE, which has primarily been designed for research purposes. Additional challenges arise from the fact that data permit is normally granted for a fixed duration after which the data is no more available unless an extension is granted.

In the MAITE project we will develop a software demonstration (PoC) without any requirement to comply with MDR. However, preparing for potential MDSW development after the MAITE project, we shall follow a systematic process to carefully document the ML model development steps. This will be achieved by the setting up software and data version management tools in the SPE to ensure full traceability between the ML model and the MDSW.

4.4 Deployment in the clinical environment

Deployment of ML based applications in the operational clinical environment involves many challenges. Technical integration challenges may be caused by the diversity of healthcare information systems between different service providers. Such challenges, require investment of resources to multiple integrations, but can usually be overcome. The most critical issue seems to be low acceptance by the professional users: development of solutions for healthcare

is in many cases technology-driven without sufficient contribution of end-users [28, 29].

The MAITE project addresses this problem by involving health and social services professionals during the full application development lifecycle. The project organizes several workshops during the research and model-development phase to understand user needs and the relevant personal health and social services usage data to be included as ML model variables.

Other potential challenges for ML application deployment are related to the GDPR and AIA regulation. The GDPR may limit the possibility of updating the application's ML model by directly using data from an EHR system. Even more critical an issue is the overall lawfulness of the application. For example, using applications for automatic profiling of individuals is not allowed by GDPR and AIA. Therefore, it is considered important to stress that ML based applications should only be used as decision support tools, leaving the final decision always to the end-user. The MAITE project addresses such problems by carrying out an in-depth investigation of the regulatory impacts and needed precautions to ensure that the application complies with applicable regulation.

5 DISCUSSION AND CONCLUSIONS

ML models are increasingly used in healthcare applications. The life cycle of ML based applications differs from conventional MDSW products, such as EHR systems. Existing studies concerning the development process of ML based medical applications have been mainly focused to the challenges of software change management of continuously updated applications. Less attention has been given to the challenges related to the exploitation of large amounts of personal data in the application development. However, the related data protection requirements have a remarkable impact on the application life cycle.

In this paper, we have outlined three main phases of application development. Each phase is characterized by its specific relation with regulation. In the research phase, regulation (GDPR) especially concerns privacy protection, while in the software development phase the main objective of regulation (MDR) is to ensure safety and performance. In the application deployment phase, regulation is typically national and focused to ensure secure integration of the application to the operational service environment.

Using our ongoing MAITE project as an example, we have analysed challenges in ML application development and we have presented best practices to overcome them. In the research phase, we propose a certified SPE to be used for data processing. This approach minimizes privacy risks and helps to reassure the data controller of appropriate data usage. We also propose privacy-preserving methods, such as differential privacy, to be applied for protecting the ML model. In order to meet the traceability requirements of MDSW, we recommend systematic version control and data set management processes to be applied already in the early phase of ML model development taking place in a closed SPE environment. This approach will help the transfer of the ML model developed in the research phase into the software development phase. To overcome the deployment challenges, we propose early involvement of

end-users already in the research phase as well as involvement of legal experts to ensure that the application being developed complies with regulation.

6 SUMMARY

This paper analyses the impact of regulation on the development of ML applications for healthcare. We especially focus on the challenges related to the use of sensitive personal data in the ML model development. We outline best practices for ensuring safe personal data processing and usage of the ML models as a basis for medical software development. We also highlight the importance for end-user involvement and legal evaluation at early development stage as a precondition for successful application deployment.

7 REFERENCES

- [1] Muehlematter UJ, Daniore P, Vokinger KN. Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015–20): a comparative analysis. *Lancet Digit Health* 2021; 3: e195–e203.
- [2] Subrahmanya VG, Shetty DK, Patil V., et al. The role of data science in healthcare advancements: applications, benefits, and future prospects. *Irish J Med Sci (1971 -);* 2021.
- [3] WHO. Ethics and Governance of Artificial Intelligence for Health: WHO guidance. *World Health Organ* 2021: 1–148.
- [4] Keutzer L, Simonsson USH. Medical Device Apps: An Introduction to Regulatory Affairs for Developers. *JMIR mHealth uHealth* 2020; 8.
- [5] Kane DW, Hohman MM, Cerami EG, et al. Agile methods in biomedical software development: a multi-site experience report. *BMC Bioinformatics* 2006; 7.
- [6] Granlund T, Stirbu · Vlad, Mikkonen · Tommi. Towards Regulatory-Compliant MLOps: Oravio's Journey from a Machine Learning Experiment to a Deployed Certified Medical Product. *SN Comput Sci* 2021; 2: 1–14.
- [7] Hsieh M-T, Huang K-C, Hsieh C-Y, et al. Validation of ICD-10-CM Diagnosis Codes for Identification of Patients with Acute Hemorrhagic Stroke in a National Health Insurance Claims Database. *Clin Epidemiol* 2021; 13: 43–51.
- [8] Artificial Intelligence in Medical Device Legislation | Futurium, <https://futurium.ec.europa.eu/en/european-ai-alliance/document/artificial-intelligence-medical-device-legislation> (accessed 27 June 2022).
- [9] Lähteenmäki J, Vuorinen A-L, Pajula J, et al. Integrating data from multiple Finnish biobanks and national health-care registers for retrospective studies: Practical experiences. *Scand J Public Health* 2021; 140349482110044.
- [10] European Health Data Space | Public Health, https://ec.europa.eu/health/ehealth/dataspace_en (accessed 27 June 2022).

- [11] Beskow LM, Dombeck CB, Thompson CP, et al. Informed consent for biobanking: consensus-based guidelines for adequate comprehension. *Genet Med* 2015; 17: 226–233.
- [12] Oxley PR, Ruffing J, Campion TR, et al. Design and Implementation of a Secure Computing Environment for Analysis of Sensitive Data at an Academic Medical Center. *AMIA Annu Symp Proc* 2018: 857.
- [13] Feng J, Emerson S, Simon N. Approval policies for modifications to machine learning-based software as a medical device: A study of bio-creep. *Biometrics* 2021; 77.
- [14] Babic B, Gerke S, Evgeniou T, et al. Algorithms on regulatory lockdown in medicine. *Science* 2019; 366: 1202–1204.
- [15] Minssen T, Gerke S, Aboy M, et al. Regulatory responses to medical machine learning. *J Law Biosci* 2020; 7: 1–18.
- [16] Gilbert S, Fenech M, Hirsch M, et al. Algorithm Change Protocols in the Regulation of Adaptive Machine Learning–Based Medical Devices. *J Med Internet Res* 2021; 23.
- [17] Artificial Intelligence Act, <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206> (accessed 27 June 2022).
- [18] Harp S, Carpenter T, Hatcliff J. A Reference Architecture for Secure Medical Devices. *Biomed Instrum Technol* 2018; 52: 357–365.
- [19] Frenz C. OWASP Secure Medical Device Deployment Standard, <https://owasp.org/www-pdf-archive/SecureMedicalDeviceDeployment.pdf> (accessed 27 June 2022).
- [20] Jormanainen V, Reponen J. CAF and CAMM analyses on the first 10 years of national Kanta services in Finland. *Finnish J eHealth eWelfare* 2020; 12: 302–315.
- [21] Overview of the national laws on electronic health records in the EU Member States (2016), https://ec.europa.eu/health/other-pages/basic-page/overview-national-laws-electronic-health-records-eu-member-states-2016_en (accessed 27 June 2022).
- [22] Leppla L, Hobelsberger S, Rockstein D, et al. Implementation Science Meets Software Development to Create eHealth Components for an Integrated Care Model for Allogeneic Stem Cell Transplantation Facilitated by eHealth: The SMILe Study as an Example. *J Nurs Scholarsh* 2021; 53: 35–45.
- [23] Lysaght T, Lim HY, Xafis V, et al. AI-Assisted Decision-making in Healthcare. *Asian Bioeth Rev* 2019; 11: 299–314.
- [24] Pritchard D, Petrilla A, Hallinan S, et al. What contributes most to high health care costs? Health care spending in high resource patients. *J Manag Care Spec Pharm* 2016; 22: 102–109.
- [25] Kapseli remote access - Findata, <https://findata.fi/en/kapseli/> (accessed 27 June 2022).
- [26] Al-Rubaie M, Chang JM. Privacy-Preserving Machine Learning: Threats and Solutions. *IEEE Secur Priv* 2019; 17: 49–58.
- [27] Bae H, Jang J, Jung D, et al. Security and Privacy Issues in Deep Learning. *SN Comput Sci* 2020; 1: 253.
- [28] McIntosh C, Conroy L, Tjong MC, et al. Clinical integration of machine learning for curative-intent radiation treatment of patients with prostate cancer. *Nat Med* 2021; 27: 999–1005.
- [29] El Naqa I. Prospective clinical deployment of machine learning in radiation oncology. *Nat Rev Clin Oncol* 2021; 18: 605–606.

8 ACKNOWLEDGEMENT

The paper originates from research work carried out in AHMED project (Agile and holistic medical software development) and MAITE project (Data-driven identification of elderly individuals with future need for multi-sectoral services). The authors want to thank Business Finland, Ministry of social affairs and health and VTT Technical Research Centre of Finland for their financial contributions in the projects and the support of Päijät-Häme Joint Authority for Health and Wellbeing in defining the research plan and data contents in the MAITE use case.

A temporal analysis of depression related tweets - a case study in Finland

Jonas Tana¹, Bao Cai² and Leonardo Espinosa-Leal²

¹Arcada University of Applied Sciences, Graduate School and Research, Helsinki, Finland jonas.tana@arcada.fi

² Arcada University of Applied Sciences, Graduate School and Research, Helsinki, Finland

Abstract

Depression is one of the most burdensome diseases in the world. A problem that depression presents, is the fact that it is connected with a high rate of unwillingness to seek professional help, and therefore many aspects of depression go unreported, affecting our understanding of it. Nowadays, individuals turn to online platforms for help and support, which creates vast amounts of data. This infodemiological study utilised data from Twitter to identify temporal patterns of behaviours related to depression in Finland. The findings of this study can be used to improve the impact of public health measures in relation to depression.

Keywords

depression, mental health, social media, twitter, infodemiology

1 INTRODUCTION

Mental health related health challenges are on the rise in today's society. One of these challenges, depression, which is one of the most common mental health diseases, has been projected to become the leading cause of disability in the next 20 years. In fact, it has already been ranked as one of the most burdensome diseases in the world by the World Health Organization (WHO) [1]. According to the WHO, depression is both a symptom and an illness, and characterized by sadness, loss of interest or pleasure, feelings of guilt or low self-worth, disturbed sleep or appetite, feelings of tiredness and a poor ability to concentrate [2]. Not only can depression be debilitating and cause impairments in an individual's everyday life, depression can also affect chronic health conditions, such as cancer, diabetes and obesity, as well as cardiovascular diseases [1]. Depression typically has a broad spectrum and can range from anxiety and milder forms to episodic or chronic severe depression. Depression can also be cyclic in its occurrence, and factors like temporal aspects and natural rhythms, such as the shift between night and day or the seasonal changes, can influence an individual's tendency to suffer from depression and depression like symptoms [3, 4 pp. 325-326, 5, 6 pp. 1-3, 7]. These kind of periodic variations in depression, and more generally in mood and mental health are familiar for most individuals, both at a circadian level (24-hour cycles), circaseptan (weekly cycles) and on a seasonal level (yearly cycles), especially in the southern and northern hemispheres, where changes in external conditions are evident. Reasons for these variations in mood are, apart from the environmental changes, likely endocrinological as well as social [8, 4 pp. 325-326]. However well known, from a public health perspective, there is a challenge in monitoring these variations. A reason for this is that more traditional methods, such as surveys, are gathered at a specific time

once a year or even less often, and thus lack the ability to identify trends, changes, and variations in health status within smaller time periods and intervals [1]. Depression incidence is also usually studied based on data from registers in healthcare institutions and organisations [9]. A challenge in relation to this, is that depression is connected with a high rate of unwillingness to seek professional help and can therefore go unreported. Whereas the first step toward appropriate treatment would be to seek professional help, individuals have been shown more likely to turn to online platforms and resources for help because of the stigma and barriers to care associated with depression [10, 11, 12, 13]. This presents a large gap in our knowledge in relation to depression and its variations. These large temporal gaps can hinder the development of effective and timely intervention programs and delay public health officials from acting on changes in health status [1, 9]. In the case of depression this is highly relevant, as a delay in initial treatment contact is associated with worse outcomes, despite effective interventions [12, 14].

The use of online generated user data may provide a method to overcome some of these gaps. As already mentioned, people suffering from depression are likely to turn to online platforms, and today, this online health information behaviour is extensive [11]. The vast amounts of data this generates, in search engines, on websites and in social media, not only allows us to study behaviours in relation to depression that have previously gone unreported, but also makes it possible to capture them in real-time and at a fine-grained temporal scale [1]. The use of data from online communication to study health related behaviours has been on the rise in the past decades, and has been called infodemiology, which can be defined as *the science of distribution of and determinants of information in an electronic medium, specifically the Internet, or in a population, with the ultimate aim to inform public health*

The 18th Scandinavian Conference on Health informatics, Tromsø, Norway, August 22-24, 2022. Organized by UiT The Arctic University of Norway. Conference Proceedings published by Linköping University Electronic Press at <https://doi.org/10.3384/ecp187>. © The Author(s). This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>

and public policy [15, p. 155]. The underlying assumption in infodemiology is, that in the age of the Internet, there is a relationship between population health on one hand, and information and communication patterns on the internet, on the other [16]. Therefore, changes in peoples' health status or behaviour, as well as public attention and attitudes are echoed in changes in online information and communication patterns [15]. The infodemiology approach has been effectively utilised to study temporal variation of online health behaviour in relation to depression, both in search engines and on discussion forums [11, 17]. Especially social media, where individuals are increasingly sharing their thoughts, emotions, and health concerns, can serve as a resourceful medium for mining information about the public's mental health and behaviours in a non-invasive manner on a large scale over time and space [13, 18].

Of all social media platforms, Twitter, the microblogging service, has been the most utilised platform for conducting research on expressed emotions [18]. Tweets have in previous research been analysed in relation to a myriad of health-related topics, ranging from influenza outbreaks and vaccinations to obesity, physical activity and drinking problems [19]. Mental health related tweets have also been analysed in various studies, and previous studies have shown, that people post about their depression and their treatment for depression on Twitter [1]. In relation to temporal variations, the expression of emotions on Twitter have also been studied, with different aims, ranging from the detection of emotional contagion and change in public opinions to identifying mental disorders and measuring population mood before, during and after natural disasters [18]. A few previous studies have found circadian and seasonal patterns in the content of emotionally loaded wordings in Twitter messages. However, the majority of studies have focused on simple frequency analysis, or content analysis as well as classification by machine learning approaches, and disregarded the temporal variations and patterns [19, 20]. Thus, these previous investigations do not examine the full potential of the platform. Moreover, to the authors' best knowledge, no studies to date have analysed depression related tweets limited to Finland. According to Official Statistics of Finland [21], in 2017, 11 percent of the population aged 15 or over used Twitter. In early 2022, Twitter was reported to have 759.5 thousand users in Finland, which would indicate that 15.8 percent of the population aged 15 or above in Finland used Twitter in early 2022 [22]. Even if the amount of Twitter users in Finland is limited, and only comprises of a small segment of the adult population, it can provide useful information, and complement previous infodemiological studies, on depression related online health information behaviour in Finland. Moreover, Finland, with its northern location and extreme variations in external conditions such as seasonal temperature and daylight, is an interesting subject to study health related phenomena from a temporal aspect [23, 24].

Therefore, the aim of this preliminary study is to identify temporal patterns and variation, as well as periods of heightened interest, in mental health topics, in this case depression, on Twitter in Finland.

2 METHOD

For this study, we collected and analysed timestamped and geo-located tweets identified with hashtags and terms relating to depression in Finnish (a list of included words can be found in Appendix 1). The terms were identified via the Google Trends top related query terms function, which resulted in 35 terms. It is worth noting that, even though the language and geographic location were set to Finnish and Finland, respectively, some of the keywords used in the search were English, simply because it is not uncommon to mix languages, especially with scientific vocabulary involved. We limited the tweets to include only tweets that originated in Finland geographically. UTC time was converted to the local time of the country in which tweets originated (Finland), after which we computed Hour and Day as numeric variables. We used the European convention to order the days of the week (leading to values of 0 for Sunday and 6 for Saturday).

Tweets were collected via the Twitter API version 2, specifically the `/2/tweet/search/all` endpoint of the Twitter API. The study utilised a subset of features the endpoint provided. A query is the main search feature containing the keywords, geographic location, and the language setting for the tweets. In addition, the date range was used to specify tweets' creation time, tweet, and users' fields to control the amount of data received and expansions to bypass the tweet's shortened form. A script was made to programmatically collect the tweets which contained any word in a set of keywords and were posted during a certain timeframe. The script accepts, at minimum, plain text keywords or a keyword text file, start and end time, and the location to output the file containing the tweets. In addition, geographic location, language preference, and data fields can also be passed to the script to control the receiving contents.

Aside from the text and the ids associated with each data point, the timestamp when the data point was created, the type of the data point, and the author id were also recorded. No personally identifiable information (PII) was collected in this study. The time frame of the data collection was from 2017 to 2021, including three categories: *quoted*, *replied_to*, and *tweet*. The complete set included 5114 data points, but we only studied direct tweets; therefore, we are limited to 2751 data points in this study. In the plots, the data is normalised yearly, and the average is calculated over the total number of years, five in this paper.

Statistical calculations and data analysis was performed using python programming language version 3.8.13, including the libraries Pandas (version 1.2.3) and Scipy (version 1.5.2). Plots were done using Matplotlib (version 3.5.1).

3 RESULTS

The A total of 2751 ($n=2751$) tweets were identified containing one or more of the chosen keywords in relation to depression during the chose five-year period, between 2017 and 2021. Out of sample of 2751 tweets, we found 1030 unique users. Of these unique users, for instance, 690 users posted one tweet, 172 users posted two tweets, and 49 users posted three tweets. Therefore, most of the unique users (999) posted less than 10 tweets in the dataset. We found one user that posted 249 tweets. Statistical

significance of the data distributions depicted in Figures 2-5 were obtained using the standard Kolmogorov–Smirnov test available in the Scipy python package. The obtained p-values are consigned in Table 1. All results reject the null hypothesis ($p < 0.05$).

Year	Monthly (Fig. 2)	Yearly (Fig. 3)	Weekly (Fig. 4)	Hourly (Fig. 5)
2017	0.0015	1.26e-12	0.022	7.36e-06
2018	0.0017	1.40e-12	0.023	4.34e-06
2019	0.0016	1.41e-12	0.024	7.44e-06
2020	0.0018	7.49e-13	0.018	4.33e-06
2021	0.0016	7.85e-13	0.018	4.26e-06

Table 1. Kolmogorov-Smirnov p-values for each series of data per year and distribution as presented in Figures 2-5.

As can be seen in Figure 1, the yearly distribution of depression related tweets shows a peak in 2019, with 678 tweets. In 2021 again, the number of tweets mentioning depression is at its lowest, with 388 tweets.

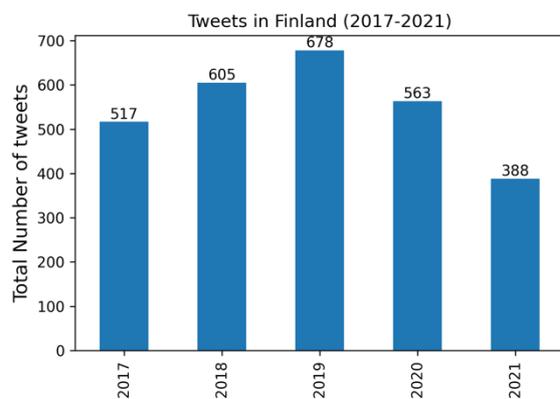


Figure 1. Yearly distribution of depression related tweets in Finland between 2017-2021.

On a seasonal scale, our analysis shows that tweets that mention depression follow a bimodal curve, with higher activity during spring and autumn. On an average level, the autumn peak is more significant compared to the spring peak. However, as can be seen in Figure 2, there is an exception for the year 2020, where a clear spring peak can be identified, and where activity is significantly higher during February. This February peak in 2020 is followed by a high activity in the end of the year 2019, specifically in November 2019.

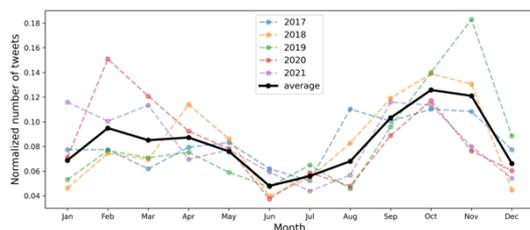


Figure 2. Monthly distribution of tweets that mention depression in Finland during 2017-2021.

A more detailed analysis of the seasonal variations on a weekly level, as illustrated in Figure 3, shows that the November peak in 2019 is located more towards the end of November, and the February 2020 peak again, is located in mid-February, shortly after the WHO declared a public

health emergency of international concern in relation to the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) that was rapidly spreading worldwide.

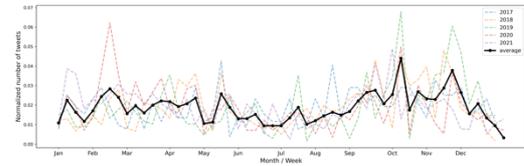


Figure 3. Yearly distribution of tweets mentioning depression on a weekly scale in Finland for the years 2017-2021.

On a circaseptan, or weekly, scale, tweets containing references to depression related terms, show a higher distribution towards the beginning of the week, reaching a small peak on Wednesday and Thursday, from where onwards activity rapidly starts to decline. The lowest number of tweets that mention depression related terms can be displayed on Sunday. The weekly distribution of tweets is presented in Figure 4.

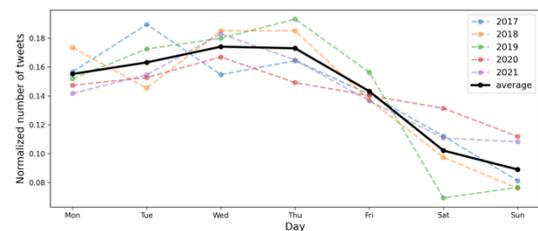


Figure 4. Weekly distribution of tweets containing depression related terms for the years 2017-2021.

The diurnal variation of tweets that contain depression related content is presented in Figure 5. Our analysis reveals a clear unimodal curve, where activity starts to rise in early morning from 03:00 onwards and reaching a peak at between 07:00 and 08:00 in the morning. Activity then starts to decline in the evening, from 19:00 onwards. Activity then is at its lowest between 22:00 and 03:00.

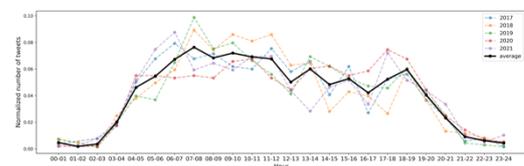


Figure 5. Distribution of tweets that contain depression related content on a diurnal (24-hour) scale for the years 2017-2021.

4 DISCUSSION

Our results indicate that tweets in Finnish containing depression related content follow temporal patterns, on a seasonal (yearly), circaseptan (weekly), as well as circadian (24-hour) level. The bimodal curve identified for tweets that contain depression related content in Finnish is similar

to previous infodemiological studies in relation to depression in Finland. Peaks in spring and autumn have previously been identified for search engine [24] and discussion forum activity [11] in relation to depression. Moreover, these seasonal findings, with peaks during spring and autumn, also correspond with previous findings of seasonality in diagnosed depression, hospital admissions and suicide rates in the Scandinavian countries and more broadly in the Northern Hemisphere [11]. The late November peak in activity is also similar to Dzogang et al. [8], who found that negative affect in messages on Twitter is significantly periodic and over-expressed in late November. The significantly higher peak in activity in depression related tweets during February 2020 identified in this study is most likely related to the Covid-19 pandemic and does validate previous findings that have indicated a significant increase in mental health symptoms early in the pandemic [25]. A more detailed content analysis of the tweets for these peak time-periods is already planned, with the aim to identify reasons for heightened activity.

On a circaseptan level, the findings in this study are somewhat contradictory to previous findings of online health behaviour in relation to depression in Finland, where peaks have been identified during Sundays and Mondays [11, 24]. Sundays and Mondays have also been identified as days when lower mood or depression symptoms are more prevalent [26]. From a broader perspective, health contemplations in general have been shown to peak in the beginning of the week and decline towards the end of the week [14]. Moreover, Twitter usage has in some studies been shown to decline towards weekends, which could explain the weekend trough in activity [27, 28].

As for the diurnal variations, activity on Twitter reveals a different temporal pattern compared to search engine queries in relation to depression in Finland, where significant peaks are visible during night-time, and troughs in activity during daytime [17]. The same night-time peaks have also been found in discussion forums for discussions related to depression [11]. However, a similar early morning peak to the one identified in this study, in tweets containing words representing negative affect has been identified by Golder & Macy [29] globally, in four English speaking regions. They identified a peak between 03:00 and 04:00 in the morning. Moreover, Ten Thij et al. [30] found, that people suffering from depression showed higher levels of “rumination” and “self-reflection” between 03:00 and 06:00 in the morning. The rising activity during early-morning, between 03:00 and 06:00, could be an indicator for disturbed sleep quality and sleep patterns, something that is a common symptom for depression [2].

An explanation for the differing results on a circaseptan as well as diurnal level, compared to the temporal patterns exhibited in search engines and discussion forums in Finland, could be the somewhat different use of Twitter compared to search engines and discussion forums. On Twitter, the largest user group consists of professional or business users, often more active during office hours, while the smallest user group is personal users. On Twitter, the personal, or casual users, are also the users that have a low or mild social engagement [31].

As can be seen, our findings indicate that collecting tweets offers an empirically based, objective and valid way of identifying and revealing temporal patterns and periods of heightened activity on Twitter in relation to depression. Data from Twitter can therefore help create a broader picture of online behaviours in relation to depression. The findings in this study also shows that comprehensive trend analysis of data from social media can reveal important insights that can be useful for timing more effective interventions and disseminate credible information related to depression [19]. For instance, detecting a rise in levels of depression related content could trigger social media platforms, in this case Twitter, to recommend internet-based intervention services, such as online cognitive behavioural therapy, to the users [13]. This would allow for early or timely optimal interventions, which have been shown to be of relevance in treatment response [14]. Moreover, this kind of infodemiological research can be extended to almost any health condition, and by collecting and analysing publicly available data, in near-real time, public health officials can be provided with early warning signals. This again can help provide important information to intervene by designing and implementing timely public health campaigns [15, 18]. It however needs to be noted, that this method of data collection and analysis are not meant to be a replacement for traditional public health surveillance systems or clinical diagnoses of depression, but as a compliment to them. As a substantial part of the population act on health-related issues without the involvement of health professionals, utilizing novel ways to detect these behaviours can complement more traditional data gathering methods, and aid in taking appropriate and needful public health measures [1, 15, 32].

5 LIMITATIONS

Our study does present some limitations. Firstly, there is no way to ensure that all tweets mentioning depression or related terms are strictly related to depression. A solution to this would be better filtering of the data and techniques, such as sentiment analysis to further refine the sample, could improve precision [19]. However, as the aim of this preliminary study is merely to identify temporal patterns in depression related online behaviour, future studies will utilize more precise tools to conduct further analysis on content of the tweets. Another limitation is the amount of Twitter users in Finland. With a relatively small percentage of users, the results cannot be generalised to a wider population, and the population observed on Twitter may display activity at times that are not corresponding with a broader population. Nevertheless, as more and more people are starting to engage in health-related behaviour on different online platforms, there is a need to examine these to be able to gain a holistic perspective on these behaviours. This again, is necessary in order to provide effective interventions, in this case for people that might need help with mental health related issues.

6 CONCLUSIONS

The aim of this study was to identify temporal variations and patterns of tweets containing depression related content in Finland. This preliminary study is the first of its kind in Finland to analyse the temporal variations and patterns of tweets that relate to depression. The results are somewhat

contradictory to previous findings from similar infodemiology studies, which advocates more research within this field, to gain a broad perspective of online behaviours in relation to depression. Generally, research on harnessing online data for understanding online health behaviours is still in its infancy, and as so often, more research is needed to harness the full potential that the analysis of this relatively new data generated on online media platforms can offer to tackle public health concerns [19]. The amount of online engagement in matters that relate to health keep rising on existing, as well as new platforms, which results in more and more user generated data. Therefore, it is necessary to pursue research on online platforms in order to improve representativeness of health behaviours, and as a result of that, the impact of public health measures.

7 REFERENCES

- [1] De Choudhury, M., Counts, S., Horvitz, E. "Social media as a measurement tool of depression in populations." In *Proceedings of the 5th annual ACM web science conference*, pp. 47-56. 2013.
- [2] Marcus, M., Yasamy, M. T., van Ommeren, M., Chisholm, D., Saxena, S. "Depression: A global public health concern." WHO Department of Mental Health and Substance Abuse, Vol. 1, pp. 6-8. 2012.
- [3] Antypa, N., Vogelzangs, N., Meesters, Y., Schoevers, R., Penninx, B. W. "Chronotype associations with depression and anxiety disorders in a large cohort study." *Depression and Anxiety*, Vol. 33, Issue 1, pp. 75-83. 2016.
- [4] Dunlap J.C., Loros J.J., DeCoursey P.J. (eds.). "Chronobiology: biological timekeeping." Sunderland MA: Sinauer Associates, 2004.
- [5] Maes, M., Meltzer, H. Y., Suy, E., De Meyer, F. "Seasonality in severity of depression: relationships to suicide and homicide occurrence." *Acta Psychiatrica Scandinavica*, Vol. 88, Issue 3, pp. 156-161. 1993.
- [6] Roenneberg, T. "Internal time: Chronotypes, social jet lag, and why you're so tired." Harvard University Press. 2012.
- [7] WHO. "Depression fact sheet."2018. Available at: <https://www.who.int/news-room/fact-sheets/detail/depression> > [Accessed 02 May 2022].
- [8] Dzogang, F., Lansdall-Welfare, T., Cristianini, N. "Seasonal fluctuations in collective mood revealed by Wikipedia searches and Twitter posts." In *2016 IEEE 16th International Conference on Data Mining Workshops*, pp. 931-937. 2016.
- [9] Luo, J., Du, J., Tao, C., Xu, H., Zhang, Y. "Exploring temporal patterns of suicidal behavior on Twitter." In *2018 IEEE International Conference on Healthcare Informatics Workshop*, pp. 55-56. 2018.
- [10] Oliphant, T. "The information practices of people living with depression: constructing credibility and authority." Doctoral dissertation, The School of Graduate and Postdoctoral Studies, University of Western Ontario, London, Ontario, Canada, 2010.
- [11] Tana, J., Eirola, E., Eriksson-Backa, K. "Exploring temporal variations of depression-related health information behaviour in a discussion forum: the case of Suomi24." *Information Research*, Vol 25, Issue 2, paper 854. 2020.
- [12] Yoshikawa, E., Taniguchi, T., Nakamura-Taira, N., Ishiguro, S. Matsumura, H. "Factors associated with unwillingness to seek professional help for depression: a web-based survey." *BMC Research Notes*, Vol. 10, Issue 1, pp.1-6. 2017
- [13] Zhang, Y., Lyu, H., Liu, Y., Zhang, X., Wang, Y., Luo, J. "Monitoring depression trends on Twitter during the COVID-19 pandemic: Observational study." *JMIR infodemiology*, Vol 1, Issue 1, e26769. 2021.
- [14] Ayers, J. W., Althouse, B. M., Johnson, M., Dredze, M., Cohen, J. E. "What's the healthiest day?: Circaseptan (weekly) rhythms in healthy considerations." *American Journal of Preventive Medicine*, Vol 47, Issue 1, pp. 73-76. 2014a.
- [15] Eysenbach, G. "Infodemiology and infoveillance: tracking online health information and cyberbehavior for public health." *American Journal of Preventive Medicine*, Vol 40, Issue 5, pp. 154-S158. 2011.
- [16] Eysenbach, G. "Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet." *Journal of Medical Internet Research*, Vol. 11, Issue 1, e11. 2009.
- [17] Tana, J., Kettunen, J., Eirola, E., Paakkonen, H. "Diurnal variations of depression-related health information seeking: case study in Finland using Google Trends data." *JMIR Mental Health*, Vol. 5, Issue 2, e43. 2018.
- [18] Mayor, E., Bietti, L. M. "Twitter, time and emotions." *Royal Society open science*, Vol. 8, 201900. 2021.
- [19] McClellan, C., Ali, M. M., Mutter, R., Kroutil, L., Landwehr, J. "Using social media to monitor mental health discussions– evidence from Twitter." *Journal of the American Medical Informatics Association*, Vol 24, Issue 3, pp. 496-502. 2017.
- [20] Lampos, V., Lansdall-Welfare, T., Araya, R., Cristianini, N. "Analysing mood patterns in the United Kingdom through Twitter content." *arXiv preprint*, arXiv:1304.5507. 2013.
- [21] Official Statistics of Finland (OSF): "Using social media service, those aged 10 or over (2017) %, in Participation in leisure activities" [online]. Helsinki: Statistics Finland. Available at: http://www.stat.fi/til/vpa/meta_en.html > [Accessed 02 May 2022].
- [22] Kemp, S. "Digital 2022: Finland." [online]. 2022. Available at: <https://datareportal.com/reports/digital-2022-finland> > [Accessed 03 May 2022].
- [23] Basnet, S., Merikanto, I., Lahti, T., Männistö, S., Laatikainen, T., Vartiainen, E. and Partonen, T. "Seasonal variations in mood and behavior associate with common chronic diseases and symptoms in a population-based study." *Psychiatry Research*, Vol. 238, pp. 181-188. 2016.

- [24] Tana, J. "An infodemiological study using search engine query data to explore the temporal variations of depression in Finland." *Finnish Journal of eHealth and eWelfare*, Vol. 10, Issue 1, pp. 133-142. 2018.
- [25] Robinson, E., Sutin, A. R., Daly, M., Jones, A. "A systematic review and meta-analysis of longitudinal cohort studies comparing mental health before versus during the COVID-19 pandemic in 2020." *Journal of affective disorders*, Vol. 296, pp. 567-576. 2022.
- [26] Tumen, S., Zeydanli, T. "Day-of-the-week effects in subjective wellbeing: does selectivity matter?." *Social Indicators Research*, Vol. 119, Issue 1, pp. 139-162. 2014.
- [27] García-Palomares, J. C., Salas-Olmedo, M. H., Moya-Gomez, B., Condeco-Melhorado, A., Gutierrez, J. "City dynamics through Twitter: Relationships between land use and spatiotemporal demographics." *Cities*, Vol. 72, pp. 310-319. 2018.
- [28] Longley, P. A., Adnan, M., Lansley, G. "The geotemporal demographics of Twitter usage." *Environment and Planning*, Vol. 47, Issue 2, pp. 465-484. 2015.
- [29] Golder, S. A., Macy, M. W. "Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures." *Science*, Vol. 333, Issue 6051, pp. 1878-1881. 2011.
- [30] Ten Thij, M., Bathina, K., Rutter, L. A., Lorenzo-Luaces, L., van de Leemput, I. A., Scheffer, M. Bollen, J. "Depression alters the circadian pattern of online activity." *Scientific reports*, Vol. 10, Issue 1, pp. 1-10. 2020.
- [31] Uddin MM, Imran M, Sajjad H. "Understanding types of users on Twitter." arXiv preprint arXiv:1406.1335. 2014.
- [32] Ayers, J. W., Althouse, B. M., & Dredze, M. "Could behavioral medicine lead the web data revolution?" *Jama*, Vol. 311, Issue 14, pp. 1399-1400. 2014b.

Appendix 1 – terms related to depression used to identify tweets.

Masennus (eng. depression)

masennus oireet (eng. depression symptoms)

masennustesti (eng. depression test)

masennus testi (eng. depression test)

synnytyksen jälkeinen masennus (eng. postpartum depression)

masennus hoito (eng. depression treatment)

ahdistus (eng. anxiety)

keskivaikea masennus (eng. moderate depression)

psykoottinen masennus (eng. psychotic depression)

lapsen masennus (eng. childrens depression)

vakava masennus (eng. severe depression)

depression (eng. depression)

vaikea masennus (eng. serious depression)

nuorten masennus (eng. youth depression)

raskaus masennus (eng. pregnancy depression)

väsymys (eng. fatigue)

lievä masennus (eng. mild depression)

masennuslääkkeet (eng. depression medication)

mielenterveys (eng. mental health)

nuoren masennus (eng. adolescent depression)

masennus blogi (eng. depression blog)

itsemurha (eng. suicide)

masennus keskustelu (eng. depression discussion)

psykoosi (eng. psychosis)

masennuksen hoito (eng. depression treatment)

masennus itsehoito (eng. depression self

krooninen masennus (eng. chronic depression)

kaksisuuntainen mielialahäiriö (eng. bipolar disorder)

depression test

depression symptoms

manic depression

postpartum depression

crippling depression

clinical depression

high functioning depression

Data collection and smart nudging to promote physical activity and a healthy lifestyle using wearable devices

Seshathiri Dhanasekaran¹, Anders Andersen¹, Randi Karlsen¹, Anne Håkansson¹, and André Henriksen¹

¹UiT The arctic university of Norway, Department of Computer Science, Tromsø, Norway, seshathiri.dhanasekaran@uit.no

Abstract

Nudge principles and techniques can motivate and improve personal health through emerging digital devices, such as activity trackers. Tracking people's health and well-being using such devices have earned widespread interest. These devices can continuously capture and analyze health-related data from individuals and communities in their everyday environment. Providing context-aware nudges can help individuals to self-manage and improve their health. In this study, we discuss how a consumer-based activity tracker can be used to track different variables for physical activity (PA) and how it has the potential to be an important source of data for future smart nudging.

Keywords

Digital nudge; physical activity; artificial intelligence; lifestyle; smartwatch

1 INTRODUCTION

A nudge was defined by Thaler and Sunstein in 2008 as "any aspect of the choice architecture that alters people's behaviour predictably without forbidding any options or significantly changing their economic incentives" [1]. Nudges were initially explored in offline decision-making, focusing mainly on personal health or wealth decisions.

The concept of *digital nudging* was introduced in 2016 when the notion of nudging was transferred to digital user interfaces. The term was coined by Weinmass et al. [2], who defined it as the "*use of user-interface design elements to guide people's choices or influence user's inputs in online decision environments.*" Others have later revised this, and Meske and Potthoff [3] further expanded the definition in 2017 to include free decision-making and defined digital nudging as "*a subtle form of using design, information, and interactive elements to guide user behaviour in digital environments, without restricting the individual's freedom of choice.*"

Further, in 2019, Karlsen and Andersen [4] defined *smart nudging* as "*digital nudging, where the guidance of user behaviour is tailored to the current situation of each user.*"

Smart nudging requires a user profile with a broad scope of connected data. The data is investigated before a customized nudge is structured. User acceptance of a customized nudge likely has a higher probability of succeeding (i.e., the user approves the nudge and adheres to the recommendation) than a non-customized nudge [4].

Current evolving digital technologies and devices, such as smartwatches, activity trackers, and smartphones, can provide continuous and long-term collection and analysis of behavioural data. This includes health-related data types like physical activity (PA), pulse, body temperature, stress levels, and sleep, and contextual data (e.g., location by global positioning systems [GPS]) from which other data types can be inferred (e.g., weather). These devices thus

provide an opportunity for providing a continuous feedback loop to provide on-time nudges, to allow people to better self-manage their health and make informed choices [5].

From a public health perspective, by collecting this type of data from a large number of the population, these devices can potentially provide insights into different demographic groups, with access to near real-time data collection and evaluation at the population level, without being directly involved with the individual [6].

Recent developments in wearable sensor technology have taken such implementation closer to existence [7, 8]. Adding context-aware nudges to these digital health devices can help individuals identify and self-manage their PA levels to lead a healthy lifestyle.

The World Health Organization recommends that adults perform at least 150-300 minutes of moderate PA each week [9]. Globally, 25% of adults do not reach this goal [9]. Inactive people have a 20%-30% higher risk of death, and it is estimated that five million deaths could be avoided each year if people were more physically active [9].

Understanding an individual's context is complex, and smart device data must be precise to get valuable insights [10]. Natural human behavior consists of multiple simultaneous circumstances [11]. For example, people may be running indoors, outdoors, and on running tracks. The geographical location data can assist in identifying the user's context in this situation.

We propose a system that provides nudges related to the user's context and interests [4]. A system with context-aware data sources such as location data, time, weather, PA, and heart rate data, will help tailor nudges suitable to the user's context. For example, we could present a nudge to a user living close to a ski resort with a previous interest in skiing, suggesting when they can ski in the mountains using weather and location data for context and information about the activity levels for the user the previous days.

The 18th Scandinavian Conference on Health informatics, Tromsø, Norway, August 22-24, 2022. Organized by UiT The Arctic University of Norway. Conference Proceedings published by Linköping University Electronic Press at <https://doi.org/10.3384/ecp187>. © The Author(s). This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>

To create such a smart nudging system, we must first understand what type of data is available in current digital smart technology and explore how that data can be used.

Therefore, the objective of this study was to use PA data from consumer smartwatches to understand the context of the consumer and prepare for a future smart nudging solution for leading a healthy lifestyle.

2 RELATED WORKS

A pilot study by Haga et al. [11] was conducted in Canada to promote PA among college students with smartwatches. One hundred seventy-five students completed surveys on stress, diet, PA, and behaviors during weekends. The study did not test the effect of the intervention, so they could not assess the impact of these technologies on health and wellness outcomes [11].

A study by Cherubini et al. [10] aimed to find the correlation between tangible rewards and PA. The study reveals that tangible rewards and motivational messages decrease intrinsic motivation and thus their related PA. One of the primary outcomes of this work is that they learned that tangible rewards do not help establish lasting healthy routines. They observed no significant difference in the number of steps walked during the experiment, whether they offered participants money or not to perform this activity [10].

A study by Hafner et al. [12] suggests that incentivizing PA can lead to increased activity levels. In addition to the overall rise in the amount of activity monitored, the gain also appears to be correlated with higher levels of intensive activity over time, measured in the average number of days of intensive activity per month [12].

Finally, Mozgai et al. [13] proposed a mHealth application that provided optimized and interactive digital content through a mobile application. Novel adaptive logic algorithms used behavioral change techniques like a virtual human coach. The system would be aware of context and personal usage patterns with additional data sources, e.g., user mobility and calendar data.

3 METHOD

Recruitment

We used convenience sampling to recruit six volunteers. Participants had to be 18 years or older, owned and operated an Apple Watch (Apple Inc., CA, USA), and willing to share watch recorded PA data. Participants were recruited among people living in the northern part of Norway, north of the arctic circle.

Equipment

The first Apple Watch was released in 2015. Since then, eight generations have been released, and it is currently one of the most popular smartwatches available. In 2021, the Apple Watch had a 30% market share worldwide [14]. Depending on generation, the watch comes in several sizes (38-45mm), colors, metal finish, memory- and storage capacity, display quality, connectivity capabilities, and sensor support.

The fourth generation Apple Watch was released in 2018. It included a multitude of sensors, including a Global positioning system (GPS), tri-axial accelerometer, gyroscope, altimeter, compass, electrical heart sensors

(ECG), and optical pulse sensor (i.e., photoplethysmograph).

Fuller et al. [15] concluded in a 2020 systematic review of smartwatch validation studies that the Apple Watch (with Samsung) had less measurement variability than other brands when estimating step counts and heart rate. The same study also concluded that although energy expenditure on average was overestimated by 58%, it was one of the most accurate brands. A recent paper by Kwon et al. [16] further concluded that the Apple Watch had a mean absolute percentage error (MAPE) of only 1% for moderate-to-vigorous PA and a MAPE of 4% for activity energy expenditure. A MAPE below 10% is generally used as a threshold for an acceptable MAPE when comparing smartwatch data in free-living conditions [17, 18].

Data extraction and variables

We asked participants to export their Apple Watch collected data from the Apple Health web solution to access the data. The exported data contains daily data on PA, workout details, and device information encoded as one Extensible Markup Language (XML) file per participant. The collected data has aggregated the steps count of a person from mobile phones and smartwatches. We excluded mobile phone steps and only analyzed step data retrieved from the smartwatch.

We extracted relevant variables from the source data using *Python* 3.9 [19] with *NumPy* 1.21 [20]. We only extracted data between January 2020 and December 2021. We stored variables in comma-separated value (CSV) files, divided into daily variables for *Active energy*, *Exercise minutes*, *Steps*, and *Distance*.

Active energy is understood as PA energy expenditure, given as Kcal. *Exercise minutes* are considered minutes of light-, moderate-, and vigorous PA combined. *Distance* is the sum of walking- and running distance, given in kilometers. *Steps* are the sum of daily steps counted on the smartwatch device.

Additional variables exist, including Heart rate, heart rate variability (HRV), oxygen saturation (SpO₂), peak oxygen uptake (VO₂Max), stand minutes, and sleep data. These additional variables were not used in the current study but can be relevant for data analysis in future smart nudging.

Statistical analysis

We used *Python* 3.9 [19] with *Pandas* 1.4.2 [21, 22] and *Matplotlib* 3.5.2 [23] packages to visualize and analyze the PA data for the following variables: *Active energy*, *Exercise minutes*, *Steps*, and *Distance*.

For each combination of selected variables, we calculated Pearson's correlation coefficients. The strength of the associations is assessed using cutoffs suggested by Evans [24], i.e., very weak, less than 0.2; weak, 0.2-0.4; moderate, 0.4-0.6; strong, 0.6-0.8; and very strong, greater than 0.8. Correlations are presented as scatter plots. Value distributions for each variable are shown as bar plots. Pearson correlations and variable distributions are combined and presented in a *pair* plot. Pearson's correlations are further explained as a heat map.

For Steps and Exercise Minutes, we created *time series* plots for mean daily values to evaluate change in these variables throughout the two years for all the participants.

4 RESULTS

Participant characteristics

The mean age (in 2022) for included participants was 36.3 years (SD=11.7). We collected two years of PA data for each participant, from January 2020 to December 2021. Included participants used the 4th generation Apple Watch or newer. All participants signed informed consent.

Variable correlation and value distribution

Here we present how the four selected variables correlate. (i.e., Active energy [kcal], Exercise minutes [light-, moderate-, and vigorous PA], Steps, and Distance [km]).

Figure 1 gives a *pair* plot showing the Pearson's correlation coefficient between each variable. We used the mean value of every month's activity of all participants as data for the variables. The value distribution for each variable is also provided, shown as natural divided time in months over two years. A corresponding heat map with correlation values is given in Figure 2.

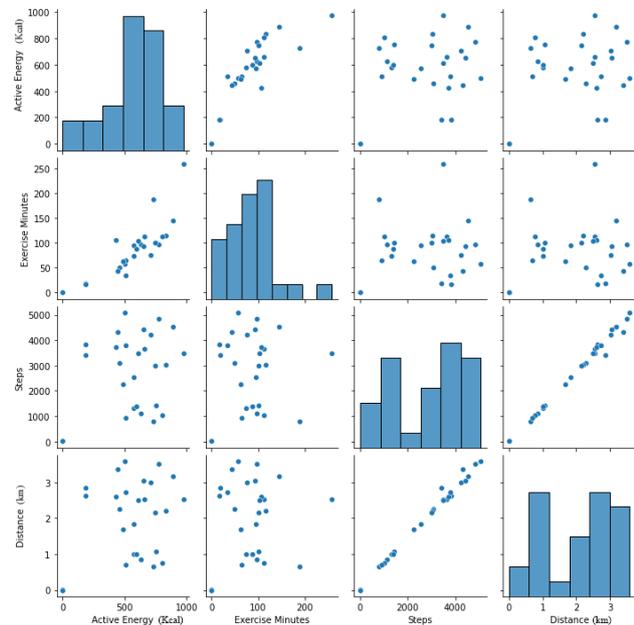


Figure 1. Pair plot with correlations and value distribution for each variable under study.

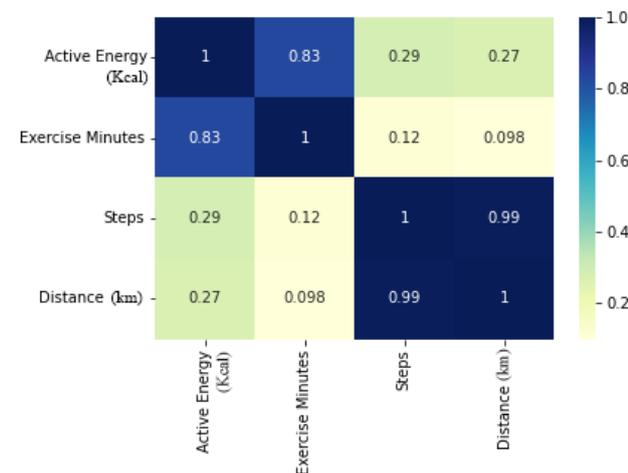


Figure 2. Correlation heat map for each variable under study.

In the pair plot (Figure 1), we checked the correlation between the variables with monthly mean value as data. Each scatter point denotes a monthly mean data of the variables.

Steps have a very strong correlation with Distance (0.99), a very weak correlation with Exercise Minutes (0.12), and a weak correlation with Active Energy (0.29). Exercise Minutes have a very weak correlation with Distance (0.098) and a very strong correlation with Active Energy (0.83). Active Energy has a weak correlation with Distance (0.27).

Change in step counts

Figure 3 gives a time series visualization plot for *Step* data for 2020 and 2021, using mean value of all participants.

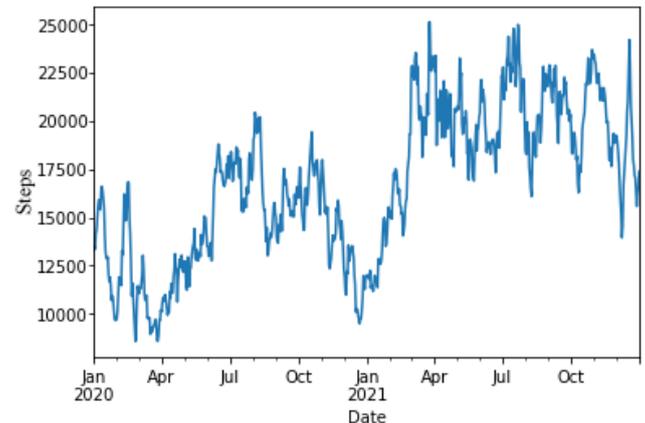


Figure 3. Change in daily mean steps between January 2020 and December 2021

The average step count data from 2020 and 2021 shows a higher step count between June and September compared to the remaining months. The lowest average daily Step count was found in March 2021 (2890 steps). The highest average daily Step count was in April 2022 (13750 steps).

The Step count data for 2021 shows a lower level of recorded steps during the year's initial months and an increased number of steps from April to December, especially between April and August.

Change in Exercise Minutes

Figure 4 gives a time series visualization plot for the *Exercise Minutes* data for 2020 and 2021, using mean value of all participants.

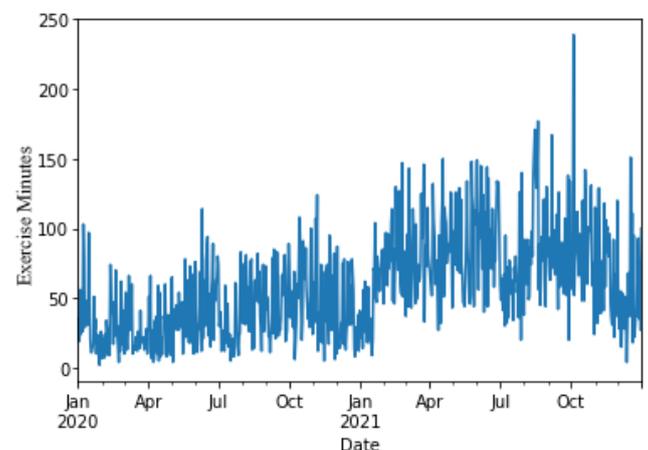


Figure 4. Mean Exercise Minutes for the two years

The mean *Exercise Minutes* was lowest from March 12, 2020, to April 10, 2020; this time period was also the Norwegian COVID-19 lockdown period. The *Exercise Minutes* did not diverge much from the yearly mean during the COVID-19 lockdown period. From December 2020 to January 2021, participants' *Exercise Minutes* were reduced.

5 DISCUSSION

In this study we collected and analyzed PA data for 2020 and 2021 from six participants wearing an Apple Watch, recruited in the Arctic region. We analyzed different variables for PA collected by the watch.

The correlation plots show that only Distance and Step count, *Exercise Minutes* and *Activity Energy* are linear dependent. Remaining combination of variables does not have a clear linear dependency.

We observed a decline in *Steps* and *Exercise Minutes* during the COVID-19 lockdown period (from March 2020). After June 2020, people seemed to find ways to be active. A potential reason for the reduction of performed physical activity in January and February 2021, may be that these are the two coldest months of the winter season. Furthermore, although not a lockdown, there were COVID-19 restriction during this period.

A sharp increase in people's *Exercise Minutes* after the COVID-19 restrictions were lifted, i.e., after September 25, 2021, was also observed. However, three weeks later, *Exercise Minutes* had returned to normal levels (i.e., before September 25).

Physical activity findings from the present study is in accordance with a previous study by Henriksen et al. [6], where two years of PA data were collected using consumer-based activity trackers. Results showed that change in PA due to the Norwegian COVID-19 lockdown was distinct and clearly detected.

The Tromsø study, the longest-running population study in Norway [25] was initialized in 1974 with the overarching goal of combating the cardiovascular disease epidemic in the northern part of Norway. The seventh survey of The Tromsø study saw a steady decrease in cardiovascular risk variables while obesity continued to rise [26]. In this study, and similar population studies, data on physical activity are traditionally collected using questionnaires and research-based accelerometers. The present study suggests future population studies could investigate new technologies and tactics to improve data gathering efficiency and reduce participant burden. Smartwatches and activity trackers could be investigated as a potential source of PA data, adding to already existing methods for PA data collection.

The advantage of the data collection method in the present study, is that it shows that *Exercise Minutes* (and other variables) can be recorded with a smart watch, and that daily levels of PA can be measured. However, evaluating PA using *Steps* and *Activity Energy* alone is not sufficient.

Depending on which activity tracker a participant wears, a multitude of other variables can be extracted. For the Apple Watch, in addition to PA, various sleep and heart rate variables exist, including: pulse, heart rate variability (HRV), oxygen saturation (SpO₂), peak oxygen uptake (VO₂Max), Electrocardiograms (ECG), time in bed, average sleep minutes, and more.

For participants already owning an activity tracker, data can be collected without adding much burden to participants.

Conclusions

Understanding the context of the user is critical for smart nudging. But, we faced the toughest pandemic and lockdown in the form of COVID-19 effect on our data [26].

People found new ways of performing workouts and physical activity during COVID-19, and people are interested in following the same even after removing the restrictions due to their constraints and comfort.

We are working on data analysis techniques to predict the activity type with the impact of context to assess how context-aware nudges can work better for them. We also plan to evaluate the effect of weather conditions and calendar data on physical activity in the future to tailor smart nudges for the users. Input from smartwatch data is an essential source for future smart nudging in this context. We will use this study and methods to analyze the context impact on physical activity and increase physical activity with context-awareness.

6 SUMMARY

To understand users' context and PAs in the Arctic regions, we collected and analyzed smartwatch activity for 2020 and 2021. We enrolled six participants. Our finding indicates that PA patterns changed due to COVID-19, albeit temporarily. Despite governmental restrictions, participants found other means of performing PA, even though they could not leave their house as much as usual. This study focuses on understanding the context of the users to assess and prepare nudges for healthy lifestyle patterns.

7 REFERENCES

- [1] Thaler, R.H. and Sunstein, C.R. (2009), *Nudge: Improving decisions about health, wealth, and happiness* (Penguin Books).
- [2] Weinmann, M., Schneider, C., and vom Brocke, J., "Digital Nudging", *SSRN Electronic Journal*, 2015
- [3] Meske, C. and Potthoff, T. (2017), 'The DINU-Model - a process model for the design of nudges', *25th European conference on information systems (ECIS)* (Guimarães, Portugal: Research-in-Progress Papers), 2587-97.
- [4] Karlsen, R. and Andersen, A., "Recommendations with a Nudge", *Technologies*, 7 (2), 2019
- [5] Shaw, R.J., Bonnet, J.P., Modarai, F., George, A., and Shahsahebi, M., "Mobile health technology for personalized primary care medicine", *Am J Med*, 128 (6), 555-7, 2015
- [6] Henriksen, A., Johannessen, E., Hartvigsen, G., Grimsgaard, S., and Hopstock, L.A., "Consumer-Based Activity Trackers as a Tool for Physical Activity Monitoring in Epidemiological Studies During the COVID-19 Pandemic: Development and Usability Study", *JMIR Public Health Surveill*, 7 (4), e23806, 2021
- [7] Meske, C., Amojó, I., Poncette, A.-S., and Balzer, F. (2019), 'The Potential Role of Digital Nudging in the Digital Transformation of the Healthcare Industry',

Design, User Experience, and Usability. Application Domains (Springer International Publishing).

- [8] Reeder, B. and David, A., "Health at hand: A systematic review of smart watch uses for health and wellness", *J Biomed Inform*, 63, 269-76, 2016
- [9] World Health Organization 'Physical activity', <<https://www.who.int/news-room/fact-sheets/detail/physical-activity>>, accessed 2022-01-20.
- [10] Cherubini, M., Villalobos-Zuñiga, G., Boldi, M.-O., and Bonazzi, R., "The Unexpected Downside of Paying or Sending Messages to People to Make Them Walk", *ACM Transactions on Computer-Human Interaction*, 27 (2), 1-44, 2020
- [11] Haga, S.B., Shaw, R., Kneifel, C., Bond, S.J., and Ginsburg, G.S., "Promoting Wellness Through Mobile Health Technology in a College Student Population: Protocol Development and Pilot Study", *JMIR Res Protoc*, 9 (4), e16474, 2020
- [12] Hafner, M., Pollard, J., and Van Stolk, C., "Incentives and Physical Activity: An Assessment of the Association Between Vitality's Active Rewards with Apple Watch Benefit and Sustained Physical Activity Improvements", *Rand Health Q*, 9 (1), 4, 2020
- [13] Mozzai, S., Hartholt, A., and Rizzo, A., "An Adaptive Agent-Based Interface for Personalized Health Interventions", *IUI '20: Proceedings of the 25th International Conference on Intelligent User Interfaces Companion*, 118-19, 2020
- [14] GmbH, S. (2022), 'Smartwatch market share worldwide from 2020 to 2021, by vendor', <<https://www.statista.com/statistics/1296818/smartwatch-market-share/>>, accessed 05-23.
- [15] Fuller, D., Colwell, E., Low, J., Orychock, K., Tobin, M.A., Simango, B., Buote, R., Van Heerden, D., Luan, H., Cullen, K., Slade, L., and Taylor, N.G.A., "Reliability and Validity of Commercially Available Wearable Devices for Measuring Steps, Energy Expenditure, and Heart Rate: Systematic Review", *JMIR Mhealth Uhealth*, 8 (9), e18694, 2020
- [16] Kwon, S., Kim, Y., Bai, Y., Burns, R.D., Brusseau, T.A., and Byun, W., "Validation of the Apple Watch for Estimating Moderate-to-Vigorous Physical Activity and Activity Energy Expenditure in School-Aged Children", *Sensors (Basel)*, 21 (19), 2021
- [17] Nelson, M.B., Kaminsky, L.A., Dickin, D.C., and Montoye, A.H., "Validity of Consumer-Based Physical Activity Monitors for Specific Activity Types", *Medicine & Science in Sports & Exercise*, 48 (8), 1619-28, 2016
- [18] Roos, L., Taube, W., Beeler, N., and Wyss, T., "Validity of sports watches when estimating energy expenditure during running", *BMC Sports Sci Med Rehabil*, 9, 22, 2017
- [19] Python Software Foundation (2022), 'Python', (3.9 edn.: Guido van Rossum).
- [20] The NumPy development team (2022), 'NumPy', (1.21.0 edn.: Sebastian Berg, Ralf Gommers, Charles Harris, Stephan Hoyer, Melissa Weber Mendonça, Inessa Pawson, Matti Picus, Stéfan van der Walt, Eric Wieser).
- [21] McKinney, W. (2010), 'Data structures for statistical computing in Python', in Stefan van der Walt and Jarrod Millman (ed.), *Proceedings of the 9th Python in Science Conference*, 56-61.
- [22] The pandas development team (2020), 'pandas-dev/pandas: Pandas', (1.4.2 edn.: Zenodo).
- [23] The matplotlib development team (2022), 'Matplotlib', (3.5.2 edn.: John Hunter, Darren Dale, Eric Firing, Michael Droettboom).
- [24] Evans, J.D. (1996), *Straightforward statistics for the behavioral sciences*. (Brooks/Cole Publishing).
- [25] Jacobsen, B.K., Eggen, A.E., Mathiesen, E.B., Wilsgaard, T., and Njolstad, I., "Cohort profile: the Tromso Study", *Int J Epidemiol*, 41 (4), 961-7, 2012
- [26] Hopstock, L.A., Grimsgaard, S., Johansen, H., Kanstad, K., Wilsgaard, T., and Eggen, A.E., "The seventh survey of the Tromso Study (Tromso7) 2015-2016: study design, data collection, attendance, and prevalence of risk factors and disease in a multipurpose population-based health survey", *Scand J Public Health*, 14034948221092294, 2022

8 ACKNOWLEDGEMENT

The authors would like to thank the participants for sharing their data.

Maintaining Data Quality at the hospital department level

The data work of medical secretaries

Casper Knudsen¹ and Pernille Bertelsen¹

¹Aalborg University, Department of Planning, Aalborg, Denmark ckn@plan.aau.dk

Abstract

This article explores the collaborative work of maintaining data quality of a major health administrative database as it is carried out by medical secretaries in the role of ‘registration responsible medical secretaries’. The article reports on ongoing socio-technical study of local, on-the-ground data work in 5 Danish hospital departments. We argue the medical secretaries make important and skillful contributions to data quality at department level, including identifying and correcting errors, implementing changes to the coding practice, and maintenance of data input quality at the department level requiring a high level of context sensitivity.

Keywords

Data work, medical coding, health care, medical secretaries

1 INTRODUCTION

In this paper, we explore what we call the context sensitive collaborative work of data registration and data maintenance at five hospital departments in Denmark.

We frame this data work as being socio-technical as the electronic health records (EHR) and other health informatics technologies, together with rules and regulation and staff competencies for doing data work and clinical work, is highly situated and contextual. To understand the driving forces behind the registration and quality assurance work on patient trajectories done by the medical secretaries at the department level, we first introduce some of the important parameters guiding high health data quality in Denmark.

1.1 Data-driven health care and data work

As data has risen as the new “oil” of the information economy [1], along with widespread digitalization and the growth of data-intensive resourcing in healthcare [2, 3] the focus on achieving data-driven healthcare management is increasingly evident in both academic literature [4, 5] as well as the health care strategies worldwide, as well as in the strategies of Danish health authorities [6, 7]. As ambitions for data-driven potentials are rising with advances in data-powered technologies such as artificial intelligence (AI), so is the pressure on health care professionals to deliver high-quality data. Data is not simply pumped out of the ground as the oil metaphor suggests but requires work [8]. This focus is reflected in the growing body of literature concerned with exploring the work required for sustaining the data-driven health care systems [8–13] recognizing that large-scale data is a product of the work of many people and professions [14], leads to new professions [8] and changes the task portfolio and relationship of existing professions [14, 15]. It remains pertinent to investigate the on-the-ground data work of

healthcare professionals, both clinical and non-clinical such as the medical secretaries in focus of this article. We aim to contribute to this field by foregrounding some of the important and skilled data work happening backstage in Danish hospitals to maintain data quality of a major health database.

1.2 Digitalization and data-driven healthcare in the Danish context

From an international perspective, Denmark is often regarded as a frontrunner in digital health care and boasts a long history of successful national standardization efforts with numerous national health information technology strategies starting in 1995 building toward an increasingly coherent data infrastructure [16, 17]. Today this enables a wide range of data to be shared between GPs, specialists, pharmacies, municipal health services, and public hospitals, as well as with the patient through patient-facing infrastructures such as Sundhed.dk. Denmark has a well-registered population, a digitalized system of performance measurement in the health services [2], and a recent joint initiative of the Danish health authorities aim to make Denmark an international leader in the use of health data for treatment, research, public management, and innovation [7]. The aim is to further strengthen the access to and use of health data building onto, what is already referred to as the ‘epidemiologists dream’ [2, 17, 18] building onto ‘...a large network of population-based medical databases, which routinely collect high-quality data as a by-product of health care provision’ [17].

In this article, the primary context of the data quality work in focus is the data quality work of medical secretaries related to data on patient hospital encounters reported into the administrative medical database The Danish National Patient Registry (DNPR).

The 18th Scandinavian Conference on Health informatics, Tromsø, Norway, August 22-24, 2022. Organized by UiT The Arctic University of Norway. Conference Proceedings published by Linköping University Electronic Press at <https://doi.org/10.3384/ecp187>. © The Author(s). This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>

1.3 The Danish National Patient Registry

The DNPR which has been in place since 1977 [19] is one of the primary administrative health databases in the Danish context. DNPR collects all information on examinations and treatments in hospitals including e.g. primary and secondary discharge diagnoses, dates of contact, surgical procedures with dates from inpatient, outpatient as well as emergency encounters [17]. The Register provides data for different health registers, research, disease monitoring, and treatments [20] and is one of the most widely used registries for epidemiological research [17]. It is also a key database for performance and finance management at the hospital, regional, and state levels, as the financing model for Danish public hospitals includes both block grants, as well as activity-based subsidies (based on DRG), which utilize data reported to the DNPR for the settlements. One example of a political objective, which to a high degree shapes performance monitoring is the policy of ‘extended free hospital choice’ under which citizens may choose freely among all public or private hospitals if a region cannot deliver a diagnostic examination within a 30-day timeframe [17]. Creating and maintaining the necessary, correctly coded data to enable such performance measurement and fulfillment of patient rights is a key aspect of the data quality work performed by medical secretaries which is in focus in this article.

Data for the DNPR derives from registration in the patient administrative systems utilizing an integrated classification system, which combines entirely Danish classification systems based on international standards such as versions of the WHO's International Classification of Diseases ICD-10 and the Nordic operation classification systems (NCSP) [20, 21].

In 2019 a major upgrade (DNPR3) to the database was carried out implementing a shift to trajectory-oriented registration, bundling contacts, diagnoses, procedure, and result registration in relation to a clinical disease trajectory for each patient. Hence is possible for patients to have 2 or more ongoing trajectories at the same time, which is needed among patients with co-morbidity.

Data for monitoring interventions and outcomes as recommended in national clinical guidelines is an important objective for a comprehensive national data structure. Moving from the level of political intentions to the organizational level, as well as the concrete registration, and quality assurance of health care data, four levels and entities are of significance.

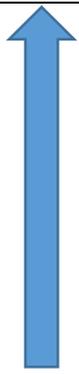
Organizational structure	Data registration body	
National	Danish Health Data Authority	
Regional	Business intelligence and/central IT departments	
Hospital	Registration unit and consultants	
Hospital Department	Registration responsible Physicians and Medical secretaries/health administrative coordinators	

Figure 1. Health data registration organization levels: National, regional, hospital, and department

At the hospital department level, a registration responsible physician and a registration responsible medical secretary (RRMS) is appointed by department management to make sure data work is done according to the national guidelines. At the hospital level, a registration responsible unit and registration responsible consultants are monitoring the aggregated data work and data compilation. At the regional level data from all departments and hospitals are managed, quality assessed, and used by the regional politicians and managers to monitor regional health-related activities, but also forwarded to the Danish Health Data Authority (DHDA) where health data from all 5 regions are quality assessed, aggregated, used for research, and visualized.

The specific responsibilities and guidelines for the execution of the RRMS role in relation to the wider registration organization are formalized in a region-wide role description for the RRMS. There are no formal certification requirements connected with the role of the RRMS beyond those required for medical secretaries in general and the specific organization of the role is decided locally depending on the department context.

1.4 Data work of medical secretaries

The study, from which this article emanates, particularly seeks to make visible the *data work* of medical secretaries in Danish hospitals (see [15]). Medical secretaries in Danish hospitals undertake a broad array of patient-oriented and health-administrative tasks as well as support to clinicians’ work [22, 23]. The profession of medical secretaries is actively working to acquire new tasks and remain relevant in modern digitalized health care and has been taking on new tasks following from an increasingly digitalized work setting [14, 24, 25]; one example being the work of data quality assurance of patient data registration as in focus here. In 2021 the education for medical secretary has been replaced with a one of ‘Health Administrative Coordinator’ placing a higher emphasis on their role in the administration of a more complex, digitalized healthcare setting.

Since the profession began gaining in numbers in the Danish hospital context in the 1950s medical secretaries were closely tied to the doctors and thus the clinical context. They have still today maintained a decentral department distribution where they physically and organizationally are located close to clinical practice. Several studies emphasize this as a key factor, which enables them to “often act as the organizational ‘glue’ or connecting thread between other professional groups at the hospital’ [26] and that “secretaries are deeply involved in diagnostic work through the eligible administration of patients in the collaborative electronic information systems’ [27] and thus are positioned at the intersection of clinical and administrative work (ibid).

Like that of other clerical workers, the work of medical secretaries has often been relegated to the background [23], [28], regarded as mere routine work rather than knowledge work and thus targeted for automation [29, 30]. This article seeks to contribute to the growing body of literature foregrounding the work behind the high-quality data powering the ‘epidemiologists’ dream’ and AI, described above, by investigating the work of medical secretaries in

Danish hospital departments as they work to maintain the quality of data for a major patient trajectory database, the DNPR.

Entering into this, we ask the following research question:

Which practices and skills are employed by registration responsible medical secretaries for maintaining the quality of patient trajectory registration data at the department level?

2 METHODS

This paper reports insights from ethnographic fieldwork carried out by the first author from February to March 2022, at five different hospital departments in the region of Northern Jutland in Denmark. The region spans from small, local hospitals to a major university hospital. This study includes departments in both regional and the central university hospital, and departments ranging from 7 to 30 secretaries. The departments included were chosen based on prior knowledge from an ongoing study of the data work of medical secretaries in Danish hospitals (see [15]) as well as snowballing to identify RRMSs deemed highly skilled by their peers. Hence, the medical secretaries who were interviewed are all experienced in the use of the ICT systems, which was a crucial element when enquiring into their skills and competencies as professionals.

Five RRMS were interviewed for each 30-60 minutes after which a focused shadowing was carried out by engaging in a form of apprenticeship peer training session. This was carried out by instructing the RRMS to introduce the tasks, main systems, etc to the observer, the first author as if he was a new apprentice creating a space for both the RRMC to present their expertise in context and for the interviewer to enquire into specific elements of practice. The shadowing happened at the desk of the RRMC being a natural workplace connected with the role. Three of the sessions lasted 2 hours, two lasted 6 hours split over two days.

In addition, one interview (45 minutes) was made with a member of the Registration Unit as well as one interview (60 minutes) with a head of medical secretaries in one of the departments.

All interviews were transcribed in full. During observation, handwritten notes were jotted down and written into full notes immediately after supported by focused transcription of the sound recording from the observation sequence.

Analysis of the data largely followed the process of Grounded Theory [31] constructing themes by shifting between open, horizontal coding, and vertical, consolidating coding and subsequent focused coding. The presentation of data and insights in this paper is the product of the discussions between the two authors.

3 ANALYSIS

In the following, we present characteristics of the work of quality assurance in the work of the RRMS as they emerged through the fieldwork and subsequential analysis. Initially, we touch on the necessary competencies or mindset required for the role as RRMS (3.1) after which we outline the process and tools related to the work on error lists. Finally, we outline two strategies applied in the work on error lists, namely ‘detective work’ (3.3) and ‘data quality educator’ (3.4), and show how these rely on a context

sensitivity based on the intertwined situated skill and knowledge from both the administrative, clinical, and organizational context of the department.

3.1 The registration mindset – being RRMS

As mentioned earlier, there are no formal requirements or certifications tied to the role of the RRMS. Two of the RRMS in this study had taken specific courses oriented toward health data registration (e.g. data processing and controlling), but obtaining the role as RRMS rather seems to be based on experience and a certain type of mindset. As described by a head medical secretary:

*“To be a registration responsible really is demanding. It simply requires a huge insight into the hospital, and it also requires a lot of experience. *RRMS* is really - she is mega experienced. And likewise - the thing about going in-depth, wondering and the thing about things simply having to be right - you have to have a lot of... a pride, a professional pride about it. So, she's very good at it. You need to have someone like her in the departments. Also, because it gets... complicated sometimes.” (Head of medical secretaries)*

When talking about the skills needed for performing the work, it was unanimously agreed that it took a certain type of personality. One secretary refers to herself as having a ‘registration brain’ (RRMS 5), while others jokingly suggested having ‘a bit of OCD’ and finding joy in cleaning up and ‘getting things completely right’. During the in-situ interviews the RRMSs would at times enthusiastically celebrate upon opening a list, which was empty or had fewer errors than previous or expected such as the quote below, in which an RRMS opened an error list which she had been particularly attentive to for a period, but had not checked for several days:

“Wow, now have look! It's empty. Oh my God, this is the first time in a long time – that is amazing!” (RRMS 2)

While most of the RRMS explained how they got the role somewhat coincidentally rather than as a deliberate career choice, they stress the particular attitudes – or the ‘registration brain’ – as a key characteristic to thrive in the role.

3.2 The work on error lists

The main part of the work of the RRMS is centered around quality and error lists, which are “a tool for medical secretaries to perform ongoing quality assurance of registrations” [32].

The error lists, which are uniform across all hospitals in the region, are automated data extracts supplied by the regional Business Intelligence Unit containing logical errors from the clinical registration of patient trajectories. New extracts are published once every 24 hours and accessible through a folder in a shared drive in the form of lists in PDF and Excel files. The error lists are sorted in thematic subfolders and the files are separated by error type (eg ‘Potentially missing clinical decision’).

Though, most of the lists are accessible through different avenues (eg. the patient-administrative system, a business intelligence solution, and websites from the national health data authority) the lists are accessed through a folder in a shared drive, referred to as the ‘department folder’.

Most of the lists are permanent, while others are in place to enable specific events; in the case of this fieldwork one of these events was preparation for migration to a new EHR.

The role description, see [32], specifies the frequency to which each list should be reviewed, however in practice these are not followed. Rather the pace of the routine relies on the size of the department, concrete registration deadlines related to certain lists, and the overall activity in the department.

In the observed departments, the time spent on the role varies depending on the size and context of the department. In one large department with a coordinating role for other departments in a regional hospital, the role was managed by two persons as the primary function. In a small department, the role was assigned to the head medical secretary, who assessed that she spent a few hours per week on the work on error lists specifically.

A central aspect of the work on monitoring and clearing errors off the error lists requires the RRMS to maintain an overview of the status of the lists over time. Due to the system of lists being updated daily (around midnight), there is a delay in the feedback as to whether the corrections made to the trajectory data results are correct, as this will only be visible on the next day's error list. Additionally, one registration issue often shows up on several error lists, meaning that the correction of one issue will often solve issues figuring on multiple lists. Analyzing these connections requires insights both into the logic of the registration system as well as the intricacies of the patient administrative systems in general. Failing to foresee the connections can lead to the RRMS 'searching for an error, which isn't there anymore' (RRMS 4).

3.3 Data detective work – investigating and correcting errors

A majority of the errors on the lists are, despite being time-consuming due to being many, simple in the sense that they require little analysis and are often uniform and easily corrected, eg by adding or changing a single code. At the level of the individual complex error, several of the RR secretaries in this study refer to this as being 'detective work' or as taking a 'Sherlock Holmes' approach. This is to describe the process of going from the decontextualized error on the error list, eg 'Missing A-Diagnosis', to investigating through the available patient data across different systems (eg booking system, EHR, patient-administrative system, physician task lists) and different types of data (eg medical notes such as s or admission and discharge letters, test answers, bookings) and to identify the cause of the error and what needs to be done by whom to correct it.

'It's a lot about tracing down what has been going on' (RRMS 1)

The 'detective work' description illustrates the ability to identify the cause of the errors on the list by combining a deep contextual understanding of the IT systems in use, the registration regime, the role of registration in department/hospital administration and the organization and medical specialty.

As described above, the work on errors often spans a period and involves awaiting other people or processes to contribute:

'RRMS is working on a case on the error list of cases in which inquiries have outrun the 30-day period. The RRMS is 'trying to chase down the clinical decision', she says, to establish whether a clinical decision has been made by meticulously browsing through the relevant systems. In the notes module, she identifies a medical note (a sound file) pending transcription, which fits the timespan to potentially contain a clinical decision made within the 30-day period. Hence, not able to assign the correct code, the RRMS adds a 'priority' mark to accelerate the transcription process. In the patient-administrative system, she notes down 'Note #date#' as an indication for herself, that transcription is awaiting. The error remains on the list and will appear again after 24 hours until solved.' [Field notes]

In another example an error concerns a referral, which has been simultaneously canceled/closed and referred to inquiry and therefore conflicts with the need to assign the start date of a patient's 30-day inquiry period:

'To solve the issue, she has to 'dig through earlier referrals', which she opens in Clinical (EHR), where she – as opposed to the patient administrative system – can read through the visitation information; who did what when in the visitation history. By reading through history, she can conclude that the cancellation must have been added by mistake by another department. She sends the case to the department with a note on suggested processing.' (Field notes)

In this case, the RRMS show a high level of interactional expertise in the clinical field they are in in addition to a deep knowledge of the related coding regime, which is required to be able to effectively navigate complex patient trajectories and construct these in the form of correct registration. This combination of knowledge has been particularly necessary after the introduction of the DNPR3 and the related trajectory-oriented coding regime as many physicians and regular medical secretaries still tend to fall into using the former activity-oriented coding regime.

The coding expertise is also visible in how, the RRMSs engage in *refining* the coding of trajectories:

*'*RRMS* is showing the workflow of correcting an error on the 'Missing A-Diagnosis' list. She points out that most (non-RR) secretaries would probably just see that 'this is ambulant' and see that it is a control visit after an emergency room visit. She, however, likes to go back in the patient history (in the patient administrative system) to see earlier activities in the trajectories of the patient to be able to connect the diagnosis to the correct place in the trajectory as well as add the right fracture diagnosis as a 'plus-diagnosis'' (From field notes)*

This speaks to the 'professional pride' or urge to 'get things right' mentioned earlier.

A considerable number of errors, the RRMSs agree, stem from poor integrations between the booking system and the patient-administrative system, where registrations do not properly synchronize between the systems. An example of this is that it is possible to create a booking in the booking

system without the contact being registered to the patient trajectory (in the PAS), which prompts an error. To identify how to fix the error the RRMS has ‘to go hunting’, as one secretary put it, in patient data (physicians’ notes, discharge letters, etc.) across different systems to establish what kind of contact, if any, took place. One example is identifying in the free text of a note from the physician that the patient will be contacted by a nurse.

‘That is a really, I mean, that’s really annoying work to clean up, because you have to rewind the whole thing to see if ‘is this one supposed to be there or is it not and if not, when and to which contact is it supposed to be assigned?’ (RRMS, 1)

In sum, the RRMSs – as Sherlock Holmes in a crime scene – navigate the complex contexts of patient trajectory data from the points of single, decontextualized logical errors in the registration as they appear on error lists through a range of skills and knowledge situated in both clinical, administrative and organizational aspects of the department context.

3.4 Data quality educator

In the observation and interviews of this study, the focus on supporting the ability of the medical secretaries in the department to deliver high-quality registrations unsurprisingly comes out very clearly as a key concern for all the RRMS.

The work of building the registration capacity of the department’s secretaries covers aspects of identifying and correcting patterns of error in the registration at the group and individual levels as well as contextualizing and communicating changes to the registration practice.

One of the formalized responsibilities of the RRMS is to “Ensure that news regarding registration is known and manageable by medical secretaries” [32]. Through the Registration Unit, the RRMSs regularly receive notifications (mostly by e-mail) of changes to the registration practice. The RRMS then decides the necessary steps to roll out the new practice and whether it necessitates any changes to registration workflows. Depending on the type and complexity of the change, the size and organization of the department, and other contextual factors, the RRMS produces, initiates, and monitors the implementation of these changes at the department level. An example of a complex shift is the upgrade from DNPR2 to DNPR3, mentioned earlier, which constituted a major shift in the registration logic for secretaries and clinicians alike. Though this upgrade was rolled out in 2019, the RRMS often mention dealing with residue from the former activity-oriented registration logic (eg. Examinations that are logged without connection to an active trajectory).

The RRMSs in this study emphasize the identification and acting on patterns of error in the registration of the department is an integral part of managing data quality at the department level. The RRMSs describe how their choice of action regarding an error – whether to fix the error or send it back to the secretary who made the error – is a balancing of when to step into the role of the educator vs simply correcting the error. This balancing is based on whether the error is systematic and part of a pattern either with the individual secretary or the group, as well as an

assessment of the cause of the error; e.g. whether it is due to a wrongful understanding of the correct registration practice, based on the registration systems (such as issues from poor integrations) or organizational (eg delays in transcription, illness or lack of personnel in own or other departments, etc).

They stress how considerations of upholding relations with other secretaries and physicians are central to their decision of when to choose the educator role. In a recent paper, Jensen [33] point out how diplomacy skills, ie “creat(ing) and uphold(ing) good relations (...) is a fundamental and under-recognized aspect of transplant data practices” (p 9). This echoes the considerations of the RRMS as they balance their strategies between educating the department and knowing that what they are asking often is often regarded as annoying.

Thus, to enforce the continuously changing registration practice and maintain department registration quality, the RRMSs apply a deep contextual knowledge of organizational as well as administrative domains. In doing so a key ability of the RRMS is to identify patterns in registration errors at the department level and successfully use this in building the registration capacity of secretaries and clinicians in the department while balancing their relations in the collaboration with their colleagues whose registration they are correcting.

4 DISCUSSION

As with most other clerical functions, a major part of the work of medical secretaries happens ‘backstage’, in the back office behind a desk, looking at a screen, while physicians and nurses are much more visible. As discussed in several classical texts within Computer-Supported Cooperative Work (CSCW) and related fields [28, 30] the invisibility of clerical workers and considering their work as routine work tends to lead to ambitions to automate them away.

As Holten-Møller [25] foresees, while many discussions of automation are ongoing in the context of data and registration in health care, “the future of AI and automation in hospitals seems to have little or no place for clerical work’ (ibid). What we are seeing might be that “data work may simply be shifting hands’ (ibid).

In Denmark, we recently saw with the implementation of the EPIC EHR system in two Danish regions in 2017, how the business case of the new EHR systems were been partly financed through a planned redundancy of medical secretaries [34]. Here the ambitions followed the general trend of realizing a double aim of cutting costs and achieving real-time data by authorizing physicians to the coding of patient trajectories at the bedside [4]. The regions in question, however, ended up re-hiring most of the medical secretaries to work on correcting error lists to maintain data quality as the quality of registrations plummeted and frustrations of the physicians grew [34]. The work didn’t disappear – it merely shifted hands and took a new form.

The role of the RRMS discussed in this article illustrate data work needed to maintain data quality and how a group of non-clinical health care professionals does this work by applying situational knowledge and skills from the clinical, administrative, and organizational context of a hospital

department. As a potential future role in a context of increasing automation of coding work as new generation EHRs are being implemented, shedding light on the qualities of on-the-ground data quality work becomes crucial in understanding how non-clinical workers in health care with a deep knowledge of the departmental context contribute to realizing the increasing ambitions of data-driven health care.

5 CONCLUSION

This paper has highlighted how medical secretaries as non-clinical workers in Danish hospitals, contribute to data quality of an major national health database by correcting data registration errors made by clinicians, fellow medical secretaries, and/or EHR systems at the hospital departments. Their work requires highly situated skills and competencies within data registration practices, on-the-ground knowledge, and an interest in close follow-up and implementation of new data registration guidelines for the health care organization.

We have shown how the work of identifying and correcting errors, implementing changes to the coding practice, and maintenance of data input quality at the department level is carried out with a high level of context sensitivity through strategies as ‘detectives’ (analysing and correcting errors) and ‘educators’ (identifying and acting on error patterns in department coding practice) while simultaneously balancing the diplomacy of correcting the errors of colleagues to uphold a good working relationship.

As a profession, medical secretaries contribute at regional and national levels to quality health care data and thus to meet the goals in the national strategies for data-informed patient trajectory, treatment, research, public management, and innovation.

6 REFERENCES

- [1] K. van Es og M. T. Schäfer, *The Datafied Society. Studying Culture through Data*. Amsterdam University Press, 2017. doi: 10.5117/9789462981362.
- [2] K. Hoeyer, “Denmark at a Crossroad? Intensified Data Sourcing in a Research Radical Country”, Cham: Springer International Publishing, 2016, s. 73–93. doi: 10.1007/978-3-319-33525-4_4.
- [3] L. F. Hogle, “Data-intensive resourcing in healthcare”, *BioSocieties*, bd. 11, nr. 3, s. 372–393, aug. 2016, doi: 10.1057/s41292-016-0004-5.
- [4] R. S. Evans, “Electronic Health Records: Then, Now, and in the Future”, *Yearbook of medical informatics*, bd. 25, nr. S 01, s. S48–S61, 2016, doi: 10.15265/IYS-2016-s006.
- [5] L. B. Madsen, *Data-driven healthcare: how analytics and BI are transforming the industry*. John Wiley and Sons, 2014. [Online]. Tilgængelig hos: <http://cds.cern.ch/record/2010643>
- [6] S. ældreministeriet, “Sundhed i fremtiden. Ansvarlig brug af data”, 2018. [Online]. Tilgængelig hos: <https://sum.dk/publikationer/2018/juni/sundhed-i-fremtiden>
- [7] Sundhedsdatastyrelsen, “Vision - strategisk samarbejde for bedre brug af sundhedsdata”. Sundhedsdatastyrelsen, 2021. [Online]. Tilgængelig hos: https://sundhedsdatastyrelsen.dk/-/media/sds/filer/strategi-og-projekter/bedre-brug-af-sundhedsdata/vision_bede_brug_sundhedsdata.pdf?la=da
- [8] C. Bossen, K. H. Pine, F. Cabitza, G. Ellingsen, og E. M. Piras, “Data work in healthcare: An Introduction”, *Health informatics journal*, bd. 25, nr. 3, s. 465–474, 2019, doi: 10.1177/1460458219864730.
- [9] C. Bjørnstad og G. Ellingsen, “Data work: A condition for integrations in health care”, *Health informatics journal*, bd. 25, nr. 3, s. 526–535, 2019, doi: 10.1177/1460458219833114.
- [10] C. Bossen, K. Pine, G. Ellingsen, og F. Cabitza, “Data-work in Healthcare: The New Work Ecologies of Healthcare Infrastructures”, i *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion*, New York, NY, USA, feb. 2016, s. 509–514. doi: 10.1145/2818052.2855505.
- [11] C. Bossen, Y. Chen, og K. H. Pine, “The emergence of new data work occupations in healthcare: The case of medical scribes”, *International journal of medical informatics (Shannon, Ireland)*, bd. 123, s. 76–83, 2019, doi: 10.1016/j.ijmedinf.2019.01.001.
- [12] A. M. Pedersen og C. Bossen, “Data Work in Healthcare: An Ethnography of a BI Unit”, 2021, doi: 10.18420/ihc2021_006.
- [13] E. Håland og L. Melby, “Coding for quality? Accountability work in standardised cancer patient pathways (CPPs)”, *Health (London, England: 1997)*, s. 13634593211013882, apr. 2021, doi: 10.1177/13634593211013882.
- [14] N. M. Holten, K. H. Pine, T. R. Nielsen, og C. Bossen, “Who does the work of data?”, *Interactions*, 2020, doi: 10.1145/3386389.
- [15] C. Knudsen og P. S. Bertelsen, “Data Work in Health Care: The Case of Medical Secretaries”, *Studies in health technology and informatics*, bd. 286, s. 60–64, nov. 2021, doi: 10.3233/SHTI210638.
- [16] C. Nøhr, S. Vingtoft, S. K. Andersen, og P. Bertelsen, *Digitalisering af det danske sundhedsvæsen: Øjenvidneberetninger fra nøgleaktører*. Syddansk Universitetsforlag: Syddansk Universitetsforlag, 2019. [Online]. Tilgængelig hos: <https://portal.findresearcher.sdu.dk/da/publications/40fa7044-f1e9-4865-999d-05bfc974262e>
- [17] M. Schmidt *m.fl.*, “<p>The Danish health care system and epidemiological research: from health care contacts to database records</p>”, *CLEP*, bd. 11, s. 563–591, jul. 2019, doi: 10.2147/CLEP.S179083.
- [18] L. Frank, “The Epidemiologist’s Dream: Denmark”, *Science*, bd. 301, nr. 5630, s. 163–163, jul. 2003, doi: 10.1126/science.301.5630.163.

- [19] M. Nguyen-Nielsen, E. Svensson, I. Vogel, V. Ehrenstein, og L. Sunde, “Existing data sources for clinical epidemiology: Danish registries for studies of medical genetic diseases”, *CLEP*, bd. 5, nr. 1, s. 249–262, aug. 2013, doi: 10.2147/CLEP.S45228.
- [20] H. H. Bjerregaard og L. Hejndorf, *Styr på sundhedsdata - fundamentet for et datadrevet sundhedsvæsen*. Komponent - Kommunernes Udviklingscenter, 2022.
- [21] Sundhedsdatastyrelsen, “SKS Klassifikationer - Sundhedsdatastyrelsen”. <https://sundhedsdatastyrelsen.dk/da/rammer-og-retningslinjer/om-klassifikationer/sks-klassifikationer> (set 19. maj 2022).
- [22] P. Bertelsen, *Hvad laver lægesekretærene, når de ikke er sekretærer for lægerne?* Aalborg: Virtuelt Center for Sundhedsinformatik, 2005.
- [23] P. Bertelsen, “Vil elektroniske patientjournaler overflødiggøre lægesekretærer på de danske sygehuse?”, *Tidsskrift for Arbejdsliv*, bd. 9, nr. 3, 2007, doi: 10.7146/tfa.v9i3.108615.
- [24] C. Bossen, L. G. Jensen, og F. W. Udsen, “Boundary-object trimming: On the invisibility of medical secretaries’ care of records in healthcare infrastructures”, *Computer Supported Cooperative Work (CSCW)*, bd. 23, nr. 1, s. 75–110, 2014.
- [25] N. Møller, “The future of clerical work is precarious”, *Interactions (New York, N.Y.)*, bd. 25, nr. 4, s. 75–77, 27. juni 2018. doi: 10.1145/3231028.
- [26] P. Bertelsen og C. Nøhr, “The work practice of medical secretaries and the implementation of electronic health records in Denmark.”, *Health Information Management Journal*, bd. 34, nr. 4, 2005.
- [27] N. L. H. Møller og S. Vikkelsø, “The Clinical Work of Secretaries: Exploring the Intersection of Administrative and Clinical Work in the Diagnosing Process”, London: Springer London, 2012, s. 33–47. doi: 10.1007/978-1-4471-4093-1_3.
- [28] S. Star og A. Strauss, “Layers of Silence, Arenas of Voice: The Ecology of Visible and Invisible Work”, *Computer supported cooperative work*, bd. 8, nr. 1, s. 9–30, mar. 1999, doi: 10.1023/A:1008651105359.
- [29] L. Suchman, “Making work visible”, *Communications of the ACM*, bd. 38, nr. 9, s. 56–64, 1995.
- [30] L. Suchman, “Making a case: knowledge and routine work in document production.”, 2000.
- [31] K. Charmaz, *Constructing grounded theory*, 2nd edition. London ; Thousand Oaks, Calif: Sage, 2014.
- [32] Registreringsorganisationen i Region Nordjylland, “Velkommen i Registreringsorganisationen i Region Nordjylland”. Region Nordjylland, 20. september 2018.
- [33] A. M. B. Jensen, “Making it happen: data practices and the power of diplomacy among Danish organ transplant coordinators”, *BioSocieties*, 2022, doi: 10.1057/s41292-021-00267-z.
- [34] N. Bentzon, J. Rosenberg, J. Bansler, J. Søgaard, B. Dahlerup, og B. Haarder, *Destruktiv digitalisering :*

en debatbog om Sundhedsplatformen 2016-2021, 1. udg. Kbh: FADL, 2021.

7 ACKNOWLEDGEMENTS

This research is supported by the Independent Research Fund Denmark.

Evaluation of LIME and SHAP in Explaining Automatic ICD-10 Classifications of Swedish Gastrointestinal Discharge Summaries

Alexander Dolk¹, Hjalmar Davidsen¹, Hercules Dalianis^{1,2}, Thomas Vakili¹

¹Department of Computer and Systems Sciences (DSV) Stockholm University, Kista, Sweden

²Norwegian Centre for E-health Research, Tromsø, Norway

alexander.dolk@hotmail.com, hjalmar.davidsen@gmail.com,

hercules.dalianis@ehealthresearch.no, thomas.vakili@dsv.su.se

Abstract

A computer-assisted coding tool could alleviate the burden on medical staff to assign ICD diagnosis codes to discharge summaries by utilising deep learning models to generate recommendations. However, the opaque nature of deep learning models makes it hard for humans to trust them. In this study, the explainable AI models LIME and SHAP have been applied to the clinical language model SweDeClin-BERT to explain ICD-10 codes assigned to Swedish gastrointestinal discharge summaries. The explanations have been evaluated by eight medical experts, showing a statistically higher significant difference in explainable performance for SHAP compared to LIME.

Keywords

ICD-10 diagnosis code, natural language processing, eXplainable AI, multi-label text classification.

1 INTRODUCTION

The International Classification of Diseases (ICD) has been used globally for over a century to classify information in patient records [1]. Using the ICD coding system, reported conditions in patient records are converted into medical codes. The coded patient records are then used for administrative and research purposes. The ICD coding system has been revised multiple times. Currently, the tenth version (ICD-10) is the most widely used edition.

The ICD framework is important as it is a common way of recording diseases, enabling health practitioners within and between countries to share their data. The ICD promotes the compilation and storage of medical data for decision-making and analysis. Currently, the ICD is used by all member states of the World Health Organization and has been translated into 43 languages [2].

Human coders are prone to making errors when assigning ICD-10 codes. For example, one study [3] found that 20 percent of the main diagnoses in Swedish discharge summaries were incorrectly coded.

A computer-assisted coding (CAC) tool for ICD-10 coding that utilises artificial intelligence (AI) can give recommendations to physicians on possible ICD-10 codes for a discharge summary. In addition, it can also validate already assigned ICD-10 diagnosis codes. The use of such tools has the potential to increase the efficiency of the health care system.

The development of a CAC-tool for ICD-10 coding is highly needed in the medical field. This is part of the ClinCode project, at the Norwegian Centre for E-health Research [4].

The use of artificial intelligence (AI) and machine learning has in recent years become widespread. Novel strategies like deep learning (DL) models have demonstrated great results for a multitude of regression and natural language processing (NLP) tasks [5][6]. Nonetheless, DL models are opaque in nature. It is often impossible for humans to understand why DL models make particular predictions. This is an issue as it makes it hard for humans to trust the predictions of DL models [7]. To remedy this problem, the field of explainable artificial intelligence (XAI) has recently emerged [8]. The purpose of XAI is to provide methods that can explain the prediction of AI models.

In this article, the XAI models Local Interpretable Model-agnostic Explanations (LIME) [9], and SHapley Additive exPlanations (SHAP) [10] have been applied post hoc to the classification model **Swedish De-identified Clinical BERT** (SweDeClin-BERT) [11], to explain ICD-10 classifications of Swedish gastrointestinal discharge summaries. SweDeClin-BERT is a derivation of the model **KB-BERT**, a **Bidirectional Encoder Representations from Transformers** (BERT) model [12] developed by the National Library of Sweden (KB). The explanations by LIME and SHAP have then been evaluated by medical doctors and ICD coding experts through a questionnaire, resulting in a comparison of the model's explainable performance. Specifically, the models have been compared for the factors of user trust, explanation satisfaction and perceived usability.

While there have been previous evaluations of LIME and SHAP with medical data, only one peer-reviewed study has been found which applies SHAP with Swedish medical data [13]. The domain dependence of data when evaluating

The 18th Scandinavian Conference on Health informatics, Tromsø, Norway, August 22-24, 2022. Organized by UiT The Arctic University of Norway. Conference Proceedings published by Linköping University Electronic Press at <https://doi.org/10.3384/ecp187>. © The Author(s). This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>

AI models motivates the need for this article, as it uses medical data labelled with ICD-10 codes in Swedish. To the extent of our knowledge, this paper is the first that evaluates LIME and SHAP on Swedish medical data labelled with ICD-10 codes. From a Human-Computer Interaction (HCI) perspective, it is also valuable to get feedback from respondents of the population that will be future users of a CAC-tool for ICD-10 coding. It is valuable as the performance of XAI models is highly subjective depending on the user group.

2 RELATED RESEARCH

This article builds upon previous research [14], which has evaluated the deep learning model KB-BERT against a range of baseline models for the task of multi-labelling Swedish gastrointestinal discharge summaries with ICD-codes. The results of that article showed that a fine-tuned version of KB-BERT achieved an F_1 -micro of 0.80 and F_1 -macro of 0.58 on grouped ICD-10 codes. However, when tested on the full 263 ICD codes, the KB-BERT model underperformed against the baseline models. In the article, it was recommended to further study the possibility of including explainability mechanisms in a CAC-tool for ICD-10 diagnosis coding, which this article aims to do.

Previous research [15] has evaluated the model eXtreme Gradient Boosting (XGBoost) against the models Random Forest (RF) and Support Vector Machines (SVM) in the ability to predict sarcasm in natural text using punchline utterance and context. In the article, it was found that XGBoost achieved a higher F_1 -score than RF and SVM when using only utterance as well as when using both utterance and context. In the study, LIME and SHAP were used to give explanations for individual predictions. By using LIME and SHAP, the study could show that the models can explain that word importance is vital to correctly predict sarcasm in dialogues.

In another article [16], a user study was performed to evaluate the performance of an XAI system called HealthXAI. HealthXAI had the purpose of predicting cognitive decline from early symptoms. In the study, participants performed evaluations on the three factors of User Trust and Reliance (UTR), Explanation Satisfaction (ES) and Human-Machine Task Performance (HMTTP).

Eight neurologists (clinicians) participated in the study who were well versed with technology and experts in cognitive decline. The explanations provided by HealthXAI were evaluated through a questionnaire using Likert scale answers. The study showed that HealthXAI with explanations performed better for all three factors than without explanations. Furthermore, the participants were very positive toward the explanations by HealthXAI for all three metrics.

In a related study [17], a proposed model aimed at explaining local multi-label classifications in NLP was compared to LIME and CXPlain. The models were evaluated through a user study and found that users could complete tasks faster with recommendations from the proposed model than with LIME. Additionally, Hamming score was used to evaluate the models, which is the fraction of correctly predicted labels out of all labels. On one dataset, LIME achieved 91%, the proposed model 90.75%

and CXPlain 81.67%. On another dataset, LIME achieved 66.08%, the proposed model 65.23% and CXPlain 52.95%.

A previous study [13] compared an attention-based Recurrent Neural Network (RNN) to a basic RNN on which SHAP has been applied in the ability to give local and global explanations of Adverse Drug Events (ADE) in Swedish medical records. In the study, users assessed the explanations by the attention-based RNN and SHAP. Also, the Top-k Jaccard Index was used to assess the explanations by comparing the index of the models to those of medical experts. The medical experts in the study thought that SHAP gave more efficient explanations to show how features additively contribute to predictions. As such, SHAP was deemed most suitable for real-time scenarios where efficiency is important.

As apparent by the related research described above, there is previous research that has investigated the explainable performance of LIME and SHAP. However, there have been no comparisons of LIME and SHAP in their ability to explain ICD-10 classifications of Swedish gastrointestinal discharge summaries.

3 METHODOLOGY

3.1 Hypotheses

As stated, the aim of this study is to compare LIME and SHAP for the factors of user trust, explanation satisfaction and perceived usability. We hypothesise a difference between LIME and SHAP in terms of the three aforementioned factors for explaining ICD-10 classifications of Swedish gastrointestinal discharge summaries made by SweDeClin-BERT.

3.2 Selection of XAI Approaches

There is a multitude of XAI models that could be evaluated in explaining ICD-10 classifications. In a recent systematic review [8], 137 papers proposing XAI models were reviewed. However, not all the models can be considered for our study. For this study, the XAI models need to be local and post hoc. This means that they can explain individual classifications and be applied to existing prediction models respectively [18][19]. These two factors are necessary for an XAI model to be implemented in a CAC-tool for ICD-10 coding. The XAI must be able to explain individual classifications of ICD-10 codes and need to have the versatility of being applied to powerful classification models like BERT. This reduces the 137 models reviewed in [8] to 51 models. Further delimitation has been made by ranking the 51 remaining models by citations on Google Scholar. The model Gradient-weighted Class Activation Mapping (Grad-CAM) with 8,134 citations can be disregarded as it is intended for computer vision. This leaves LIME and SHAP are the most relevant models, with 8,430 and 6,432 citations, respectively, as of 2022-04-01. We use the number of citations to judge which models are most used and use this as a proxy for relevance.

3.3 Collection of Data for ICD-10 Classification

The data used in this study consist of Swedish gastrointestinal discharge summaries contained in the second version of the Stockholm EPR Gastro ICD-10

Corpus (ICD-10 Corpus)¹. The ICD-10 Corpus is part of the research infrastructure Health Bank at DSV/Stockholm University. The Health Bank² contains Swedish patient records from over 2 million patients from Karolinska University Hospital, encompassing the years 2006 to 2014 [20]. All data in our study have been de-identified and hereafter called Stockholm EPR Gastro ICD-10 Pseudo Corpus or ICD-10 Pseudo Corpus for short.

The ICD-10 Pseudo Corpus consists of 6,014 gastrointestinal discharge summaries. The dataset has a heavily imbalanced distribution of ICD-10 codes and this can be seen in Figure 1. To have a meaningful evaluation of LIME and SHAP, the predictions of ICD-10 codes being explained need to be made from a high-quality classifier. To mitigate the negative impact of the imbalanced data, a subset selection of discharge summaries has been made that have at least one of 18 selected ICD-10 codes out of the 263 original ones. While this approach might not be appropriate for an end-product application, it allows us to simulate a scenario where LIME and SHAP are applied to a model that has learned the underlying patterns of the data. This enables LIME and SHAP to also learn the underlying patterns of the data as they try to approximate the prediction function of the classification model. The subset selection of discharge summaries means that there are at least 100 discharge summaries for each of the 18 selected ICD-10 codes. The other discharge summaries, which do not have one of the 18 ICD-10 codes, have been removed from the subset. The subset consists of 3,636 samples. The distribution of ICD-10 codes in the subset and the ICD-10 codes are visible in Figure 2.

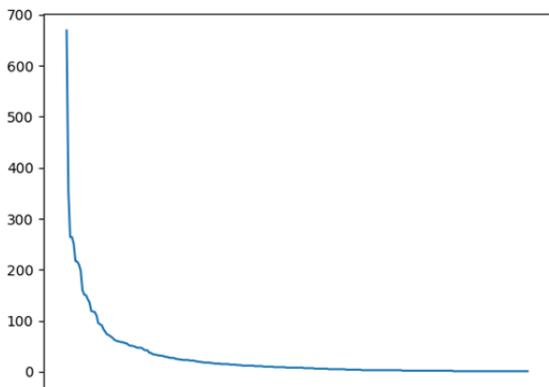


Figure 1. ICD-10 code distribution in the original dataset of ICD-10 Pseudo corpus. There are in total 263 unique ICD-10 codes on the X-axis.

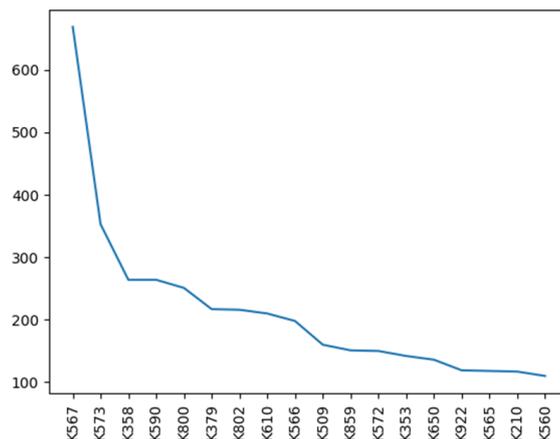


Figure 2. ICD-10 code distribution in the subset of 18 ICD-10 diagnosis codes.

As is typical in machine learning [21], the data has been split into a training set, validation set and test set. The training set consists of 2,617 samples (72%), the validation set of 655 samples (18%), and the test set of 364 samples (10%).

3.4 Implementation of SweDeClin-BERT

SweDeClin-BERT [11] has been used as the classification model in this study, a model based on the general Swedish KB-BERT [22] that has been further pre-trained on pseudonymised clinical text from the Health Bank. Pseudonymised means sensitive personal information has been identified in the text and replaced with surrogate values. SweDeClin-BERT is, therefore, a privacy preserving clinical language model for Swedish.

In our study SweDeClin-BERT has been fine-tuned for the downstream task of labelling discharge summaries with ICD-10 codes, using the aforementioned dataset of 3,636 discharge summaries. The fine-tuning has been done with the hyperparameters described in Table 1. To determine the number of epochs to train for, 5-fold cross-validation has been performed. The validation loss can be seen in Figure 3, resulting in the decision to fine-tune the model for nine epochs.

Hyperparameter Name	Value
Batch size	2
Learning rate	2e-5
Gradient accumulation steps	16
Number of warmup steps	155
Weight decay	0.01

Table 1. Hyperparameters for fine-tuning SweDeClin-BERT

¹ This research has been approved by the Regional Ethical Review Board in Stockholm under permission no. 2007/1625-31/5.

² Health Bank, <http://dsv.su.se/healthbank>

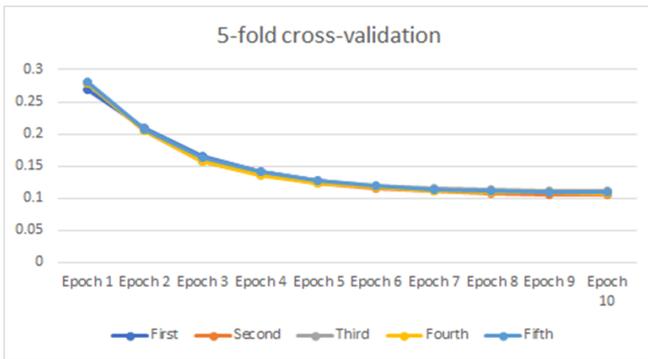


Figure 3. Validation loss during the five folds of the 5-fold cross validation

All hyperparameters except the number of epochs are based upon earlier studies [12]. Since the goal of this study is not to create an optimal classification model, no hyperparameter optimisation has been performed. The code used to implement the model can be found in our Github repository³.

3.5 Evaluation Results of SweDeClin-BERT

As SweDeClin-BERT returns prediction probabilities for each of the labels to a discharge summary, labelling has been considered true when having a prediction probability of 0.5 or higher. All evaluation results have been rounded to two decimals. The evaluation of SweDeClin-BERT on the test set has returned a mean accuracy of 0.97, as well as the mean results in Table 2. Micro and macro averaging captures different things. Micro averaging gives equal weight to every sample in a dataset, while macro gives equal weight to every class [23].

	Precision	Recall	F ₁
Weighted	0.95	0.97	0.96
Micro averaged	0.97	0.97	0.97
Macro averaged	0.75	0.76	0.75

Table 2. Evaluation results of SweDeClin-BERT

3.6 Implementation of LIME and SHAP

In Figures 4 and 5, as well as Figure 6, one of the identified discharge summaries used in the questionnaire can be seen explained by LIME [24] and SHAP [25], respectively.

When LIME was implemented for this study, ten features were used and 100 samples. As such, ten is the greatest number of features for an explanation and 100 is the size of

Text with highlighted words

Ciproxin 500 mg x 2 samt Flagyl 400 mg x 3 samt tabl Primperan 10 mg x 1 v.b samt Omeprazol 20 mg x 1., Kvinna med KOL, angina, hyperkolesterolemi samt ryggbesvär inkommer från Fanna Lasarett med CT buk konstaterad **divertikulit**. Inlägges fastande med dropp. Initialt temp, CRP samt LPK stegring. Tarmvila samt insatt på peroral antibiotika. Initialt inlagd på ASIH. Flyttas över till B62 med förbättrat AT. Infektionsparametrar i nedåtgående. Tempfri samt börjar flyta fritt. Vid utskrivning mår pat väsentligen bättre. Afebril. Fritt flytande kost samt ordinerar följa denna veckan ut samt försiktigt införa fast föda. Skrivs ut till hemmet med lugnande besked samt tabl antibiotika Ciproxin 500 mg x 2 samt Flagyl 400 mg x 3 samt tabl Primperan 10 mg x 1 v.b samt Omeprazol 20 mg x 1, som beh till tidigare refluxesofagitbesvär. Skriver även en remiss för poliklinisk uppföljning med coloskopi. Pat informera som att ha regelbundna avföringsvanor, tillföra fiberrik kost samt att vid framtida känningar av uppseglande **divertikulit** självmant gå över till flytande kost / tarmvila. I övrigt vid kraftig förvärring åter till akuten. Utskrives till hemmet. Åter v.b

Figure 5. Example of LIME explanation in Swedish – features highlighted in text

the neighbourhood of closest samples used to learn the linear model [26]. Ten features are the default value of the library, while 100 samples have been chosen due to computational limitations. On the left-hand side of Figure 4, LIME lists the prediction probabilities for the most probable ICD codes for the particular discharge summary. On the right-hand side, LIME lists the features that have the strongest influence for classifying the discharge summary with a certain ICD code. In Figure 5, the features can be seen highlighted in the discharge summary. The feature with the darkest colour has the highest impact. In Figure 6, the same discharge summary can be seen explained by SHAP. It has been classified with the ICD-code K859, with $f_{K859}(inputs)$ value of 1.96426. $f_{outputclass}(inputs)$ is the output from the model for the original output [27]. The base value for K859 is -1.03703, which means that it is the average prediction for that label [28]. Similarly to the visualisation by LIME, a darker colour indicates a more impactful feature. The blue features impact the classification negatively, while the red features impact it positively. In Figure 6, the ten most important features have been toggled to show their SHAP value. In the SHAP tool, more features can be toggled at will.

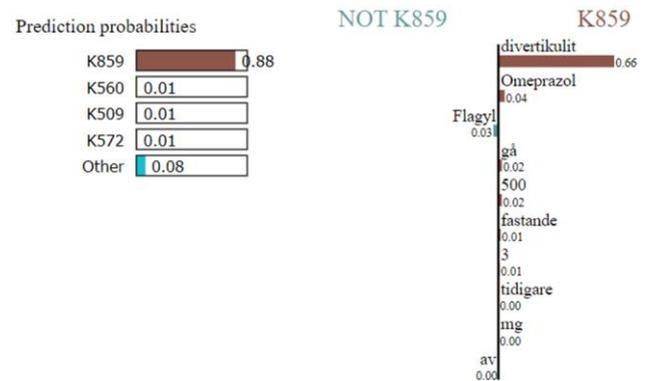


Figure 4. Example of LIME explanation in Swedish - prediction probabilities and features

3.7 Questionnaire Design

15 randomly selected discharge summaries from the test set have been included in the questionnaire, where the explanations for their predicted ICD-10 codes are explained by LIME and SHAP. This set of discharge summaries includes samples whose ICD-10 codes have been correctly predicted and ones that have been incorrectly predicted. This choice has been made to not give a misrepresentative view of the AI model's performance. Only the ten most

³ <https://github.com/Alex01234/MastersThesis>



Figure 6. Example of SHAP explanation in Swedish

impactful features for the most probable ICD-10 code from LIME and SHAP's explanations have been included in the questionnaire. The explanations by LIME and SHAP are evaluated by the respondents through the questionnaire in Google Forms.

As the explanations by LIME and SHAP are visually very different, the explanations have been harmonised in order to reduce potential design preference bias by the respondents. See examples of harmonised explanations in Figures 7 and 8, as well as Figures 9 and 10, for LIME and SHAP, respectively. In the harmonised explanations, the positively contributing features to an ICD-10 classification have a green colour. The negatively contributing features have a red colour. Again, the gastrointestinal discharge summaries are in Swedish. For reference, contrast the harmonised explanations with their original counterparts seen in section 3.6.

Prediction probability of K859 is 0.88, 10 most significant features:

NOT K859	K859	Weights
	divertikulit	0.66
	Omeprazol	0.04
Flagyl		0.03
	gå	0.02
	500	0.02
	fastande	0.01
	3	0.01
	tidigare	0.00
	mg	0.00
av		0.00

Figure 7. Harmonised explanation in Swedish by LIME - features and weights

Ciproxin 500 mg x 2 samt Flagyl 400 mg x 3 samt tabl Primperan 10 mg x 1 v.b samt Omeprazol 20 mg x 1, Kvinna med KOL, angina, hyperkolesterolemi samt ryggbesvär inkommer från Fanna Lasarett med CT buk konstaterad divertikulit. Inlägges fastande med dropp. Initialt temp, CRP samt LPK stegring. Tarmvila samt insatt på peroral antibiotika. Initialt inlagd på ASIH. Flyttas över till B62 med förbättrat AT. Infektionsparametrar i nedåtgående. Tempfri samt börjar flyta fritt. Vid utskrivning mår pat väsentligen bättre. Afebril. Fritt flytande kost samt ordinerar följa denna veckan ut samt försiktigt införa fast föda. Skrivs ut till hemmet med lugnande besked samt tabl antibiotika Ciproxin 500 mg x 2 samt Flagyl 400 mg x 3 samt tabl Primperan 10 mg x 1 v.b samt Omeprazol 20 mg x 1, som beh till tidigare refluxesofagitbesvär. Skriver även en remiss för poliklinisk uppföljning med coloskopi. Pat informera som att ha regelbundna avföringsvanor, tillföra fiberrik kost samt att vid framtida känningar av uppseglade divertikulit självmant gå över till flytande kost / tarmvila. I övrigt vid kraftig förvärring åter till akuten. Utskrives till hemmet. Åter v.b

Figure 8. Harmonised explanation in Swedish by LIME - features highlighted in text

Base value = -1.03703
f_{K859}(inputs) = 1.96426

NOT K859	K859	Weights
	buk konstaterad divertikulit.	2.746
	av uppseglade divertikulit självmant	1.424
refluxesofagitbesvär. Skriver även		-0.903
	Inlägges fastande med dropp.	0.796
en remiss för poliklinisk		-0.297
b samt Omeprazol 20		-0.257
fast föda. Skrivs		-0.224
ut till hemmet med		-0.218
x 2 samt Flagyl		-0.165
antibiotika Ciproxin 500 mg		-0.161

Figure 9. Harmonised explanation in Swedish by SHAP - features and weights

Ciproxin 500 mg x 2 samt Flagyl 400 mg x 3 samt tabl Primperan 10 mg x 1 v.b samt Omeprazol 20 mg x 1, Kvinna med KOL, angina, hyperkolesterolemi samt ryggbesvär inkommer från Fanna Lasarett med CT buk konstaterad divertikulit. Inlägges fastande med dropp. Initialt temp, CRP samt LPK stegring. Tarmvila samt insatt på peroral antibiotika. Initialt inlagd på ASIH. Flyttas över till B62 med förbättrat AT. Infektionsparametrar i nedåtgående. Tempfri samt börjar flyta fritt. Vid utskrivning mår pat väsentligen bättre. Afebril. Fritt flytande kost samt ordinerar följa denna veckan ut samt försiktigt införa fast föda. Skrivs ut till hemmet med lugnande besked samt tabl antibiotika Ciproxin 500 mg x 2 samt Flagyl 400 mg x 3 samt tabl Primperan 10 mg x 1 v.b samt Omeprazol 20 mg x 1, som beh till tidigare refluxesofagitbesvär. Skriver även en remiss för poliklinisk uppföljning med coloskopi. Pat informera som att ha regelbundna avföringsvanor, tillföra fiberrik kost samt att vid framtida känningar av uppseglade divertikulit självmant gå över till flytande kost / tarmvila. I övrigt vid kraftig förvärring åter till akuten. Utskrives till hemmet. Åter v.b

Figure 10. Harmonised explanation in Swedish by SHAP - features highlighted in text

All 30 explanations (15 discharge summaries explained by both LIME and SHAP) have three questions attached to them to evaluate the explanations on the factors of user trust, explanation satisfaction and perceived usability. The three questions are as follows:

- On a scale from 1 to 5, how trustworthy do you find the explanation of sample x to be?
- On a scale from 1 to 5, how satisfied are you with the explanation of sample x?
- On a scale from 1 to 5, how useful would you find the explanation of sample x to be, if used as a recommendation to classify the discharge summary?

4 RESULTS

Answers to the questionnaire have been collected from eight respondents, where seven are medical doctors/physicians, and one is a professional ICD-coder. All respondents have experience with ICD-coding. The collected data through the questionnaire has resulted in 120 data points for each of the factors of user trust, explanation

satisfaction and perceived usability for both models. The 120 data points come from the number of respondents multiplied by the number of discharge summaries in the questionnaire.

Paired t-tests have been performed to compare the score between LIME and SHAP for the aforementioned factors. Normally, some assumptions need to hold for a paired t-test [29]. These assumptions can, however, be disregarded when using Likert scale data [30].

4.1 Test for User Trust

A paired t-test has been conducted with the user trust for LIME and the user trust for SHAP, instantiated in the variables LIME_UT and SHAP_UT, respectively. The results can be seen in Tables 3 and 4. The p -value is 0.012, meaning that the difference of the means between LIME_UT and SHAP_UT is statistically different from zero at $\alpha = 0.05$ level of significance. The mean user trust for SHAP ($M = 2.99$, $SD = 1.553$) is higher than the mean user trust for LIME ($M = 2.47$, $SD = 1.478$), $t(119) = -2.544$, $p = 0.012$.

	Mean	N	Std. Deviation	Std. Error Mean
LIME_UT	2.47	120	1.478	.135
SHAP_UT	2.99	120	1.553	.142

Table 3. Paired samples statistics of LIME_UT and SHAP_UT

	LIME_UT - SHAP_UT
Mean	-.525
Std. Deviation	2.260
Std. Error Mean	.206
95% Confidence Interval of the Difference – Lower	-.934
95% Confidence Interval of the Difference – Upper	-.116
t	-2.544
df	119
Significance – One-Sided p	.006
Significance – Two-Sided p	.012

Table 4. Paired samples test of LIME_UT and SHAP_UT

4.2 Test for Explanation Satisfaction

A paired t-test has also been conducted with the explanation satisfaction for LIME and the explanation for SHAP, instantiated in the variables LIME_ES and SHAP_ES, respectively. The results can be seen in Tables 5 and 6. The p -value is 0.002, meaning that the difference of the means between the variables LIME_ES and SHAP_ES is statistically different from zero at $\alpha = 0.05$ level of significance. The mean explanation satisfaction for SHAP ($M = 3.04$, $SD = 1.434$) is higher than the mean explanation satisfaction for LIME ($M = 2.44$, $SD = 1.377$), $t(119) = -3.191$, $p = 0.002$.

	Mean	N	Std. Deviation	Std. Error Mean
LIME_ES	2.44	120	1.377	.126
SHAP_ES	3.04	120	1.434	.131

Table 5. Paired samples statistics of LIME_ES and SHAP_ES

	LIME_ES - SHAP_ES
Mean	-.600
Std. Deviation	2.060
Std. Error Mean	.188
95% Confidence Interval of the Difference – Lower	-.972
95% Confidence Interval of the Difference – Upper	-.228
t	-3.191
df	119
Significance – One-Sided p	<.001
Significance – Two-Sided p	.002

Table 6. Paired samples test of LIME_ES and SHAP_ES

4.3 Test for Perceived Usability

A paired t-test has been carried out with the perceived usability for LIME and the perceived usability for SHAP, instantiated in the variables LIME_PU and SHAP_PU, respectively. The results can be seen in Tables 7 and 8. The p -value is 0.005, meaning that the difference of the means between the variables LIME_PU and SHAP_PU is statistically different from zero at $\alpha = 0.05$ level of significance. The mean perceived usability for SHAP ($M = 2.99$, $SD = 1.569$) is higher than the mean perceived usability for LIME ($M = 2.39$, $SD = 1.485$), $t(119) = -2.855$, $p = 0.005$.

	Mean	N	Std. Deviation	Std. Error Mean
LIME_PU	2.39	120	1.485	.136
SHAP_PU	2.99	120	1.569	.143

Table 7. Paired samples statistics of LIME_PU and SHAP_PU

	LIME_PU - SHAP_PU
Mean	-.600
Std. Deviation	2.302
Std. Error Mean	.210
95% Confidence Interval of the Difference – Lower	-1.016
95% Confidence Interval of the Difference – Upper	-.184
t	-2.855
df	119
Significance – One-Sided p	.003
Significance – Two-Sided p	.005

Table 8. Paired samples test of LIME_PU and SHAP_PU

4.4 Analysis

SHAP has a higher mean value than LIME for all three factors investigated in our study. As evident from the results, SHAP has a mean value of around 3.0 for all factors, while LIME has a mean value of ca 2.4 for all factors. It could therefore be worthwhile to further study the explainable performance of SHAP. This could be particularly interesting since the explanations by SHAP have been harmonised with the explanations by LIME in this article. Further studies could evaluate SHAP in its original format, where the full capability of the model can be utilised.

5 DISCUSSION

This article has aimed to evaluate relevant XAI models that could be incorporated into a CAC-tool for ICD-10 coding. The most relevant XAI models for this purpose have been judged to be LIME and SHAP. While there exist previous studies that have evaluated LIME and SHAP on NLP tasks, there have been few studies evaluating them in a Swedish context. Previous research has given indications of the promising potential of LIME and SHAP. One example is LIME and SHAP's ability to show that word importance is crucial to predicting sarcasm in dialogues. Another one is that medical experts think that SHAP gives efficient explanations of how features additively contribute to explanations when explaining ADE in Swedish medical records. However, this paper is the first one to compare the ability of LIME and SHAP to explain the assignment of ICD-10 diagnosis codes to Swedish discharge summaries.

The main limitation of our study is its generalisability. A non-probabilistic approach has been applied to recruit respondents to generate an exploratory sample. This means that the opinions of the respondents may not be representative of the whole research population. Furthermore, the evaluations of LIME and SHAP are heavily dependent on the performance of the underlying model on which it is applied, in this case, SweDeClin-BERT. The evaluations are also dependent on the data used, in this case, Swedish gastrointestinal discharge summaries. This is since LIME and SHAP try to approximate the prediction function of the model to which they are applied. If a CAC-tool for ICD-10 coding is to be constructed in the future, there are many things that will have to be optimised in comparison to what has been done in this article. For example, a balanced dataset will have to be used, contrary to the dataset used in this article. Additionally, the hyperparameters of the classification model will have to be optimised during training. Once satisfactory predictive performance on all ICD-10 codes has been established, the results of this study can be used as decision support on which XAI model to incorporate into the CAC-tool. Naturally, the limitations of this study have to be kept in mind when making such a decision.

Future research is recommended to conduct a similar survey with a larger sample, which could have greater generalisability for the whole research population. Future research could also be done that applies LIME and SHAP on a classification model that has been optimised, using the considerations in the previous paragraph. The data used could also be extended to not only include gastrointestinal discharge summaries. Other kinds of discharge summaries,

as well as other medical records than only discharge summaries, could be used to train the underlying classification model on which an XAI model is applied. If this increases the predictive capability of the underlying classification model, it will likely improve the explanations by the applied XAI model. Furthermore, a qualitative study is recommended with the original SHAP tool, where visualisations are unaltered (as in Figure 6 rather than Figure 9 and 10). Then all the aspects of SHAP's visualisations can be evaluated to gain knowledge of which aspects of the explanations are valuable for a future CAC-tool. This could be especially interesting, as medical experts in previous research have indicated that SHAP gives efficient explanations of how features additively contribute to predictions. Such a qualitative study does not have to be limited to SHAP only, as it could be worthwhile to investigate the full capability of LIME as well.

In this article, explanations are only given for the most likely predicted ICD-10 code. However, a CAC-tool might include a longer list of suggested codes. A qualitative study where recommendations are given for multiple ICD-10 codes could also uncover interesting findings.

6 CONCLUSIONS

This article has examined and compared the explainable performance of LIME and SHAP in their ability to explain ICD-10 classifications of Swedish gastrointestinal discharge summaries. The classification model SweDeClin-BERT has been fine-tuned for the task of labelling discharge summaries with ICD-10 codes. LIME and SHAP have then been applied to SweDeClin-BERT to generate explanations for SweDeClin-BERT's classifications. 15 discharge summaries have been randomly chosen from the test set of data and visualised in a questionnaire. In the questionnaire, the ten most impactful features for the most probable ICD-10 code as deemed by LIME and SHAP have been visualised. The visualisations have been harmonised to reduce design preference bias. Eight answers have been collected from respondents experienced in ICD-10 coding, who have evaluated the explanations by LIME and SHAP by the factors of user trust, explanation satisfaction and perceived usability. The results of paired t-tests show that there is a statistically significant difference between LIME and SHAP for the mean value of all factors. SHAP has a higher mean value than LIME for all three factors.

7 REFERENCES

- [1] World Health Organization. 2022. International Statistical Classification of Diseases and Related Health Problems (ICD). <https://www.who.int/standards/classifications/classification-of-diseases> (Accessed 2022-05-27)
- [2] World Health Organization. 2022. Importance of ICD. <https://www.who.int/standards/classifications/frequently-asked-questions/importance-of-icd> (Accessed 2022-07-02)
- [3] Jacobsson A., Serde, L. 2013. Kodningskvalitet i patientregistret (In Swedish). <https://www.socialstyrelsen.se/globalassets/sharepoint-dokument/artikelkatalog/statistik/2013-3-10.pdf> (Accessed 2022-05-27)

- [4] Norwegian Centre for E-health Research. 2022. ClinCode - Computer-Assisted Clinical ICD-10 Coding for improving efficiency and quality in healthcare. <https://ehealthresearch.no/en/projects/clincode-computer-assisted-clinical-icd-10-coding-for-improving-efficiency-and-quality-in-healthcare> (Accessed 2022-07-02)
- [5] Pu, Y., Apel, DB., Liu, V. Mitri, H. 'Machine learning methods for rockburst prediction-state-of-the-art review', in *International Journal of Mining Science and Technology*. Vol. 29, Issue. 4, pp. 565 – 570, 2019.
- [6] Young, T., Hazarika, D., Poria, S., Cambria, E. 'Recent trends in deep learning based natural language processing', in *IEEE Computational Intelligence Magazine*. Vol. 13, Issue. 3, pp. 55-75, 2018.
- [7] Asan, O., Bayrak, A. E., Choudhury, A. 'Artificial intelligence and human trust in healthcare: focus on clinicians', in *Journal of medical Internet research*. Vol. 22, Issue. 6, e15154, 2020.
- [8] Islam, MR., Ahmed, MU., Barua, S., Begum, S. 'A Systematic Review of Explainable Artificial Intelligence in Terms of Different Application Domains and Tasks', in *Appl. Sci*. Vol. 12, Issue. 3, p. 1353, 2022.
- [9] Ribeiro, MT., Singh, S., Guestrin, C. "'Why should I trust you?'" Explaining the predictions of any classifier', in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 1135-1144, 2016.
- [10] Lundberg, SM., Lee, SI. 'A unified approach to interpreting model predictions', in *Advances in neural information processing systems*, 30, 2017.
- [11] Vakili, T., Lamproudis, A., Henriksson, A., Dalianis, H. 'Downstream Task Performance of BERT Models Pre-Trained Using Automatically De-Identified Clinical Data', accepted to *LREC 2022*, 2022.
- [12] Devlin, J., Chang, MW., Lee, K., Toutanova, K. 'Bert: Pre-training of deep bidirectional transformers for language understanding', *arXiv preprint arXiv:1810.04805*, 2018.
- [13] Rebane, J., Samsten, I., Pantelidis, P., Papapetrou, P. 'Assessing the Clinical Validity of Attention-based and SHAP Temporal Explanations for Adverse Drug Event Predictions', in *2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS)*, pp. 235-240, 2021.
- [14] Remmer, S., Lamproudis, A., Dalianis, H. 'Multi-label Diagnosis Classification of Swedish Discharge Summaries-ICD-10 Code Assignment Using KB-BERT', in *RANLP 2021: Recent Advances in Natural Language Processing, 1-3 Sept 2021, Varna, Bulgaria*, pp. 1158-1166, 2021.
- [15] Kumar, A., Dikshit, S., Albuquerque, VHC. 'Explainable Artificial Intelligence for Sarcasm Detection in Dialogues', in *Wireless Communications and Mobile Computing, 2021*, 2021.
- [16] Khodabandehloo, E., Riboni, D., Alimohammadi, A. 'HealthXAI: Collaborative and explainable AI for supporting early diagnosis of cognitive decline', in *Future Generation Computer Systems*, Vol. 116, pp. 168-189, 2021.
- [17] Singla, K., Biswas, S. 'Machine learning explainability method for the multi-label classification model', in *2021 IEEE 15th International Conference on Semantic Computing (ICSC)*, pp. 337-340, 2021.
- [18] Vilone, G., Longo L. 'Explainable artificial intelligence: a systematic review', *arXiv preprint arXiv:2006.00093*, 2020.
- [19] Vilone, G., Longo L. 'Classification of explainable artificial intelligence methods through their output formats', in *Machine Learning and Knowledge Extraction*, Vol. 3, Issue. 3, pp. 615-661, 2021.
- [20] Dalianis, H., Henriksson, A., Kvist, M., Velupillai, S., Weegar, R. 'HEALTH BANK-A Workbench for Data Science Applications in Healthcare', in *CAiSE Industry Track*. Vol. 1381, pp. 1-18, 2015.
- [21] James, Gareth. Witten, Daniela. Hastie, Trevor. Tibshirani, Robert. *An Introduction to Statistical Learning*. Springer, New York, p. 198, 2013.
- [22] Malmsten, M., Börjeson, L., Haffenden, C. 'Playing with Words at the National Library of Sweden-- Making a Swedish BERT', *arXiv preprint arXiv:2007.01658*, 2020.
- [23] Yang, Y. 'An evaluation of statistical approaches to text categorization', in *Information retrieval*. Vol. 1, Issue. 1, pp.69-90, 1999.
- [24] Ribeiro, MT. 2021. marcotcr/lime. <https://github.com/marcotcr/lime> (Accessed 2022-04-21)
- [25] Lundberg, SM. 2022. SHAP. <https://github.com/slundberg/shap> (Accessed 2022-03-30)
- [26] Ribeiro, MT. 2016. lime package. <https://lime-ml.readthedocs.io/en/latest/lime.html> (Accessed 2022-04-22)
- [27] Lundberg, S. 2018. Using custom functions and tokenizers. https://shap.readthedocs.io/en/latest/example_notebooks/extension_examples/sentiment_analysis/Using%20custom%20functions%20and%20tokenizers.html (Accessed 2022-05-28)
- [28] Molnar, C. SHAP (SHapley Additive exPlanations). In *Interpretable Machine Learning*, 2022. <https://christophm.github.io/interpretable-ml-book/shap.html> (Accessed 2022-03-30)
- [29] Newcastle University. n.d. Paired Samples T Test (Dependent Samples T test). <https://services.ncl.ac.uk/itservice/research/dataanalysis/simpletests/ttests/pairedsamplesttestdependentsamplestest/> (Accessed 2022-04-17)
- [30] Norman, G. 'Likert scales, levels of measurement and the "laws" of statistics', in *Advances in health sciences education*, Vol. 15, Issue. 5, pp. 625-632, 2010.

8 ACKNOWLEDGEMENT

This study has been partially funded by the Norwegian Research Council through the ClinCode project, number 318098.

We would also like to thank our eight respondents for their work.

The Influence of NegEx on ICD-10 Code Prediction in Swedish: How is the Performance of BERT and SVM Models Affected by Negations?

Andrius Budrionis^{1,2}, Taridzo Chomutare¹, Therese Olsen Svenning¹ and Hercules Dalianis^{1,3}

¹Norwegian Centre for E-health Research, Tromsø, Norway Andrius.Budrionis@ehealthresearch.no

²Faculty of Science and Technology, UiT The Arctic University of Norway, Tromsø, Norway

³Department of Computer and Systems Sciences (DSV), Stockholm University, Kista, Sweden

Abstract

Clinical text contains many negated concepts since the physician excludes irrelevant symptoms when reasoning and concluding about the diagnosis. This study investigates the machine interpretation of negated symptoms and diagnoses using a rule-based negation detector and its influence on downstream text classification task. The study focuses on the effect of negated concepts and NegEx preprocessing on classifier performance for predicting ICD-10 gastro surgical codes assigned to discharge summaries. Based on the experiments, NegEx preprocessing resulted in a slight performance improvement for traditional machine learning model (SVM) and had no effect on the performance of the deep learning model KB/BERT.

Keywords

Clinical text, negation, NegEx, Swedish, BERT, ICD-10 diagnosis codes

1 INTRODUCTION

Physician's reasoning to find the correct diagnosis of a patient are often trying to exclude symptoms until the hopefully correct diagnosis is concluded. This leads to the patient's record containing many negations excluding various clinical concepts [1].

Negated terms make it difficult for machines to interpret natural language in patient records. One of the first approaches to identify negated symptoms and diagnoses was the development of NegEx, [2]. This relatively simple rule-based algorithm showed acceptable performance in identifying negated clinical symptoms and diagnosis. The output of NegEx (negation tags) could be an important additional feature for various Natural Language Processing (NLP) tasks.

Numerous approaches for detecting negations in free text have been proposed. In addition to the aforementioned NegEx [2], more advanced models were developed taking advantage of language semantics [3] [4]. Ettinger [5] performed fine-tuning of BERT on manually annotated data sets containing both negations and non-negations. The findings showed that BERT was not able to distinguish the negations in the text within acceptable accuracy. On contrary, Lin et al. found that fine-tuning BERT on clinical text that contains annotated negations led to BERT learning to predict negations [6].

The influence of negated symptoms and diagnosis existing in the text and the output of negation detectors on the downstream modelling tasks has only be studied to a limited extent [7] [8] [9]. It is unclear how NegEx preprocessing contributes to the overall performance of text classification models and what use cases or machine learning models benefit from negation tagging or removal from the text.

The 18th Scandinavian Conference on Health informatics, Tromsø, Norway, August 22-24, 2022. Organized by UiT The Arctic University of Norway. Conference Proceedings published by Linköping University Electronic Press at <https://doi.org/10.3384/ecp187>. © The Author(s). This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>

Remmer et al. [10] carried out classification experiments for predicting groups of ICD-10 diagnosis codes assigned to Swedish discharge summaries. The researchers used both traditional machine learning methods and the deep learning BERT model. The BERT model outperformed the traditional models. Since clinical text contains many negations, specifically for Swedish clinical text 13.5% of the sentences or expressions were negated [11], it would be valuable to know how these affect the classification results and also if there are methods to cope with negations. This paper studies the effect of NegEx on ICD-10 code prediction task.

2 RELATED RESEARCH

Existing research on the effect of NegEx on the downstream modelling tasks is fragmented and limited to a few publications, mostly focusing on traditional machine learning algorithms, such as Support Vector Machines (SVM) used for sentiment analysis. Sharif et al. reported significant increase in accuracy, precision and recall predicting sentiments in customer reviews after text was preprocessed by a negation detector [7]. Similar findings were reported for Naïve Bayes, Artificial Neural Network (ANN), and Recurrent Neural Network (RNN) models used for sentiment analysis. The largest positive effect of negation tagging was observed in RNN models [8]. Kaddoura et al. demonstrated that treating negations in Facebook posts resulted in 20% increase in F1-score in sentiment analysis [9]. Considering the number and importance of negations in medical narrative, similar improvements in clinical NLP tasks could be expected.

3 DATA AND METHODS

3.1 Data

The Stockholm EPR Gastro ICD-10 Pseudo Corpus was used in the experiments. It contains discharge summaries and their manually assigned ICD-10 gastro related diagnosis codes. Additional details on the dataset can be found in [10]. The specific corpus variant used in this paper is called Pseudo Corpus since it has been de-identified with regard to Protected Health Information, PHI, and the identified PHIs have been replaced with realistic pseudonyms [12].

The deidentification and pseudonymisation system, called HB-Deid, was used for cleaning the text from sensitive details. It detects the following PHI classes: First Name, Last Name, Age, Location, Health Care Unit, Date, Phone Number, Organisation and Social Security Number and replaces them with realistic pseudonyms or surrogates. HB-Deid is based on Conditional Random Fields algorithm and rule-based preprocessing step to find missed phone numbers and social security numbers through regular expressions. The final step in HB-Deid is the Pseudonymiser that replaces the identified entities with realistic surrogates. After deidentification the Stockholm EPR Gastro ICD-10 Pseudo Corpus can be shared with academic community. This research has been approved by the Swedish Ethical Review Authority under permission no 2021-03758.

The dataset consists of 6,002 discharge summaries from 4,985 unique patients, 813,154 tokens in total. 263 distinct class labels (ICD-10 codes) are present in the text.

3.2 ICD-10 blocks

Many classes (ICD-10 codes) were represented by very few examples. To make modelling task easier, the label space was condensed into ICD-10 code blocks combining multiple labels into a single class (K00-K14, K20-K31, K35-K38, K40-K46, K50-K52, K55-K64, K65-K67, K70-K77, K80-K87, K90-K93).

The blocks are logical partitions of the gastrointestinal domain, starting from the oral cavity to the rest of the digestive system, and are well-recognised in the medical field. Classifying notes into the blocks can be useful in clinical practice, as well as part of a pipeline to classify into more granular ICD-10 codes.

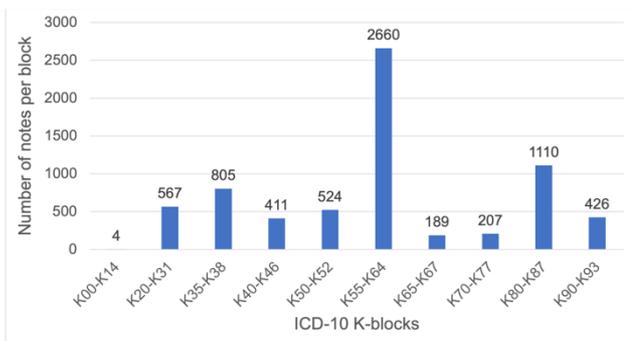


Figure 1. The distribution of number of discharge summaries per ICD-block.

The distribution of discharge summaries per block are shown in Figure 1. The majority of discharge summaries contains only one code block label; the maximum number of code block labels per discharge summary is four.

3.3 Methods

To study the effect of negations in clinical text on the performance of ICD-10 code block classification using different classifiers, the clinical notes were preprocessed using a Swedish negation detector (NegEx). The Swedish NegEx has a performance of 75.2% precision and a recall of 81.9% applied on Swedish clinical text [13].

Two versions of NegEx outputs were used in downstream modelling tasks: negated symptoms and diagnoses were either tagged or removed (referred to as tagged and removed negations in the remainder of this paper). The original (without NegEx preprocessing) dataset was used in baseline experiments.

Two types of classifiers, representing traditional and state-of-the-art machine learning models used in NLP were selected. For the traditional models, a SVM classifier (scikitlearn v1.0.2) was trained using TF-IDF vectors from clinical notes. 10-fold cross-validation was used when training the model; 10% of data was held out from training and used for testing purposes only. Data was split at a record level.

For the state-of-the-art model, a KB/BERT, a Swedish general language model pretrained on newspapers, Swedish Wikipedia and government documents, was used [14]. KB/BERT was finetuned on the preprocessed datasets (tagged and removed negations) and tested on the holdout dataset. The parameters used with the pytorch model are shown in Table 1.

Parameter	Value
Test size	0.2
K-fold	10
Random state	42
Epochs	15
Batch size train	6
Batch size test	6
Gradient accumulation	8
Learning rate	3e-5
Warm up	400
Threshold	0.3

Table 1. The parameters used to fine tune KB/BERT model.

4 RESULTS

Experiment results are summarised in Table 2 and show some interesting trends. Tagging and removal of negated symptoms and diagnoses resulted in a small performance boost for the SVM model in comparison to the baseline. While this improvement in performance was minor, it may be sufficient for considering NegEx preprocessing as a technique for training an optimal model.

SVM code blocks

	Precision	Recall	F1-score
Original dataset	0.89	0.61	0.73
Negations tagged	0.90	0.63	0.74
Negations removed	0.90	0.63	0.74

KB/BERT code blocks

	Precision	Recall	F1-score
Original dataset	0.80	0.81	0.81
Negations tagged	0.81	0.80	0.81
Negations removed	0.80	0.81	0.81

Table 2. Model performance on holdout dataset. (In micro average precision, recall and F1-score respectively).

Performance of KB/BERT did not follow the same trend. Tagging and removing negations in the fine-tuning dataset did not affect model performance in terms micro average of F1-score. Per class performance is presented in Table 3.

4.1 NegEx post-processing challenges

So far we have discussed pre-processing of data for training purposes. Another option is to use negation in post-processing, which has the potential to yield improved performance. One approach to post-processing involves analysing the negated concepts to determine the likely associated ICD-code or block. For instance, one of the most frequently negated terms is "abscess" (see Table 4), which is associated with two blocks within the Gastrointestinal ICD-10 space, namely, ['K00-K14', 'K55-K64']. Comparing predicted results against these known ICD-10 associations can help exclude possibly wrong predictions. However, in this case, there are multiple factors that presented significant challenges for post-processing, and in this subsection we discuss the top four factors.

Block	ICD-10 description	F1-score original		F1-score tagged		F1-score removed	
		SVM	BERT	SVM	BERT	SVM	BERT
K00-K14	Oral cavity, salivary glands and jaws	0.00	0.00	0.00	0.00	0.00	0.00
K20-K31	Esophagus, stomach and duodenum	0.32	0.68	0.56	0.70	0.42	0.68
K35-K38	Disease of appendix	0.91	0.95	0.93	0.95	0.92	0.95
K40-K46	Hernia	0.39	0.70	0.36	0.70	0.39	0.72
K50-K52	Non-infectious inflame. Intestine	0.32	0.72	0.37	0.73	0.26	0.73
K55-K64	Other diseases of the intestine	0.85	0.87	0.86	0.86	0.87	0.86
K65-K67	Disease of the peritoneum	0.00	0.12	0.00	0.20	0.00	0.15
K70-K77	Diseases of the liver	0.18	0.49	0.00	0.49	0.00	0.52
K80-K87	Gallbladder, bile ducts and pancreas	0.83	0.92	0.85	0.92	0.81	0.91
K90-K93	Other diseases of the digestive system	0.36	0.54	0.24	0.52	0.38	0.53
Micro avg		0.73	0.81	0.74	0.81	0.74	0.81

Table 3. Model performance on held out dataset using SVM and KB BERT using 10-fold cross validation. The summary of all classes is in micro F1-score.

The first challenge is related to the ICD-10 level or hierarchy, where the ICD-10 block is a high-level reference, as opposed to a lower level 4-char code such as K56.7 or simply K567. Whereas it is possible to exclude lower-level codes if the associated concept is negated, it is not logical to exclude blocks, since a block contains multiple related ICD-10 codes.

Term	Similar ICD-10 concept	Associated block(s)	Frequency
abscess	abscess [abscess]	'K00-K14', 'K55-K64'	9
kolecystit	kolecystit [cholecystitis]	'K80-K87'	7
peritonit	peritonit [peritonitis]	'K65-K67', 'K35-K38'	6
kräkning	Kräkningar [Vomiting]	'K90-K93'	4
divertikulit	Divertikel [diverticulum]	'K20-K31', 'K55-K64', 'K35-K38'	3

Table 4. Test partition's top 5 negated terms and their associated ICD-10 concepts and blocks.

The second factor is related to where the negated terms appear in the wider context that includes non-gastrointestinal blocks. Since our label-space is focused on Gastrointestinal blocks, some negated concepts will fall outside the relevant vocabulary or scope. For instance, "leukocytos" [leukocytes] is associated more with D72--"Other disorders of white blood cells", rather than gastro K-codes. Therefore, negation of this term likely will not affect the performance for the limited gastro case. However, we found that these out-of-vocabulary negations were comparatively small, making up approximately 10% all of negations in the test set.

The third factor is somewhat related to the second factor in that there were only a few records with negations, compared to the number of records in the test set. There were only 46 records with a total of 52 negated terms, representing approximately 7.7% of the test set. This factor also helps partially explain the generally muted performance gains after tagging negations.

The final factor relates to the performance of NegEx on the Swedish clinical text. Since NegEx does not understand complex semantic meaning of sentences, some negated terms will actually be a positive diagnosis. NegEx performance on Swedish data presents peculiar challenges that reduces the effect of any post processing. We discuss concrete examples of NegEx failures in the next subsection.

4.2 Error analysis

Here follows an error analysis on the performance of the rule-based NegEx on some of the clinical text.

First a correct negation tagging is shown below, where NegEx marks up "leukocytes" as a negated concept.

"Pat inlägges fastande. Labmässigt noterar man CRP 47, som sjunker till 8. Ingen <NEGATED>leukocytes</NEGATED>".

(In Eng.) "Pat is admitted fasting. CRP of 47 was noted in the lab, which drops to 8. No <NEGATED>leukocytes</NEGATED>".

Linking words such as 'but' are prone to misinterpretation. 'Not something but something else' should confirm the last condition, not exclude it as shown in the example below.

"genomgår gastroskopi som inte visar någon <NEGATED>främmande kropp </NEGATED>, däremot en <NEGATED>esofagit grad 3 </NEGATED>".

(In Eng.) "undergoes gastroscopy which shows no <NEGATED>foreign body </NEGATED>, but a <NEGATED>esophagitis grade 3 </NEGATED>".

Double trigger words may also lead to wrong negations. In the following example two trigger words which should null each-other out, meaning that there is in fact a condition present, gets interpreted as the opposite.

"man med smärta i buken till vänster sedan en vecka, där man inte kan utesluta <NEGATED>divertikulit </NEGATED>".

(In Eng.) "man with abdominal pain to the left for a week, where one cannot rule out <NEGATED>diverticulitis </NEGATED>".

5 DISCUSSION

Our findings indicate that the presence of negated symptoms and diagnoses in clinical text may have varying effect on the performance of ICD-10 code prediction tasks.

While traditional machine learning model (SVM) trained using bag-of-words vector representations of clinical text experienced minor benefits of NegEx preprocessing (tagging and removal), the performance of the state-of-the-art model (KB/BERT) was not affected.

These findings may be explained by the way text is represented for the classification algorithms. A bag-of-words representation is not capable of capturing any semantic relationships between text tokens, and NegEx preprocessing enables it to differentiate between negated and non-negated concepts. Removal of negated symptoms and diagnoses discards some "noise" in the training data resulting in a small increase in classifier performance.

KB/BERT uses an underlying Swedish language model learned from a large amount of text. This language model is used when tokenizing the text and transforming tokens into embeddings capturing various language properties. These capabilities help KB/BERT differentiate between negated and non-negated clinical terms and use negations as additional features when making predictions. Therefore, tagging and removing negated symptoms and diagnoses has not affected classifier performance. KB/BERT managed to interpret the negations without using any preprocessing with NegEx and showed better results than the traditional machine learning methods combined with NegEx.

5.1 Limitations

The choice of classification algorithms comes as the main limitation of this paper. While both traditional and deep-learning-based models were studied in the experiments, the findings cannot be generalized for all algorithms in these classes. A great variety of classification algorithms belonging to both classes calls for extensive experiments using a well-grounded selection process taking properties of specific algorithms into account. SVM and KB/BERT were chosen due to their popularity for text classification tasks, therefore results presented in this paper should only be considered as preliminary findings calling for more research.

Rule based NegEx has a relatively high error rate and a more robust approach should be explored for preprocessing clinical text. More sophisticated implementations of negation detectors result in higher accuracy that may affect the performance of downstream modelling tasks [3] [4].

6 CONCLUSIONS

Regardless of the limitations of NegEx (see section 4.2 for more details), the performed experiments show the following trends. Model based on bag-of-words text representation benefited from NegEx preprocessing, resulting in increasing performance. On the contrary, performance gains were absent for the advanced model, capturing the semantic relationships between text tokens.

Post-processing was complicated by the non-atomic nature of the label space, where the ICD blocks contain multiple individual ICD codes, making it difficult to exclude any negated labels.

7 LANGUAGE RESOURCE AVAILABLE

The pseudonymised dataset, the Stockholm EPR Gastro ICD-10 Pseudo Corpus, is available for research for academic researchers after signing a confidentiality agreement, please contact Hercules Dalianis, email: hercules@dsv.su.se.

8 REFERENCES

- [1] Epstein, R. M. "How Doctors Think." *Journal of Clinical Investigation*, Vol. 117, No. 10, 2007, pp. 2738–2738. <https://doi.org/10.1172/JCI33149>.
- [2] Chapman, W. W., Bridewell, W., Hanbury, P., Cooper, G. F., and Buchanan, B. G. "A Simple Algorithm for Identifying Negated Findings and Diseases in Discharge Summaries." *Journal of Biomedical Informatics*, Vol. 34, No. 5, 2001, pp. 301–310. <https://doi.org/10.1006/jbin.2001.1029>.
- [3] Montenegro, O., Pabon, O. S., and De Pinerez R., R. E. G. "A Deep Learning Approach for Negation Detection from Product Reviews Written in Spanish." Presented at the 2021 XLVII Latin American Computing Conference (CLEI), Cartago, Costa Rica, 2021.
- [4] Khandelwal, A., and Britto, B. K. "Multitask Learning of Negation and Speculation Using Transformers." Presented at the Proceedings of the 11th International Workshop on Health Text Mining and Information Analysis, Online, 2020.
- [5] Ettinger, A. "What BERT Is Not: Lessons from a New Suite of Psycholinguistic Diagnostics for Language Models." *Transactions of the Association for Computational Linguistics*, Vol. 8, 2020, pp. 34–48. https://doi.org/10.1162/tacl_a_00298.
- [6] Lin, C., Bethard, S., Dligach, D., Sadeque, F., Savova, G., and Miller, T. A. "Does BERT Need Domain Adaptation for Clinical Negation Detection?" *Journal of the American Medical Informatics Association*, Vol. 27, No. 4, 2020, pp. 584–591. <https://doi.org/10.1093/jamia/ocaa001>.
- [7] Sharif, W., Samsudin, N. A., Deris, M. M., and Naseem, R. "Effect of Negation in Sentiment Analysis." Presented at the 2016 Sixth International Conference on Innovative Computing Technology (INTECH), Dublin, Ireland, 2016.

- [8] Mukherjee, P., Badr, Y., Doppalapudi, S., Srinivasan, S. M., Sangwan, R. S., and Sharma, R. "Effect of Negation in Sentences on Sentiment Analysis and Polarity Detection." *Procedia Computer Science*, Vol. 185, 2021, pp. 370–379. <https://doi.org/10.1016/j.procs.2021.05.038>.
- [9] Kaddoura, S., Itani, M., and Roast, C. "Analyzing the Effect of Negation in Sentiment Polarity of Facebook Dialectal Arabic Text." *Applied Sciences*, Vol. 11, No. 11, 2021, p. 4768. <https://doi.org/10.3390/app11114768>.
- [10] Remmer, S., Lamproudis, A., and Dalianis, H. "Multi-Label Diagnosis Classification of Swedish Discharge Summaries – ICD-10 Code Assignment Using KB-BERT." Presented at the The International Conference on Recent Advances in Natural Language Processing (RANLP 2021), 2021.
- [11] Dalianis, H., and Velupillai, S. "How Certain Are Clinical Assessments? Annotating Swedish Clinical Text for (Un)Certainties, Speculations and Negations." Presented at the the Seventh International Conference on Language Resources and Evaluation (LREC'10), Valletta, Malta, 2010.
- [12] Berg, H., and Dalianis, H. "Augmenting a De-Identification System for Swedish Clinical Text Using Open Resources and Deep Learning." Presented at the Workshop on NLP and Pseudonymisation, Turku, Finland, 2019.
- [13] Skeppstedt, M. "Negation Detection in Swedish Clinical Text: An Adaption of NegEx to Swedish." *Journal of Biomedical Semantics*, Vol. 2 Suppl 3, 2011, p. S3. <https://doi.org/10.1186/2041-1480-2-S3-S3>.
- [14] Malmsten, M., Börjeson, L., and Haffenden, C. "Playing with Words at the National Library of Sweden -- Making a Swedish BERT." *arXiv:2007.01658 [cs]*, 2020.

9 ACKNOWLEDGEMENT

This work was partially funded by the ClinCode project at the Norwegian Centre for E-health Research. Grant number 318098, Research Council of Norway.

Autostrata

Improved Automatic Stratification for Coarsened Exact Matching

Jo Inge Arnes¹, Alexander Hapfelmeier², and Alexander Horsch³

¹UiT The Arctic University of Norway, Department of Computer Science, Tromsø, Norway, jo.i.arnes@uit.no

²Technical University of Munich, Institute of AI and Informatics in Medicine, München, Germany

³UiT The Arctic University of Norway, Department of Computer Science, Tromsø, Norway

Abstract

We commonly adjust for confounding factors in analytical observational epidemiology to reduce biases that distort the results. Stratification and matching are standard methods for reducing confounder bias. Coarsened exact matching (CEM) is a recent method using stratification to coarsen variables into categorical variables to enable exact matching of exposed and nonexposed subjects. CEM's standard approach to stratifying variables is histogram binning. However, histogram binning creates strata of uniform widths and does not distinguish between exposed and nonexposed. We present Autostrata, a novel algorithmic approach to stratification producing improved results in CEM and providing more control to the researcher.

Keywords

Analytic epidemiology, confounder bias, stratification, coarsened exact matching, algorithms

1 INTRODUCTION

Epidemiologists conduct analytical observational studies [1] to investigate associations between exposures and outcomes. Instead of assigning a treatment or exposure to the participants of a randomized experiment [2], we rely on observations of the subjects in their usual environment with minimal interference. There are many established ways of designing observational studies, from cross-sectional, cohort, and case-control studies to more complex prospective cohorts with several nested case-control and cross-sectional designs [3, 4, 5]. A common theme for these is awareness of biases. Confounding factors [6, 7, 8] are a common source of bias that can, if measured, be adjusted for in the analysis [9, p. 1020]. Stratification [10], for example, can control for confounding by dividing study subjects into groups based on observed confounders. Iacus et al. [11] present the *coarsened exact matching* (CEM) method that adjusts for bias by turning confounder covariates into categorical variables through stratification, which we can then use to match comparable subjects exactly. Blackwell et al. [12] introduce a Stata (<https://www.stata.com>) implementation of CEM, and Iacus et al. [13] provide an implementation for R (<https://www.r-project.org>). In addition, a web page with an overview of implementations for other programming languages and platforms is available (<https://gking.harvard.edu/cem>). The same webpage also informs that CEM is officially qualified for scientific use by the U.S. Food and Drug Administration. The CEM implementations let users create strata manually or use automatic stratification. The built-in automatic stratification creates uniform width bins by applying general rules of thumb for constructing histograms. The three binning algorithms included in both Stata and R are Sturges' rule [14], Scott's rule [15], and Freedman-Diaconis' rule [16]. Additionally, Stata includes an implementation of Shimazaki-Shinomoto's rule [17].

Blackwell et al. [12, p. 534] demonstrate that manually defining strata based on domain knowledge can sometimes give better results than the current automatic approach. In their example, the manually defined strata are less imbalanced while giving a higher number of matched units. However, according to King et al. [18, p. 439], researcher biases are highly likely to affect qualitative choices even when researchers attempt to avoid them. 'The literature makes clear that the way to avoid these biases is to remove researcher discretion as much as possible,' following King. On the other hand, the general histogram binning rules do not support the specific challenges of stratifying confounders:

- The histogram binning algorithms do not distinguish between different groups of units and include no concept of matching.
- They do not take into account multivariate imbalance between groups.
- The strata have uniform widths, i.e., all strata for a covariate have the same width.
- The researcher cannot in advance give parameters to influence the stratification process.

Against this background, we researched and developed a novel algorithmic approach to the stratification problem that addresses the shortcomings above. We implemented the algorithm and experimentally compared it to CEM's built-in histogram binning with good results.

We conclude the introduction with a brief example of Autostrata's applicability to health-related studies. For instance, say we want to study if coffee consumption is associated with a beneficial effect on the risk of liver cancer. In the respective observational study, we must be cautious of possible systematic differences between the compared groups, such as smoking habits. Failing to adjust for these differences can challenge the validity of the results. Autostrata improves such adjustments when using

The 18th Scandinavian Conference on Health Informatics, Tromsø, Norway, August 22-24, 2022. Organized by UiT The Arctic University of Norway. Conference Proceedings published by Linköping University Electronic Press at <https://doi.org/10.3384/ecp187>. © The Author(s). This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>

CEM. The method creates more precise results and keeps more study participants included in the analysis.

After the introduction, the structure of the paper is as follows: First, we provide essential background for understanding the problem. Next, we describe our approach and algorithm. We then present the experiments and results, followed by a discussion. Last, we briefly touch upon related work before concluding.

1.1 A note on terminology

The paper primarily uses the terms *treated and controls* instead of *exposed and nonexposed* due to their use in CEM and the general causal inference literature. In addition, although case-control studies are different from cohort studies that focus on exposed and unexposed, they are interchangeable in this paper because we concentrate on the stratification of confounder covariates in isolation from these differences. Further, we use the general statistical term units instead of subjects, individuals, or study participants often seen in epidemiology.

2 BACKGROUND

Before presenting our approach, we provide the background necessary to understand the challenges of the stratification problem.

2.1 Confounding

We often assess whether the risk of a health event (outcome) is increased or decreased among an exposed or treated group compared to a control group. To quantify the relationship between an exposure or treatment and the outcome, we calculate risk ratios, odds ratios, or other measures. However, other factors not directly under investigation can skew the results or even lead us to the opposite conclusion of what is correct. Figure 1 illustrates how confounding factors influence both the exposure and the outcome. Note that the confounder is not in the direct causal pathway between the two. Also, a relevant property of confounders is that the compared groups have differently distributed values for the confounder covariate. If the confounders are measured and included in the dataset, we can adjust for confounders during analysis, which is the purpose of the stratification discussed in this paper. It is worth noting that according to Wacholder et al. [9, p. 1020], the use of stratification or matching can, in effect, adjust for unknown or unmeasured confounders through reduced variability because this variability is measured conditionally on the levels of other studied variables.

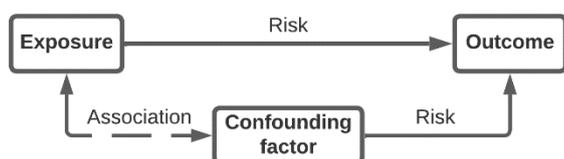


Figure 1 shows an exposure that is associated with a risk of an outcome. The confounding factor is associated with both the exposure and the outcome without being in the direct causal pathway of the two.

2.2 Counterfactuals and imbalance

The Neyman-Rubin causal model (RCM) [19] is one of the notable influences on the understanding of causal inference in observational studies. According to the model, to estimate the effect of a treatment on an outcome, we should

ideally compare the treated subjects with the same subjects without treatment. Except for the treatment, all other conditions must be the same, including the time. The latter is a counterfactual and is impossible to observe. We instead compare to relatively similar, untreated controls. However, the treated and controls in our sample are often systematically different or imbalanced for the confounding factors, which leads to bias. Lowering this imbalance between treated and controls to make them more similar is thus a strategy to reduce the bias.

2.3 Coarsened exact matching

As earlier explained, the confounder covariates are distributed differently for the compared groups. Thus, we can view the bias as stemming from an imbalance in the data. Coarsened exact matching (CEM) [11] is a method for adjusting confounder bias as a preprocessing step before analysis. It belongs to a class of monotonic imbalance bounding (MIB) methods, enabling the researcher to set a maximum imbalance between treated and controls for the confounder covariates or reduce the maximum imbalance for a covariate independently of others. The theoretical foundation of CEM is outside the scope of this paper, but its use is relatively straightforward.

We partition the confounder covariates into subintervals. Each subinterval then represents a single value of a categorical variable. For example, a covariate for years of education can be partitioned into subintervals representing the highest level of education instead. In CEM, this is called coarsening and opens for simple, exact matching of similar treated and control units. It additionally helps balance the sample by pruning treated and control units without suitable matches. The coarsening is temporary and not passed to subsequent analysis steps.

The described coarsening corresponds to stratification. We stratify each covariate, and each treated and control unit will then belong to a multi-dimensional stratum. Although the current CEM software packages use uniform width histogram binning for automatic stratification, CEM as a method is not restricted to strata of uniform widths. For example, manual stratification and non-uniform widths are supported. Autostrata is an alternative approach to automatically stratifying covariates, which constructs strata of non-uniform widths.

2.4 Imbalance and unmatched trade-off

The most commonly described imbalance measure for CEM involves the relative difference between the number of treated and control units per stratum. However, the software packages use an imbalance measure based on a per stratum difference in means between the covariate values for the two groups as default. This is similar to what Appendix B of [11, p. 34] describes. We thus base our approach on the latter.

As shown in Figure 2, two strata with the same number of treated and control units can have a different internal imbalance because the covariate means are different for the groups. Nevertheless, the maximum imbalance is bounded by the stratum widths because the differences cannot be greater than the widths. Therefore, the narrower the stratum is, the lower its maximum imbalance. The lowest maximum imbalance is when each stratum only has a single unit or equal-valued units. A stratum with only one type of unit contributes zero to the imbalance, while multiple equal-valued units have an imbalance of zero. The challenge is that there is a trade-off.

CEM prunes unmatched units from the sample. If all units in a stratum are from the same group, these units are unmatched and discarded. Recall that the confounder covariates for treated and control units have different distributions. Hence, various degrees of overlap and densities will be found along the covariate axes, restricting how narrow a stratum containing both types of units can be. As we decrease the maximum imbalance, the number of unmatched units generally increases, and vice versa. Autostrata aims to lower this trade-off.

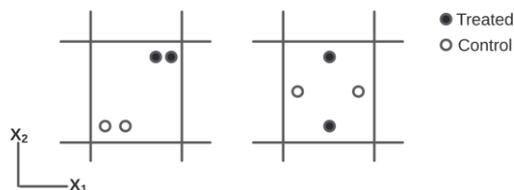


Figure 2 illustrates two strata for covariates x_1 and x_2 . Both strata have two treated and two controls, but the left stratum has a higher mean difference. Also, the maximum difference is bounded by the width between the stratum edges.

2.5 Stratification problem properties

Before concluding the background section, we describe a few properties of the stratification problem relevant to solving it algorithmically.

First, the number of relevant stratum edges is finite. The reason is that a stratum edge for a covariate can be placed anywhere between two adjacent observations without changing stratum memberships. If an observation coincides with an edge, it belongs to the higher stratum. The exact position of an edge does not matter, only that it separates two adjacent observations for the given covariate. Neither do multiple stratum edges between two neighboring observations change any memberships. Further, if two or more observations have equal values for a covariate, they cannot be separated by adding stratum edges for the given covariate. Conclusively, the maximum number of relevant stratum edges equals the number of distinct values per covariate.

Second, the number of possible combinations of the stratum edges, from including no edge to including all edges, grows exponentially with the number of distinct covariate values, i.e., the problem space is non-polynomial.

Figure 3 shows all possible combinations of stratum edges for four distinct values, organized as a tree of nodes. The number of new stratifications that can be made by adding one stratum edge to a given stratification is illustrated in Figure 4.

We can deduce the number of different stratifications possible for a covariate. Let S be the set of possible stratifications for a covariate with n distinct values. Then the cardinality, $|S|$, is:

$$|S| = 1 + (n - 1) + \sum_{i=2}^{n-1} 2^{i-2}(n - i) = 2^{n-1}$$

Given m covariates, the total number of combinations, $|S_{tot}|$, becomes:

$$|S_{tot}| = \prod_{i=1}^m |S_i| = \prod_{i=1}^m 2^{n_i-1}$$

For cases where all $n_i = n$ are equal:

$$|S_{tot}| = |S|^m = 2^{m(n-1)}$$

Thus, the *state space* of the problem grows exponentially with increasing numbers of distinct values and covariates. Furthermore, considering that each stratification can contain relatively many multi-dimensional strata and that we must compute imbalance measures and the number of unmatched units for each stratification, it quickly becomes computationally infeasible to perform a brute-force search through all combinations to find an optimal solution with the resources typically available to researchers.

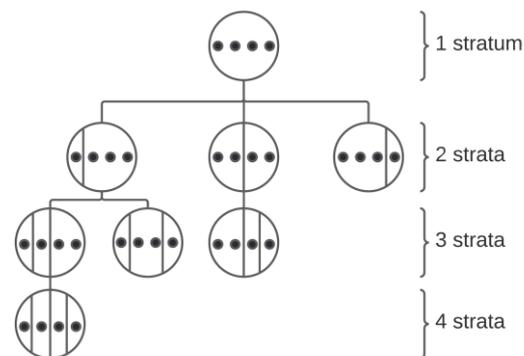


Figure 3. All possible stratifications of a covariate with four distinct observed values. The four values are illustrated as black dots within the tree nodes, and the stratum edges as vertical lines between the dots.

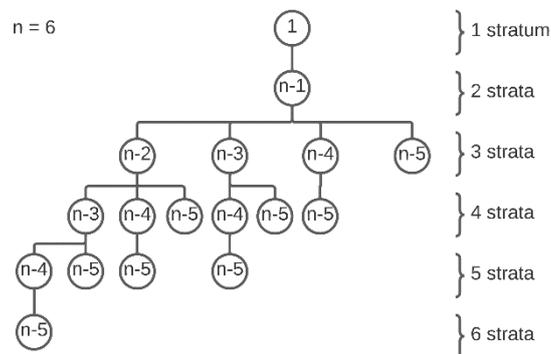


Figure 4. This tree illustrates a pattern in the number of different stratifications that can be made as we move from a given parent to a child node by adding a new stratum edge, as in Figure 3. In this case, the number of distinct values is $n=6$.

3 AUTOSTRATA

We now present Autostrata, a novel algorithmic approach for improved stratification of confounder covariates for CEM. Improving CEM's standard stratification method—histogram binning—is not trivial. However, analysis results need to be as free of bias as possible to avoid them from being invalid. Often, the imbalance is higher than we wanted, the number of unmatched units is high, or both. Autostrata aims to lower the trade-off between the imbalance and the number of unmatched units. Figure 5 shows a comparison of histogram binning and Autostrata.

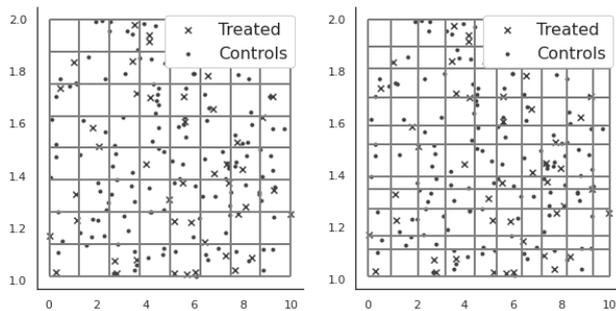


Figure 5 shows two stratifications for the same two-covariate dataset. The left plot is from histogram binning, and the right is from Autostrata. Each grid cell is a two-dimensional stratum. On the left, the strata have uniform widths. On the right, the strata widths are non-uniform.

3.1 Overall approach

This section gives an overall description of the Autostrata approach and explains its reasoning.

The generic histogram binning rules used in CEM work surprisingly well for stratification in our context. Therefore, understanding the underlying reasons is invaluable to improving the results: Any stratum containing both treated and controls is valid. Also, the sample’s total maximum imbalance will be lower if the strata are narrower. To construct strata spanning over a mixture of treated and control units, regions of common support must be present for the sample, i.e., there must be some overlap in the distributions for treated and controls. The treated and controls in regions with sparse or no overlap are further apart and more dissimilar than units in denser and more overlapping regions. Because we usually have a reasonable common support level, the uniform width strata will readily contain both treated and control units. Further, units in the sparser and less overlapping regions are more likely to be pruned, as they should. These factors contribute to why histogram binning works well. Conclusively, knowing these factors makes it reasonable to assume that much of the potential for improvement is in the regions where the distributions for treated and controls overlap most.

Autostrata’s strategy is to construct narrow strata while keeping the number of unmatched units low. The strata can be of varying widths. Having narrower strata on average is equivalent to more strata. We thus start with an initial stratification state where all possible stratum edges for all covariates are included (see section 2.5). This state represents the narrowest stratification that is relevant. All units will be in a stratum containing only a single unit or same-valued units. From there, we iteratively remove one edge at a time. This edge can belong to any of the covariate dimensions.

In its simplest form, the algorithm does not consider widths but removes edges one by one until the number of unmatched units is as low as requested by an input parameter. The main selection criterion for removing an edge, per iteration step, is the edge that gives the most significant reduction in unmatched units when removed. Removing a stratum edge for one dimension (covariate) merges one or more strata divided by stratum edges for other dimensions. Merging strata for a given covariate results in strata that are wider, so the increase in the average width of the strata for a covariate is strictly monotonic.

The crux of the algorithm is: For each stratum edge that we remove from the initial state, the average maximum

imbalance increases. If the algorithm reaches the requested maximum number of unmatched in fewer steps, i.e., by removing fewer edges, the *average maximum imbalance* will be lower than if more steps are spent. Thus, to reduce the number of iterations needed to reach the goal number of unmatched, for each iteration, we remove the edge that gives the greatest reduction in the number of unmatched, after assessing all currently remaining edges in any dimension. If several equally good options are found, the one giving the narrowest width is chosen. In Section 3.2, we describe how the widths for different covariates are scaled to be comparable.

Autostrata also provides the researcher with input parameters for more control over the resulting stratification:

- The maximum wanted numbers of unmatched treated and controls
- The maximum allowed widths between stratum edges per covariate

The researcher can specify maximum numbers of unmatched treated and controls as two separate input parameters. The stratification process will continue until reaching both numbers or until the point when there is no closer solution. For example, suppose the stratification algorithm reaches one of the requested maximum numbers of unmatched for either treated or controls. It will then continue until reaching the requested number of unmatched for the other group. It continues iterating, and the numbers can continue to improve for both treated and controls. Section 3.2 describes how Autostrata incorporates weights to account for the difference in the requested maximum numbers of unmatched treated and controls while iterating.

Further, Autostrata has a parameter for the maximum allowed stratum width per covariate, and it will not create strata wider than the given widths. If widths are not of importance, a large or infinite value can be given as input instead. The background for the maximum width parameter is that researchers may want to set a maximum difference, *caliper*, between treated and controls for the covariates—for example, max five years age difference or five points difference for a given performance score. In addition, setting a maximum width restricts the maximum imbalance. Another reason to set widths, which concerns the algorithm, is to prevent a single or a few strata from expanding too much while leaving others unchanged. Broader strata have a higher potential imbalance. It is possible to imagine that, on average, a large stratum combined with many narrow ones may somewhat cancel each other out imbalance-wise, but it is probably not what we want. A large stratum will still have a greater risk of being imbalanced. Lastly, we can use the widths produced by CEM’s histogram binning as input to Autostrata. Histogram binning only supports uniform width strata, but Autostrata can use these widths as the maximum allowed when defining strata of non-uniform widths.

3.2 Heuristics

In section 3.1, we gave an introduction to the overall approach. Autostrata is an algorithmic approach to stratifying covariates that starts with an initial state where all stratum edges are present and iteratively removes one edge at a time until the end criterion is met or no further improvements are found. Here, we describe the heuristics in more detail.

When we remove a stratum edge along the direction of one dimension (covariate), two and two strata become merged

to form new, wider strata. Removal of an edge usually results in more than two strata being merged because there are also edges along the other dimensions separating the covariate values into distinct strata. If two neighboring strata contain only treated and only controls, respectively, merging the two strata results in a stratum with a mix of both types. These units are no longer unmatched and, thus, not pruned from the sample.

Autostrata has two criteria for choosing which stratum edge to remove for each iteration. The first criterion has the highest priority, and the second criterion applies only to alternatives with equally good values for the first. The two criteria are:

1. Choose the greatest relative increase in matched treated and controls if the stratum edge is removed
2. Choose the stratum with the narrowest width

Instead of using the increase in matched units directly, Autostrata uses a weighted measure for increase. Let Δ_t and Δ_c be the increase in the number of matched treated and control units, respectively, when we remove a given stratum edge. The relative increase, Δ_{rel} , is then:

$$\Delta_{rel} = w_t \Delta_t + w_c \Delta_c$$

, where w_t and w_c are weights. The weight for the treated group, $w_t = w(t)$, and control group, $w_c = w(c)$, is found as follows:

$$w(g) = \begin{cases} \frac{m_g^{cur} - m_g^{max}}{n_g - m_g^{max}}, & m_g^{cur} - m_g^{max} \geq 0 \\ 0, & m_g^{cur} - m_g^{max} < 0 \end{cases}$$

, where g is the group, m_g^{cur} is the number of currently unmatched units for the group, m_g^{max} is the requested maximum number of unmatched for the group, and n_g is the total number of units from the group in the sample. Here, we also assume that $n_g > m_g^{max}$.

The purpose of the weights is threefold:

1. If one group is represented less than the other, each new matched unit from the group should weigh more.
2. The researcher can set parameters for how many unmatched (pruned) treated and controls are acceptable. The difference $n_g - m_g^{max}$ takes into account that the gap between available and discardable units can differ between groups.
3. If Autostrata has reached the goal for the number of unmatched units for one group, an increase in the other groups should weigh more when choosing an edge to remove. As one group comes closer to the goal, reducing the number of unmatched for the other group is prioritized higher. The difference $m_g^{cur} - m_g^{max}$ is the remaining units to match for the given group.

Width is the second selection criterion for edge removal. The widths must be scaled because Autostrata compares stratum edges from all covariates per iteration. We compute a scale factor by removing outliers and taking the min-max difference. Observations having a standard score, $|z| \geq 3$, are outliers. The data can be scaled once as an initial step. In that case, the maximum widths must be scaled as well. Also, we must restore the resulting stratum edges to the original scale. For clarity, the pseudocode in Listing 1 does not scale the data until needed.

3.3 Algorithm

Here we present the algorithm in pseudocode form. The pseudocode is at an abstraction level sufficient to implement the algorithm. However, we omit implementation details and performance enhancements that do not contribute to the understanding. Listing 1 presents the algorithm in pseudocode form, and Table 1 describes the variables used in the listing.

Variable	Meaning
tr and ct	The covariate values for the treated and the control units
Δ_{best}	The best relative increase in matched units for the current iteration
Δ_{cur}	The relative increase in matched units for currently assessed edge
Δ_t and Δ_c	The increase in the number of matched treated and controls for assessed edge
m_t^{cur} and m_c^{cur}	The current number of unmatched treated and controls
m_t^{max} and m_c^{max}	The requested maximum number of unmatched treated and controls
covariates	The covariates (dimensions)
cov	The current covariate
edges	The current set of edges, including the outer left- and rightmost edge per covariate
n _{edg}	The number of edges in the current set of edges
edges _{cov}	The current set of edges for the current covariate, <i>excluding</i> the outer left and right edges
e _{cur}	The currently assessed edge
e _{sel}	The currently best edge for the iteration and candidate for selection
e _l and e _h	e _{cur} 's lower and higher adjacent edges
width _{cur}	The scaled widths of merged strata if we remove the currently assessed edge
width _{sel}	The scaled widths of strata if removing the iteration's current candidate for best edge
widths _{max} and width _{max}	The set of maximum allowed stratum widths, and the maximum width for the current covariate

Table 1. The pseudocode variables and their meaning

Autostrata Algorithm	
1	Input: tr, ct, widths _{max} , m_t^{max} , m_c^{max}
2	Output: edges
3	Initialization of variables:
4	edges _{cov} \leftarrow one edge per distinct covariate value
5	m_t^{cur} , $m_c^{cur} \leftarrow$ calculate the initial number of unmatched treated and controls
6	Stratification:

```

7  while ( $m_t^{cur} > m_t^{max}$  or  $m_c^{cur} > m_c^{max}$ )
   and ( $n_{edg} > 0$ ) do
8     $\Delta_{best} \leftarrow -1$ 
9     $width_{sel} \leftarrow \infty$ 
10    $e_{sel} \leftarrow \text{nil}$ 
11   for cov in covariates do
12     for  $e_{cur}$  in  $edges_{Scov}$  do
13       get  $e_l$  and  $e_h$ 
14        $width_{unscaled} \leftarrow |e_h - e_l|$ 
15       if  $width_{unscaled} > width_{max}$  then
16         continue // stratum too wide
17        $width_{cur} \leftarrow scaled\_width(e_l, e_h)$ 
18        $\Delta_t, \Delta_c \leftarrow$  the difference in numbers of
unmatched (for the multi-dimensional
strata) between  $e_l$  and  $e_h$  before and
after removing  $e_{cur}$ 
19        $\Delta_{cur} \leftarrow relative\_increase(\Delta_t, \Delta_c)$ 
20       if ( $\Delta_{cur} > \Delta_{best}$ ) or ( $\Delta_{cur} == \Delta_{best}$ 
and  $width_{cur} < width_{sel}$ ) then
21          $\Delta_{best} \leftarrow \Delta_{cur}$ 
22          $width_{sel} \leftarrow width_{cur}$ 
23          $e_{sel} \leftarrow e_{cur}$ 
24       end // if
25     end // for  $e_{cur}$ 
26   end // for cov
27   if  $e_{sel} == \text{nil}$  then
28     break // no more improvements found
29   else
30     remove  $e_{sel}$  from edges
31     update  $m_t^{cur}$  and  $m_c^{cur}$ 
32   end
33   if  $m_t^{cur} \leq m_t^{max}$  and  $m_c^{cur} \leq m_c^{max}$  then
34     break // goal reached
35   end // while
36   return edges

```

Listing 1. Pseudocode for the algorithm

3.4 Implementation

A version of the algorithm corresponding to Listing 1 was implemented in Python 3.9 (<https://python.org>), with some added performance enhancements. For example, we utilize Numba (<https://numba.pydata.org>) for counting unmatched units in strata, yielding a speedup [20, p. 125] of 2.25 for the algorithm as a whole when stratifying for Dataset 3 in Table 2 on an Intel i7-8850H CPU with 12 logical cores. A far more significant performance enhancement is achieved by caching already computed results for each stratum. The same strata are visited repeatedly during the iterations, and the algorithm finishes 17.67 times faster for Dataset 1 in Table 2 when reusing already computed results. Further, strata not affected by removing a given edge are not visited unnecessarily. Lastly, only relevant units are included in computations regarding subsets of strata.

Still, there is plenty of room to enhance performance. Many of the algorithm’s computational tasks can be performed independently, e.g., the difference in unmatched units if a given edge is removed. Such independent

computations that are well suited for parallelization are often termed embarrassingly parallel [21, p. 79-98]. A systematic approach to parallelizing algorithms is found in Foster’s methodology [22]. In addition to parallelization, we can enhance the performance by designing data structures for efficient access to frequently used data and extensively reusing previously computed results in the algorithm’s iterations. For clarity, we concentrate on the basic algorithm in this paper, leaving the suggested performance enhancements to future work.

The accompanying source code for the paper is available on GitHub (<https://github.com/jo-inge-arnes/autostrata>).

4 EXPERIMENTS AND RESULTS

4.1 Datasets

A generator for synthetic data was implemented that lets us draw random samples from a composition of distributions for treated and controls. Figure 6–Figure 8 show the datasets as violin and swarm plots, and Table 2 shows the number of units and the mixed distributions for the datasets.

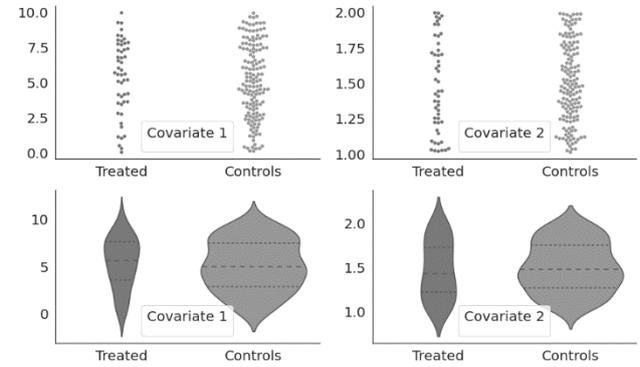


Figure 6 shows the swarm and violin plots of Dataset 1 with two uniformly distributed covariates.

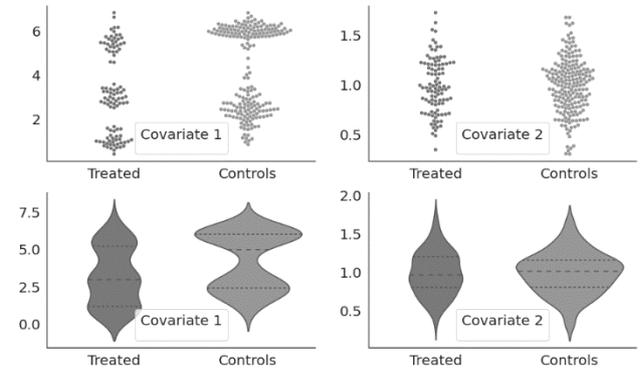


Figure 7 shows the swarm and violin plots of Dataset 2 with a mixture of Gaussians.

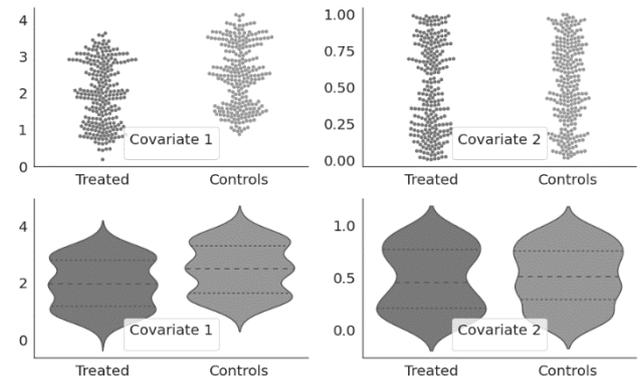


Figure 8. Dataset 3 has a mixture of Gaussians for Covariate 1 and uniform distribution for Covariate 2.

	Treated	Controls
Dataset 1	50 units	150 units
Cov. 1	U(0, 10)	U(0, 10)
Cov. 2	U(1, 2)	U(1, 2)
Dataset 2	100 units	200 units
Cov. 1	$N(1, \frac{1}{3.5})$	$N(2.5, \frac{2.5}{3.5})$
	$N(2, \frac{1}{3.5})$	$N(6, \frac{1}{3.5})$
	$N(5.5, \frac{1.5}{3.5})$	–
Cov. 2	$N(0.95, \frac{1}{3.5})$	$N(1.0, \frac{1}{3.5})$
Dataset 3	250 units	250 units
Cov. 1	$N(1, \frac{1}{3.5})$	$N(1.5, \frac{1}{3.5})$
	$N(2, \frac{1}{3.5})$	$N(2.5, \frac{1}{3.5})$
	$N(3, \frac{1}{3.5})$	$N(3.5, \frac{1}{3.5})$
Cov. 2	U(0, 1)	U(0, 1)

Table 2 shows the number of units and the mixed distributions for the datasets. U(min, max) stands for uniform and $N(\mu, \sigma)$ for normal distribution.

4.2 Experiments

To automate the experiments, we wrote Python and R scripts. The role of the R scripts is to call the CEM library. A reference manual for the CEM library is available online (<https://CRAN.R-project.org/package=cem>). In the code for the experiments, rpy2 (<https://rpy2.github.io>) is used to bridge between Python and R.

The experiments are as follows:

1. We call CEM to get pre-stratification scores and statistics for the given dataset.
2. Next, CEM is used to stratify the covariates by applying Scott’s rule for histogram binning. It also computes the number of unmatched units, imbalance scores, and other statistics.
3. We then pass CEM’s outputted number of unmatched units and stratum widths to Autostrata.
4. Autostrata stratifies the covariates.
5. Autostrata’s outputted stratum edges are given as input to CEM, which uses them to stratify and compute statistics equivalent to step 2.

Two experiments are conducted per dataset. They differ only in how the results are passed to Autostrata in Step 3:

Input type	Input parameters
P1	The numbers of unmatched treated and controls from histogram binning are passed as m_t^{max} and m_c^{max} and the bin widths are passed as $widths_{max}$.
P2	The m_t^{max} and m_c^{max} values are as in P1, but $widths_{max}$ values are set to infinity.

Table 3. Input parameters. See Table 1 for variables

4.3 Results

Table 4 shows the experiment results.

DS1, DS2, and DS3 are headers for the results of the three datasets. The top column headers stand for ‘results before stratification’ (Before), ‘stratification with histogram binning’ (Hist.), and the input types P1 and P2 from Table 3. ‘Res.’ is an abbreviation for results, and ‘Imp.’ is the percent improvement compared to histogram binning.

The row labels denote multivariate imbalance measure (MIM), total unmatched (UM_{TOT}), unmatched treated (UM_{TR}), and unmatched controls (UM_{CT}).

TOI is the percent improvement in the trade-off, which is the sum of the improvements for UM_{TOT} and MIM.

	Before	Hist.	P1		P2	
	Res.	Res.	Res.	Imp.	Res.	Imp.
DS1						
MIM	0.240	0.199	0.167	16%	0.187	6%
UM_{TOT}	0	65	80	-23%	54	17%
UM_{TR}	0	4	5	-25%	3	25%
UM_{CT}	0	61	75	-23%	51	16%
TOI	–	–	–	-7%	–	23%
DS2						
MIM	0.465	0.273	0.266	3%	0.335	-23%
UM_{TOT}	0	88	84	5%	72	18%
UM_{TR}	0	16	24	-50%	16	0%
UM_{CT}	0	72	60	17%	56	22%
TOI	–	–	–	7%	–	-5%
DS3						
MIM	0.348	0.290	0.174	40%	0.280	3%
UM_{TOT}	0	53	92	-74%	39	26%
UM_{TR}	0	21	53	-	21	0%
				152%		
UM_{CT}	0	32	39	-22%	18	44%
TOI	–	–	–	-34%	–	30%

Table 4. Results from experiments. Best TOI results per dataset are in bold and thicker cell borders.

5 DISCUSSION

Table 4 shows that both imbalance and the total number of unmatched units are lower for Autostrata for all three datasets. The input parameter type P2 gave the best results for DS1 and DS3, while P1 gave the best for DS2. The difference is that P2 sets the maximum allowed stratum widths to infinity, which effectively disables the parameter. By visually comparing the swarm plots in Figure 6–Figure 8, we see the difference between DS2 and the other two: DS2 has several regions with minimal overlap between treated and controls. As Section 3.1 explains, finding narrow strata with mixed types of units is easier in regions with high overlap. Therefore, restricting the widths is usually not necessary in such regions. Autostrata also works well for sparser overlap, but as illustrated by the experiment for DS2, setting maximum widths is more important.

Autostrata competed with CEM’s best effort in the experiments, and we passed parameters not necessarily ideal for non-uniform widths. It is possible to adjust these parameters manually or programmatically, but for objectivity, we use the unchanged output from CEM as input to Autostrata.

Lastly, Autostrata can be used stand-alone. A researcher can decide the acceptable differences between treated and controls based on domain knowledge. The researcher can also request a maximum number of unmatched units. Autostrata thus provides researchers with more up-front control. After stratification, the researcher can input the stratum edges to the CEM software as manual cutpoints. A

combination is even possible, where Autostrata stratifies a subset of the covariates given to CEM.

6 RELATED WORK

Aikens, R.C. et al. [23] present *Stratamatch*, a method for stratification of covariates for CEM. Only datasets from a minimum of 5 000 up to millions of observations are recommended. The method divides the dataset into training (pilot) and analysis sets, and the resulting strata are close to equal-sized. The size must be manually decided.

Jackson, B. et al. [24] present an algorithm for optimal data partitioning on an interval that Scargle, J.D. et al. [25] apply for astronomical time series. The algorithm supports custom fitness functions, and we tried defining a function. However, a common issue is the unwanted case of one subinterval per value; thus, the researcher must choose an expected number of subintervals. Also, while theoretically possible to extend for multivariate data, the algorithm is primarily univariate.

7 CONCLUSION

We have presented Autostrata, an algorithmic approach to stratifying confounder covariates. Autostrata shows improved results compared to the standard CEM stratification. In addition, it provides the researcher with parameters for controlling the stratification. Autostrata can be used stand-alone.

8 REFERENCE

- [1] Ranganathan, P., Aggarwal, R. “Study Designs: Part 3 – Analytical Observational Studies” in *Perspectives in Clinical Research*, Vol. 10, Issue 2, pp. 91-94. 2019.
- [2] Hariton, E., Locascio, J.J. “Randomised Controlled Trials – The Gold Standard for Effectiveness Research” in *BJOG: An International Journal of Obstetrics & Gynaecology*, Vol. 125, Issue 13, pp. 1716-1716. 2018.
- [3] Arnes, J.I., Bongo, L.A. “The Beauty of Complex Designs” in *Advancing Systems Epidemiology in Cancer: Exploring Trajectories of Gene Expression*, pp. 23-47. Scandinavian University Press, 2020.
- [4] Kim, S. “Case-Cohort Studies vs Nested Case-Control Studies” in *Datum Newsletter Division of Biostatistics*, Vol. 22, Issue 1, pp. 1-2. 2016.
- [5] Ngo, L.H., et al. “Methodologic Considerations in the Design and Analysis of Nested Case-Control Studies: Association Between Cytokines and Postoperative Delirium” in *BMC Medical Research Methodology*, Vol. 17, Issue 1, pp. 88. 2017.
- [6] Alexander, L.K., et al. “Confounding Bias, Part I” in *ERIC Notebook*, Vol. 11. 2015.
- [7] Alexander, L.K., et al. “Confounding Bias, Part II and Effect Measure Modification” in *ERIC Notebook*, Vol. 12. 2015.
- [8] Howards, P.P. “An Overview of Confounding. Part 1: The Concept and How to Address It” in *Acta Obstetrica et Gynecologica Scandinavica*, Vol. 97, Issue 4, pp. 394-399. 2018.
- [9] Wacholder, S., et al. “Selection of Controls in Case-Control Studies: I. Principles” in *American Journal of Epidemiology*, Vol. 135, Issue 9, pp. 1019-1028. 1992.
- [10] Tripepi, G., et al. “Stratification for Confounding – Part 1: The Mantel-Haenszel Formula” in *Nephron*

Clinical Practice, Vol. 116, Issue 4, pp. 317-321. 2010.

- [11] Iacus, S.M., King, G., Porro, G. “Matching for Causal Inference Without Balance Checking” in *SSRN Electronic Journal*. 2008.
- [12] Blackwell, M., et al. “cem: Coarsened Exact Matching in Stata” in *The Stata Journal*, Vol. 9, Issue 4, pp. 524-546. 2009.
- [13] Iacus, S.M., King, G., Porro, G. “cem: Software for Coarsened Exact Matching” in *Journal of Statistical Software*, Vol. 30, Issue 9, pp. 1-27. 2009.
- [14] Sturges, H.A. “The Choice of a Class Interval” in *Journal of the American Statistical Association*, Vol. 21, Issue 153, pp. 65-66. 1926.
- [15] Scott, D.W. “On Optimal and Data-Based Histograms” in *Biometrika*, Vol. 66, Issue 3, pp. 605-610. 1979.
- [16] Freedman, D., Diaconis, P. “On the Histogram as a Density Estimator: L2 theory” in *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, Vol. 57, Issue 4, pp. 453-476. 1981.
- [17] Shimazaki, H., Shinomoto, S. “A Method for Selecting the Bin Size of a Time Histogram” in *Neural Computation*, Vol. 19, Issue 6, pp. 1503-1527. 2007.
- [18] King, G., Nielsen, R. “Why Propensity Scores Should Not Be Used for Matching” in *Political Analysis*, Vol. 27, Issue 4, pp. 435-454. 2019.
- [19] Sekhon, J.S. “The Neyman-Rubin Model of Causal Inference and Estimation via Matching Methods” in *The Oxford Handbook of Political Methodology*, Vol. 2, pp. 1-32. 2008.
- [20] Pacheco, P.S. *An Introduction to Parallel Programming*. Morgan Kaufmann, Burlington, MA, 2011.
- [21] Wilkinson, B., Allen, M. *Parallel Programming: Techniques and Applications Using Networked Workstations and Parallel Computers*. Pearson/Prentice Hall, Upper Saddle River, N.J, 2005.
- [22] Foster, I. *Designing and Building Parallel Programs*. Addison-Wesley, Reading, MA, 1995. Also available from: <https://www.mcs.anl.gov/~itf/dbpp/> Accessed 2022-06-26.
- [23] Aikens, R.C., et al. “stratamatch: Prognostic Score Stratification Using a Pilot Design” in *arXiv preprint arXiv:2001.02775*. 2020.
- [24] Jackson, B., et al. “An Algorithm for Optimal Partitioning of Data on an Interval” in *IEEE Signal Processing Letters*, Vol. 12, Issue 2, pp. 105-108. 2005.
- [25] Scargle, J.D., et al. “Studies in Astronomical Time Series Analysis. VI. Bayesian Block Representations” in *Astrophysical Journal*, Vol. 764, Issue 2, pp. 167. 2013.

9 ACKNOWLEDGEMENT

We want to thank Marc Weitz for inspiring the use of swarm and violin plots.

Designing, implementing, and testing a modern electronic clinical study management system

The Hubro system

Miroslav Muzny^{1,2}, Meghan Bradway¹, Håvard Blixgård¹, Eirik Årsand^{1,2}

¹ Norwegian Centre for E-health Research, University Hospital of North Norway, Tromsø, Norway, miroslav.muzny@ehealthresearch.no

²University of Tromsø – The Arctic University of Norway, Department of Computer Science, Tromsø, Norway

Abstract

Clinical trials need to adapt to the rapid development of today's digital health technologies. The fast phase these technologies are changing today, make the clinical study administration demanding. To meet this challenge, new and more efficient platforms for performing clinical trials in this domain need to be designed. Since the process of following up such trials is very time-consuming, it calls for revisiting several of the methods for performing both randomized, and other clinical trials. We present system for electronic management of clinical studies that addresses many of the time-consuming challenges, which additionally address many of the quality assurance aspects. We also present results from testing the system in two studies with 50 and 8 participants.

Keywords

RCT, clinical study, Kubernetes, GDPR, privacy, automatic data collection

1 INTRODUCTION

Randomized control trials (RCTs) are still the gold standard for the evaluation of health interventions, including evaluating technologies being introduced into healthcare providers' clinical practice and patient care. However, because clinical trials operate off of an established standard of meticulous and manufactured testing environments, they are often unable to produce results on schedule (delayed 86% of the time), demonstrate efficacy (57% were not able to) [1] or provide patient-relevant outcomes [2]. This is especially relevant in studies involving the rapid developing digital health technologies, e.g. mobile apps, wearables and sensors for recording and self-registering of patient data.

Today's tools for managing RCTs, and clinical trials in general, are simply too slow and ill equipped to evaluate such technologies. Clinical trials need to adapt; the most time-consuming phases of trials, i.e., recruitment, participant management and data-collection, need to be streamlined. In fact, companies such as CRO Analytics [3], CenterWatch [4] and the FDA [5] have echoed that a clinical trial's adherence to their stated objectives and schedule – as a justification of funding - requires that investigators become more adaptive to real-world impacts upon their trial [6].

An increasing number of clinical trials put a strong accent on remote sensing data, self-management, and patient engagement through their smartphones. To facilitate this, researchers need to provide more advanced and direct participant support including guidance and troubleshooting compared with traditional clinical trials. Since the process of such support is very time-consuming and often difficult

to assure the quality of all adopted channels, it calls for revisiting methods for performing clinical trials, including the variety of types of studies, from feasibility studies to randomized control trials.

Two critical issues with studies involving personal smartphones and sensors are privacy and data security. The apps and software involved must fulfil several legislative requirements, where a major one was issued in July 2020 by a judgment made by the Court of Justice of the European Union (CJEU), commonly known as Schrems II [7]. This invalidated the EU-US Privacy shield and enforced the integration of additional privacy-focused measures. In Norway, Schrems II has been a showstopper for the national project *Health Analytics Platform* [8].

From the perspective of software infrastructure, Schrems II's judgment restricts which cloud vendors can be used to deliver compute-, or storage services. Even though many of these providers have widespread local data centers, their utilization is limited by the fact that a use of these services may liaise with auxiliary services in the technology stack of the cloud provider (e.g. authentication service) that may send metadata to US-based servers.

The range of public cloud providers is enormous, and the spectrum of their managed services significantly facilitates integration of the system in various aspects. The convenience of use may also act as a double-edged sword due to violation of regulations like Schrems II. Therefore, it is necessary to validate compliance with legislative requirements not only with regards to GDPR, but also with other regulations that may influence the software architecture design and implementation.

The 18th Scandinavian Conference on Health informatics, Tromsø, Norway, August 22-24, 2022. Organized by UiT The Arctic University of Norway. Conference Proceedings published by Linköping University Electronic Press at <https://doi.org/10.3384/ecp187>. © The Author(s). This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>

For study administration, researchers typically have to use several different disjointed tools for each stage of the project. Because these tools are siloed, they do not communicate directly, leaving the researcher to separately access, format and translate the information between systems. An example from our own research portfolio demonstrates this point well. In 2010-2013, our research team was part of the Norwegian arm of the REgIONs of Europe WorkINg together for HEALTH (RENEWING HEALTH) EU-project, a three-armed prospective randomized controlled trial (RCT) with data collection from questionnaires, a mobile phone app, and qualitative interviews [9]. Individuals with type 2 diabetes (T2D) were recruited through their general practitioners (GPs) using paper handouts for informed consent, participation and study information. The intervention involved participants using a diabetes app on a smartphone to self-manage their diet, physical activity, goals and blood glucose over the course of a year. Randomization was performed using a separate system, the WebCRF (Case Report Form), through an external group. Participants were trained on the use of the app during in-person meetings. The data that participants registered in the app and their usage logs were automatically and continuously and stored into a secure server. Self-reported outcomes were collected via standardized questionnaires in paper formats, and clinical measures were manually reported by GPs from medical records, documented in a paper-based study protocol. This is all in addition to patients' feedback and healthcare perspectives collected from the GPs via interviews [9]. The main challenges were the coordination of using all these different systems; we needed to open a key-locked cabinet to access the names and addresses when needing to send out invitations to meetings, and when we wanted to connect the different extracted data, from a look-up table of the different IDs of the different data sources. We also had to

put up a separate server solution to log the user-recorded data, and their usage logs. In some cases, automatic data transfer was not completed, e.g. when users were not online at the right times for scheduled data transfer. We, therefore, required patient participants to meet in-person to download their data via cable. These processes and work-arounds took much time and was a source of errors in organizing the participant information and analysis.

The challenges that we as researchers experienced in this project helped to inspire the work described in this paper.

In this paper, we describe the iterative design, development and testing of a study administration system called the Hubro platform. The purpose of this work was to develop a single tool that would allow a researcher to administer all stages of a study using one platform instead of multiple. The aim of such a system would be to increase the efficiency, reduce errors, and give a better overview and control to the researchers performing studies involving digital health technologies. As well as an ease way to follow-up participants in the different stages of the clinical trial.

2 METHODS

We aimed to design a system that enables the researchers to do all necessary tasks in a clinical trial, including RCT, more efficiently and from one user interface – called Hubro.

Hubro was designed and developed primarily to address the need to perform studies dynamically and fully electronically. These tasks include recruitment, obtaining informed consent, randomization, distributing polls/questionnaires, distributing any intervention like an app or other software, follow-up of participants, data gathering, data analysis and closure of the study (Figure 1).

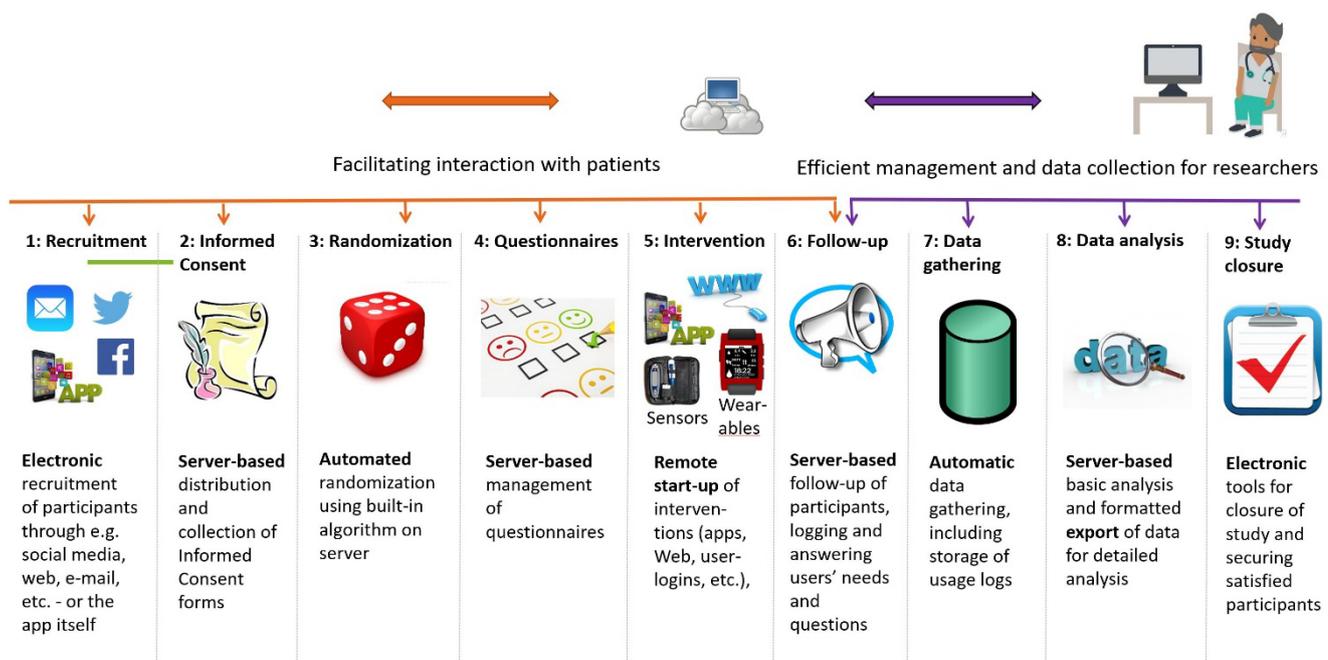


Figure 1 The introduced system for more efficient study management for clinical trials – the Hubro platform.

The original intention was to develop a secure questionnaire data collection platform that would honor requirements by the Regional committees for medical and health research ethics (REK) for handling of personal data.

2.1 Platform design and development

We used an iterative and fluid development process involving developers and researchers with experience conducting intervention studies using digital health technologies. During the design and implementation stage, meetings were conducted between developers and researchers to 1) share past experiences with intervention study administration, 2) identify bottlenecks and 3) brainstorm solutions for the research-users' needs for conducting efficient and effective digital health intervention studies. In addition, it was a comprehensive process with communication with the Regional Ethical Committee (REK), and the Norwegian Data Protection Authority (Datatilsynet), to find solutions in order to get acknowledgement for the platform.

2.2 Platform evaluation

In the evaluation stage, the system was practically tested by using it to administer two studies using mobile health (mHealth) smartphone applications for diabetes self-management. Information from researchers who conducted the studies, was collected via a questionnaire that queried specifics about how the system was used, impressions of the system, challenges, and suggested improvements. Table 1 details the questions included.

Number	Question
1	Can you describe your role/competence when using the Hubro system?
2	What functions are you missing?
3	About which functions do you think should be implemented better/in a more user-friendly way?
4	Have you experienced any problems with the system?
5	Have you used the possibility to manage the study from the smartphone?
6	Do you find the user interface intuitive?
7	Have you got any positive/negative feedback from the study participants?

Table 1. Questionnaire to research staff who used the Hubro system during two intervention studies

Over the course of the studies, e-mails were sent to developers and informal in-house meetings were conducted to trouble-shoot use of the system in practice. Free-text responses from the questionnaires and experiences reported over the course of the studies by the research-users regarding the identified missing functions, functions that need improvement and problems with the Hubro system were summarized separately. Thematic qualitative analysis was used to identify common themes of information.

3 RESULTS

3.1 Results of the iterative development of Hubro

During the design and development process, the researchers shared with us their experiences, which lead to the addition of user-centered requirements. Users expressed the need to collect not only questionnaire data but also study participants' own self-reported data (e.g. health-related data that were recorded automatically or manually into an mobile phone application) and usage data (i.e. the data that are collected automatically and that describe patient's interaction with the application). During the development process, we as researchers identified the need to randomize and distribute patients into treatment groups and send targeted messages via e-mail or via the mHealth application itself. At the end of the development process, the software portfolio consisted of multiple self-developed components, but also included multiple 3rd party open-source software components, which are used to perform specific part of data collection process and for which re-implementation would be out-of-the-scope of the project (e.g. the questionnaire data collection software LimeSurvey [10], and the usage data collection software Piwik/Matomo [11]).

Since multiple independent services gathered the personal health information (PHI) in Hubro, it would be problematic to implement a link table operating over all these services. Instead, the approach we chose was to create an alternate identity linked with the real identity, specific study and treatment group. Therefore, implementation of a linkage table in Hubro is based on two directory services (LDAP servers) and a translation service between them. This solution was approved by REK and The Norwegian Data Protection Authority (DPA) [12] and was used across the whole system in two clinical trials at the University Hospital of North Norway (UNN).

The implementation was done solely using in-house developers, who facilitated the development process by continuously revalidating implementation of relevant user stories, worked out in close collaboration with the researchers. The system itself does not depend on any external 3rd party service and is packaged as a self-contained bundle. Consequently, the system can run on a self-serviced IT infrastructure, but also in a cloud environment if necessary.

The Hubro system consists of multiple loosely coupled applications and server components, making it a candidate for provisioning through the SaaS model [13] i.e. a software distribution model in which applications are hosted by a vendor or service provider.

To rule out potential legislative non-compliance, we have designed the system integrating essential services, not relaying on any services from the public cloud vendors. We have based the solution on Kubernetes, which is a popular open-source container orchestration platform, (Figure 2). Use of Kubernetes simplifies the deployment and provisioning process, as it becomes a standard control plane for scheduling and managing jobs in highly distributed environments. By using vanilla Kubernetes cluster, we also prevent vendor lock-in with a particular cloud provider.

Pseudo anonymization of data is secured by a translation function, that performs ad hoc mapping of anonymous identities with e-mail addresses-based list of registered participants. Translation function is accessible from all

microservices within the cluster, and the access needs to be confirmed by password or by using a service principal i.e., managed identity.

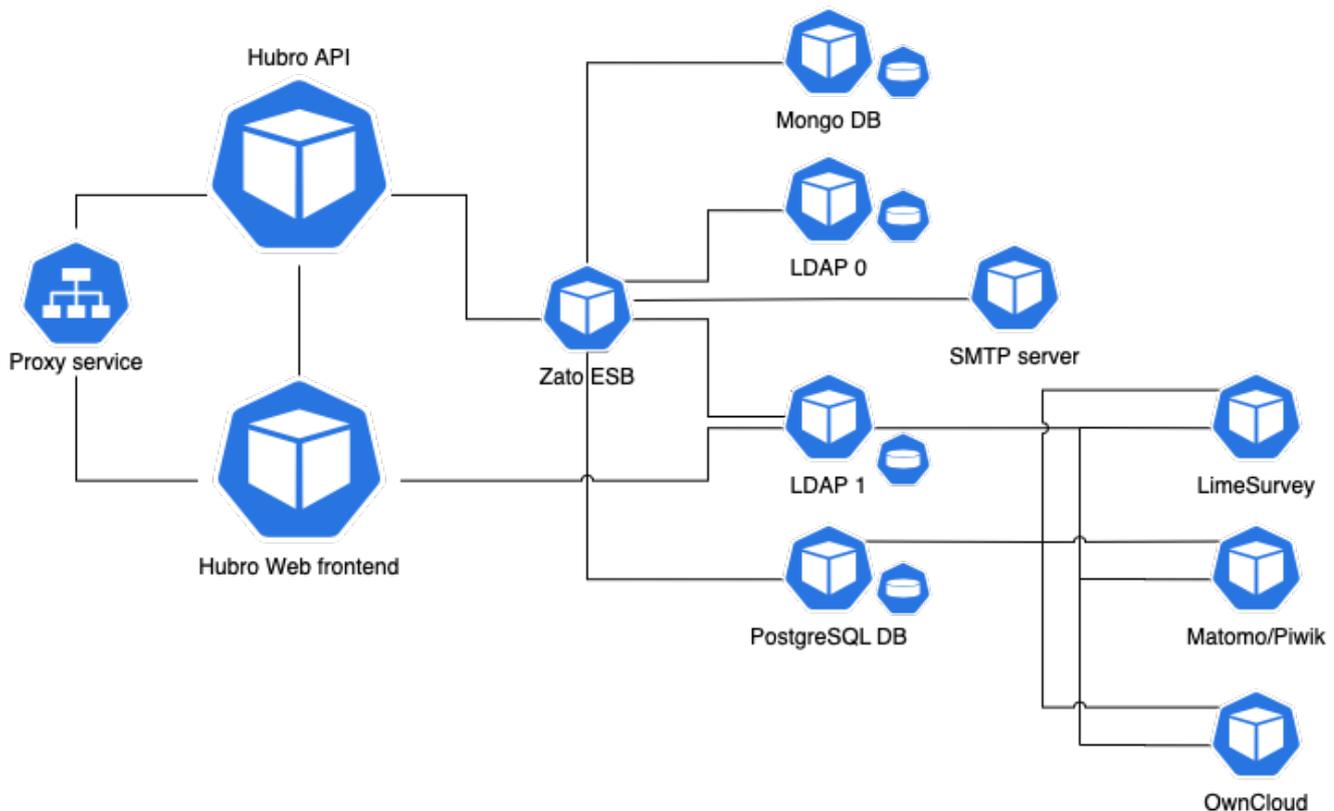


Figure 2 Kubernetes cluster diagram visualizing individual Docker containers that are sealing off services of Hubro.

One of the primary goals was to avoid an exhausting time dedication to study management as the number of participants increases. Therefore, we designed the user interface in a way, that the researcher can discover actions which may require his attention with a glance, using logical color coding (red, yellow and green). We introduced an informational column for each participant, that summarizes the following information:

- date and time when the app installed on a participant's phone checked for new messages

- indication, whether there are any unfilled questionnaires
- date and time when the participant was enrolled in the study
- date and time when the data was uploaded to the Hubro server

A general overview of participants' statuses is depicted in Figure 3.

Project group: Group 1 - Control

Group's questionnaires

A. Spørreskjemapakke ved oppstart av studien
B. Spørreskjemapakke etter 3 måneder - Gruppe 1

Group's file areas

No file areas in project group

Group's members

Name	Consent form: 0. Informert samtykke (Participant admin)	Questionnaire: A. Spørreskjemapakke ved oppstart av studien (Participant admin) (Invite remaining)	Questionnaire: B. Spørreskjemapakke etter 3 måneder - Gruppe 1 (Participant admin) (Invite remaining)		
anon-53852 Added 26. mars 2017 20:49	Completed - 27. mars 2017 12:30 View response	Completed - 27. mars 2017 22:38 View response	Not participating	Add user	Actions ▾
anon-45795 Added 26. mars 2017 19:46	Completed - 26. mars 2017 21:01 View response	Not completed - Last reminder sent: 29. mars 2017 18:23 Send reminder	Not participating	Add user	Actions ▾
anon-28253 Added 24. mars 2017 12:22	Completed - 25. mars 2017 10:42 View response	Completed - 25. mars 2017 12:53 View response	Not participating	Add user	Actions ▾
anon-42634 Added 23. mars 2017 10:18	Completed - 23. mars 2017 11:32 View response	Completed - 24. mars 2017 08:26 View response	Not participating	Add user	Actions ▾

Figure 3 Treatment group management as viewed by study manager in the Hubro Web front-end.

One of the aims was to provide an easy-to-use administration interface, eventually usable also from a smartphone, on-the-go if needed. Based on these requirements, we have chosen a web application as a user interface for the Hubro.

The user interface of the web application is responsive, i.e. it scales nicely on small device screen sizes. This way, the study manager can interact with the system dynamically through the smartphone or tablet at every place with internet access. This way, the researcher can operate the system on-the-go e.g. randomize users or send messages.

Messages in the Hubro system can be distributed through various channels. So far, the Hubro system supports distribution of information via an ordinary e-mail, or/and through the integrated REST API, from where the messages can be pulled by individual applications connected to the platform. For the messages delivered via REST API, we have not implemented an option for a direct response, and therefore the messages are only one-way function. Lack of option for direct response is however supplemented by a possibility to include formatted text enriched by HTML tags. This way, the study manager, can include images or website links containing additional information or interactive elements such as forms.

3.2 Results of the Hubro platform practical implementation

We were able to design a full functioning system that fulfilled most required tasks for a clinical trial, as presented in Figure 1. We received all the necessary acknowledgments from the national data protection

authority DPA, the regional ethical authority REK, and the local security and privacy team at the University Hospital of North Norway (UNN) to run two clinical trials.

The first trial in which Hubro was used included 50 patient participants and the second included 8 participants. In each study, participants were asked to test new functions of mHealth tools, especially the Diabetes Diary self-management application.

The first study [14] lasted for 12 months, where the study manager estimated the following amount of time on each task per participant: Informed consent delivery and collection (2-minutes); Randomization (1-minute); Delivery of the Initial questionnaire (1-minute); App distribution (4-minutes); Mid-study questionnaire (1-minute); and Final questionnaire (1-minutes). Minutes spent logging into the system, checking participant status, sending questionnaire reminders and other tasks were estimated to triple these times. In total, 30-minutes per user. The second study [15] was conducted for a 6 months period, and time usage for this study is currently being estimated in an ongoing follow-up study [16].

3.3 Results of evaluation Research-users' experiences using Hubro

The questionnaire about research-users' experiences was distributed to three researchers who had directly worked with the Hubro system, throughout the two studies.

- 1) The questionnaire has been filled in by 3 respondents –PI/project manager, researcher and system developer. PI/project manager and researcher have been involved

in all phases of the two studies, while the system developer has provided technical support (service availability assurance and data extraction support).

- 2) Based on the feedback from the researchers, the following new functionalities and system updates would be appreciated: integration of reminders to the researchers; integration with more advanced recruitment capabilities other than those currently implemented; recruitment scheduling and monitoring; notifications; more secure implementation of the messaging function.
- 3) The previously mentioned connection between the systems' user-ID (anon-ID) and data in third-party systems (Piwik/Matomo, LimeSurvey) was confusing when performing various operations such as fetching user data or exporting usage logs. Two respondents raised this issue. Another concern was expressed about the utilization of the third-party tools – the app usage analytics platform (Piwik/Matomo) and the survey platform (LimeSurvey), specifically questioning their choice due to their complexity and low user-friendliness. These tools are complex, designed to suit a wide range of use cases, and therefore they are coupled with a variety of options and settings, that might be confusing for researchers/study administrators, who only need to use a specific subset of these functions. Although, within the narrowed portfolio of alternatives, these tools stand out among other choices in terms of integration possibilities.
- 4) Researchers mentioned the following optimizations to minimize manual efforts and frequent check-ups in the following ways - a better process of keeping track of the recruited patients, i.e. placing them into groups; more understandable user-interface; easier distribution of questionnaires and tools (applications); more streamlined process of questionnaire completion checks, and a user interface to interact directly with the LDAP servers (e.g. removing participants).
- 5) From the researcher's point of view, it was suggested to implement a two-way messaging function, which would mitigate the necessity to use an ordinary e-mail as a primary communication channel from patients to researchers. Also, the user interface for user's look-up based on their e-mail addresses was perceived as a candidate for improvement, as it required a manual effort in terms of e-mail client utilization as discussed within Question 2. Terms *Texts* and *Messages*, used in the Hubro system, were seen as potentially misleading and confusing for an uneducated researcher using the system. From the patient's perspective, a case was reported when a patient was able to only fill half of the questionnaire, due to accidentally clicking on send button. Re-submitting the un-finished questionnaire (i.e. an edit function) should thus be implemented.
- 6) PI/project manager was able to use the Hubro functions to recruit and randomize study participants through the smartphone, e.g. when travelling. The Hubro system and its third-party components are web-based and responsive so that they can be used with mobile devices such as tablets or smartphones.
- 7) Mixed opinions have been reported on the user-friendliness of Hubro's user interface. All agreed that it was a complex and steep learning curve for both

main third-party components, the Piwik/Matomo and LimeSurvey. The complexity of these tools was discussed in question 3.

Researchers stated that no negative feedback has been reported about Hubro from the study participants.

In addition, we present a summary of identified missing functions, functions that need improvement and problems with the Hubro system in Table 2.

Response categories	Summarized responses
Appreciated functions	<ul style="list-style-type: none"> • Reliability • Cost-effectivity • Convenience of participant management
Functions that should be implemented in a more user-friendly way	<ul style="list-style-type: none"> • Easier user-interface • Integration of reminders • Integration with more advanced recruitment capabilities other than those currently implemented • Recruitment scheduling and monitoring notifications • More secure implementation of the messaging function
Problems with the Hubro system	<ul style="list-style-type: none"> • User's look-up (deanonymization) based on their e-mail addresses requires much manual effort • Should allow editing and re-submission of questionnaires

Table 2 Summary of identified missing functions, functions that need improvement and problems with the Hubro system.

As mentioned, most of the functionalities were implemented and used. However, data analysis was only performed to a limited degree using Hubro. Data extraction, organization and analysis was performed manually through a collaboration between a researcher and a developer. This took several months (not effective time, but to coordinate) and data needed to be shared via a separate secure research platform. The secure data-sharing platform required additional and un-planned set-up of the files and granting access. The bulk of the un-planned work included the researcher learning three different formats of anonymous data as well as how to reformat, translate and analyse them using a more well-known program such as Microsoft Excel. This reinforced the challenges as well as opportunities of developing the Hubro system further, to support also more of the last parts of clinical trials, analysis and conclusions.

4 DISCUSSION

Considering the nature of health data, rolling the Hubro system out as a service to end-users might be troublesome, due to compliance matters. Also, a naïve, on-premise distribution might be complicated because of many onboarded components resulting in a significant workload on IT system operators. The way for seamless delivery and

operation could be a use of self-sealed container-based architecture orchestration through the management system like Kubernetes, bringing in the best of both worlds – an easily manageable and configurable system running on hardware located within the premises of the institution, or within a cloud computing environment.

The concept of a user ID and anonymous ID was introduced to provide study participants with a way to quickly join the study, i.e. by providing a 5 characters identifier (user ID) that was connected fully qualified GUID-based identifier (anonymous ID) in the database. Both identifiers have to be used when performing basic operations such as, e.g. extracting user data from the database, deanonymizing the user, which introduces an additional level of manual effort. In future, this concept should be revisited and implemented in a more user-friendly manner.

It is also important to note, that Hubro, currently, neither collects nor generates documents formally required by regulatory authorities or compliance protocols i.e. Trial Master File (TMF) or its electronic equivalent (eTMF), that must be maintained alongside the electronic study management system separately in a dedicated repository. Similarly, the Hubro keeps a very limited audit trail of specific actions, that require access to the link table, or the implemented link function respectively (deanonymization).

5 CONCLUSION

The motivation behind the need to move to new study management methods is driven by the fact that traditional methods for clinical interventions may provide outdated results, given the fast pace of technological advances. Therefore, it is necessary to react on current trends and evaluate these changes promptly. Inadequate old methods are used on new types of interventions that involve the processing of vast amount of data, dynamic support and troubleshooting of potential participant's technical problems during the study and inclusion of various types of new technology. Commonly used clinical trial study management systems are too scattered and time-consuming, to provide results in a reasonable time as they are not systematically addressing the identified challenges.

The system we designed proved to improve areas of recruitment, enrollment, engagement, and retention of participants into an RCT and a clinical trial. Use of the system objectively improved the time spent managing individual patients in the study. Also, the primary investigator (PI) stated the feedback from audience was overly positive when presenting the system to other researchers at conferences, workshops, and other occasion. Although there are several other products available out on the market, none of them is integrating such a generic toolset, aiding the process of patient recruitment, randomization, survey and consent distribution, patient follow-up, data processing and trial closure. Inconclusive clinical trials, because of clinical trials' disrupted nature, may result in a delay of development of optimal treatment practice, and/or an additional financial burden.

6 REFERENCES

1. Fogel, D.B., Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: a review. *Contemporary clinical trials communications*, 2018. **11**: p. 156-164.
2. Heneghan, C., B. Goldacre, and K.R. Mahtani, *Why clinical trial outcomes fail to translate into benefits for patients*. *Trials*, 2017. **18**(1): p. 1-7.
3. Peter Malamis Precision Medicine Deserves Precision Trial Optimization. 2016.
4. CenterWatch TrialNetworks adds visual intelligence to clinical trial optimization system. 2014.
5. *Adaptive Design Clinical Trials for Drugs and Biologics Guidance for Industry*, Food and Drug Administration, Editor. 2019, Center for Biologics Evaluation and Research and Center for Drug Evaluation and Research: fda.gov.
6. Jennifer Goldsmith, *It's Time to Transform Clinical Trial Operations*. *Applied Clinical Trials*, 2016. **25**(8).
7. Mildebrath, H. *The CJEU judgment in the Schrems II case*. 2020 29. [2022; 2]. Available from: [https://www.europarl.europa.eu/RegData/etudes/ATAG/2020/652073/EPRS_ATA\(2020\)652073_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/ATAG/2020/652073/EPRS_ATA(2020)652073_EN.pdf).
8. The Directorate of e-health. *Setter arbeidet med Helseanalyseplattformen på pause*. [Web page] 2021 [cited 2022 29.06]; Available from: <https://www.ehelse.no/aktuelt/setter-arbeidet-med-helseanalyseplattformen-pa-pause>.
9. Ribu, L., et al., Low-intensity self-management intervention for persons with type 2 diabetes using a mobile phone-based diabetes diary, with and without health counseling and motivational interviewing: protocol for a randomized controlled trial. *JMIR Res Protoc*, 2013. **2**(2): p. e34.
10. LimeSurvey. *LimeSurvey: The Open Source Survey Application*. 2022; Available from: <https://www.limesurvey.org>.
11. Matomo, *Piwik is now Matomo – Announcement*. 2018, matomo.org: matomo.org.
12. Datatilsynet. *Datatilsynet*. 2022; Available from: <https://www.datatilsynet.no/en/about-us/>.
13. Cusumano, M., *Cloud computing and SaaS as new computing platforms*. *Communications of the ACM*, 2010. **53**(4): p. 27-29.
14. Årsand, E. *Tailoring Type 2 Diabetes Self-Management*. 2016; Available from: <http://ehealthresearch.no/en/projects/tailoring-type-2-diabetes-self-management>.
15. Bradway, M., et al., Measuring the Effects of Sharing Mobile Health Data During Diabetes Consultations: Protocol for a Mixed Method Study. *JMIR research protocols*, 2020. **9**(2): p. e16657.

16. Norwegian Centre for E-health Research. The need for updated evaluation approaches for e-health and m-health interventions - a dynamic concept for more efficient trials. 2022.

Who are the “Hard-to-Reach” groups in chronic-health and health technology research?

A scoping review

Meghan Bradway¹, Henriette Lauvhaug Nybakke^{1,2}, Stine Agnete Ingebrigtsen^{1,2}, and Kari Dyb^{1,2}

¹ Norwegian Centre for E-health Research, University Hospital of North Norway, Tromsø, Norway, Meghan.Bradway@ehealthresearch.no

² University of Tromsø – The Arctic University of Norway, Department of Social Science, Tromsø, Norway

Abstract

Many in health and technology research opt to focus on those who are already “engaged”, not those who are considered “hard-to-reach” or “unreached”. This exacerbates the digital divide and inequity in healthcare. We report findings of a scoping review of literature in PubMed/Medline from 2000-2022. 90 of 795 articles were identified based upon 1) the health researchers’ recruitment of unreached groups with a chronic condition for active participation in a study and 2) clear definitions of who they considered to be unreached. Findings support the need for interdisciplinary and community-level involvement to reach and include unreached groups in health studies.

Keywords

Chronic, digital health, unreached, underserved, study participation

1 INTRODUCTION

The words we use affect how we perceive and react to the world. The term “hard-to-reach” was first used in the 1970’s to describe the police officer’s stigmatizing perceptions of gay, lesbian, and bisexual individuals [1]. It was popularized in social marketing, referring to those who are more time and financially expensive to engage in whatever social intervention is offered [2]. In today’s healthcare setting, the connotations we apply are still and similarly stigmatizing; terms such as “hard-to-reach” or “unengaged” are used to primarily describe those who are non-compliant with or non-receptive to treatment [3]. This implies a lack of effort or blame, which affects how researchers and healthcare providers approach these groups. However, the situation is much more complex than a person simply choosing not to follow a doctor’s orders. Reasons for not engaging in the care they need could be within or outside out of their control, and everywhere in between. This is reflected in the different ways these terms are used to describe different groups within the population.

Within the context of chronic conditions - broadly defined as “continuing or occurring again and again for a long time” by Bernell and Howard [4] - the consequences of belonging to a “hard-to-reach” group are more cumulative compared to those of someone with an acute health condition. If health services and resources are not accessible to someone with a chronic condition, symptoms may be experienced in the short-term as well as more severe complications that are more costly to both the individual and the healthcare system, in the long-term. The number of preventable hospitalizations for those with an acute condition are half the number than for those with chronic condition in the US. Potentially preventable hospitalizations have also been cited as highly associated with social indicators of health [5]. These numbers also vary considerably by country [6],

suggesting a community-level impact to preventative care and treatment.

If health research does not represent individuals living with chronic conditions who are hard-to-reach, here-to-for referred to as unreached, the consequences are interventions that do not address the needs and contexts of these unreached groups.

In this paper, we present the results of a scoping review of literature describing those whom health research considers unreached, who also have a chronic condition. To the best of our knowledge, a review that focuses on how researchers describe unreached groups has not been performed. Recently, factors from environmental to societal infrastructure have changed rapidly. Climate, technological and socio-political activity changes affect our resources, social interactions and priorities, including how we interact with the healthcare system [7]. Therefore, it is prudent to look at who we consider to be unreached in chronic care and what contributes to them being unreached, to properly inform intervention development, testing and implementation. This work is completed as part of the project Watching the Risk Factors (WARIFA): Artificial intelligence and the prevention of chronic conditions [8].

2 METHODS

2.1 Literature search strategy

We followed the PRISMA-ScR checklist to perform the scoping review [9]. In March 2022, PubMed (including Medline) was searched using the following terms: (Recruit* OR Participa*) AND (“Hard-to-reach” OR “Difficult to reach” OR Hidden OR Underserved OR Disadvantaged OR Marginalized OR Unengaged) AND (Population* OR Group*) AND (Healthcare OR “healthcare services” OR health) AND (intervention OR survey OR study OR trial) NOT (Adolescent OR child OR children OR infant OR youth). Due to the broad, “catch-all” search strategy, one database was initially used, with the intention of involving

The 18th Scandinavian Conference on Health informatics, Tromsø, Norway, August 22-24, 2022. Organized by UiT The Arctic University of Norway. Conference Proceedings published by Linköping University Electronic Press at <https://doi.org/10.3384/ecp187>. © The Author(s). This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>

another if saturation, i.e., of descriptions used to describe unreached groups, was not reached.

2.2 Article review and inclusion

All references were uploaded into Rayyan [10]. A trial review of 10 titles and abstracts was conducted between three of the co-authors (MB, HLN, and SAI) to determine interrater agreement (Fleiss' kappa, κ_f) [11]. The interrater agreement was strong ($\kappa_f=0.85$). Disagreements and uncertainties regarding inclusion were resolved before data-extraction from full-texts.

Articles were included if they described any type of study that 1) specifically recruited unreached groups with chronic conditions to actively participate, 2) were conducted from 2000-2022, 3) were written in English, 4) included only adults, and 5) explicitly described why they described their recruited groups as unreached. Reviews and protocols without results were excluded, as well as articles that did not describe recruitment strategies specifically for the group they define as unreached. A protocol is not registered but exists and can be made available upon request. Quality assessment of the articles was not performed because this review focused only on the definitions or characteristics used to describe the unreached populations.

2.3 Data extraction

The following data were extracted from each article: year, country, study type, intervention type, health condition, group(s) recruited, number recruited and definition of that/those group(s) (i.e., description of why the target group was considered unreached).

“Definitions” of unreached groups were considered by the authors to be an explanation or characteristics of why certain groups of people were not receiving the care that they needed or were not achieving their health goals. The exact text used to describe these groups and their definitions were cut and pasted into a common document and underwent a thematic content analysis.

3 RESULTS

Of the 795 articles identified in the literature, 158 were included for data extraction. In this paper, we focus on the 90 articles that recruited groups with chronic conditions, who were described as unreached (Figure 1). We present a summary of the terms the researchers used to describe the unreached groups as well as factors that answered the question “why is this group(s) not receiving the care that they need?”. Note that some terms, e.g., socioeconomic status or rural residence/geography, may appear in both the group type and definition, based on the descriptions given by the authors.

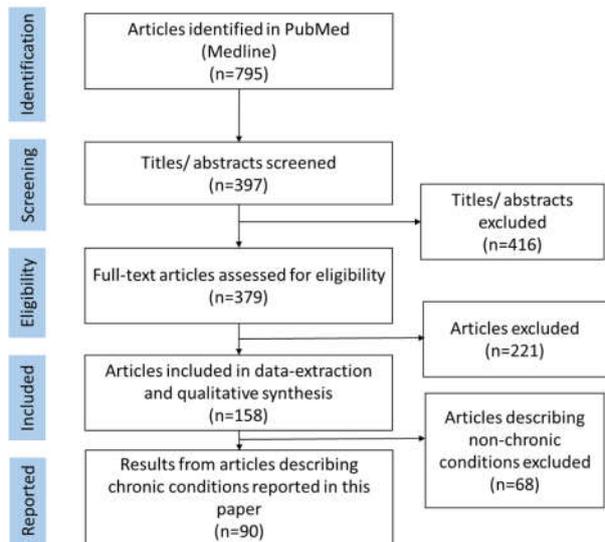


Figure 1 PRISMA flow diagram.

3.1 Articles focusing on chronic conditions

More than 76.6% of the articles that focused on chronic conditions were from one country, the United States ($n=69$), followed by Australia ($n=4$) and the UK ($n=4$). Nearly 66.6% (60/90) of the articles described randomized controlled trials (RCTs), followed by feasibility studies (6/90) and quasi-experimental studies (6/90). In terms of the intervention types, to which groups of unreached individuals with chronic conditions were recruited, half were programs (45/90), i.e., interventions coordinated via healthcare services or run by healthcare personnel, followed by telehealth/eHealth interventions (15/90), i.e. those that used technology initiated by/controlled by healthcare personnel, and screenings (10/90), i.e. those that offered screening services offered by healthcare professionals, primarily for cancer. Only three studies described mHealth interventions, i.e., those that used self-management technologies initiated by end-users. Cancer was discussed in 28.8% (26/90) of the articles, followed by diabetes (22/90), mental health conditions (21/90), and cardiovascular conditions (13/90).

3.2 Recruited group types

The results of the qualitative analysis of the text used to describe the groups recruited to the studies are summarized in Table 1. One group may be described by more than one category, indicating the complexity of group needs or deprivations. The most common descriptions of groups were those described as *People from racial/ethnic minority groups* ($n=42$), followed by *Residents of deprived, medically underserved, or rural areas* ($n=38$), and *people of Low socioeconomic status and/or uninsured* ($n=30$), and *Women* ($n=13$).

Health condition category	Recruited groups by category [reference]
Cancer	<p>Low socioeconomic status and/or uninsured [12-21]</p> <p>People from racial/ethnic minority groups [13-15, 18, 19, 22-31]</p> <p>People who are not compliant/up to date with health recommendations [18, 20]</p> <p>Residents of deprived, medically underserved, or rural areas [32-37]</p> <p>Seniors [14]</p> <p>Women [13, 16, 21, 34, 35]</p>
Cardiovascular conditions	<p>Immigrants, migrant/transient workers, refugees, or manual laborers [38]</p> <p>Low socioeconomic status and/or uninsured [39-43]</p> <p>People from racial/ethnic minority groups [38, 40, 44]</p> <p>Residents of deprived, medically underserved, or rural areas [40, 41, 45-49]</p> <p>Women [39, 45, 47]</p>
Chronic conditions (non-specific)	<p>Immigrants, migrant/transient workers, refugees, or manual laborers [50]</p> <p>People from racial/ethnic minority groups [50]</p> <p>Residents of deprived, medically underserved, or rural areas [51]</p>
Chronic obstructive pulmonary disease (COPD)	<p>Criminal offenders and/or people who use illicit drugs [52]</p>
Diabetes (non-specific)	<p>Low socioeconomic status and/or uninsured [41, 53-56]</p> <p>People from racial/ethnic minority groups [53, 54, 56, 57]</p> <p>People who are socially isolated or unengaged [54]</p> <p>People with underserved chronic condition [58]</p> <p>Residents of deprived, medically underserved or rural areas [41, 56, 57, 59-63]</p> <p>Women [53]</p>
Diabetes (Type 2)	<p>Immigrants, migrant/transient workers, refugees, or manual laborers [64]</p> <p>Low socioeconomic status and/or uninsured [64-66]</p> <p>People from racial/ethnic minority groups [64, 66-71]</p> <p>Residents of deprived, medically underserved or rural areas [49, 64-66, 69, 71, 72]</p>

Disability	<p>People who are socially isolated or unengaged [73]</p> <p>People who have been institutionalized, disabled and/or are dependent on others [73, 74]</p> <p>Women [73]</p> <p>Criminal offenders and/or people who use illicit drugs [75, 76]</p> <p>People from racial/ethnic minority groups [77]</p> <p>People who are not compliant/up to date with health recommendations [78]</p> <p>People without permanent homes [79]</p>
Hepatitis	<p>Veterans [78]</p>
Kidney disease	<p>People from racial/ethnic minority groups [80]</p> <p>Residents of deprived, medically underserved, or rural areas [80]</p> <p>Criminal offenders and/or people who use illicit drugs [81, 82]</p> <p>Immigrants, migrant/transient workers, refugees, or manual laborers [83]</p> <p>Low socioeconomic status and/or uninsured [56, 81, 83-87]</p> <p>People from racial/ethnic minority groups [56, 82, 88-94]</p> <p>People who are socially isolated or unengaged [83, 84]</p> <p>People who have been institutionalized, disabled and/or are dependent on others [95, 96]</p> <p>People with (serious) mental health conditions [82, 97-100]</p>
Mental health conditions	<p>Residents of deprived, medically underserved, or rural areas [56]</p> <p>Women [84, 85, 98]</p>
Multiple sclerosis	<p>People with underserved chronic condition [101]</p> <p>Criminal offenders and/or people who use illicit drugs [82]</p>
Alcohol, tobacco, or substance use/abuse	<p>People from racial/ethnic minority groups [82]</p> <p>People with (serious) mental health conditions [82]</p>

Table 1 Types of recruited unreachable groups or populations, by health condition.

3.3 Definitions

The most common definitions of unreachable groups with chronic conditions were based upon limitations due to *Healthcare system infrastructure*, *Socioeconomic status/factors*, and *Engagement in healthcare system*.

The definition category of *Healthcare system infrastructure* included the following factors; access to and availability of culturally appropriate and relevant healthcare services/resources, complexity of medical vocabulary and care pathways, healthcare quality and treatment options as well as racism experienced during health consultations from

providers and care coordination and continuity. *Socioeconomic status/factors* included income, education, and insurance as well as migrant-related work, and stressors of living in poverty. *Engagement in healthcare system* included factors such as adherence to recommendations, patient-provider interactions, and relationships as well as screening, other prevention and treatment seeking behaviours and use of healthcare resources.

Also of note were the community, social and political-level factors. *Community resources and setting* included factors such as inherent availability of community resources, socioeconomic status of the community, food security, place of residence or geography and history of community industrial downsizing. *Psychosocial/cultural factors* included culturally based beliefs toward health, e.g., cultural norms that do not support health recommendations, shame, stigma, social isolation, and social capital, i.e., an individual's level of support and participation in a community. *Socio-political factors* included devaluation of their culture, immigrant or minority status, marginalization, discrimination, and risk of deportation.

Of the more internal definition categories were *Personal skills/capacity, and logistic constraints* which included responsibilities of parenthood and caregiver status as well as skills to prepare fresh food, if available, and language. *Perception of/history with treatment/healthcare* was often mentioned as a barrier, citing such factors as knowledge of the healthcare system and level of experience using the healthcare services/resources, trust, and history of healthcare misconduct as well as fear of hospitalization, treatments, and diagnosis. The complete set of factors upon which definitions were based are listed in Table 2.

Categories of factors upon which definitions were based	References
Community resources and setting	[32, 36, 40-42, 48-50, 55, 56, 63, 65, 67-69, 71, 72, 74, 83, 84, 89]
Digital divide	[41, 60]
Engagement in healthcare system	[12, 13, 15, 17-19, 21, 23-27, 29, 31, 33, 36, 38, 42, 45, 50, 52, 57, 58, 64, 67, 75-79, 82, 83, 85, 87, 88, 90-97, 100]
Health beliefs and knowledge	[12, 21, 23, 26, 29, 31, 32, 38, 40, 41, 48, 57, 64, 67, 75, 77, 85, 87, 91]
Health history, status/capacity	[12, 14, 32, 37, 40, 43, 55, 58, 59, 63, 67, 70, 73, 74, 79, 84, 86, 97, 100]
Healthcare system infrastructure	[13, 15, 17, 21-25, 28, 29, 31-37, 39-41, 46-48, 51, 52, 55-62, 64-68, 70-73, 75-82, 84-87, 90-94, 96-99, 101]
Perception of/history with treatment/healthcare	[16, 22, 28, 38, 39, 48, 51, 64, 65, 75, 77, 82, 83, 86, 94, 97, 101],
Personal skills/capacity, logistical constraints	[12, 22-24, 29-32, 37, 41, 44, 49-51, 53, 54, 64, 65, 67, 68, 70, 74, 83, 85, 92, 95, 97, 101]

Psychosocial/cultural factors	[23, 25, 28-31, 39, 43, 47, 50, 55, 58, 64, 67, 73-77, 79, 82-87, 89-92, 94, 97],
Safety/security	[55, 78, 79, 86]
Socioeconomic status/factors	[12-18, 20-26, 28, 39-46, 49, 53-56, 63-67, 69, 70, 73, 74, 78, 79, 81-87, 92, 97, 99]
Socio-political factors	[13, 19, 22, 23, 27, 28, 30, 40, 41, 49, 51, 54-56, 64, 66, 67, 70, 73, 81-84, 86, 87, 90, 94, 99]

Table 2 Factors upon which definitions of unreached groups or populations are based.

4 DISCUSSION

Of the total 795 articles, we identified 90 that described the recruitment and active participation of those who were considered unreached and had chronic conditions. The most common reasons for being described as unreached were related to supply of healthcare system resources, socioeconomic factors related to a community or individual, and individuals' engagement in healthcare.

The variety of unique reasons given by the authors for why certain groups with chronic conditions were unreached was quite frankly disconcerting. Several articles cited a scepticism or lack of trust with the healthcare system due to a history of mistreatment of people like themselves [22, 65], e.g., the 1932-1972 Tuskegee syphilis experiments [102]. Unfortunately, this mistrust persists today due also to personal experiences with racism and stigma during encounters with the healthcare system or from healthcare providers [82].

While the promise of telehealth is greater access and use of healthcare resources amongst those with chronic conditions, the contextual barriers described in this review, including socioeconomic status, where they reside, and logistical challenges could stop them from using, or continue to use, a technology intervention [103]. Those in need exist within the general population, yet most studies include those who are already engaged in their health and care. If we in health research wish to pursue digital health interventions for the general population, we first must consider barriers, including cost and resources needed to support the use of such interventions – from recruitment of diverse informants to the implementation and continuity of the intervention [104].

The strengths of this review were the inclusive search strategy. We aimed not to limit the type of paper based on a certain classification of a “chronic” condition. This allowed us to explore a greater breadth of factors associated with a group not receiving the care they need.

The limitations of this review include human error. We do acknowledge the possibility that a factor was excluded in the extraction of the definitions from the article. We did attempt to minimize this possibility by copying and pasting article text for qualitative analysis instead of paraphrasing. Also, the introductions were primarily used to identify the definitions. Therefore, information presented in other sections would have been omitted.

5 CONCLUSION

As researchers and healthcare providers, we need to know not only why certain groups are not receiving the care that they need but also how those reasons came to be – the

history behind them. It was made clear by the articles that described personal and systemic history of healthcare misconduct, racism, stigma, and social exclusion, that perception is indeed everything. Awareness of the root causes of reluctance in addition to the systemic, political, or infrastructural barriers to seeking or receiving care is the only way in which we develop and effectively implement interventions for those with chronic conditions, including health technologies. Results from this work will contribute to WARIFA in terms of recruitment strategies and understanding the impact and barriers to reaching different populations for health technologies, specifically artificial intelligence.

6 REFERENCES

1. Logie, C., *Introduction*, in *Working with Excluded Populations in HIV: Hard to Reach or Out of Sight?* Springer International Publishing: Cham. p. 1-27. 2021,
2. Brackertz, N., *Who is hard to reach and why*. Institute of Social Research Working Paper, Swinburne University of Technology Institute of Social Research, Victoria. 2007.
3. United States. Congress. House. Committee on Appropriations. Subcommittee on Departments of Labor, et al., *Departments of Labor and Health, Education, and Welfare Appropriations for 1976: Hearings Before a Subcommittee of the Committee on Appropriations, House of Representatives*. U.S. Government Printing Office. 1974
4. Bernell, S. and S.W. Howard, *Use your words carefully: what is a chronic disease?* *Frontiers in public health*, **4**: p. 159. 2016.
5. Johnston, J., et al., *Study of potentially preventable hospitalisations (PPH) for chronic conditions: what proportion are preventable and what factors are associated with preventable PPH?* *BMJ open*, **10**(11): p. e038415. 2020.
6. OECD/European Union, *Health at a Glance: Europe 2020: State of Health in the EU Cycle*: Paris. 2020,
7. Walker, H.R. and M.L. Litchman, *Diabetes identity: A mechanism of social change*. *Qualitative Health Research*, **31**(5): p. 913-925. 2021.
8. *Watching the risk factors: Artificial intelligence and the prevention of chronic conditions*. European Commission: Norway, Romania, Spain. 2020,
9. *PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation*. *Annals of Internal Medicine*, **169**(7): p. 467-473. 2018.
10. Ouzzani, M., et al., *Rayyan-a web and mobile app for systematic reviews*. *Syst Rev*, **5**(1): p. 210. 2016.
11. Belur, J., et al., *Interrater reliability in systematic review methodology: exploring variation in coder decision-making*. *Sociological methods & research*, **50**(2): p. 837-865. 2021.
12. Smith, S.K., et al., *A decision aid to support informed choices about bowel cancer screening among adults with low education: randomised controlled trial*. *Bmj*, **341**: p. c5370. 2010.
13. Yi, J.K. and K.N. Luong, *Apartment-based breast cancer education program for low income Vietnamese American women*. *J Community Health*, **30**(5): p. 345-53. 2005.
14. Owusu, C., et al., *IMPROVE, a community-based exercise intervention versus support group to improve functional and health outcomes among older African American and non-Hispanic White breast cancer survivors from diverse socioeconomic backgrounds: Rationale, design and methods*. *Contemp Clin Trials*, **92**: p. 106001. 2020.
15. Doorenbos, A.Z., et al., *A randomized controlled calendar mail-out to increase cancer screening among urban American Indian and Alaska Native patients*. *J Cancer Educ*, **26**(3): p. 549-54. 2011.
16. Ahmed, N.U., et al., *Randomized controlled trial of mammography intervention in insured very low-income women*. *Cancer Epidemiol Biomarkers Prev*, **19**(7): p. 1790-8. 2010.
17. Gupta, S., et al., *Comparative effectiveness of fecal immunochemical test outreach, colonoscopy outreach, and usual care for boosting colorectal cancer screening among the underserved: a randomized clinical trial*. *JAMA Intern Med*, **173**(18): p. 1725-32. 2013.
18. Green, B.B., et al., *Financial Incentives to Increase Colorectal Cancer Screening Uptake and Decrease Disparities: A Randomized Clinical Trial*. *JAMA Netw Open*, **2**(7): p. e196570. 2019.
19. Basch, C.E., et al., *A Randomized Trial to Compare Alternative Educational Interventions to Increase Colorectal Cancer Screening in a Hard-to-Reach Urban Minority Population with Health Insurance*. *J Community Health*, **40**(5): p. 975-83. 2015.
20. Berkowitz, S.A., et al., *Building Equity Improvement into Quality Improvement: Reducing Socioeconomic Disparities in Colorectal Cancer Screening as Part of Population Health Management*. *J Gen Intern Med*, **30**(7): p. 942-9. 2015.
21. Roland, K.B., et al., *Changes in Knowledge and Beliefs About Human Papillomavirus and Cervical Cancer Screening Intervals in Low-Income Women After an Educational Intervention*. *J Prim Care Community Health*, **7**(2): p. 88-95. 2016.
22. Fischer, S.M., et al., *Apoyo con Cariño: Strategies to Promote Recruiting, Enrolling, and Retaining Latinos in a Cancer Clinical Trial*. *J Natl Compr Canc Netw*, **15**(11): p. 1392-1399. 2017.
23. Ma, G.X., et al., *Community-based colorectal cancer intervention in underserved Korean Americans*. *Cancer Epidemiol*, **33**(5): p. 381-6. 2009.
24. Lee, H.Y., et al., *Motivating underserved Vietnamese Americans to obtain colorectal cancer screening: evaluation of a culturally tailored DVD intervention*. *Asian Pac J Cancer Prev*, **15**(4): p. 1791-6. 2014.
25. Larkey, L.K., et al., *A cancer screening intervention for underserved Latina women by lay educators*. *J Womens Health (Larchmt)*, **21**(5): p. 557-66. 2012.

26. Williams, K.P., et al., *Kin KeeperSM: design and baseline characteristics of a community-based randomized controlled trial promoting cancer screening in Black, Latina, and Arab women*. *Contemp Clin Trials*, **34**(2): p. 312-9. 2013.
27. Wang, X., et al., *Evidence-based intervention to reduce access barriers to cervical cancer screening among underserved Chinese American women*. *J Womens Health (Larchmt)*, **19**(3): p. 463-9. 2010.
28. Calderón-Mora, J., et al., *Group Versus Individual Culturally Tailored and Theory-Based Education to Promote Cervical Cancer Screening Among the Underserved Hispanics: A Cluster Randomized Trial*. *Am J Health Promot*, **34**(1): p. 15-24. 2020.
29. Fang, C.Y., et al., *Addressing multilevel barriers to cervical cancer screening in Korean American women: A randomized trial of a community-based intervention*. *Cancer*, **123**(6): p. 1018-1026. 2017.
30. Juon, H.S., et al., *Effect of a liver cancer education program on hepatitis B screening among Asian Americans in the Baltimore-Washington metropolitan area, 2009-2010*. *Prev Chronic Dis*, **11**: p. 130258. 2014.
31. Kobetz, E., et al., *A randomized trial of mailed HPV self-sampling for cervical cancer screening among ethnic minority women in South Florida*. *Cancer Causes Control*, **29**(9): p. 793-801. 2018.
32. Kolovou, V., et al., *Recruitment and retention of participants from socioeconomically deprived communities: lessons from the Awareness and Beliefs About Cancer (ABACus3) Randomised Controlled Trial*. *BMC Med Res Methodol*, **20**(1): p. 272. 2020.
33. Price, J. and J. Brunet, *Feasibility and acceptability of a telehealth behavior change intervention for promoting physical activity and fruit and vegetable consumption among rural-living young adult cancer survivors*. *J Psychosoc Oncol*, **39**(6): p. 715-733. 2021.
34. Dorfman, C.S., et al., *Development and pilot testing of an mHealth behavioral cancer pain protocol for medically underserved communities*. *J Psychosoc Oncol*, **37**(3): p. 335-349. 2019.
35. Katz, M.L., et al., *Do cervical cancer screening rates increase in association with an intervention designed to increase mammography usage?* *J Womens Health (Larchmt)*, **16**(1): p. 24-35. 2007.
36. Preston, M.A., et al., *Colorectal cancer screening in rural and poor-resourced communities*. *Am J Surg*, **216**(2): p. 245-250. 2018.
37. Zernicke, K.A., et al., *A randomized wait-list controlled trial of feasibility and efficacy of an online mindfulness-based cancer recovery program: the eTherapy for cancer applying mindfulness trial*. *Psychosom Med*, **76**(4): p. 257-67. 2014.
38. Kim, M.T., et al., *Constructing and testing a self-help intervention program for high blood pressure control in Korean American seniors--a pilot study*. *J Cardiovasc Nurs*, **21**(2): p. 77-84. 2006.
39. Khare, M.M., et al., *A lifestyle approach to reducing cardiovascular risk factors in underserved women: design and methods of the Illinois WISEWOMAN Program*. *J Womens Health (Larchmt)*, **18**(3): p. 409-19. 2009.
40. Abbott, L., et al., *Exploring the intervention effect moderators of a cardiovascular health promotion study among rural African-Americans*. *Public Health Nurs*, **35**(2): p. 126-134. 2018.
41. Safford, M.M., et al., *Peer Coaches to Improve Diabetes Outcomes in Rural Alabama: A Cluster Randomized Trial*. *Ann Fam Med*, **13 Suppl 1**(Suppl 1): p. S18-26. 2015.
42. Abbas, S.Z., et al., *The effectiveness of using the workplace to identify and address modifiable health risk factors in deprived populations*. *Occup Environ Med*, **72**(9): p. 664-9. 2015.
43. Mildestvedt, T. and E. Meland, *Examining the "Matthew Effect" on the motivation and ability to make lifestyle changes in 217 heart rehabilitation patients*. *Scand J Public Health*, **35**(2): p. 140-7. 2007.
44. Han, H.R., et al., *Implementation and success of nurse telephone counseling in linguistically isolated Korean American patients with high blood pressure*. *Patient Educ Couns*, **80**(1): p. 130-4. 2010.
45. Greenberg, K.L., et al., *Cardiovascular health literacy and patient-physician communication intervention in women from disadvantaged communities*. *Eur J Prev Cardiol*, **26**(16): p. 1762-1770. 2019.
46. Samuel-Hodge, C.D., et al., *Strengthening community-clinical linkages to reduce cardiovascular disease risk in rural NC: feasibility phase of the CHANGE study*. *BMC Public Health*, **20**(1): p. 264. 2020.
47. Seguin-Fowler, R.A., et al., *The Strong Hearts, Healthy Communities Program 2.0: An RCT Examining Effects on Simple 7*. *Am J Prev Med*, **59**(1): p. 32-40. 2020.
48. Zha, P., et al., *Utilizing a Mobile Health Intervention to Manage Hypertension in an Underserved Community*. *West J Nurs Res*, **42**(3): p. 201-209. 2020.
49. Eakin, E., et al., *Telephone counseling for physical activity and diet in primary care patients*. *Am J Prev Med*, **36**(2): p. 142-9. 2009.
50. Lu, Y., et al., *The evaluation of a culturally appropriate, community-based lifestyle intervention program for elderly Chinese immigrants with chronic diseases: a pilot study*. *J Public Health (Oxf)*, **36**(1): p. 149-55. 2014.
51. Eakin, E.G., et al., *Recruitment and retention of Latinos in a primary care-based physical activity and diet trial: The Resources for Health study*. *Health Educ Res*, **22**(3): p. 361-71. 2007.
52. Burhan, H., et al., *Screening Heroin Smokers Attending Community Drug Services for COPD*. *Chest*, **155**(2): p. 279-287. 2019.
53. Ruggiero, L., S. Oros, and Y.K. Choi, *Community-based translation of the diabetes prevention program's lifestyle intervention in an underserved Latino population*. *Diabetes Educ*, **37**(4): p. 564-72. 2011.

54. Santoyo-Olsson, J., et al., *An innovative multiphased strategy to recruit underserved adults into a randomized trial of a community-based diabetes risk reduction program*. *Gerontologist*, **51 Suppl 1**(Suppl 1): p. S82-93. 2011.
55. Nelson, K., et al., *Peer Support for Achieving Independence in Diabetes (Peer-AID): design, methods and baseline characteristics of a randomized controlled trial of community health worker assisted diabetes self-management support*. *Contemp Clin Trials*, **38**(2): p. 361-9. 2014.
56. Wolff, L.S., et al., *The Effect of Integrating Primary Care and Mental Health Services on Diabetes and Depression: A Multi-site Impact Evaluation on the US-Mexico Border*. *Med Care*, **59**(1): p. 67-76. 2021.
57. Simmons, D. and J.A. Voyle, *Reaching hard-to-reach, high-risk populations: piloting a health promotion and diabetes disease prevention programme on an urban marae in New Zealand*. *Health Promot Int*, **18**(1): p. 41-50. 2003.
58. Lake, A.J., et al., *Effect of a tailored leaflet to promote diabetic retinopathy screening among young adults with type 2 diabetes: a randomised controlled trial*. *BMC Ophthalmol*, **20**(1): p. 80. 2020.
59. Liou, J.K., et al., *Shared care combined with telecare improves glycemic control of diabetic patients in a rural underserved community*. *Telemed J E Health*, **20**(2): p. 175-8. 2014.
60. Shea, S., *The Informatics for Diabetes and Education Telemedicine (IDEATel) project*. *Trans Am Clin Climatol Assoc*, **118**: p. 289-304. 2007.
61. Johnson, M., et al., *Evaluation of an Academic-Community Partnership to Implement MTM Services in Rural Communities to Improve Pharmaceutical Care for Patients with Diabetes and/or Hypertension*. *J Manag Care Spec Pharm*, **24**(2): p. 132-141. 2018.
62. Parra-Medina, D., et al., *Successful recruitment and retention strategies for a randomized weight management trial for people with diabetes living in rural, medically underserved counties of South Carolina: the POWER study*. *J Am Diet Assoc*, **104**(1): p. 70-5. 2004.
63. Piatt, G.A., et al., *Translating the chronic care model into the community: results from a randomized controlled trial of a multifaceted diabetes care intervention*. *Diabetes Care*, **29**(4): p. 811-7. 2006.
64. Timm, L., et al., *Early detection of type 2 diabetes in socioeconomically disadvantaged areas in Stockholm - comparing reach of community and facility-based screening*. *Glob Health Action*, **13**(1): p. 1795439. 2020.
65. Andrae, S.J., et al., *Recruitment of a rural, southern, predominantly African-American population into a diabetes self-management trial*. *Contemp Clin Trials*, **33**(3): p. 499-506. 2012.
66. Ramal, E., A. Champlin, and K. Bahjri, *Impact of a Plant-Based Diet and Support on Mitigating Type 2 Diabetes Mellitus in Latinos Living in Medically Underserved Areas*. *Am J Health Promot*, **32**(3): p. 753-762. 2018.
67. Kim, M.T., et al., *The Effect of a Community-Based Self-Help Intervention: Korean Americans With Type 2 Diabetes*. *Am J Prev Med*, **49**(5): p. 726-737. 2015.
68. Hawley, C.N., et al., *Cooking for Health: a healthy food budgeting, purchasing, and cooking skills randomized controlled trial to improve diet among American Indians with type 2 diabetes*. *BMC Public Health*, **21**(1): p. 356. 2021.
69. Vieira, E.R., et al., *Effects of Exercise and Diet on Body Composition and Physical Function in Older Hispanics with Type 2 Diabetes*. *Int J Environ Res Public Health*, **18**(15). 2021.
70. Spencer, M.S., et al., *Diabetes-specific emotional distress among African Americans and Hispanics with type 2 diabetes*. *J Health Care Poor Underserved*, **17**(2 Suppl): p. 88-105. 2006.
71. Davis, R.M., et al., *TeleHealth improves diabetes self-management in an underserved community: diabetes TeleCare*. *Diabetes Care*, **33**(8): p. 1712-7. 2010.
72. Jessee, B.T. and C.M. Rutledge, *Effectiveness of nurse practitioner coordinated team group visits for type 2 diabetes in medically underserved Appalachia*. *J Am Acad Nurse Pract*, **24**(12): p. 735-43. 2012.
73. Gudlavalleti, M.V., et al., *Access to health care and employment status of people with disabilities in South India, the SIDE (South India Disability Evidence) study*. *BMC Public Health*, **14**: p. 1125. 2014.
74. Melville, C.A., et al., *Effectiveness of a walking programme to support adults with intellectual disabilities to increase physical activity: walk well cluster-randomised controlled trial*. *Int J Behav Nutr Phys Act*, **12**: p. 125. 2015.
75. Brunings, P., et al., *"It's a big part of our lives": A qualitative study defining quality of hepatitis C care from the patient's perspective*. *Gastroenterol Nurs*, **36**(4): p. 249-57. 2013.
76. B, E.W., et al., *Life projects: the transformative potential of direct-acting antiviral treatment for hepatitis C among people who inject drugs*. *Int J Drug Policy*, **72**: p. 138-145. 2019.
77. Ma, G.X., et al., *Efficacy of a community-based participatory and multilevel intervention to enhance hepatitis B virus screening and vaccination in underserved Korean Americans*. *Cancer*, **124**(5): p. 973-982. 2018.
78. Groessl, E.J., et al., *The hepatitis C self-management programme: a randomized controlled trial*. *J Viral Hepat*, **18**(5): p. 358-68. 2011.
79. Bajis, S., et al., *Hepatitis C virus testing, liver disease assessment and direct-acting antiviral treatment uptake and outcomes in a service for people who are homeless in Sydney, Australia: The LiveRLife homelessness study*. *J Viral Hepat*, **26**(8): p. 969-979. 2019.
80. Nelson, R.G., et al., *Home-Based Kidney Care, Patient Activation, and Risk Factors for CKD Progression in Zuni Indians: A Randomized,*

- Controlled Clinical Trial*. Clin J Am Soc Nephrol, **13**(12): p. 1801-1809. 2018.
81. Henry, S.K., M.M. Grant, and K.L. Cropsey, *Determining the optimal clinical cutoff on the CES-D for depression in a community corrections sample*. J Affect Disord, **234**: p. 270-275. 2018.
 82. Jordan, A., et al., *A feasibility study providing substance use treatment in the Black church*. J Subst Abuse Treat, **124**: p. 108218. 2021.
 83. Goodkind, J.R., et al., *Randomized Controlled Trial of a Multilevel Intervention to Address Social Determinants of Refugee Mental Health*. Am J Community Psychol, **65**(3-4): p. 272-289. 2020.
 84. Mathias, K., et al., *"We've got through hard times before: acute mental distress and coping among disadvantaged groups during COVID-19 lockdown in North India - a qualitative study"*. Int J Equity Health, **19**(1): p. 224. 2020.
 85. van der Waerden, J.E., et al., *A randomized controlled trial of combined exercise and psycho-education for low-SES women: short- and long-term outcomes in the reduction of stress and depressive symptoms*. Soc Sci Med, **91**: p. 84-93. 2013.
 86. Burmaster, K.B., et al., *Impact of a private sector living wage intervention on depressive symptoms among apparel workers in the Dominican Republic: a quasi-experimental study*. BMJ Open, **5**(8): p. e007336. 2015.
 87. Kim, S.E., et al., *Does telephone care management help Medicaid beneficiaries with depression?* Am J Manag Care, **17**(10): p. e375-82. 2011.
 88. Grilo, C.M. and M.A. White, *Orlistat with behavioral weight loss for obesity with versus without binge eating disorder: randomized placebo-controlled trial at a community mental health center serving educationally and economically disadvantaged Latino/as*. Behav Res Ther, **51**(3): p. 167-75. 2013.
 89. Ong, M.K., et al., *A Community-Partnered, Participatory, Cluster-Randomized Study of Depression Care Quality Improvement: Three-Year Outcomes*. Psychiatr Serv, **68**(12): p. 1262-1270. 2017.
 90. Collado, A., et al., *Mental health stigma in depressed Latinos over the course of therapy: Results from a randomized controlled trial*. J Clin Psychol, **75**(7): p. 1179-1187. 2019.
 91. Gitlin, L.N., et al., *A community-integrated home based depression intervention for older African Americans: [corrected] description of the Beat the Blues randomized trial and intervention costs*. BMC Geriatr, **12**: p. 4. 2012.
 92. Shea, M., et al., *Mexican American women's perspectives on a culturally adapted cognitive-behavioral therapy guided self-help program for binge eating*. Psychol Serv, **13**(1): p. 31-41. 2016.
 93. Lovell, K., et al., *Development and evaluation of culturally sensitive psychosocial interventions for under-served people in primary care*. BMC Psychiatry, **14**: p. 217. 2014.
 94. Tondora, J., et al., *A clinical trial of peer-based culturally responsive person-centered care for psychosis for African Americans and Latinos*. Clin Trials, **7**(4): p. 368-79. 2010.
 95. Winter, L. and L.N. Gitlin, *Evaluation of a telephone-based support group intervention for female caregivers of community-dwelling individuals with dementia*. Am J Alzheimers Dis Other Demen, **21**(6): p. 391-7. 2006.
 96. Rubin, A.S., et al., *Effects on processes and costs of care associated with the addition of an internist to an inpatient psychiatry team*. Psychiatr Serv, **56**(4): p. 463-7. 2005.
 97. Jaroszewski, A.C., R.R. Morris, and M.K. Nock, *Randomized controlled trial of an online machine learning-driven risk assessment and intervention platform for increasing the use of crisis services*. J Consult Clin Psychol, **87**(4): p. 370-379. 2019.
 98. Grote, N.K., et al., *A Randomized Trial of Collaborative Care for Perinatal Depression in Socioeconomically Disadvantaged Women: The Impact of Comorbid Posttraumatic Stress Disorder*. J Clin Psychiatry, **77**(11): p. 1527-1537. 2016.
 99. Fortney, J.C., et al., *Comparison of Teleintegrated Care and Telereferral Care for Treating Complex Psychiatric Disorders in Primary Care: A Pragmatic Randomized Comparative Effectiveness Trial*. JAMA Psychiatry, **78**(11): p. 1189-1199. 2021.
 100. Ward-Ciesielski, E.F., et al., *Comparing brief interventions for suicidal individuals not engaged in treatment: A randomized clinical trial*. J Affect Disord, **222**: p. 153-161. 2017.
 101. Rimmer, J.H., et al., *Rationale and design of the tele-exercise and multiple sclerosis (TEAMS) study: A comparative effectiveness trial between a clinic- and home-based telerehabilitation intervention for adults with multiple sclerosis (MS) living in the deep south*. Contemp Clin Trials, **71**: p. 186-193. 2018.
 102. Brandt, A.M., *Racism and research: the case of the Tuskegee Syphilis Study*. Hastings center report: p. 21-29. 1978.
 103. Beratarrechea, A., et al., *The impact of mobile health interventions on chronic disease outcomes in developing countries: a systematic review*. Telemedicine and e-Health, **20**(1): p. 75-82. 2014.
 104. McMillan, J.R. and C. Conlon, *The ethics of research related to health care in developing countries*. Journal of Medical Ethics, **30**(2): p. 204-206. 2004.

7 ACKNOWLEDGEMENTS

We would like to thank the project under which this work was completed, Watching the Risk Factors (WARIFA): Artificial intelligence and the prevention of chronic conditions (Horizon 2020 grant agreement No 101017385). We also thank those who helped in reviewing the titles and abstracts of the 795 articles.

Abstracts

The Concept of Conditional Method Agreement Trees with Single Measurements per Subject

Conditional Method Agreement Trees

Siranush Karapetyan¹ and Alexander Hapfelmeier^{1,2}

¹Institute of General Practice and Health Services Research, School of Medicine, Technical University of Munich, Munich, Germany siranush.karapetyan@mri.tum.de

²Institute of AI and Informatics in Medicine, School of Medicine, Technical University of Munich, Munich, Germany

Abstract

The concept of conditional method agreement is introduced and respective solutions are proposed to define homogeneous subgroups in terms of mean and variance of the differences in measurements.

Keywords

Method agreement, Bland-Altman plot, recursive partitioning

1 INTRODUCTION

A well-established method for assessing the agreement between two measurement methods is the Bland-Altman plot [1], which illustrates the differences against the mean values of the measurements. Bland and Altman mainly discuss the question whether the agreement between the methods depends on the magnitude of the measured values and rely on the idea, that agreement is identical for all patient groups. However, the methods may not be equally accurate or comparable in every situation and for every patient. We assume that some factors, such as patient characteristics and measurement settings, may affect the ability of the methods to produce accurate measurements and thus alter the agreement between the methods.

In the present work, we therefore aim to investigate the agreement conditional on patient characteristics by introducing the concept of conditional method agreement and propose a corresponding analysis method for the case of single measurements per subject. The new method, called Conditional Method Agreement Tree (COAT), is a simple combination of recursive partitioning and Bland-Altman plot. It takes advantage of the fact that agreement can be modeled by correctly specified tree-based models, which allows the definition of subgroups with different agreement in terms of the mean and variance of the measurements. The ability of the methods to define the subgroups is investigated in a simulation study. Application to real data demonstrates the relevance to research.

2 MODELLING APPROACHES

To define homogeneous subgroups concerning the mean and the variance of the differences, the following tree-based algorithms are used: conditional inference tree (Ctree) [2], distributional tree (Disttree) [3] and model-based recursive partitioning (MOB) [4]. The general algorithm of all these methods is similar, and in order to partition the data based on the potential split variables, they all proceed in two

steps, namely, selecting the split variables based on the association of some goodness-of-fit (e.g. model scores) with each potential split variable, and choosing the split point so that the goodness-of-fit is maximized in the resulting subgroups. This procedure is repeated until a certain stopping criterion is reached, e.g., a minimum number of observations in terminal nodes or no more significant associations are detected. Finally, the agreement in each resulting subgroup is evaluated using a Bland-Altman plot. The aforementioned models differ only in the implementation of the individual steps.

2.1 Ctree

The decision to split is based on the test statistic which measures association between the outcome and the split variable, in our case: $T_j = \text{vec}(\sum_{i=1}^n Z_{ji} (Y_i, (Y_i - \bar{Y})^2)^T)$, where Z_j is a potential split variable and Y_i is the outcome. Such a transformation of the outcome $(Y_i, (Y_i - \bar{Y})^2)^T$ allows the observations to be split based not only on the association between split variables and mean differences, but also on the basis of the association with the variance of differences.

2.2 Disttree

Another approach is to use Disttree to model the parameters of the prespecified distribution of the differences. In the present case the normal distribution can be used to model the average difference between methods and the corresponding variance. Test statistic for split decision is given by $T_j = \text{vec}(\sum_{i=1}^n Z_{ji} s(\hat{\theta}, Y_i))$, where $s(\hat{\theta}, Y_i)$ are the model scores and $\hat{\theta} = (\hat{\mu}, \hat{\sigma})$ parameter estimates.

2.3 MOB

Like Disttree, MOB is also based on a parametric model and corresponding score. Here we consider the normal distribution and define the residual variance as another model parameter guiding the splitting decision. The latter is based on the empirical fluctuation process given by

The 18th Scandinavian Conference on Health informatics, Tromsø, Norway, August 22-24, 2022. Organized by UiT The Arctic University of Norway. Conference Proceedings published by Linköping University Electronic Press at <https://doi.org/10.3384/ecp187>. © The Author(s). This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>

$W_j(t) = \hat{V}^{-1/2} n^{-1/2} \sum_{i=1}^{n_t} s_{\pi(Z_{ji})}$, which measures the deviations of the model scores from their zero mean when the potential split variable is ordered.

3 METHODS

To investigate the performance of the models in terms of identifying the true split variable and the existing subgroups a simulation study is conducted. A total of 10000 simulations are performed, considering sample sizes of $n = (50, 100, 200, 400, 500)$. The differences are generated from a normal distribution $Y \sim N(\mu, \sigma^2)$. The scenarios differ in how the mean and variance are defined. In the first three scenarios, we vary the mean once, the variance once, and finally both based on one split variable. Scenarios 4 and 5 are more complex, as both mean and variance vary based on two split variables. We evaluate the models based on the Adjusted Rand Index (ARI), the measure of similarity between the true and predicted tree structure. We additionally examine a null case to investigate the tendency of methods to detect false subgroups when there are actually none.

To demonstrate the relevance of the new method in epidemiological research, we apply it to real data collected in a validation study [5] and provided by our project partner. In this study, 50 participants were recruited to simultaneously wear different accelerometers for an entire day and night. Daily variables were then created from the accelerometer measurements, including light physical activity (PA), moderate PA, vigorous PA, and activity EE (AEE). Additional information on age, sex, height, and weight was used to test our assumption of whether participant characteristics affect accelerometer agreement.

4 RESULTS

4.1 Simulation study

The results of the simulation study show that with the increasing sample size and in the more complex environment, the models are able to detect the true subgroups. In particular, for scenarios 2-5 and in the case of a higher sample size (400-500 observations), the ARI takes values between 0.65 and 0.95 (values close to 1 indicate the best fit), with the MOB being inferior to the other methods.

In a null case Ctree and Distree falsely detect subgroups from 4.1-5.4% of cases. On the contrary, MOB yields error rates of only 0.3-3.1% for smaller sample sizes (50-100 observations). However, for larger sample sizes (200, 400, 500 observations), it performs slightly worse, detecting false subgroups between 5.6% and 5.7% of cases.

4.2 Application

Figure 1 is an illustration of the agreement of measurements evaluated with the new method. As can be seen, COAT is able to identify the subgroups with different agreement in terms of mean and variance. It identifies age as a relevant covariate. From the tree structure, it is seen that better agreement is obtained for older patients (age>41).

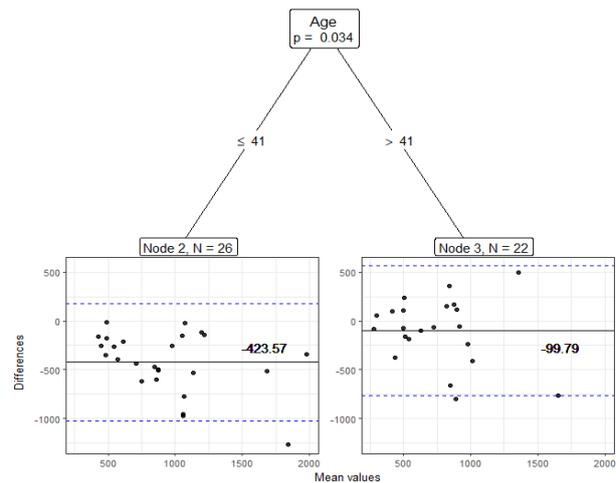


Figure 1. Conditional Method Agreement Tree using Model-Based Recursive Partitioning for AEE measured with ActiGraph Hip (triaxial) and Actiheart upper.

5 DISCUSSION

With COAT, we are able to estimate conditional agreement and agreement dependence on covariates in a single framework. The simulation study shows that the models are able to identify the true split variable and the underlying tree structure. All three tree-based methods were able to detect heterogeneity between the mean and variance of the agreement. In addition, the real data example shows that COAT is well suited to define subgroups and evaluate conditional agreement in a unified methodological framework. However, it should be noted that the new method is rather exploratory in its nature and evaluation on the real data example is not performed.

6 REFERENCES

- [1] D. G. Altman and J. M. Bland. Measurement in medicine: the analysis of method comparison studies. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 32(3):307–317, 1983.
- [2] T. Hothorn, K. Hornik, and A. Zeileis. Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical statistics*, 15(3):651–674, 2006.
- [3] L. Schlosser, T. Hothorn, R. Stauffer, and A. Zeileis. Distributional regression forests for probabilistic precipitation forecasting in complex terrain. *The Annals of Applied Statistics*, 13(3):1564–1589, 2019.
- [4] A. Zeileis, T. Hothorn, and K. Hornik. Model-based recursive partitioning. *Journal of Computational and Graphical Statistics*, 17(2):492–514, 2008.
- [5] A. Henriksen, S. Grimsgaard, A. Horsch, G. Hartvigsen, L. Hopstock, et al. Validity of the polar m430 activity monitor in freelifving conditions: validation study. *JMIR formative research*, 3(3):e14438, 2019.

7 ACKNOWLEDGEMENT

We thank the project partners for their contribution to this study and André Henriksen for providing the accelerometer data. This study is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project number (447467169).

Time for using Machine Learning for Basal Insulin Dose Guidance for People with Type 2 Diabetes?

Preliminary Results from a Systematic Review

Camilla Heisel Nyholm Thomsen¹, Stine Hangaard^{1,2}, Thomas Kronborg^{1,2}, Peter Vestergaard^{2,3,4}, Ole Hejlesen¹, and Morten Hasselstrøm Jensen^{1,2}

¹ Aalborg University, Department of Health Science and Technology, Aalborg, Denmark, chnt@hst.aau.dk

² Steno Diabetes Center North Denmark, Aalborg, Denmark

³ Aalborg University, Department of Clinical Medicine, Aalborg, Denmark

⁴ Aalborg University Hospital, Department of Endocrinology, Aalborg, Denmark

Abstract

The present systematic review aims to provide an overview and categorization of dose guidance methods that support basal insulin titration for people with type 2 diabetes. At the time of writing, quality assessment of the included articles is ongoing.

Keywords

Type 2 Diabetes, dose guidance, basal insulin, systematic review

1 INTRODUCTION

Optimal glycaemic control in people with type 2 diabetes (T2D) is crucial to prevent diabetes-related complications [1]. In many cases, treatment with basal insulin is necessary 5-10 years after diagnosis to reach glycaemic targets [2]. Titration of insulin to determine the optimal dose is a difficult and time-consuming task associated with clinical inertia [3]. Thus, 60% of people treated with insulin do not reach glycaemic targets [4].

In recent years, the development of dose guidance to support insulin titration has been of rapidly growing interest. Despite this rapidly growing interest and the fact that it has been a field of interest for several decades, no systematic review within the area has been found. Therefore, this systematic review aims to provide an overview and categorize dose guidance methods that support basal insulin titration of people with T2D by characteristics, effect, and user experience.

2 METHOD

The present systematic review will be reported according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analyses) guidelines [5]. The review is registered at PROSPERO, an international prospective register of systematic reviews, as ID number CRD42021289364 [6].

The review considers studies that include adult (>18 years) subjects with a diagnosis of T2D treated with basal insulin alone or in combination with oral antihyperglycemic agents. Studies that evaluate dose guidance methods for basal insulin titration are considered. Furthermore, any outcomes related to glycaemic control and in any setting (e.g. general practitioners, the patient's home, and specialized diabetes units) are considered. An exhaustive search was performed in PubMed, Embase, Scopus, and

IEEE. Primary peer-reviewed full-text studies in English, Norwegian, Danish, and Swedish were screened for inclusion except for animal studies, expert opinions, and case studies. Data extraction will include study population, study design, method used for dose guidance, setting, study length, and reported outcome. Quality assessment of the included studies will be done using Joanna Briggs Institute's critical appraisal tools in accordance with the study design of the included studies [7].

3 RESULTS

The systematic search identified 4,245 potential studies in the selected databases. After the removal of duplicates, 3,211 studies remained. Following the screening of title and abstract, 270 potential studies were selected for full-text screening, which resulted in the inclusion of 32 studies. Citation and reference search was performed based on the 32 included studies. This led to the inclusion of 4 additional studies, which resulted in a total of 36 included studies. An overview of the study designs of included studies is shown in table 1.

Study design	Frequency
Randomized controlled trial	21
Quasi-experimental study	6
Simulation study	4
Mixed method study	3
Cohort study	1
Qualitative study	1

Table 1 is an overview of the study designs of the 36 included articles.

Preliminary results show the use of paper-based titration algorithms, telehealth solutions, and mathematical models in basal insulin dose guidance for people with T2D.

The 18th Scandinavian Conference on Health informatics, Tromsø, Norway, August 22-24, 2022. Organized by UiT The Arctic University of Norway. Conference Proceedings published by Linköping University Electronic Press at <https://doi.org/10.3384/ecp187>. © The Author(s). This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>

Telehealth solutions cover digital implementation of a paper-based titration algorithm, shared decision-making platforms, and communication platforms between health care providers and people with T2D.

Figure 1 shows the percentwise distribution of the methods for basal insulin dose guidance identified in the included studies.

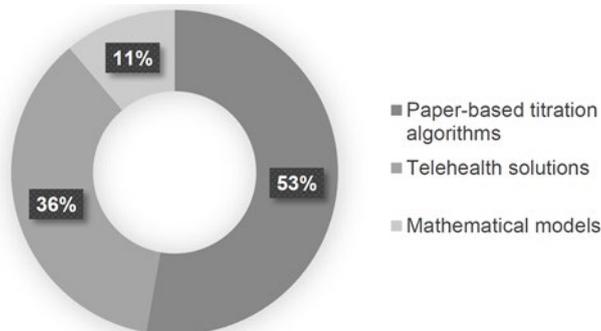


Figure 1 shows the distribution of methods used for basal insulin dose guidance identified in the included studies.

Among these identified methods for basal insulin dose guidance, both solutions aimed directly at people with T2D, and healthcare professionals are found. Though the majority of the methods are aimed at healthcare professionals.

At the time of writing, quality assessment of the included studies is ongoing.

4 DISCUSSION

The systematic review is strengthened by the broad scope since various dose guidance methods, outcomes, settings, and study designs are considered. However, this may complicate the comparison of the studies due to heterogeneity.

Based on the preliminary results, the majority of studies investigate non-technical basal insulin dose guidance in the form of paper-based titration algorithms. The use of digital solutions for basal insulin dose guidance is limited to simple telehealth solutions and one instance of a mathematical model embedded in a digital solution.

In recent years, mathematical models in the form of compartment models and recursive least square-based extremum seeking control methods have been investigated for basal insulin dose guidance for people with T2D, but the research is still very limited. From the preliminary results, it is an interesting finding that data-driven methods such as machine learning are not found among the methods used for basal insulin dose guidance for people with T2D in the existing literature. Such methods are often used to explore different research fields. This elucidates an apparent gap within the field regarding data-driven methods such as machine learning, and to some extent mathematical models. The use of such methods may bring insight to the field, which could potentially pioneer future research.

This apparent lacking use of data-driven methods in the field of basal insulin dose guidance for people with T2D could be a result of the heterogeneity of this population caused by factors such as varying pancreatic insulin production and insulin sensitivity. This complicates the modelling of blood glucose for people with T2D.

Furthermore, limited information about blood glucose levels is usually available for people with T2D since glucometers are most commonly used for measuring blood glucose. In comparison use of continuous glucose monitoring for measuring blood glucose is more widespread in the treatment of people with type 1 diabetes enabling more complex insight into the blood glucose levels. Therefore, the modelling of the effect of insulin on the blood glucose in people with T2D is challenged by the heterogeneity of the population and the limited insight into the effects on blood glucose levels due to the current standard for measuring blood glucose for these individuals. Similar challenges have been recognized by researchers investigating the use of mathematical models in the development of basal insulin dose guidance aimed at people with T2D.

5 CONCLUSION

The present systematic review will inform investigators of methods used to develop basal insulin dose guidance for people with T2D, and categorize the methods used to provide insight into gaps within the field. Preliminary results show the use of paper-based titration algorithms, telehealth solutions, and mathematical models. The use of mathematical models is limited. Furthermore, preliminary results elucidate an apparent gap in the use of data-driven methods such as machine learning for basal insulin dose guidance for people with T2D. Further results will follow.

6 REFERENCES

- [1] Chatterjee, S., Khunti, K., Davies M.J., "Type 2 diabetes" in *The Lancet* Vol. 389, Issue 10085, pp. 2239-2251, 2017.
- [2] Feinglos, M.N., Bethel, M.A., "Type 2 Diabetes Mellitus" in *Humana Press*, pp. 169-181, 2008.
- [3] Khunti, K., Gomes, M.B., Pocock, S., Shestakova, M.V., Pintat, S., Fenici, P., Hammar, N., Medina, J., "Therapeutic inertia in the treatment of hyperglycemia in patients with type 2 diabetes: A systematic review" in *Diabetes, Obesity and Metabolism* Vol. 20, Issue 2, pp. 427-437, 2018.
- [4] Aradóttir, T.B., Bengtsson, H., Jensen, M.L., Poulsen, N.K., Boiroux, D., Jensen, L.L., Schmidt, S., Nørgaard, K., "Feasibility of a new approach to initiate insulin in type 2 diabetes" in *Diabetes Science and Technology* Vol. 15, Issue 2, pp. 339-345, 2021.
- [5] Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutrijn, I., Hoffmann, T.C., Mulrow, C.D., Moher, D., "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews" in *BMJ* Vol. 372, pp. 372, 2021.
- [6] Thomsen, C.H.N., Hangaard, S., Kronborg, T., Hejlesen, O., Vestergaard, P., Jensen, M.H., "Dose guidance systems for basal insulin titration of patients with Type 2 Diabetes: A protocol of a systematic review" in PROSPERO: International prospective register of systematic reviews, 2021, CRD42021289364, Available from: https://www.crd.york.ac.uk/prosperto/display_record.php?ID=CRD42021289364
- [7] Aromataris, E., Munn, Z., "JBI Manual for Evidence Synthesis" in *JBI*, pp. 217-270, 2020.

The Adherence to Oral Second Line Antidiabetic Medication in People with Type 2 Diabetes

A Protocol for a Systematic Review

Nynne Sophie Holdt-Caspersen^{1,2}, Stine Hangaard Casper^{2,3}, Peter Vestergaard^{3,4,5}, Claus Dethlefsen^{1,6}, and Morten Hasselstrøm Jensen^{2,3}

¹Novo Nordisk, Biostatistics, Aalborg, Denmark

²Aalborg University, Health Science and Technology, Aalborg, Denmark

³Aalborg University Hospital, Steno Diabetes Center North Denmark, Aalborg, Denmark

⁴Aalborg University Hospital, Endocrinology, Aalborg, Denmark

⁵Aalborg University, Clinical Medicine, Aalborg, Denmark

⁶Aalborg University, Mathematical Sciences, Aalborg, Denmark

Abstract

The treatment goal in type 2 diabetes (T2D) is to keep the blood glucose within a target range. In the early stages of diabetes, the treatment may consist entirely of oral antidiabetics (OAD), which may be preferable over injectables for some people. However, previously published systematic reviews found a low OAD adherence and that the adherence differs between different OAD types. These systematic reviews were performed before 2015, and several new OADs have been marketed since then. Thus, a systematic review will be undertaken to review the newest studies on adherence to oral second line antidiabetics in people with T2D.

Keywords

Type 2 diabetes, Adherence, Oral antidiabetics, Systematic review, Protocol

1 INTRODUCTION

In 2021, an estimated total of 536.6 million people had diabetes [1], with approximately 90% of these cases being type 2 diabetes (T2D) [2]. Diabetes results in increased blood glucose, which leads to long-term complications [3]. Therefore, the treatment goal in diabetes is to keep the blood glucose within a target range to delay or avoid long-term complications [3]. The recommended initial treatment, referred to as first line treatment, for T2D is lifestyle interventions and treatment with metformin [3]. When the first line treatment fails to keep the blood glucose below the upper limit of the target range due to disease progression, additional antidiabetic medication, referred to as second line treatment, is added to the treatment regimen [3]. The recommended second line treatment for people with T2D and no comorbidities are glucagon-like peptide-1 receptor agonist (GLP-1RA), Sodium-glucose cotransporter-2 (SGLT-2) inhibitors, dipeptidyl peptidase-4 (DPP-4) inhibitors, sulfonylurea, and thiazolidinedione (TZD) [3].

However, the general adherence to antidiabetic medication amongst people with T2D is low [4-6], which may result in suboptimal treatment effect and decreased glycaemic control. Several factors may affect the treatment adherence e.g., some patients may prefer oral medication over injectables. As several of the second line antidiabetic medication is available as OADs, the treatment regimen in

the early stages of diabetes may consist entirely of OADs, which may be preferable for some patients. Iglay, et al. [5], Krass, et al. [6], and McGovern, et al. [7] all performed a systematic review and meta-analysis in which the adherence to different OAD drug classes were compared. These studies found that the adherence differed across the different OAD drug classes. The systematic reviews of Iglay, et al. [5], Krass, et al. [6], and McGovern, et al. [7] were performed prior to 2015, and several new OADs, such as oral semaglutide [8], have been marketed since then. Therefore, a systematic review including new literature will be undertaken with the aim to systematically review the adherence to oral second line antidiabetics in people with T2D.

2 METHODS

A systematic literature search will be performed to identify articles on adherence to oral second line antidiabetics in people with T2D. The search will be structured as a block search, which will consist of three independent blocks: 'T2D', 'Second line antidiabetics', and 'Adherence'. Each block will consist of relevant synonyms, both as free text and thesaurus term, if one exists. The search will be performed in the following databases: CINAHL, Cochrane, Embase, PsychInfo, PubMed, and Scopus.

The 18th Scandinavian Conference on Health informatics, Tromsø, Norway, August 22-24, 2022. Organized by UiT The Arctic University of Norway. Conference Proceedings published by Linköping University Electronic Press at <https://doi.org/10.3384/ecp187>. © The Author(s). This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>

Articles are included if they present original research on adherence to oral second line antidiabetics in people with T2D. The study population must be adults (≥ 18 years) with T2D and no other comorbidities living in developed countries. The study must be set during everyday circumstances i.e., not during Ramadan, the COVID-19 pandemic, or other special circumstances. Furthermore, the articles must be full text articles written in English. The selection process of relevant articles will be reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA). The quality of the included articles will be evaluated using Joanna Briggs Institute's (JBI's) critical appraisal tool.

Information extracted from the included articles will be details on the study design (control group, trial design, and trial duration), the participants (number of participants, demographics, and treatment regimen), adherence (adherence level and how adherence was measured), and potential adherence barriers.

The adherence level will be investigated with respect to the measurement method of adherence and the country of which the trial is set by dividing the extracted information into subsets.

3 RESULTS

The information extracted from the included articles will be presented in a systematic review. The reported level of adherence to oral second line antidiabetics will be presented in tables by subsets and summarised as a narrative synthesis.

4 DISCUSSION

The overview of adherence to different types of oral second line antidiabetic medication will be an addition to the previously obtained knowledge on the adherence to OADs in people with T2D. This updated overview will provide new knowledge on the patients' challenges to adhere to OADs, which may be used to investigate how these challenges might be overcome. A possible solution could be to use this knowledge to develop a decision support system to guide people with T2D on how to increase their adherence to OADs.

Furthermore, this review will provide an opportunity to compare the adherence rate of recently market OADs and OADs considered in the previously published systematic reviews.

The strength of the proposed systematic review is the broad scope of the search with no limitation on the study design, as this strategy ensures a relatively high recall of articles. The broad search may also be a limitation, as it may be difficult to compare adherence across studies with different study designs.

5 CONCLUSION

The systematic review will present an overview of the adherence to oral second line antidiabetic medications, including drugs marketed after 2015.

6 REFERENCES

- [1] H. Sun *et al.*, "IDF diabetes Atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045," *Diabetes research and clinical practice*, p. 109119, 2021.
- [2] G. Xu *et al.*, "Prevalence of diagnosed type 1 and type 2 diabetes among US adults in 2016 and 2017: population based study," *Bmj*, vol. 362, 2018.
- [3] M. J. Davies *et al.*, "Management of Hyperglycemia in Type 2 Diabetes, 2018. A Consensus Report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD)," *Diabetes Care*, vol. 41, no. 12, pp. 2669-2701, 2018, doi: 10.2337/dci18-0033.
- [4] J. A. Cramer, "A systematic review of adherence with medications for diabetes," *Diabetes care*, vol. 27, no. 5, pp. 1218-1224, 2004.
- [5] K. Iglay *et al.*, "Meta-analysis of studies examining medication adherence, persistence, and discontinuation of oral antihyperglycemic agents in type 2 diabetes," *Current medical research and opinion*, vol. 31, no. 7, pp. 1283-1296, 2015.
- [6] I. Krass, P. Schieback, and T. Dhippayom, "Adherence to diabetes medication: a systematic review," *Diabetic Medicine*, vol. 32, no. 6, pp. 725-737, 2015.
- [7] A. McGovern, Z. Tippu, W. Hinton, N. Munro, M. Whyte, and S. de Lusignan, "Comparison of medication adherence and persistence in type 2 diabetes: A systematic review and meta-analysis," *Diabetes, Obesity and Metabolism*, vol. 20, no. 4, pp. 1040-1043, 2018.
- [8] U.S. Food and Drug Administration. "Rybelsus: Highlights of Prescribing Information." https://www.accessdata.fda.gov/drugsatfda_docs/label/2019/213051s000lbl.pdf (accessed 07-JAN-2022, 2022).

7 ACKNOWLEDGEMENT

This work was supported by a research grant from the Danish Diabetes Academy, which is funded by the Novo Nordisk Foundation, grant number NNF17SA00.

Three Dimensional Perspective for Designing Healthcare Services

Gilmini Dantanarayana¹ and Prasad Jayaweera²

¹ University of Ruhuna, Matara, Sri Lanka gilmini@dcs.ruh.ac.lk

²University of Sri Jayewardenepura, Sri Lanka pja@sjp.ac.lk

Abstract

Successful e-Health system engineering mainly depends on accurate and complete modelling of the HealthCare (HC) processes. These HC service processes are governed by a variety of regulations and rules enforced by many distinct authorities. Availability of these governance directives mostly on paper based-medium makes a bunch of information logistics and related issues in HC processes of e-Health systems. Further, even captured such directives are getting often buried in lower technical realizations making its impossible not only real-time adoption but also manipulation on long run by respective non-technical higher authorities. In order to rectify and to facilitate these stakeholders' requirements for assistance with their access to these governing layers of e-Health solutions, in this work, we have proposed a three-dimensional ontological framework. This framework is expected to provide a complete and sound platform first to identify and then to develop eHealth solutions in compliance with those governing directives. In addition, it will ensure making convenient access with the non-technical higher healthcare authorities to monitoring and to governing HC processes. Proposed framework consists of three dimensions; 1) HC process activity dimension, 2) HC responsibility dimension and 3) HC directive enforcement dimension. The proposed approach facilitates the separation of concerns in HC governing perspectives in e-health solution development.

Keywords

e-Health, Healthcare Service Process, governing perspectives

1 INTRODUCTION

The evolution of E-Health system development can be classified into three main eras subjected to the overall requirement. In early 20th century the development effort mostly on technical message structure standardizations and (Health Level 7) HL7 versions have improved accordingly [1]. Later the development effort was motivated towards capturing not only technical level standardizations, but also strategic level realizations by means of semantic interoperability among heterogeneous applications [2][3]. According to the present studies, the evolving nature of complex requirements in HC context leads to the requirement of systematic mapping among the technical level requirements and strategic level governing requirements of the care pathways [4][5].

Also, many consequences have arisen due to this problem of less consideration of strategic level guidelines in HC [6]. Even though the requirement engineering of current e-Health systems follows the higher-level guidelines, the realizations of policy level requirements are hardcoded with the solution. Therefore, it is difficult to guide how to realize the governing aspects in the ultimate e-health solution. Thus, it is important to monitor and control governing perspectives of the organization in a separate layer while maintaining the integration with operation level requirements. Even though some higher level frameworks introduce the importance of maintaining governing perspectives separately, a systematic approach

with appropriate tools to address this requirement is lacking [6]. This research work provides a methodology to capture governing perspectives with three-dimensional framework. In order to align service oriented principals, HL7-OMG Healthcare Service Specification Project has proposed a trustworthy and policy-driven approach focuses on the capabilities and conformance criteria for HC services [4]. However, the step-by-step approach to realize the complex higher level requirement in e-Health solutions is unavailable. A recent study highlights the applicability of some OMG group standards to model clinical pathways and further analysis is required to derive a complete approach of policy based and rule-based service designing and deployment [5]. Therefore, this research provides a three-dimensional approach to realize rules in HC context as a separate conceptual model. The proposed study focuses on the care pathway guidelines to fulfil the requirement of rule-based service designing using appropriate modelling ontology. The remainder of this paper is organized as follows. Section 2 discusses the proposed three dimensional approach of HC rule realization with few subsections. Finally, Section 3 concludes the paper and presents directions for future work.

2 PROPOSED METHODOLOGY

The proposed approach is three dimensional; 1) HC process activity dimension, 2) HC responsibility

The 18th Scandinavian Conference on Health informatics, Tromsø, Norway, August 22-24, 2022. Organized by UiT The Arctic University of Norway. Conference Proceedings published by Linköping University Electronic Press at <https://doi.org/10.3384/ecp187>. © The Author(s). This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>

dimension and 3) HC directive enforcement dimension. The graph with three dimensions in Fig. 1 presented the choreography of the proposed ontology.

2.1 HC Process Activity Dimension

In the proposed three-dimensional approach, it is required to identify the service process activities as a first phase. Clinical work flow modelling is fundamental phase in service process identification where several existing tools and methodologies such as (Business Process Modelling Notation) BPMN can be applied [7] [8]. As major generic activities in HC care pathway, patient registration, diagnosis and treatment have been considered in this study.

2.2 HC responsibility dimension:

The corresponding responsibility spectrum for a certain action can be derived with the Healthcare Responsibility Assignment Matrix (HCRAM). Responsibility dimension covers the identification of multiple types of responsibility perspectives involved with certain process activity. Responsibility Assignment Matrix (RAM) facilitates to analyse and identify roles of the actors in different tasks [8]. Multi-actor involvement for accomplishing single task where each actor assigned with specific duty is more common in clinical context. Also the authorship variation with some delegations occurs corresponding to some tasks. This collaborative nature of the role task relationship types should be identified and modelled explicitly. Four essential matrices identified along with HC tasks are; Perform (P), Consult (C), Accountable (A), Informed (I).

2.3 HC Directive Enforcement Dimension

The rules that covered the specific action and the responsibilities could be identified as different enforcement levels. Enforcement levels have defined in the BMM is used and adopted to represent enforcements in HC context [9].

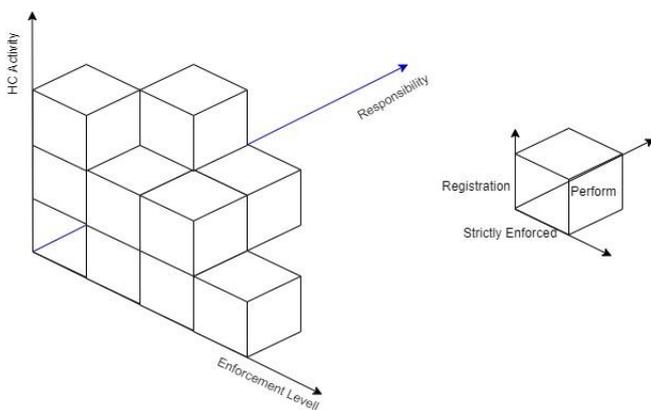


Figure 1 Three-Dimensional rule realization approach

3 CONCLUSION

The three-dimensional approach is proposed to identify and represent service rules in the healthcare domain. In activity dimension HC process perspectives have been captured and presented in terms of BPMN. Then in the

responsibility dimension, the responsibility perspectives have been modeled by utilizing HCRAM with four specific responsibility parameters. Enforcement level perspective, the third dimension is used to model service rules required to cover identified responsibility perspectives in the previous dimension. The work reported here introduced a conceptual framework in an endeavour to develop a complete and sound healthcare rule realization approach. The proposed work attempted to provide an extension to existed frameworks such as HL7/OMG that demanded as platform-neutral and network-accessible software service.

- [1] A. M. Iqbal, "LNCS 6719 - An OWL-DL Ontology for the HL7 Reference Information Model," *Lncs*, vol. 6719, pp. 168–175, 2011.
- [2] Y. F. Zhang, Y. Tian, T. S. Zhou, K. Araki, and J. S. Li, "Integrating HL7 RIM and ontology for unified knowledge and data representation in clinical decision support systems," *Comput. Methods Programs Biomed.*, vol. 123, pp. 94–108, 2016, doi: 10.1016/j.cmpb.2015.09.020.
- [3] J. C. Mcclay, P. J. Park, M. G. Janczewski, and L. H. Langford, "Standard for improving emergency information interoperability: the HL7 data elements for emergency department systems," pp. 529–535, 2015, doi: 10.1093/jamia/ocu040.
- [4] K. Kawamoto, A. Honey, and K. Rubin, "The HL7-OMG Healthcare Services Specification Project: Motivation, Methodology, and Deliverables for Enabling a Semantically Interoperable Service-oriented Architecture for Healthcare," *J. Am. Med. Informatics Assoc.*, vol. 16, no. 6, pp. 874–881, 2009, doi: 10.1197/jamia.M3123.
- [5] OMG Healthcare Domain Taskforce, "The BPM+ Field Guide to Shareable Clinical Pathways.," p. 125, 2020, [Online]. Available: <https://www.omg.org/cgi-bin/doc?health/2020-01-01>.
- [6] P. Ruotsalainen and B. Blobel, "Health information systems in the digital health ecosystem—problems and solutions for ethics, trust and privacy," *Int. J. Environ. Res. Public Health*, vol. 17, no. 9, May 2020, doi: 10.3390/ijerph17093006.
- [7] F. Zerbato, B. Oliboni, C. Combi, M. Campos, and J. M. Juarez, "BPMN-based representation and comparison of clinical pathways for catheter-related bloodstream infections," *Proc. - 2015 IEEE Int. Conf. Healthc. Informatics, ICHI 2015*, pp. 346–355, 2015, doi: 10.1109/ICHI.2015.49.
- [8] C. Cabanillas, M. Resinas, and A. Ruiz-Cortés, "A template-based approach for responsibility management in executable business processes," *Enterp. Inf. Syst.*, vol. 12, no. 5, pp. 550–586, 2018, doi: 10.1080/17517575.2017.1390166.
- [9] O. M. G. D. Number and M. C. Files, "Business Motivation Model," no. May, 2014.

Impact of the Covid-19 pandemic on use of Video consultations in a Swedish Primary care setting

Lovisa Jäderlund Hagstedt¹, Maria Hägglund^{2,3}, Göran Petersson⁴, Nadia Davoody¹, and Helena Hvitfeldt²

¹Health Informatics Center, Department of Learning, Informatics, Management and Ethics, Karolinska Institutet, Stockholm, Sweden Lovisa.jaderlund.hagstedt@tiohundra.se

²Healthcare Sciences and E-health, Department of Women's and Children's Health, Uppsala University, Sweden

³Medical Engineering and Physics, Uppsala University Hospital, Sweden

⁴eHealth Institute, Department of Medicine and Optometry, Linnaeus University, Kalmar, Sweden

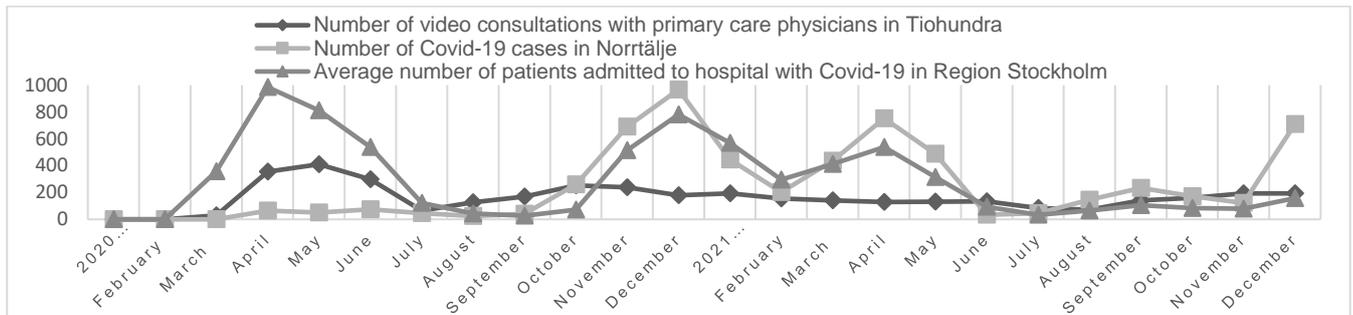


Figure 1. Video consultations and spread of Covid-19

1 INTRODUCTION

During the past decade the use of video consultations in healthcare have increased exponentially in Sweden [1, 2] as well as worldwide [3]. In Sweden, private online healthcare providers took the lead in this development, but in recent years many public health organizations in Sweden have developed their own solutions for video consultations, such as the app *Alltid Öppet* [Always open] (Region Stockholm) [4, 5]. However, implementing and integrating video consultations in traditional primary care has proven to be difficult.

In early 2020, the Covid-19 pandemic hit the world. The risk of infection was an incentive for both patients and providers to keep patients away from primary care centers [6, 7]. Instantly there was a need for alternative methods for healthcare service delivery [3]. Suddenly the previously rather widespread resistance to video consultations and other forms of digital contact between patients and healthcare professionals vanished [8, 9].

The objective of this study is to describe how the uptake of video consultations was affected by the Covid-19 pandemic, using data from a Swedish primary care setting.

2 METHODS

The study was conducted within the primary care of Tiohundra AB in Norrtälje, which consists of five primary care centers that combine care for approximately 35 000 patients, in a rural part of Region Stockholm. In March 2020 the video consultation solution *Alltid Öppet* was introduced in the Tiohundra primary care.

A descriptive approach was applied to examine the relationship between video consultations initiated by healthcare staff and the Covid-19 pandemic. Aggregated data on video consultations with primary care physicians in

Norrtälje Tiohundra, Covid-19 cases identified by positive PCR test in Norrtälje municipality, and the average monthly number of patients admitted to hospital with Covid-19 in Region Stockholm were collected for the years 2020 and 2021. The information system Medrave M4 [10], that collects clinical data from electronic health records, was used for the data collection on the video consultations. Data on the number of PCR verified Covid-19 cases was collected using statistics from the Public Health agency of Sweden [11]. The average monthly number of patients admitted to hospital with Covid-19 in Region Stockholm was collected using the Region's statistics on Covid-19 [12].

3 RESULTS

The number of video consultations increased from zero monthly visits in January and February 2020 to a peak of 410 in May. The average number of video consultations was 177 per month in 2020 compared to 144 in 2021. Figure 1 shows the number of video consultations per month and verified Covid-19 cases in Norrtälje municipality. However, public testing of Covid-19 was not initiated until June 2020 resulting in a low number of verified Covid-19 cases in the spring of 2020. Therefore, the average number of patients admitted to hospital in Region Stockholm is probably a better predictor of Covid-19 spread in the early 2020, also Figure 1.

Video consultations were mainly used by patients older than 25 years of age. The age distribution of patients that utilized video consultations is shown in Figure 2.

4 DISCUSSION

Prior to the pandemic there had been attempts to implement video consultations in the Tiohundra primary care but due to technical problems and a lack of interest by the healthcare staff, they were never adopted into clinical

The 18th Scandinavian Conference on Health informatics, Tromsø, Norway, August 22-24, 2022. Organized by UiT The Arctic University of Norway. Conference Proceedings published by Linköping University Electronic Press at <https://doi.org/10.3384/ecp187>. © The Author(s). This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>

practice. In the spring of 2020, however, there was an urgent need for alternative ways of delivering healthcare and almost overnight, the video consultation solution *Alltid Öppet* [5] was successfully implemented and considered to be advantageous.

The adoption of video consultations in public primary care in Sweden would probably have happened eventually, even without the pandemic. However, most certainly, the pandemic provided the necessary incentive for patients as well as healthcare professionals resulting in the rapid adoption. The pandemic also provided opportunities and time to test video consultations on all kinds of medical conditions, since otherwise the alternative in many cases would have been no consultation at all. This open-minded use of video consultations presumably facilitated the adoption.

There seem to be a relationship between the use of video consultations and spread of Covid-19 in 2020, where the use of video consultations varied with the contagion waves. In 2021 the use of video consultations was more consistent over time.

Previous studies have shown that people who uses video consultations tend to be younger, healthier, having higher socioeconomic status and living in urban areas where the supply of other healthcare is plentiful, contrary to what one might expect [13, 14]. Gabrielsson-Järhult e.g., showed that approximately 70% of the patients that used video consultations were between the ages of 0-30 years [13], whereas less than 10% of the patients in our study were younger than 25 years of age. This might be explained by the fact that the older patients are at larger risk for severe Covid-19, and therefore may have been more inclined to avoid traditional healthcare during the pandemic. Another possible explanation could be that the younger population in general suffer from less serious health issues and to a larger extent chose to refrain from visiting primary care during the pandemic.

Future research should explore whether patients in the older age groups will continue to use video consultations in the long term, even when the effects of the pandemic are reduced.

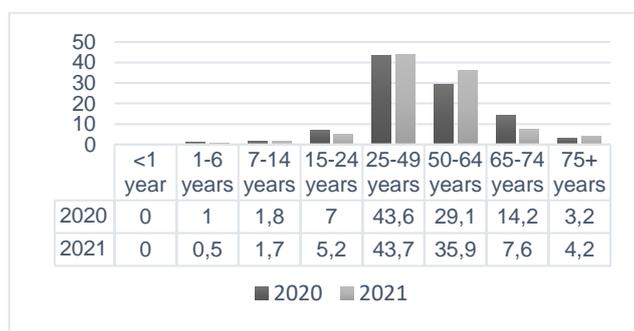


Figure 2. Age distribution and proportion of video consultations in percent

5 CONCLUSION

In summary, the Covid-19 pandemic had a direct impact on the adoption of video consultations in our primary care setting. Yet, our case study indicates that after the initial high use, it appears to have stabilized on a lower level. Further research is needed to understand whether this

finding is applicable to primary care all over Sweden, but also on the impact on healthcare professionals' work environment, patients' experiences, and quality of care.

6 REFERENCES

- [1] Cederberg J. Kraftig ökning av digital vård. (In Swedish: Sharp increase in digital care) *Läkartidningen*; 2020. <https://lakartidningen.se/aktuellt/nyheter/2020/05/kraftig-okning-av-digital-varld/>.
- [2] SKR. Snabb ökning av digitala vårdtjänster (In Swedish: Rapid increase in digital healthcare services). 2020. <https://skr.se/tjanster/press/nyheter/nyhetsarkiv/snabbokningavdigitalavardtjanster.33078.htm>.
- [3] Greenhalgh T, Wherton J, Shaw S, Morrison C. Video consultations for covid-19. *BMJ*. 2020 Mar 12;368:m998. doi: 10.1136/bmj.m998.
- [4] Cederberg J. Landstingen tar upp kampen om digital vård (in Swedish: Counties start digital care). *Läkartidningen*; 2018; pp. 115. <https://www.sls.regionstockholm.se/varld-hos-oss/alltid-oppet/>
- [5] Stokel-Walker C. Why telemedicine is here to stay. *BMJ*. 2020;371:m3603.
- [7] OECD. Strengthening the frontline: how primary health care helps systems adapt during the Covid-19 pandemic. Paris: OECD; 2021.
- [8] Hollander JE, Carr BG. Virtually Perfect? Telemedicine for Covid-19. *N Engl J Med*. 2020;382:1679–1681.
- [9] Greenhalgh T, Koh GC, Car J. Covid-19: a remote assessment in primary care. *BMJ*. 2020;368:m1182.
- [10] Medrave Software A <https://www.medrave.com/>
- [11] The Public Health Agency of Sweden, Number of covid-19 cases in the municipalities <https://experience.arcgis.com/experience/a6d20c1544f34d33b60026f45b786230>
- [12] Vårdgivarguiden, Statistics of covid-19 <https://vardgivarguiden.se/kunskapsstod/smittskydd/sjukdomar/coronavirus-2019-ncov/statistik-covid-19/>
- [13] Gabrielsson-Järhult F, Kjellström S, Josefsson KA. *Telemedicine consultations with physicians in Swedish primary care: a mixed methods study of users' experiences and care patterns*. *Scand J Prim Health Care*. 2021 Jun;39(2):204-213
- [14] Björn Ekman, Hans Thulesius, Jens Wilkens, Anna Lindgren, Olof Cronberg, Eva Arvidsson, Utilization of digital primary care in Sweden: Descriptive analysis of claims data on demographics, socioeconomic, and diagnoses. *International Journal of Medical Informatics*, Volume 127, 2019, Pages 134-140, ISSN 1386-5056, <https://doi.org/10.1016/j.ijmedinf.2019.04.016>.

7 ACKNOWLEDGEMENT

This work is funded by AFA Försäkring through the research project "ePrIm" (190210).

Exploring the validity of a contactless monitor used to measure vital parameters during sleep: A pilot study

Tanja Fredensborg Holm^{1*}, Julie Egmos¹, Katrine Holmstrup Sørensen¹, Ole Hejlesen¹,
Morten Hasselstrøm Jensen^{1,2}, and Stine Hangaard^{1,2}

¹Aalborg University, Department of Health Science and Technology, Aalborg, Denmark

²Steno Diabetes Center North Denmark, Aalborg University Hospital, Denmark

* Presenting author: tanja.holm@hotmail.com

Abstract

Study investigating the validity of a contactless monitor during sleep. Overall, the contactless monitor seems to be a valid method for monitoring vital parameters, even though parameters such as position and number of people in bed might have implications for monitoring.

Keywords

Type 1 diabetes - Nocturnal Hypoglycemia - Contactless Monitor - Validation - Vital parameters

1 INTRODUCTION

Diabetes mellitus (DM) is a major global challenge [1]. DM causes a chronic condition of hyperglycemia, which can lead to various health complications [2]. In type 1 diabetes (T1D), insulin treatment is required. However, insulin treatment induces a risk of hypoglycemia, which can be life-threatening [3]. The risk of hypoglycemia increases at night, leading to fear and reduced quality of life among people with T1D [4]. To prevent hypoglycemia, people with T1D frequently measure glucose at night [5]. Self-monitoring of blood glucose (SMBG) is a major negative burden for people with T1D [6]. Some people with T1D use a continuous glucose monitor (CGM) as an alternative to SMBG. However, CGMs also have disadvantages such as a discrepancy between glucose values measured with CGM and venous blood glucose values [7], and skin irritation [8]. Currently, there are no acceptable alternatives to CGM and SMBG for measuring blood glucose levels. However, there is a significant interest in developing new methods based on hypoglycemia induced changes in vital parameters [9]. A new contactless monitor (Sleepiz One, Sleepiz AG, Switzerland) may be able to predict hypoglycemia as it, based on radar technology, measures vital parameters that are affected hereby. However, there is no available documentation for the validity of the contactless monitor. Therefore, the aim of the present study was to investigate the validity of the contactless monitor during sleep.

2 METHODS

The study examined the contactless monitor's ability to monitor heart rate (HR) and respiration rate (RR). Furthermore, the sensitivity of the monitor related to the participants' position and the number of people in bed were assessed. Inclusion criteria were Danish or English-speaking healthy individuals aged ≥ 18 years. Exclusion criteria were pregnancy, tendonitis, carpal tunnel syndrome, implanted electronic devices, reduced circulation, epilepsy, and taking photosensitive medicine. Three women aged 25-31 years

participated in the study. Study duration was two nights, where the participant slept one night alone and one night with a partner.

HR and RR from the monitor were compared with HR from a smartwatch (Fitbit Charge 2) and manual RR counts (Figure 1).

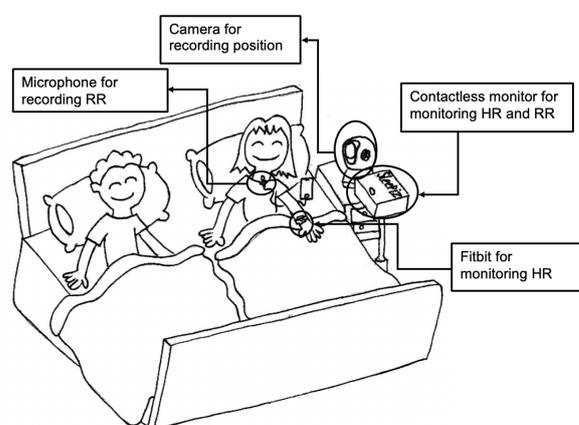


Figure 1. Illustration of experimental setup.

A correlation analysis was performed to examine the correlation between data from the contactless monitor and reference measurements for each participant. In addition, Root Mean Square Error (RMSE), Pearson's correlation coefficient (r), and 95% confidence interval (95% CI for r) were calculated.

3 RESULTS

For both HR and RR, a statistically significant correlation was found between data from the contactless monitor and reference measurements, except for one test. Correlation of HR from the monitor and smartwatch was R^2 (0.0183 - 0.6948), RMSE (1.65 - 3.87), and r (0.1354 - 0.8336). Correlation of RR from the monitor and manual RR counts was R^2 (0.5551 - 0.895),

RMSE (0.22 - 0.73), and r (0.7451 - 0.9460). In the absence of missing data, supine was the dominant position (Figure 2A and 2B). According to the number of people in bed, no unambiguous results were found (Figure 2C and 2D).

Participant	p -value			
	HR	HR	RR	RR
	Alone	> one	alone	> one
P1	0.630	0.001	0.000	0.000
P2	0.009	0.027	0.000	0.000
P3	0.000	0.011	0.001	0.000

Table 2 P -value from correlation between reference measurements and the contactless monitor for HR and RR, related to participants (P) and number of people in bed. Alone in bed (alone), more than one in bed (> one).

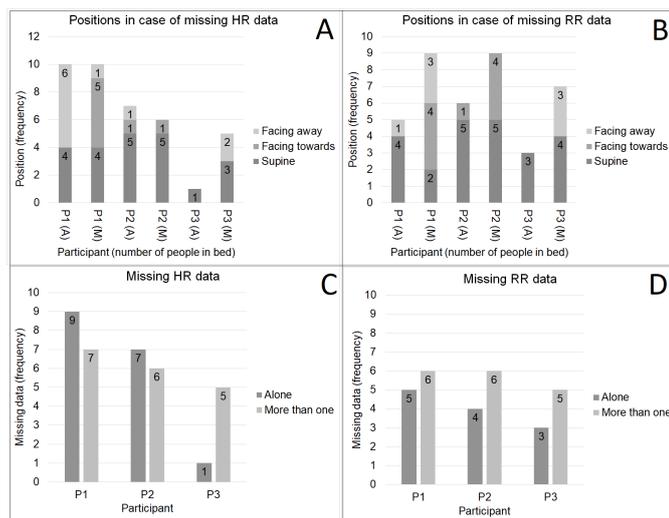


Figure 2. Sensitivity of the contactless monitor related to position and number of people in bed. Participant (P), sleeping alone (A), sleeping more than one in bed (M). Figure A,B illustrates whether the participant is facing away from the monitor, is facing towards the monitor or is in supine position in case of missing HR and RR data registration > 2 min. Figure C,D illustrates the frequency of missing data registration > 2 min. depending on whether the participant slept alone or more than one in bed.

4 DISCUSSION

The aim of the present study was to investigate the validity of the contactless monitor during sleep. HR and RR data measured with the contactless monitor correlated with data from the reference measurements. Turppa et al. found a statistically significant correlation for HR and RR measured with radar technology and reference technology, respectively ($p < 0.01$) [10]. This corresponds to the results of the present study. Turppa et al. assessed that radar technology could be used to accurately monitor vital parameters during sleep [10]. It is assessed that the contactless monitor in the present study has the same potential for monitoring vital parameters during sleep.

In conclusion, the contactless monitor seems to be a valid method for monitoring vital parameters during sleep. However,

the present study suggests that the user's position in bed may have implications for monitoring. It is not possible to determine the significance of the number of people in the bed as there are no unambiguous results.

5 LIMITATIONS

A limitation of the present study is the small and homogenous sample size of three women, which make the results less generalizable. Furthermore, the short duration of the study. Therefore, increased sample size, inclusion of both genders, and a longer study duration is recommended for future studies. Another limitation is that the missing data, may be caused by other reasons than position and number of people in bed. Therefore, future research is needed to evaluate the sensitivity of the monitor related to the position and the number of people in bed during sleep.

6 REFERENCES

- [1] International Diabetes Federation. IDF diabetes atlas [Online]. 2019. Available from: https://www.diabetesatlas.org/upload/resources/material/20200302_133351_IDFATLAS9e-final-web.pdf
- [2] Ramachandran A, Snehalatha C, Nanditha A. Classification and Diagnosis of Diabetes. I: Textbook of Diabetes, pp. 23–29. United Kingdom, 2017.
- [3] Cryer PE, Arbeláez AM. Hypoglycemia in Diabetes. I: Textbook of Diabetes, pp. 513-529. United Kingdom, 2017.
- [4] Edelman SV, Blose JS. The Impact of Nocturnal Hypoglycemia on Clinical and Cost-Related Issues in Patients With Type 1 and Type 2 Diabetes. *Diabetes Educ.* Vol. 40, pp. 269-279. 2014.
- [5] Wild D, von Maltzahn R, Brohan E, Christensen T, Clauson P, Gonder-Frederick L. A critical review of the literature on fear of hypoglycemia in diabetes: Implications for diabetes management and patient education. *Patient Education and Counseling.* Vol. 68, pp. 10-15. 2007.
- [6] Szadkowska A, Czyżewska K, Pietrzak I, Mianowska B, Jarosz-Chobot P, Myśliwiec M. Hypoglycaemia unawareness in patients with type 1 diabetes. *Pediatr Endocrinol Diabetes Me.* Vol. 24, pp. 126-134. 2018.
- [7] Ludwig V, Kulzer B, Schnell O, Heinemann L. Current Trends in Continuous Glucose Monitoring. *J Diabetes Sci Technol.* Vol. 8, pp. 390-406. 2014.
- [8] Englert K, Ruedy K, Coffey J, Caswell K, Steffen A, Levandoski L. Skin and Adhesive Issues With Continuous Glucose Monitors: A Sticky Situation. *J Diabetes Sci Technol.* Vol. 8, pp. 745-751. 2014.
- [9] Diouri O, Cigler M, Vettoretti M, Mader JK, Choudhary P, Renard E, m.fl. Hypoglycaemia detection and prediction techniques: A systematic review on the latest developments. *Diabetes/Metabolism Research and Reviews.* Vol. 37, pp. 1-19. 2021.
- [10] Turppa E, Kortelainen JM, Antropov O, Kiuru T. Vital Sign Monitoring Using FMCW Radar in Various Sleeping Scenarios. *Sensors.* Vol. 20, pp. 1-19. 2020.

Pilot Testing and Evaluation of Participatory Patient Record in Psychiatric Care

Molly Gibson^{1*}, Magne Halvorsen^{2*}, Mikkel Angelo Anchissi Joner^{2*}, Mari Lehne^{2*}, Elise Rabassa Stautland^{2*}, Gustaf Svensson^{2*}, Hanna Thevik^{2*}, Georgia Fehler^{3^}, Stefan Hochwarter^{2^}

¹Norwegian University of Science and Technology, Department of Computer Science, Norway, mollypg@stud.ntnu.no

²Norwegian University of Science and Technology, Department of Computer Science, Norway

³ Medical University of Brandenburg, Digital Health group at Center for Health Services Research, Germany

*These Authors share first authorship

^These authors share senior authorship

Abstract

The PEPPPSY (Pilot Testing and Evaluation of Participatory Patient Record in Psychiatric Care) project is a new initiative from the Digital Health group of the Centre for Health Services Research in Brandenburg. PEPPPSY stands for “Pilot Testing and Evaluation of a Participatory Patient Record in Psychiatric Care”. The project is based on the United States-based initiative, OpenNotes[1], an international movement promoting and studying transparent communication in healthcare. The project is created with open communication in mind, as it aims towards giving patients insight into their treatments by accessing their psychiatric records. This will hopefully increase trust, engagement, and understanding of both themselves and the therapeutic process.

Keywords

Health Informatics System, User Centred Design, PAEHR, mental health, patient portals, user involvement

1 INTRODUCTION

A Patient Accessible Electronic Health Record (PAEHR), which includes visit notes, laboratory results and diagnoses, is not common practice in Germany. Patients can claim their records by law, but it is a bureaucratic process, making it rare for the patients to do so[2]. Of the patients surveyed, only about 41% of the patients have claimed this right, and of all cases UPD1 handles, about 44% are cases of record-accession conflicts[2]. For psychiatric patients, there are even stronger boundaries, such as needing the permission of a clinician.

Studies indicate that patients may not remember even half of the information relayed to them by their healthcare provider[3][4]. Providing clients with real-time access to their health records will minimize the loss of important information and allow post-appointment reviews. By reading through their notes, patients can identify mistakes and ensure that their patient records are accurate[1]. Giving patients access to their therapist's notes creates mutual trust, which in turn leaves the patient feeling more in control of their health[2].

OpenNotes, and the sharing of patient records, is about transparent communication in general. What makes PEPPPSY unique is that PEPPPSY is centred around providing psychiatric care patients access to their notes made by their health care providers.

The 18th Scandinavian Conference on Health informatics, Tromsø, Norway, August 22-24, 2022. Organized by UiT The Arctic University of Norway. Conference Proceedings published by Linköping University Electronic Press at <https://doi.org/10.3384/ecp187>. © The Author(s). This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>

The Digital Health group at MHB implemented the first pilot of PEPPPSY[5]. The pilot is a web application that functions as a digital one-way message platform. It makes all notes made by health professionals available to the patient that the note is about. The further development of the PEPPPSY project was performed by a group of students at the Norwegian University of Science and Technology as part of their bachelor program.

2 METHODS

The PEPPPSY pilot[5] started in June of 2021, and is an ongoing research project that follows an explorative and qualitative approach. The research group based in Germany conducted data analysis and evaluated the pilot in practice. Nine health professionals and 14 service users participated in the study and evaluated the prototype. The study is conducted at a rural German mental health day clinic. The bachelor group based in Norway implemented changes based on the feedback gained from the study and further developed the application over a 5-month period.

The application is developed with Next.js, a hybrid static, and server rendering node framework. It also uses React, for rendering the page, and Webpack, for minifying the code. The user authentication is handled by Next-Auth, a plug-in for Next.js, and Everify for two-factor authentication. The web application has a three-tiered

client-server architecture. The separation of presentation, logic, and data storage makes it easier to develop those three independently of each other. Furthermore, modularization increases maintainability as it results in smaller fine-grained components. Using a RESTful API helps separate the presentation of, and the logic surrounding retrieval/storage of data. The resource sent to the client is conceptually different from the internal representation in the database. This allows the data storage and consumption to be developed independently, as long as the interface stays consistent. We followed an agile development process to iteratively including findings from the ongoing pilot phase in practice, and to evaluate newly introduced features.

3 RESULTS

The resulting prototype currently being used is organized based on the three main user types: Health Professionals, Patients and Administrators. For data privacy and GDPR concerns it was important to restrict users to only data they had permissions to access.

The administrator users can create, edit and delete patients and therapists as well as clinics. To maintain confidentiality, the administrators are restricted from accessing the patients' notes or any information about an individual patient. Patient users can only access information about their own notes created by health professionals, and all Health professionals in a clinic can view information and notes pertaining to patients in their own clinic. The health professionals can create and edit patients in their own clinic, as well as create and edit notes on a patient in their own clinic. The health professionals can comment on notes made by themselves or another health professional.

The key feature of the existing solution was the patients' ability to read notes that therapists had published on them. The patients could see when a note was published as well as which occupational group the author of the note belonged to. Along with the ability to publish notes on their patients, the therapists could create patient users. The administrator did not have the ability to read patients' notes, and the admin could not access any information about patients or clinics, nor any information about therapists after they had been created. Even though the project centres around providing the patients a certain amount of information concerning their care, it should not substitute the actual institution which provides said care. The patients cannot use the service provided by PEPPSY as a 24/7 on-call therapist, and the service is only available during opening hours of the clinic.

In the feedback from the current prototype being tested in Germany, a need for more two-way communication was discovered, therefore comments on notes were implemented into the solution. The user can edit and delete comments on their notes, for the purpose of creating a better communication flow between health professionals and patients. When receiving a new note or comment, the users

get a notification. Another important feature added is that administrator may also download a .CSV file that details statistics of every clinic to help with further research. This CSV file only contains clinic level information and does not include individual patients.

4 DISCUSSION AND CONCLUSION

Our implementation addressed needs and challenges uncovered during the pilot phase of the initial system. Patients requested a way to interact with their notes and to create comments. Comments on notes can facilitate discussion of subsequent therapy sessions and provide an asynchronous communication channel for both patients and therapists. Yet, the role of these comments for the therapy process needs to be clearly defined. The wish to interact with notes are in line with previous research [4].

5 REFERENCES

- [1] Wolff, Jennifer L. Darer, Jonathan D. Berger, Andrea. Clarke, Deserae. Green, Jamie A. Stametz, Rebecca A. Delbanco, Tom. Walker, Jan. *Inviting patients and care partners to read doctors' notes: OpenNotes and shared access to electronic medical records*, Journal of the American Medical Informatics Association, Volume 24, Issue e1, April 2017, Pages e166–e172, <https://doi.org/10.1093/jamia/ocw108>
- [2] Dr. Inga Münch and Marion Grote Westrick. SPOTLIGHT Healthcare: OpenNotes. 2020. url: <https://www.bertelsmannstiftung.de/de/publikationen/publikation/did/spotlight-healthcare-opennotes> (visited on 3rd May 2022).
- [3] Kessels, R P. *'Patients' memory for medical information*. In: JRSM 96.5 (2003), pp. 219–222. doi: 10.1258/jrsm.96.5.219.
- [4] Schwarz J, Bärkås A, Blease C, Collins L, Hägglund M, Markham S, Hochwarter S. *Sharing Clinical Notes and Electronic Health Records With People Affected by Mental Health Conditions: Scoping Review* JMIR Ment Health 2021;8(12):e34170
URL: <https://mental.jmir.org/2021/12/e34170>
DOI: 10.2196/34170
- [5] Hochwarter S. Fehler, Georgia. Munte, Catharina. Eisenmann, Yvonne. Heinze, Martin. Hibs, Charline. Schwarz, Julian. *Design of a Patient-Accessible Electronic Health Record System in Mental Health* Studies in Health Technology and informatics 2022;294, 583-584
URL: <https://ebooks.iospress.nl/doi/10.3233/SHTI220534>
DOI: 10.3233/SHTI220534

Quantification of Insulin Adherence in Adults with Insulin-treated Type 2 Diabetes: A Protocol for a Systematic Review

Jannie Toft Damsgaard Nørlev¹, Morten Hasselstrøm Jensen^{1,2}, Ole Hejlesen¹, Tinna Björk Aradóttir³, Nicholas Ciccone³, Stine Hangaard^{1,2}

¹Aalborg University, Health Science and Technology, Faculty of Medicine, Denmark jadano@hst.aau.dk

²Steno Diabetes Center North Denmark. Aalborg, Denmark

³Novo Nordisk A/S, Bagsværd, Denmark

Abstract

A systematic review will be conducted to update the evidence base and provide an overview of insulin adherence quantification in adults with insulin-treated type 2 diabetes, i.e., the methods used to assess insulin adherence and the cut-off points that constitute adequate insulin adherence.

Keywords

Type 2 Diabetes, Insulin Adherence, Quantification, Systematic Review

1 INTRODUCTION

The progressive nature of type 2 diabetes (T2D) necessitates insulin therapy for most people to achieve glycaemic control [1]. Although the full benefit of insulin therapy will be accomplished only if the person with T2D complies with the prescribed insulin regimen reasonably closely [2], nonadherence to insulin therapy is common in adults with T2D [3].

While nonadherence is recognized as a key contributor to poor outcomes, many clinicians feel unable to address nonadherence [2]. A study by MacEwan et al. demonstrated that assessment of adherence rates is generally not better than a random guess (53 vs. 50%) [4].

Precise measure or assessment of insulin adherence is acknowledged as an important prerequisite to improving insulin nonadherence [3] and interpreting the effects of an intervention [5]. Results and clinical outcomes from an intervention cannot be interpreted realistically without information regarding correct adherence, i.e., if therapy fails to achieve the desired outcome, there is a risk that the clinician or researcher assume that a drug failure has occurred if adherence is assessed incorrectly and nonadherence, therefore, is not discovered [5]. Yet, there is no gold-standard method to assess insulin adherence and no cut-off point or consensual standard for what constitutes adequate insulin adherence [2].

In recent years, advances in technology have brought improvements in the field of diabetes. The spectrum of new technologies spans smartphone apps, smart pens, and insulin pumps [6]. Access to these new technologies may have had an impact on how insulin adherence is assessed. Despite this, no systematic review within the area has been published since 2016 [7].

This systematic review aims to update the evidence base by including literature published from 2012 to the time of the

review and provide an overview of the methods used to assess insulin adherence and the cut-off points that constitute adequate insulin adherence in adults with insulin-treated T2D.

2 METHODS

A systematic review will be conducted and reported according to the PRISMA 2020 checklist [8]. The systematic review protocol was submitted for registration with PROSPERO on May 20, 2022 and has not yet received a registration number. The protocol will form the basis of the review. The review process is illustrated in Figure 1.

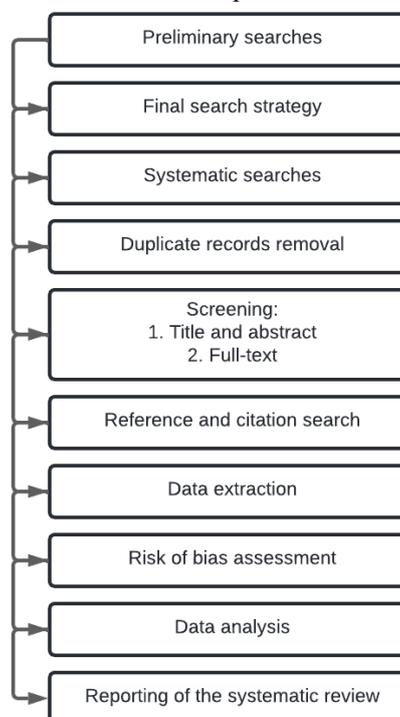


Figure 1. The review process.

The 18th Scandinavian Conference on Health informatics, Tromsø, Norway, August 22-24, 2022. Organized by UiT The Arctic University of Norway. Conference Proceedings published by Linköping University Electronic Press at <https://doi.org/10.3384/ecp187>. © The Author(s). This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>

To qualify the systematic search preliminary searches will be performed to obtain an overview of published literature and to identify relevant index terms, search terms, and keywords. A systematic search will be performed in PubMed, Embase, Cinahl, and PsycINFO. The search will include three blocks: type 2 diabetes, insulin, and adherence. Synonyms, near-synonyms, acronyms, index terms, and spellings for each keyword will be identified. Different search functions such as Boolean operators, truncation, thesaurus, phrase searching, and text word (title, abstract, keyword) will be applied to focus and structure the search. Studies published from 2012 to the time of the review and describe a method to assess insulin adherence and include details on the cut-off point will be considered. Primary full-text studies in English, Danish, Norwegian, or Swedish will be screened for inclusion, except for study protocols and animal research. Reference lists will be hand-searched and citation searching will be conducted to identify additional relevant studies within the field.

Data extraction will include details of the methods and cut-off points used to assess insulin adherence, insulin regimen, the population, and study design. The risk of bias will be assessed for each of the included studies using critical appraisal tools from the Joanna Briggs Institute (JBI) [9]. The results of the systematic search, screening, and risk of bias assessment will be reported in full in the systematic review. By the PRISMA guidelines, the screening process will be presented in a PRISMA 2020 flow diagram [8]. A narrative synthesis will be provided with the information presented in text and tables. The identified methods to assess insulin adherence will guide the organization and description of the results. Cut-off points from each study will be listed.

The systematic search will be performed by one reviewer and facilitated by a research librarian with expertise and experience in medical science and diabetes. Title and abstract screening will be done by one reviewer, while two independent reviewers will screen full-text articles. Data extraction and analysis will be performed by one reviewer. In the event of questions or doubts, the review co-authors will be consulted, and an agreement will be reached by discussion.

3 RESULTS

The results will update the evidence base by providing an overview of the reported methods to assess insulin adherence and the cut-off points used to define adequate adherence. Potential novel methods or technologies to assess insulin adherence will be included and the results will provide insight into potential gaps within the field. The results are expected to be published by the end of 2022.

4 DISCUSSION AND CONCLUSIONS

This systematic review will clarify methods used to assess insulin adherence and the cut-off points that constitute adequate insulin adherence in adults with T2D. Potentially, the results can be used to guide clinicians and researchers when selecting a method to assess insulin adherence in adults with T2D. The systematic review may also inform of new methods or technologies used to assess insulin adherence if such has been implemented. Hence, the results could potentially pioneer the implementation of future

technologies or methods for the assessment of insulin adherence in adults with T2D.

5 REFERENCES

- [1] Mahler, R.J., Adler, M.L. "Type 2 Diabetes Mellitus: Update on Diagnosis, Pathophysiology, and Treatment" in *The Journal of Clinical Endocrinology & Metabolism* Vol. 84, Issue 4, pp. 1165-1171. 1999.
- [2] Osterberg, L., Blaschke, T. "Adherence to Medication" in *N Engl J Med* Vol. 353, Issue 5, pp. 487-497. 2005.
- [3] Cramer, J.A. "A Systematic Review of Adherence With Medications for Diabetes" in *Diabetes Care* Vol. 27, Issue 5, pp. 1218-1224. 2004.
- [4] MacEwan, J.P., Silverstein, A.R., Shafrin, J., Lakdawalla, D.N., Hatch, A., Forma, F.M. "Medication Adherence Patterns Among Patients with Multiple Serious Mental and Physical Illnesses" in *Adv Ther.* Vol. 35, Issue 5, pp. 671-685. 2018.
- [5] Farmer, K.C. "Methods for Measuring and Monitoring Medication Regimen Adherence in Clinical Trials and Clinical Practice" in *Clinical Therapeutics* Vol. 21, Issue 6, pp. 1074-1090. 1999.
- [6] Adolffsson, P., Hartvig, N., Kaas, A., Møller, J., Hellman, J. "Increased time in range and fewer missed Bolus Injections after introduction of a smart connected insulin pen" in *Diabetes Technology and Therapeutics* Vol. 22, Issue 10, pp. 709-718. 2020.
- [7] Stolpe, S., Kroes, M.A., Webb, N., Wisniewski, T. "A Systematic Review of Insulin Adherence Measures in Patients with Diabetes" in *JMCP* Vol. 22, Issue 11, pp. 1224-1246. 2016.
- [8] Page, M.J., McKenzie, J.E., Bossuyt, P.M., et al. "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews" in *BJM* Vol. 372, Issue 71. 2021.
- [9] Aromataris, E., Munn, Z., "JBI Manual for Evidence Synthesis" in *JBI*, pp. 217-270. 2020.

Exploring the user-friendliness of a contactless monitoring system used for sleep monitoring: A usability study

Katrine Holmstrup Sørensen^{1*}, Tanja Fredensborg Holm¹, Julie Egmos¹, Ole Hejlesen¹,

Morten Hasselstrøm Jensen^{1,2}, and Stine Hangaard^{1,2}

¹Aalborg University, Department of Health Science and Technology, Aalborg, Denmark

²Steno Diabetes Center North Denmark, Aalborg University Hospital, Denmark

* Presenting author: katrine.holmstrup@gmail.com

Abstract

A usability study exploring the usability of a contactless monitoring system used during sleep. Overall, the participants found the contactless monitor easy and satisfying, but further improvements to the user manual and design might optimize the usability of the monitor.

Keywords

Type 1 diabetes - Nocturnal Hypoglycemia - Contactless Monitor - User-friendliness - Usability Study

1 INTRODUCTION

Diabetes is an increasing global health challenge [1]. The most common complication among people with type 1 diabetes (T1D) is hypoglycemia [2]. Strong counterregulatory responses exist if hypoglycemia occurs [3]. Activation generates hormone creation, which causes physiological changes in e.g., heart rate (HR) and respiration rate (RR) [4–6]. More than 50% of critical cases of hypoglycemia occur at night [7], leading to fear and reduced quality of life for people with diabetes and their relatives [7,8]. As a result, people with T1D frequently measure glucose at night to prevent hypoglycemia, which causes a major negative load [2]. Continuous glucose monitoring (CGM) or self-monitoring of blood glucose (SMBG) remains the decisive marker in the detection of hypoglycemia [9]. CGM is found to estimate inaccurate glucose values during hypoglycemia [10]. In comparison, SMBG depends on frequent monitoring, which can cause massive pain, scarring, and loss of sensibility [11]. Due to disadvantages and lack of access to CGMs, there is an increasing interest in methods that can predict hypoglycemia based on physiological changes [9]. The contactless monitor (Sleepiz One, Sleepiz AG, Switzerland) may be used as an alternative prediction method as it can monitor vital parameters affected by hypoglycemia. However, the contactless monitor is a newly developed technology, and the usability of the monitor is uncertain. Therefore, it is highly relevant to explore the usability of the monitor as this is an important parameter for a successful implementation of new technology [12]. Thus, the aim of the present study was to explore the usability of a contactless monitor used for sleep monitoring.

2 METHODS

The present study was a usability test, which was performed at Aalborg University, Denmark. Five healthy individuals aged ≥ 18 years were included. They were recruited through social media. Exclusion criteria were pregnancy, diagnosed cognitive challenges, vulnerability, and implemented electronic devices. The usability test was divided into two parts (Figure 1). Part one was a “thinking aloud test” followed by follow-up questions. The participants were instructed to use a user

manual to set up the contactless monitor in a laboratory setting while thinking aloud. In part two, the participants slept with the monitor at home for one night, after which a semi-structured interview was conducted. The interview guide was based on the four components: learnability, memorability, errors and satisfaction, which according to Nielsen defines usability [13]. The components were furthermore used as a starting point for an overall assessment of the user-friendliness of the contactless monitor. Data were analyzed using Kvale and Brinkman’s thematic approach [14].

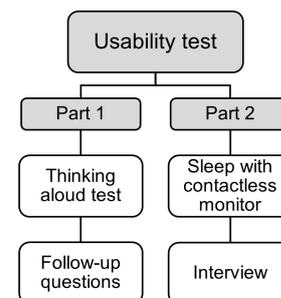


Figure 1. Overview of the two parts of the usability test.

3 RESULTS

Five participants aged 26-73 years were included in the study (two females). Four participants completed the study, one participant was excluded from part two due to illness. A total of 10 themes, related to the four components defining usability, were identified (Figure 2). Overall, the participants found that the contactless monitor was easy to use. The participants found that the monitor was comfortable to use, as it did not require any physical contact and thus was not perceived as physically uncomfortable. When the participants used the user manual they were in doubt regarding the location and position of the monitor. Furthermore, the participants found it essential to follow the user manual step-by-step to prevent errors in the setup. Overall, the participants found it easier to set up and use the contactless monitor for the second time at home. The majority of the participants stated that they did not have a feeling of surveillance when using the monitor. The

participants quickly grew accustomed to the monitor, and the monitor did therefore not affect their sleep. To increase the usability some of the participants suggested improvements related to the design and the user manual. Some participants expressed concerns about using the monitor due to the risk of radiation. All the participants were interested in using the monitor in case of a relevant disease.

Learnability	Memorability	Errors	Satisfaction
<ul style="list-style-type: none"> • Usermanual • First meeting 	<ul style="list-style-type: none"> • Easier second time 	<ul style="list-style-type: none"> • Follow the usermanual 	<ul style="list-style-type: none"> • Design • Power supply • Access to data • Anxiety • Accept • Overall assessment

Figure 2. Overview of the identified themes related to main keywords.

4 DISCUSSION

The present study aimed to explore the usability of a contactless monitor used for sleep monitoring. Overall, the participants found that the contactless monitor was easy to use. The participants found that the contactless monitor was not physically uncomfortable. Previous studies found that people preferred contactless monitoring over contact-based monitoring and that contact-based monitoring could be related to skin irritation, discomfort, and constraining movement [15,16]. All the participants agreed that the step-by-step guide in the user manual helped them set up the monitor. A previous study found that a step-by-step structure of the user manual was considered satisfying among users [17]. However, the participants had suggestions for improvements that could increase the usability of the user manual and the contactless monitor.

5 CONCLUSION

In conclusion, the contactless monitor was user-friendly, as it was easy and satisfying to use. However, the usability of the contactless monitor may be optimized by further improving the user manual and design of the monitor.

6 LIMITATIONS

A limitation of this study is the relatively small sample size of five participants, as new knowledge was still generated during the last interview, which might indicate that the number of participants were not sufficient to achieve data saturation. Therefore, increased sample size is recommended for future studies.

7 REFERENCES

[1] International Diabetes Federation. IDF diabetes atlas [Online]. 2019. Available from: https://www.diabetesatlas.org/upload/resources/material/20200302_133351_IDFATLAS9e-final-web.pdf

[2] Szadkowska A, Czyżewska K, Pietrzak I, Mianowska B, Jarosz-Chobot P, Myśliwiec M. Hypoglycaemia

unawareness in patients with type 1 diabetes. *Pediatr Endocrinol Diabetes Me.* Vol. 24, pp. 126-134. 2018.

[3] Martín-Timón I, Del Cañizo-Gómez FJ. Mechanisms of hypoglycemia unawareness and implications in diabetic patients. *World J Diabetes.* Vol. 6, pp. 912-926. 2015.

[4] Thompson EL, Ray CJ, Holmes AP, Pye RL, Wyatt CN, Coney AM, m.fl. Adrenaline release evokes hyperpnoea and an increase in ventilatory CO₂ sensitivity during hypoglycaemia: a role for the carotid body. *J Physiol.* Vol. 594, pp. 4439-4452. 2016.

[5] Ward DS, Voter WA, Karan S. The effects of hypo- and hyperglycaemia on the hypoxic ventilatory response in humans. *J Physiol.* Vol. 582, pp. 859-869. 2007.

[6] Frier BM, Heller S, McCrimmon R. *Hypoglycaemia in Clinical Diabetes.* pp. 14. United Kingdom, 2014.

[7] Edelman SV, Blose JS. The Impact of Nocturnal Hypoglycemia on Clinical and Cost-Related Issues in Patients With Type 1 and Type 2 Diabetes. *Diabetes Educ.* Vol. 40, pp. 269-279. 2014.

[8] Martyn-Nemeth P, Schwarz Farabi S, Mihailescu D, Nemeth J, Quinn L. Fear of hypoglycemia in adults with type 1 diabetes: impact of therapeutic advances and strategies for prevention - a review. *Journal of Diabetes and its Complications.* Vol. 30, pp. 167-177. 2016.

[9] Diouri O, Cigler M, Vettoretti M, Mader JK, Choudhary P, Renard E, m.fl. Hypoglycaemia detection and prediction techniques: A systematic review on the latest developments. *Diabetes/Metabolism Research and Reviews.* Vol. 37, pp. 1-19. 2021.

[10] Teo E, Hassan N, Tam W, Koh S. Effectiveness of continuous glucose monitoring in maintaining glycaemic control among people with type 1 diabetes mellitus: a systematic review of randomised controlled trials and meta-analysis. *Diabetologia.* Vol. 65, pp. 604-619. 2022.

[11] Heinemann L. Finger Pricking and Pain: A Never Ending Story. *J Diabetes Sci Technol.* Vol. 1, pp. 919-921. 2008.

[12] Ammenwerth E. Technology Acceptance Models in Health Informatics: TAM and UTAUT. *Applied Interdisciplinary Theory in Health Informatics.* pp. 64-71. 2019.

[13] Nielsen J. Usability 101: Introduction to Usability. Nielsen Norman Group [Internet]. 2012 [henvist 6. oktober 2021]; Available at: <https://www.nngroup.com/articles/usability-101-introduction-to-usability/>

[14] Kvale S, Brinkmann S. *InterViews: learning the craft of qualitative research interviewing.* Third edition. pp. 405. Los Angeles, 2015.

[15] Piantino J, Luther M, Reynolds C, Lim MM. Emfit Bed Sensor Activity Shows Strong Agreement with Wrist Actigraphy for the Assessment of Sleep in the Home Setting. *Nat Sci Sleep.* Vol. 13, pp. 1157-1166. 2021.

[16] Kebe M, Gadhafi R, Mohammad B, Sanduleanu M, Saleh H, Al-Qutayri M. Human Vital Signs Detection Methods and Potential Using Radars: A Review. *Multidisciplinary Digital Publishing Institute.* Vol. 20, pp. 1454. 2020.

[17] Møller MH. Usability Testing of User Manuals. *Communication & Language at Work.* Vol. 2, pp. 51-59. 2013.

Exploring Digital Psychosocial Follow-up for Survivors of Childhood Critical Illness

Marte Hoff Hagen

Norwegian University of Science and Technology, Department of Computer Science, Norway marte.h.hagen@ntnu.no

Abstract

This extended abstract describes the plan and status of a PhD project using a design science research approach to explore how to design digital psychosocial follow-up for survivors of childhood critical illness.

Keywords

Childhood critical illness, Long-term follow-up care, Psychosocial, Digitalization, Design science research

1 INTRODUCTION

The risk of long-term negative psychosocial consequences may be increased by intensive treatment procedures and hospitalization in childhood [1 - 3]. This risk applies especially to survivors of Anorectal malformations and Hirschsprung's disease [4 - 7], two congenital colorectal defects which require surgery and intensive treatment procedures [7, 8]. Research indicates that information technology solutions provide beneficial psychosocial outcomes to survivors of critical illness (e.g., [9, 10]), but these examples target only childhood cancer survivors. Thus, this PhD project aims to explore designing solutions for digital psychosocial follow-up for survivors of childhood critical illness in general, not only childhood cancer survivors.

2 METHODS

This PhD project uses a design science research approach [11] to investigate how an innovative information technology solution for digital psychosocial follow-up for survivors of childhood critical illness should be designed according to the current knowledge and the different stakeholders' needs. A specific solution will be developed for the target group identified as child survivors of Anorectal malformations and Hirschsprung's disease to create general design principles regarding digital psychosocial follow-up for all different survivors of childhood critical illness.

The first step of the design science research approach is to identify and motivate the problem. To identify the problem, expert interviews [12] were conducted with six health professionals at a Norwegian hospital with relevant knowledge of the current psychosocial follow-up for survivors of childhood critical illness at the hospital. The qualitative data gained in this sub-study were analyzed through thematic analysis [13].

3 RESULTS

The expert interviews resulted in a mapping of the current psychosocial follow-up for survivors of childhood critical illness at the Norwegian hospital, presented in Figure 1.

4 DISCUSSION

The expert interview findings demonstrated the importance of considering the affecting factors and technological opportunities and limitations when designing digital psychosocial follow-up for survivors of childhood critical illness. Based on these findings, it was created the following preliminary design principles for the information technology solution of digital psychosocial follow-up for survivors of childhood critical illness:

1. Work toward gender-, culture-, and residence-neutrality.
2. Enables customization to specific medical conditions and age groups.

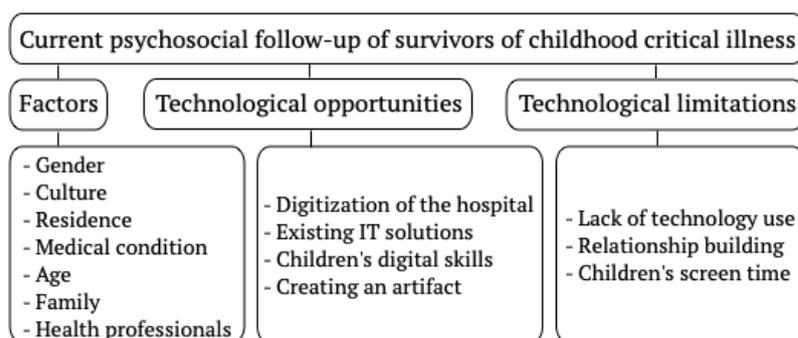


Figure 1 Mapping of the current psychosocial follow-up for survivors of childhood critical illness at the Norwegian hospital.

The 18th Scandinavian Conference on Health informatics, Tromsø, Norway, August 22-24, 2022. Organized by UiT The Arctic University of Norway. Conference Proceedings published by Linköping University Electronic Press at <https://doi.org/10.3384/ecp187>. © The Author(s). This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>

3. Include information sites for family members and health professionals.

4. Exploit the target group's technological opportunities but also consider their technological limitations.

These preliminary design principles will be further explored during the remaining PhD project period. To supplement the principles in a practical context, the next step is conducting a systematic literature review following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement [14] to identify the current existing information technology solutions. Then, a sub-study consisting of interactive workshops with the target group will combine the obtained knowledge to identify the target group's needs for the specific information technology solution. Besides, an interview sub-study with designers and developers in companies and NGOs developing relevant information technology solutions will be conducted to investigate if existing design theory could help them understand these identified needs. Moreover, the findings will be used to develop an artifact contribution - a prototype of an information technology solution for digital psychosocial follow-up for the target group. This prototype will be user-tested in two iterations. The prototype can contribute to the vulnerable children in the target group individually by giving more personalized follow-up improving life quality and socially by providing society with a better solution for survivors of childhood critical illness, reducing the burden on the psychological service. The results from all these sub-studies will finally be used to create the methodological contribution - final design principles regarding digital psychosocial follow-up for survivors of childhood critical illness.

5 REFERENCES

- [1] Ekim, A. "The Post-Intensive Care Syndrome in Children" in *Comprehensive Child and Adolescent Nursing* Vol. 43, Issue 1, pp. 15–21. Tylor & Francis 2020. <https://doi.org/10.1080/24694193.2018.1520323>
- [2] Manning, J. C., Hemingway, P., and Redsell, S. A. "Long-term psychosocial impact reported by childhood critical illness survivors: a systematic review" in *Nursing in Critical Care* Vol. 19, Issue 3, pp. 145–156. John Wiley & Son 2014. <https://doi.org/10.1111/nicc.12049>
- [3] Diseth, T. H. "Dissociation following traumatic medical treatment procedures in childhood: A longitudinal follow-up" in *Development and Psychopathology* Vol. 18, Issue 1, pp. 233–251. Cambridge University Press 2006. <https://doi.org/10.1017/S0954579406060135>
- [4] Collins, L., Collis, B., Trajanovska, M., Khanal, R., Hutson, J. M., Teague, W. J., and King, S. K. "Quality of life outcomes in children with Hirschsprung disease" in *Journal of Pediatric Surgery* Vol. 52, Issue 12, pp. 2006–2010. Elsevier 2017. <https://doi.org/10.1016/j.jpedsurg.2017.08.043>
- [5] Grano, C., Aminoff, D., Lucidi, F., and Violani, C. "Long-term disease-specific quality of life in children and adolescent patients with ARM" in *Journal of Pediatric Surgery* Vol. 47, Issue 7, pp. 1317–1322. Elsevier 2012. <https://doi.org/10.1016/j.jpedsurg.2012.01.068>
- [6] Nah, S. A., Ong, C. C. P., Saffari, S. E., Ong, L. Y., Yap, T. L., Low, Y., and Jacobsen, A. S. "Anorectal malformation & Hirschsprung's disease: A cross-sectional comparison of quality of life and bowel function to healthy controls" in *Journal of Pediatric Surgery* Vol. 53, Issue 8, pp. 1550–1554. Elsevier 2018. <https://doi.org/10.1016/j.jpedsurg.2017.08.018>
- [7] Rintala, R. J. and Pakarinen, M. P. "Outcome of anorectal malformations and Hirschsprung's disease beyond childhood" in *Seminars in Pediatric Surgery* Vol. 19, Issue 2, pp. 160–167. Elsevier 2010. <https://doi.org/10.1053/j.sempedsurg.2009.11.021>
- [8] Diseth, T. H. and Emblem, R. "Long-term psychosocial consequences of surgical congenital malformations" in *Seminars in Pediatric Surgery* Vol. 26, Issue 5, pp. 286–294. Elsevier 2017. <https://doi.org/10.1053/j.sempedsurg.2017.09.009>
- [9] Berger, C., Casagrande, L., Sudour-Bonnange, H., Massoubre, C., Dalle, J. H., Teinturier, C., Martin-Beuzart, S., Guillot, P., Lanlo, V., Schneider, M., Molin, B., Molin, M., Mounier, O., Garcin, A., Fresneau, B., Clavel, J., and Demoor-Goldschmidt, C. "Personalized Massive Open Online Course for Childhood Cancer Survivors: Behind the Scenes" in *Applied Clinical Informatics* Vol. 12, Issue 2, pp. 237–244. Thieme Medical Publishers 2021. <https://doi.org/10.1055/s-0041-1725185>
- [10] Wärnestål, P., Svedberg, P., Lindberg, S., Nygren, J. "Effects of Using Child Personas in the Development of a Digital Peer Support Service for Childhood Cancer Survivors" in *Journal of Medical Internet Research* Vol. 19, Issue 5, pp. e161. JMIR Publications 2017. <https://doi.org/10.2196/jmir.7175>
- [11] vom Brocke, J., Hevner, A., and Maedche, A. *Introduction to Design Science Research* pp. 1–13. Springer International Publishing, Cham, Germany, 2020. https://doi.org/10.1007/978-3-030-46781-4_1
- [12] Meuser, M. and Nagel, U. "The Expert Interview and Changes" in *Knowledge Production* Chapter 1, pp. 17–42. Palgrave Macmillan, Hampshire, UK, 2009.
- [13] Braun, V. and Clarke, V. "Using thematic analysis in psychology" in *Qualitative Research in Psychology* Vol. 3, Issue 2, pp. 77–101. Taylor & Francis 2006. <https://doi.org/10.1191/1478088706qp063oa>
- [14] Page M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Moher, D. "Updating guidance for reporting systematic reviews: development of the PRISMA 2020 statement" in *Journal of Clinical Epidemiology* Vol. 134, Issue June 2021. pp. 103–112, 2021. <https://doi.org/10.1016/j.jclinepi.2021.02.003>

6 ACKNOWLEDGEMENT

I would like to thank my supervisors, professor Maria Letizia Jaccheri, professor Gunnar Hartvigsen, and associate professor Sofia Papavasopoulou, for their support and guidance in this PhD project.

Automatic Report Generation for Medical Images

Muhammad Kamran¹ Mohib Ullah² Ali Shariq Imran² Muhammad Sajjad²

¹Islamia College University Peshawar, Pakistan.

²Norwegian University of Science and Technology, Norway.

Abstract

In this work, we propose an encoder-decoder-based automatic report generation system capable of generating radiology reports for chest x-rays. We tested five backbone Convolutional Neural Networks, namely VGG16, InceptionV3, Resnet50, MobileNet and NasNet mobile, to extract visual features and used Long Short-Term Long Memory (LSTM) to extract the text features from the reports. Both features are concatenated and given to a deep network for report generation. We performed experiments on publicly available Indiana University's NLMCXR dataset. We evaluated our system against different backbones and evaluated accuracy and BLEU score. The result showed that our method achieved reliable and convincing results.

Keywords

X-ray, Convolutional Neural Network, Long Short-Term Long Memory, encoder-decoder, visual features, medical reports.

1 INTRODUCTION

Medical images like radiological, X-ray, CT, and MRI are popular in the medical field due to their usefulness in diagnosis and prognosis [1]. Looking at and interpreting these images is an arduous, tedious, and time-consuming task for medical experts. In the medical field, time is of prime importance because saving patients' lives is the doctors' primary objective.

Automatic Report Generation is a task similar to image captioning. Primarily, visual features are extracted from images using computer vision techniques and the text features from corresponding captions for the images using natural language processing (NLP) and combining those visual and text features to generate a caption for the corresponding image [2]. It is widely used to assist visually challenged people with hearing problems. The idea of generating a report from medical images is inspired by recent work in caption generation for natural images using multi-modalities, i.e., natural language data and images data done by Andrej et al. [3]. They used Flickr8, Flickr30k and MSCOCO datasets for experimentation. Their model was based on a combination of Convolutional Neural Network (CNN) [4] for visual features extraction and bidirectional Recurrent Neural Networks for extracting features from text data. Yuan Xue et.al [5] proposed a novel generative model that automatically generates a complete radiology report. Their proposed model combines the Convolutional Neural Networks (CNN) with Long Short-Term Memory (LSTM) in a recurrent way. Their model cannot only generate high-level conclusive impressions but can also generate a detailed description of findings sentence by sentence to support the conclusion. The multi-modal model combines the encoder with an image and one generated sentence to build attention input so that it can guide the generation of the following sentence. In this way, the model maintains the coherence among the generated sentences. Omar et al. [6] proposed a conditioned transformer-based method to generate radiology reports

and claim that they are the first to condition the pre-trained transformer on visual and semantic features to generate medical reports. their work is divided into three stages: (1) first, they fine-tuned the pre-trained Chex-Net to predict the specific tags from the images. (2) then calculated the weighted semantic features from the predicted tag's pre-trained embeddings. (3) finally conditioned a pre-trained GPT2 model on the visual and semantic features to generate the medical reports. They analyzed the generated reports using word-lapping metrics and adding new meaningful semantic-based similarity metrics. Similarly, Changchang Yin et.al [7] came up with a novel idea based on Hierarchical Recurrent Neural Network (HRNN) and introduced a novel frame to generate accurate and diverse medical reports. Their model can detect medical abnormalities and can generate long captions simultaneously. Moreover, they suggested replacing the global feature pooling in multi-label classification CNN with multi-label pooling to improve the accuracy and robustness of CNN.

Compared to earlier work, in this paper, we develop an algorithm that takes an X-ray image at the input and generates an automatic report at the output. In a nutshell, the algorithm combines Natural language processing and computer vision to accomplish the task. The paper is organized into the following sections. The methodology and important components of the model are elaborated in section 2. The data preparation and dataset details are listed in subsection 2.3. The experiments and implementation details are given in section 3 and section 4 concludes the paper with the final remarks.

2 METHODOLOGY

The proposed framework is inspired by Marc Tanti et.al [8] encoder-decoder architecture. We used a CNN encoder to extract visual features and an LSTM network to extract text features. The output of both CNN and LSTM is concatenated and given as input to a feed-forward network. The dataset used for this study is the IU chest radiology

The 18th Scandinavian Conference on Health informatics, Tromsø, Norway, August 22-24, 2022. Organized by UiT The Arctic University of Norway. Conference Proceedings published by Linköping University Electronic Press at <https://doi.org/10.3384/ecp187>. © The Author(s). This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>

dataset with chest images with their corresponding text reports. In our method, first, we extracted features from images using pre-trained models like VGG16, INCEPTION, RESNET, MOBILE NET, and NASNET MOBILE and saved those features into a dictionary and wrote it to a pickle file. The dictionary keys are the names of the images, and values for those keys were the features from the corresponding name. Then, we pre-processed the text data so the algorithm could process it. Then we passed the pre-processed text to the LSTM, where sequential features are learnt from the whole text, and then finally, the image features and the text features are concatenated in feed forward network.

Pre-processing

We have two modalities, i.e., images and text, so we pre-processed both data separately.

Each image in the dataset is resized to 500x500 pixels to reduce the computations and maintain the dataset's consistency. Then we normalized the images and scaled the image pixels to a range of [0, 255]. For the text data, to avoid inconsistency in the dataset, we converted all the letters in the reports to their lower case. Then we removed all the punctuation and special characters because we do not need that in our reports, and they could confuse our model while training. We removed those words from the reports that were repeated only once or consisted of a single letter, e.g., "a". As our desired output report does not contain any number, we checked if there were any alphanumeric characters and removed them too. To direct the model that our sequence starts and ends from here, we added "startseq" and "endseq" in the start and end of the reports, respectively. Then we created a vocabulary of the unique words in the reports. Then we tokenized the words to pass them to the model.

Feature Extraction

Several pre-trained models are used for extracting features like VGG16, GoogleNet, ResNet, and NasNet. We tested these models to determine which pre-trained model is best for report generation.

Dataset

For training the model, we used Indiana university's publicly available Chest X-Ray dataset [9] (IU X-Ray). The dataset contains 7,470 pairs of images and text reports. Each report consists of impressions, findings, tags, comparisons, and indications.

3 EXPERIMENTS

We implemented the model on a system with Core i5-6500 Processor, 24 GB RAM and GeForce GTX 1070 Ti Graphics card. We tested different backbone models on 80 epochs, including VGG16, Inceptionv3, Rasnet50, MobileNet and NasNet Mobile. The comparative analysis based on performance is tabulated in Table 1.

Models	Training Accuracy	Validation Accuracy	BLEU score
VGG16	95	93	0.94
Inception V3	95.6	88	0.93
ResNet 50	99.5	91	0.91

MobileNet	97	91	0.23
NasNet	99	93.8	0.11

Table 1. BLEU score, training, and validation accuracy on different backbone models.

4 SUMMARY

In this study, we proposed an automatic report generation system that generates radiology reports for chest x-rays using deep learning. The system uses CNN as encoder and LSTM as decoder for report generation. The experiment results show that the suggested system is reliable and fast. This reporting system not only has the potential to reduce radiology errors but also makes the radiology practice more efficient.

Reference

- [1] Jianbo Yuan, Haofu Liao, Rui Luo, and Jiebo Luo, "Automatic radiology report generation based on multi-view image fusion and medical concept enrichment," in International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2019, pp. 721–729.
- [2] Adrian Brady, Riste'ard 'O Laoide, Peter McCarthy, and Ronan McDermott, "Discrepancy and error in radiology: concepts, causes and consequences," The Ulster medical journal, vol. 81, no. 1, pp. 3, 2012.
- [3] Andrej Karpathy and Li Fei-Fei, "Deep visual-semantic alignments for generating image descriptions," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 3128–3137.
- [4] Yann LeCun, Yoshua Bengio, et al., "Convolutional networks for images, speech, and time series," The handbook of brain theory and neural networks, vol. 3361, no. 10, pp. 1995, 1995.
- [5] Yuan Xue, Tao Xu, L Rodney Long, Zhiyun Xue, Sameer Antani, George R Thoma, and Xiaolei Huang, "Multimodal recurrent model with attention for automated radiology report generation," in International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2018, pp. 457–466.
- [6] Omar Alfarghaly, Rana Khaled, Abeer Elkorany, Maha Helal, and Aly Fahmy, "Automated radiology report generation using conditioned transformers," Informatics in Medicine Unlocked, vol. 24, pp. 100557, 2021.
- [7] Changchang Yin, Buyue Qian, Jishang Wei, Xiaoyu Li, Xianli Zhang, Yang Li, and Qinghua Zheng, "Automatic generation of medical imaging diagnostic report with hierarchical recurrent neural network," in 2019 IEEE international conference on data mining (ICDM). IEEE, 2019, pp. 728–737.
- [8] Marc Tanti, Albert Gatt, and Kenneth P Camilleri, "Where to put the image in an image caption generator," Natural Language Engineering, vol. 24, no. 3, pp. 467–489, 2018.
- [9] Dine Demner Fushman et al. "Preparing a collection of radiology examinations for distribution and retrieval," Journal of the American Medical Informatics Association, vol. 23, no. 2, pp. 304–310, 2016.

SHI 2022

Proceedings of the 18th Scandinavian Conference on Health Informatics

August 22-24, 2022
Tromsø, Norway

Published by:

Linköping University Electronic Press, Sweden Linköping Electronic
Conference Proceedings No. 187 ISBN 978-91-7929-344-4 (PDF)

URL <https://doi.org/10.3384/9789179293444>

ISSN 1650-3686 (print)

ISSN 1650-3740 (online)

© The Authors, 2022