STUDIES IN HEALTH TECHNOLOGY AND INFORMATICS 313

dHealth 2024

Proceedings of the 18th Health Informatics Meets Digital Health Conference

Editors: Dieter Hayn
Bernhard Pfeifer
Günter Schreier
Martin Baumgartner

IOS Press

The integration of technology has become key to improving patient outcomes, optimizing clinical workflows, and expanding access to healthcare. The use of large language models (LLMs) like ChatGPT is becoming more familiar and acceptable to users, and a number of research groups are now exploring the use of LLMs for various healthcare purposes. The next few years will show to what extent the huge expectations raised by LLMs will be met, and which classical health IT areas will survive this technological transformation.

This book presents the proceedings of dHealth 2024, the 18th annual conference on Health Informatics meets Digital Health, held on 7th & 8th May 2024 in Vienna, Austria. The dHealth conference series aims to provide insight into the research and application of up-to-date health IT solutions. Attracting around 300 participants each year, the series provides a platform for researchers, practitioners, decision makers and vendors to discuss innovative health informatics and eHealth solutions aimed at improving the quality and efficiency of healthcare by means of digital technology. The book includes 42 papers delivered at the conference. Topics range from the adoption of emerging technologies like LLMs, telemedicine and cloud computing, to the ethical, legal, social, and economic implications of health IT.

The book provides an up-to-date overview of ongoing research in health IT which will contribute to shaping the future of healthcare delivery, advancing digital health, improving patient outcomes, and ensuring equitable access to quality care for all, and will be of interest to all those working in the field.

ISBN 978-1-64368-516-8 (print)
ISBN 978-1-64368-517-5 (online)
ISSN 0926-9630 (print)
ISSN 1879-8365 (online)

dHEALTH 2024

Studies in Health Technology and Informatics

Internationally, health informatics is driven by developments in biomedical technologies and medical informatics research that are advancing in parallel and form one integrated world of information and communication media and result in massive amounts of health data. These components include genomics and precision medicine, machine learning, translational informatics, intelligent systems for clinicians and patients, mobile health applications, data-driven telecommunication and rehabilitative technology, sensors, intelligent home technology, EHR and patient-controlled data, and Internet of Things.

The series Studies in Health Technology and Informatics (HTI) was started in 1990 in collaboration with EU programmes that preceded the Horizon 2020 to promote biomedical and health informatics research. It has developed into a highly visible global platform for the dissemination of original research in this field, containing more than 250 volumes of high-quality works from all over the world.

The international Editorial Board selects publications with relevance and quality for the field. All contributions to the volumes in the series are peer reviewed.

Volumes in the HTI series are submitted (for evaluation) for indexing by MEDLINE/PubMed; Web of Science: Conference Proceedings Citation Index – Science (CPCI-S) and Book Citation Index – Science (BKCI-S); Google Scholar; Scopus; EMCare.

Volume 313

Recently published in this series

Vol. 312 K. Keshavjee and A. Khatami (Eds.), The Role of Digital Health Policy and Leadership – From Reactive to Proactive
Vol. 311 HTI Open Access Collection 1
Vol. 310 J. Bichel-Findlay, P. Otero, P. Scott and E. Huesing (Eds.), MEDINFO 2023 – The Future Is Accessible – Proceedings of the 19th World Congress on Medical and Health Informatics
Vol. 309 M. Giacomini, L. Stoicu-Tivadar, G. Balestra, A. Benis, S. Bonacina, A. Bottrighi, T.M. Deserno, P. Gallos, L. Lhotska, S. Marceglia, A.C. Pazos Sierra, S. Rosati and L. Sacchi (Eds.), Telehealth Ecosystems in Practice – Proceedings of the EFMI Special Topic Conference 2023

ISSN 0926-9630 (print)
ISSN 1879-8365 (online)

dHealth 2024

Proceedings of the 18th Health Informatics Meets Digital Health Conference

Edited by

Dieter Hayn

AIT Austrian Institute of Technology GmbH, Graz, Austria

Bernhard Pfeifer

UMIT TIROL – Private University for Health Sciences, Medical Informatics and Technology, Hall in Tirol, Austria

Günter Schreier

AIT Austrian Institute of Technology GmbH, Graz, Austria

and

Martin Baumgartner

AIT Austrian Institute of Technology GmbH, Graz, Austria

IOS Press

Amsterdam • Washington, DC

ISBN 978-1-64368-516-8 (print)
ISBN 978-1-64368-517-5 (online)
Library of Congress Control Number: 2024936492
doi: 10.3233/SHTI313

Publisher
IOS Press BV
Nieuwe Hemweg 6B
1013 BG Amsterdam
Netherlands
e-mail: order@iospress.nl

For book sales in the USA and Canada:
IOS Press, Inc.
6751 Tepper Drive
Clifton, VA 20124
USA
Tel.: +1 703 830 6300
Fax: +1 703 830 2300
sales@iospress.com

Preface

In the ever-evolving healthcare landscape, the integration of technology has become paramount in improving patient outcomes, optimising clinical workflows, and expanding access to care. The rise of large language models (LLMs), such as ChatGPT in 2022, has changed the view of citizens as regards artificial intelligence (AI)-based applications. This has triggered the foundation of various research groups to explore the use of large language models in healthcare, for purposes including literature review, documentation, translation, coding, decision support, and digital assistants, among others. The coming years will show to what extent the huge expectations raised by LLMs will be met, and which classical health IT areas will survive this technological transformation.

Since 2007, the dHealth conference has been providing insight into the research and application of up-to-date health IT solutions. Each year, this event attracts around 300 participants from academia, industry, government, and healthcare organisations. In keeping with its interdisciplinary mission, the dHealth conference series provides a platform for researchers, practitioners, decision makers and vendors to discuss innovative health informatics and eHealth solutions aimed at improving the quality and efficiency of healthcare by means of digital technology.

This year's proceedings of the 18th edition of dHealth represent an up-to-date mirror of ongoing research in health IT in Europe. Topics range from the adoption of emerging technologies like LLMs, telemedicine and cloud computing, to the ethical, legal, social, and economic implications of health IT. Each paper contributes to the rich tapestry of research aimed at advancing digital health and improving patient outcomes. As we navigate the complexities of modern healthcare, interdisciplinary collaboration, technological innovation, and ethical considerations will continue to shape the future of healthcare delivery, and ensure equitable access to quality care for all.

Graz, Hall in Tyrol, Vienna, May 2024

Dieter Hayn (AIT)
Bernhard Pfeifer (UMIT)
Günter Schreier (AIT)
Martin Baumgartner (AIT)

This page intentionally left blank

About the Conference

Scientific Programme Committee

Heads

Members:

Prof. Dr. Michael Marschollek, *Medical University of Hannover, Germany*
Prof. Dr. Christian Menard, *UAS Carinthia, Austria*
Prof. Dr. George Mihalas, *University of Medicine and Pharmacy, Timisoara, Romania*
Dr. Sabrina Neururer, *Hall in Tirol, Austria*
Prof. Josef Niebauer, *Ludwig Boltzmann Institute for Digital Health and Prevention, Salzburg, Austria*
Ass.-Prof. Dr. Ivana Ognjanovic, *Donja Gorica University, Podgorica, Montenegro*
Univ.-Prof. Dr. Karl-Peter Pfeiffer, *FH Joanneum, Graz, Austria*
Univ.-Prof. Dr. Ulrich Prokosch, *University of Erlangen, Germany*
Dr. Ana Rath, *INSERM, US14 – Orphanet, Paris, France*
Steven Redmond, *University College Dublin, Ireland*
Ass. Prof. Marco Roos, *Leiden University Medical Center, Netherlands*
FH-Prof. Dr. Stefan Sauermann, *UAS Technikum Wien, Vienna, Austria*
Prof. Franz Schaefer, *Heidelberg University Hospital, Germany*
Prof. Dr. Paul Schmücker, *University of Mannheim, Germany*
Dr. Jan Smeddinck, *Ludwig Boltzmann Institute for Digital Health and Prevention, Salzburg, Austria*
Univ. Doz. Dr. Günter Schreier, *AIT Graz, Austria*
Dr. Michael Shifrin, *N.N. Burdenko Neurosurgical Institute, Moscow, Russia*
Prof. Dr. Martin Staemmler, *UAS Stralsund, Germany*
Prof. Dr. Lăcrămioara Stoicu-Tivadar, *Polytechnic University of Timisoara, Romania*
Dr. Hanna Suominen, *NICTA, Canberra, Australia*
Dr. Andreas Trügler, *Know-Center, Austria*
Univ.-Prof. Dr. Zlatko Trajanoski, *Medical University of Innsbruck, Austria*
Prof. Elena Villalba, *Universidad Politecnica de Madrid, Spain*
Dr. Andreas Ziegl, *telbiomed, Graz, Austria*

Reviewers

We are extremely grateful to all reviewers for their significant contribution to the conference proceedings. Without them, dHealth 2024 would not be possible:

Josef Altmann
Elske Ammenwerth
Alireza Banaye Yazdipour
Christian Baumgartner
Martin Baumgartner
Andrea Berghold
Michelle Bindel
Andrea Cattaneo
Mihaela Crisan-Vida
Kerstin Denecke
Klaus Donsa
Jakob Doppler
Georg Duftschmid
Markus Falgenhauer
Jan Gaebel
José García Moros
Holger Gothe
Martin Haag
Werner O Hackl
Dieter Hayn
Emmanuel Helm
Mira Hercigonja-Szekeres
Christoph Hintermüller
Guido Kempter

Peter Klutke
Alexander Kollmann
Karl Kreiner
Martin Kropf
Stefan Tino Kulnik
Thomas Lux
Johannes Martinek
Christian Menard
George Mihalas
Robert Mischak
Francesco Moscato
Stefan Müller-Mielitz
Sabrina B. Neururer
Paul Panek
Bernhard Pfeifer
Hans-Ulrich Prokosch
Christoph Rinner
Angelika Rzepka
Stefan Sabutsch
Stefan Sauermann
Cornelia Schneider
Günter Schreier
Marco Schweitzer
Abbas Sheikhtaheri
Michael Shifrin
Stefan Skonetzki
Cord Spreckelsen
Martin Staemmler
Lacramioara Stoicu-Tivadar
Sai Pavan Kumar Veeranki
Andrea Vitali
Fabian Wiesmüller

This page intentionally left blank

Contents

Preface v
Dieter Hayn, Bernhard Pfeifer, Günter Schreier and Martin Baumgartner

About the Conference vii

Analysis of Critical Incident Reports Using Natural Language Processing 1
Kerstin Denecke and Helmut Paula

Vivifying Knee Rehabilitation: SquatEmUp, a Serious Game Utilizing the Wii Fit Balance Board 7
René Baranyi, Marko Zivojinovic, Dominik Hoelbling, Christoph Aigner, Werner Hoerner and Thomas Grechenig

Enhancing Dementia Risk Mitigation: Standardized Digital Monitoring in Clinical Trials and Platform Design for At-Risk Elderly Individuals 9
Hannes Hilberger, Bianca Buchgraber-Schnalzer, Simone Huber, Theresa Weitlaner, Helena Untersteiner, Carina Seerainer and Sten Hanke

Acceptance of Telepresence Robotics, Telecare and Teletherapy Among Stroke Patients, Relatives and Therapy Staff 15
Edda Currle and Sonja Haug

Large Language Model-Based Evaluation of Medical Question Answering Systems: Algorithm Development and Case Study 22
Daniel Reichenpfader, Philipp Rösslhuemer and Kerstin Denecke

Mapping the Bulgarian Diabetes Register to OMOP CDM: Application Results 28
Evgeniy Krastev, Emanuil Markov, Simeon Abanos, Ralitsa Krasteva and Dimitar Tcharaktchiev

Supporting Well-Being: Exploring the Value of a Digital Coach for Older Adults in the Transition from Work to Retirement 34
Katja Neureiter, Leen Broeckx, Markus Garschall, Kim Helsen and Johannes Kropf

Symbiosis of Technology and Ethics: Preliminary Results of an Inquiry into the Moral Dimensions in the Use of Robotic Systems in Patient Care 41
Carolin Mirbeth, Christoph Ohneberg and Inge Eberl

Evaluation of a Digital Dementia Registry's IT Architecture After a Three-Year Period in Practice: digiDEM Bayern 43
Michael Zeiler, Nikolas Dietzel, Peter L. Kolominsky-Rabas, Elmar Graessel and Hans-Ulrich Prokosch

Dual Implementation Guides in FHIR and CDA 49
Nikola Tanjga and Oliver Krauss

DiGA in Primary Care and the Influence of Patient-Related Factors – Results of a Survey of Family Doctors 55
Anna-Lena Brecher, Julia Herrgen and Tobias Mueller

Unpacking Sociotechnical Discourses on Telehealth Use and Data Protection: A Path Towards Digital Health Value Creation 62
Javad Pool, Farhad Fatehi, Salma Sharifi, Morteza Namvar and Saeed Akhlaghpour

A Sustainable Approach to Telerehabilitation in Europe: Patients Are Ready, but Caregivers Are Essential 68
Andrea Cattaneo, Andrea Vitali, Daniele Regazzoni and Caterina Rizzi

Generating Actionable Insights from Patient Medical Records and Structured Clinical Knowledge 74
Natasha Trajkovska, Michael Roiss, Sophie Bauernfeind, Mohammad Alnajdawi, Simone Sandler, Daniel Herzmanek, Matthias Winkler, Michael Haider and Oliver Krauss

Exploring the Environmental Impact of Telemedicine: A Life Cycle Assessment 81
Anna Savoldelli, Daniele Landi and Caterina Rizzi

User Personas for a "Better Design" of Nation-Wide EHRs Based on Thorough Expert Evaluation and Field Analysis: Modeling Users as Individuals Plus Family Members for an Enhanced Mapping of Healthcare Situations 87
Daniela Glatz, Alexander Zeiringer, Johannes Harms, René Baranyi, Karin Kappel and Thomas Grechenig

Should We Better Stick to Pen and Paper? An Empirical Investigation on Functionality, Privacy and Data-Security of Physiotherapy Telehealth Applications 93
Lukas Maul, Ines Kramer, Lena Rettinger and Franz Werner

Editing Physicians' Responses Using GPT-4 for Academic Research 101
Magdalena T. Weber, Jannik Schaaf, Holger Storf, Thomas O.F. Wagner, Alexandra Berger and Richard Noll

Towards an Electronic Health Prevention Record Based on HL7 FHIR and the OMOP Common Data Model 107
Markus Falgenhauer, Aaron Lauschensky, Karl Kreiner, Stefan Beyer, Kristina Reiter, Andreas Ziegl, Robert Modre-Osprian, Bernhard Pfeifer, Sabrina Neururer, Susanne Krestan, Hanna Wagner, Andreas Huber, Sandra Plaikner, Sarah Kuppelwieser, Martin Widschwendter and Günter Schreier

KIJANI: Designing a Physical Activity Promoting Collaborative Augmented Reality Game 113
Florian Schweizer, Laura Willinger, Renate Oberhoffer-Fritz, Jan Müller, Stephan Jonas and Lara Marie Reimer

Visualizing Study Profiles Using the GMDS Competency Catalog for Medical Informatics Programs 121
Elske Ammenwerth, Peter Haas and Oliver J. Bott

Comparison of WebSocket and Hypertext Transfer Protocol for Transfer of Electronic Health Records 124
Nikola Kirilov

Relationship Between Dietary Intakes and Elevated Diastolic Blood Pressure Among Children: A Cross-Sectional Study Using Wearable Devices 129
Kukiat Tudpor, Panthicha Chinnakhot, Montira Boonsorn, Wiriyaporn Lomthaisong, Worrameth Kareechum, Charnyuth Sriphuwong, Le Ke Nghiep and Niruwan Turnbull

CareNet Tyrol – Information System Success Assessment for Case & Care Management Service 135
Sabrina B. Neururer, Raphael Klein, Bernhard Pfeifer, Gabi Schiessling, Eva Schulc, Chiara Messina, Kristina Reiter, Angelika Rzepka and Peter Kastner

HerzMobil Tirol PreOp – A Multidisciplinary Telemonitoring Project for Heart Failure Patients Prior to Cardiac Surgery 141
Karin Hinterbuchner, Elisabeth Kleinheinz, Bettina Fetz, Leo Pölzl, Johannes Holfeld, Bernhard Pfeifer, Sabrina Neururer, Christian Puelacher, Can Gollmann-Tepekoeylue, Susanne Krestan, Andreas Huber and Gerhard Pölzl

Comparison of the Performance of the Fast Healthcare Interoperability Resources (FHIR) Subscription Channels 143
Nikola Kirilov

Automated Clinical Trial Cohort Definition and Evaluation with CQL and CDS-Hooks 149
Anja Schwab, Anna M. Lin and Stephan Winkler

Machine Learning-Based Prediction of Malnutrition in Surgical In-Patients: A Validation Pilot Study 156
Diether Kramer, Stefanie Jauk, Sai Veeranki, Michael Schrempf, Julia Traub, Eva Kugel, Anna Prisching, Sandra Domnanich, Maria Leopold, Peter Krisper and Gerald Sendlhofer

Feasibility of an Integrated Telemonitoring System for Home-Recorded EEG with Automated AI-Based EEG Analysis 158
Clemens Lang, Silvia Winkler, Johannes Koren, Gerhard Gritsch, David Pollreisz, Stefan Köpl, Tilmann Kluge and Christoph Baumgartner

Telemedicine for Ketogenic Dietary Treatment in Refractory Epilepsy and Inherited Metabolic Disease: State of Play and Future Perspectives 160
Alexander Höller, Stefan Welte, Anna Katharina Schönlaub, Charlotte Uhlisch, Sabine Scholl-Bürgi, Anastasia Male-Dressler, Bernhard Pfeifer and Günter Schreier

Infection Control Through Clinical Pipelines Built with Arden Syntax MLM Building Blocks 167
Leonhard Hauptfeld, Andrea Rappelsberger and Klaus-Peter Adlassnig

Understanding IT in Healthcare: Relevance and Training Needs of IT in Private Medical Practice 173
Marco Schweitzer, Werner O. Hackl, Sabrina B. Neururer and Bernhard Pfeifer

Screening Automation in Systematic Reviews: Analysis of Tools and Their Machine Learning Capabilities 179
Elias Sandner, Christian Gütl, Igor Jakovljevic and Andreas Wagner

A Standardized Treatment Pathway for Telehealth-Based Care of Chronic Wounds in Austria 186
Hannah Vinatzer, Angelika Rzepka, Barbara Binder, Viktoria Gruber, Rainer Hofmann-Wellenhof, Patrick Mayr, Peter Kastner, Alexander Höller and Günter Schreier

Will 'Computable' Clinical Guidelines Be Compatible with Personalised Care? 192
Jack Dowie, Mette Kjer Kaltoft and Vije Kumar Rajput

Exploring Opportunities for Clinical Data Warehouse Enhancement Through Data Catalog Integration 198
Andreas Walker, Werner O. Hackl and Bernhard Pfeifer

Quality and Effectiveness of AI Tools for Students and Researchers for Scientific Literature Review and Analysis 203
Martin Danler, Werner O. Hackl, Sabrina B. Neururer and Bernhard Pfeifer

Unveiling the Surge: A Comprehensive Analysis of the Stimulating Impact of the Covid-19 Pandemic on Telemedicine and Its Academic Landscape 209
Werner O. Hackl, Sabrina B. Neururer and Bernhard Pfeifer

Transforming Tele-Ophthalmology: Utilizing Cloud Computing for Remote Eye Care 215
Marco Schweitzer, Philipp Ostheimer, Arno Lins, Vito Romano, Bernhard Steger, Daniel Baumgarten and Marco Augustin

Neural Network-Based Prediction of Perceived Sleep Quality Through Wearable Device Data 221
Martin Baumgartner, Manuel Grössl, Raphaela Haumer, Katharina Poimer, Flora Prantl, Katharina Weick, Markus Falgenhauer, Stefan Beyer, Andreas Ziegl, Aaron Lauschensky, Fabian Wiesmüller, Karl Kreiner, Dieter Hayn and Günter Schreier

The Evolution of Telehealth in Heart Failure Management: The Role of Large Language Models and HerzMobil as a Potential Use Case 228
Hillary Farmer, Karl Kreiner, Thomas Schütz, Gerhard Pölzl, Christian Puelacher and Günter Schreier

Agile eHealth Usability Evaluation: A Triangulative Approach to Promoting the Usability of eHealth Systems 234
Irina Sinabell and Elske Ammenwerth

Author Index 241

dHealth 2024
D. Hayn et al. (Eds.)

doi:10.3233/SHTI240002

Analysis of Critical Incident Reports Using Natural Language Processing

Kerstin DENECKE[a,1] and Helmut PAULA[b]
[a] *Bern University of Applied Sciences, Bern, Switzerland*
[b] *Stiftung Patientensicherheit Schweiz, Zürich, Switzerland*

Abstract. A Critical Incident Reporting System (CIRS) collects anecdotal reports from employees, which serve as a vital source of information about incidents that could potentially harm patients. Objectives: To demonstrate how natural language processing (NLP) methods can help in retrieving valuable information from such incident data. Methods: We analyzed frequently occurring terms and sentiments as well as topics in data from the Swiss National CIRRNET database from 2006 to 2023 using NLP and BERTopic modelling. Results: We grouped the topics into 10 major themes out of which 6 are related to medication. Overall, they reflect the global trends in adverse events in healthcare (surgical errors, venous thromboembolism, falls). Additionally, we identified errors related to blood testing, COVID-19, handling patients with diabetes and pediatrics. 40-50% of the messages are written in a neutral tone, 30-40% in a negative tone. Conclusion: The analysis of CIRS messages using text analysis tools helped in getting insights into common sources of critical incidents in Swiss healthcare institutions. In future work, we want to study more closely the relations, for example between sentiment and topics.

Keywords. Text mining, Critical incident reporting system, Text analysis, Natural language processing

1. Introduction

The safety and well-being of patients is of paramount importance to all stakeholders in healthcare. As complex organizations, hospitals face several challenges in maintaining and improving patient safety. A key tool in this ongoing effort are Critical Incident Reporting Systems (CIRS), which serve as a vital source of information about incidents that could potentially harm patients [1], [2]. Analyzing CIRS messages can provide actionable insights into systemic weaknesses and risk factors, helping hospital administrators to implement effective preventive measures [3]. By identifying trends and predicting potential future incidents, hospitals can proactively address areas of concern before they result in adverse events.

Although the clear value of CIRS messages the wealth of this qualitative data is often under-utilized due to its unstructured nature [4]. Traditional methods of manual analysis are time consuming, potentially biased and often unable to cope with the scale and complexity of the data. The analysis of these reports using Natural Language Processing (NLP) methods offers a promising avenue for supporting this process [3], [5] since they automatically process CIRS messages, identify patterns and extract

[1] Corresponding Author: Kerstin Denecke, Bern University of Applied Sciences, Institute Patient-centered Digital Health, Quellgasse 21, Bern, Switzerland, E-Mail: kerstin.denecke@bfh.

meaningful insights [5]. Such automated analysis can lead to a more comprehensive understanding of the nature and frequency of incidents, as well as their underlying causes and in this way improving patient safety and healthcare quality. So far, research on NLP applied to CIRS messages is still rare, in particular in German-speaking countries. Tetzlaff et al. analyzed the German database "cirsmedical.de" [6] using NLP. Wang et al. applied a Support Vector Machine-based approach that considers also semantics for classifying incident types and compared it to the quality of a classifier based on a bag of words feature set [7]. In another work, they tested a convolutional neural network with word embedding to identify incidents by type and severity [8].

This research aims to use NLP to analyze CIRS reports of the Swiss national incident reporting system, providing an approach to understanding and mitigating risk in the hospital environment. Furthermore, it seeks to explore and demonstrate the value of NLP in the analysis of CIRS messages. In an era where data-driven decision-making is paramount, this research therefore aligns with the broader movement towards harnessing big data for improved outcomes in healthcare.

2. Methods and Dataset

Our dataset comprised 10'063 messages from CIRRNET starting from 21/05/2006 to 23/09/2023. CIRRNET, the Critical Incident Reporting & Reacting NETwork (https://patientensicherheit.ch/cirrnet/), is a supra-regional institution that performs a central networking function for local error reporting systems in Switzerland. All affiliated facilities can feed their local CIRS reports anonymously into the CIRRNET database and thus contribute to the comprehensive survey of patient risks. Currently, 73 healthcare institutions with their 128 locations are participating in CIRRNET on a voluntary basis. The organization Patient Safety Switzerland has been operating CIRRNET since 2006. CIRRNET differs from most CIRS networks in that all local CIRS reports are used to identify problem areas of supra-regional relevance, recommendations for improvement are developed together with experts and published in the form of Quick Alerts® by Patient Safety Switzerland. In average, 559 messages are included per year with a minimum of 84 messages in 2006 and a maximum of 1'196 messages in 2012. Between 2011 and 2014, more than 1'000 messages are available per year. Only intensive care units and anesthesia contributed their reports after the initial launch of CIRRNET. In 2011, CIRRNET was opened for all medical specialties, explaining this peak.

We analyzed the CIRS messages automatically using various NLP methods. To simplify reporting, we grouped the data into five batches: years 2006-2009, years 2010-2014, years 2015-2019, years 2020-2021, years 2022-2023. This grouping was motivated by the distribution of messages per year. Beyond, we wanted to consider the two years of COVID-10 pandemic separately, hypothesizing that at the time of the pandemic reporting of incidents differs from other years. We were also interested in the incidents reporting during the pandemic.

To analyze the content of the messages, we determined the frequent words and generated word clouds using the Python library WordCloud. Further, we identified the main topics of the messages using BERTopic analysis [9]. BERTopic uses a transformer-based method to process documents. It involves embedding documents, reducing their dimensional complexity, clustering them, creating bag-of-words models and using c-TF-IDF to create topic representations. Each incident report was also classified as positive,

negative or neutral using the Python library TextBlobDE to study the sentiment of the incident messages. By all these efforts, we wanted to identify common patterns and trends in the data, i.e. to uncover frequent types of incidents, common causes, and areas within the hospital where incidents occur most often.

Table 1. Top 20 Commonly used terms in the messages per period (presented in descending order of frequency)

2006-2009 (n=1'146)	2010-2014 (n=4'703)	2015-2019 (n=3'061)	2020-2021 (n=526)	2022-2023 (n=627)
OP	verordnet	verordnet	verordnet	Verordnet
Seite	Verordnung	Verabreicht	Station	Pflege
Bein	bemerkt	Uhr	Zimmer	OP
Operateur	Uhr	Pflege	Notfall	Verabreicht
Anästhesist	mg	Bemerkt	Kind	Erhalten
Immer	OP	Verordnung	Covid	Notfall
Ab	Tag	Medikament	Uhr	Zimmer
Intubation	Infusion	Erhalten	Pflege	Covid
Gabe	verabreicht	Arzt	Verabreicht	Verordnung
Wegen	Arzt	OP	Tag	Bemerkt
Infusion	erhalten	Informiert	Informiert	Informiert
Saal	Pflege	Labor	Verordnung	Abteilung
Sei	Station	Gegeben	Arzt	Uhr
Kammer	tbl	Abteilung	Abteilung	Station
Postoperativ	Medikamente	Bekommen	Infusion	Sei
Arzt	Erhielt	Medikamente	Bemerkt	Medikamente
Darauf	Worden	Statt	Isoliert	IPS
Einleitung	Fehler	Tag	Gemacht	Verordnungen
Falsche	Statt	Station	Wegen	Worden
informiert	gegeben	falschen	Zeit	falsch

3. Results

The messages had an average number of words per message of 52.5 words (minimum of 1 word, maximum of 645 words). 97 texts were in French, 3 in Italian and 6 in English. All other texts were in German. The top 20 commonly used terms (Table 1) reflected that a frequent term is related to the prescription of medications. The word clouds (Figure 1) together with the top 20 words demonstrated a topic shift over the five considered periods: In the first period (2006-2009), many reported incidents seemed to concern the surgery. In the second (2010-2014) and third period (2015-2019), the frequent terms addressed the medication and their prescription. In 2020-2021, incidents concerned COVID-19 and isolation. In the period 2022-2023, the main topic was again related to medication and prescription, but also incident reports related to nursing, surgery and emergency become more frequent.

The BERTopic analysis provided more in-depth insights into the topics addressed in the messages (Table 2). Five topics (Topic 3,4,6,7,8,10) were related to medication in general or even to specific medication handling problems. For example, topic 4 deals with insulin injection in the context of diabetes handling. Topic 6 concerned postoperative thromboembolism prophylaxis. Topic 1 concerned laboratory procedures related to blood analysis. In particular, exchange of names and patients when labelling the blood tests seemed to be an issue. Topic 2 dealt with children and emergencies. Topic 5 pointed to various problems of patient safety related to fall and mobility of patients. Keywords let us assume that the messages concerned patients falling out of their beds or

on their way to the toilet. Topic 7 concerned incidents related to anesthesia. Topic 9 was related to COVID-19, testing and isolation.

2006-2009 2010-2014 2015-2019

2020-2021 2022-2023

Figure 1. Word clouds generated from the incident reports from the five indicated time periods

Table 2. Topics as determined by BERTopic for the complete database

Manual labeling of topic	Topic-relevant terms as identified by BERTopic
Topic 1: Blood testing and laboratory	[labor, blutentnahme, falschen, falsche, namen, röhrchen, blut, abgenommen, be, beschriftet]
Topic 2: Pediatric emergency	[kind, station, notfall, op, kommt, uhr, ca, arzt, kam, uns]
Topic 3: Medication prescription	[medikamente, medikament, tbl, medis, mg, haldol, medi, verordnung, eingenommen, morgen]
Topic 4: Handling of diabetes patients	[insulin, bz, novorapid, gespritzt, blutzucker, glucose, diabetes, einheiten, pen, nachspritzschema]
Topic 5: Fall and mobility	[boden, bett, bettgitter, wollte, rollstuhl, aufstehen, dabei, trage, wc, gestürzt]
Topic 6: Thromboembolism prophylaxis	[clexane, 40mg, sc, verordnung, injiziert, 20mg, clex, postop, gespritzt, 40]
Topic 7: Anesthesia procedures	[intubation, adrenalin, einleitung, mg, gabe, fentanyl, tubus, propofol, minuten, problemlose]
Topic 8: Heparin treatment	[heparin, heparinperfusor, perfusor, 10000, gestartet, ie, eingestellt, lief, heparininfusion, fehler]
Topic 9: Covid 19 treatment	[covid, isoliert, abstrich, getestet, positiv, verdacht, mrsa, verlegt, test, covid19]
Topic 10: Treatment with Calium chlorid	[kcl, infusion, 24h, 40mmol, eingestellt, 500ml, 20mmol, 21mlh, ringerfundin, mlh]

Regarding sentiment (Figure 2), we can recognize that most messages were formulated in a neutral manner. They became slightly more positive since 2019. In 2008, the

percentage of positive messages was the smallest with 16.4% positively and 51.1% of neutrally formulated reports.

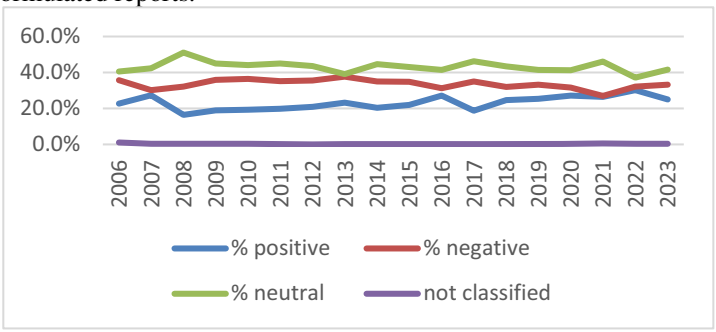


Figure 2. Sentiment of the CIRS messages

4. Discussion

Our results reflect that the most common incidents are related to medication. This corresponds to the global trends that medication-related harm affects 1 in 30 patients in healthcare settings [10], [11]. The increase in terms related to medication prescription since 2010 can be associated with the introduction of electronic prescription systems in Swiss hospitals. On a global level, surgical errors and health care-associated infections also belong to the most frequent sources of patient harm [11], [12]. While surgery occurs as frequent term also in our data, we could not find information related to infections. A reason is that CIRS reports refer to events where the relation between cause and event can be recognized. Hospital-acquired infections occur not immediately and can often not be related to a specific event. Patient falls and incidents related to venous thromboembolism in turn are known to be frequent adverse events in hospitals [13], [14] which corresponds to our results. The COVID-19 pandemic was clearly reflected in the incident reports. Interestingly, we identified the topic related to errors in handling patients with diabetes. Insulin injection schemata are complex and not integrated in electronic health records which may lead to errors. Further, multiple devices are used for diabetes management, in particular patient-owned medical devices such as insulin pens providing a source of errors.

The study results demonstrate the usefulness of applying NLP methods to CIRS reports, and show that it is beneficial to use multiple methods to identify all trends and topics and get a more complete picture. Tetzlaff et al. conducted a similar analysis using NLP applied to the German database “cirsmedical.de” [6]. They focused on their analysis rather on the reporting differences among clinical disciplines. In contrast to the work by Wang et al. we were not interested in classifying the messages into predefined categories, but to explore the dataset by the chosen methods [7], [8].

Our study has some limitations: The CIRRNET database is not a primary reporting system, but contains only reports that have been triaged and forwarded in the affiliated organizations. The dataset contained some messages in languages others than German. However, we only applied sentiment analysis methods specialized for German texts. The sentiment analysis method was not assessed regarding accuracy on our data set. The chosen method was not specifically developed for analyzing sentiments in CIRS messages and might have led to wrong classifications.

There are many additional questions we want to study in future: How does the sentiment in reports correlate with topics of incidents? Which linguistic features (e.g., specific terminologies, phrasing) are commonly associated with different types of incidents? Answers to the latter question would help in identifying linguistic markers that could automate the categorization of reports or signal the criticality of incidents. Another question of interest is: Can NLP help in identifying under-reported or misclassified incidents? We also want to closer analyze the emotional tone of the reports, which could reflect the urgency or severity of incidents. In addition, the emotional and psychological impact of critical incidents on healthcare staff is often overlooked. Sentiment analysis of CIRS messages can offer insights into relations between emotional responses associated with different incidents, helping to address staff well-being and foster a supportive work environment. In particular, this form of analysis could provide important conclusions about the willingness of employees to participate in a more constructive or negative manner and thus about the safety culture in the various organizations. It could be interesting to study whether at certain times in a year the topics of incidents are changing. Such information could be useful to predict potential future incidents based on historical data, aiding in proactive risk management.

References

[1] J. Mack, 'CIRS als Werkzeug des Qualitäts-und Risikomanagement im deutschen Versorgungssystem.', *Qualitätsmanagement in der Gesundheitsversorgung*, pp. 171–88, 2020.

[2] W. Petschnig and E. Haslinger-Baumann, 'Critical Incident Reporting System (CIRS): a fundamental component of risk management in health care systems to enhance patient safety', *Saf Health*, vol. 3, no. 1, p. 9, Dec. 2017, doi: 10.1186/s40886-017-0060-y.

[3] I. J. B. Young, S. Luz, and N. Lone, 'A systematic review of natural language processing for classification tasks in the field of incident reporting and adverse event analysis', *International Journal of Medical Informatics*, vol. 132, p. 103971, Dec. 2019, doi: 10.1016/j.ijmedinf.2019.103971.

[4] C. Neuhaus, M. Holzschuh, C. Lichtenstern, and M. St.Pierre, 'Erkenntnisse aus 10 Jahren CIRS-AINS: Eine Analyse von Nutzerverhalten und Ausblick auf neue Herausforderungen', *Anaesthesist*, vol. 69, no. 11, pp. 793–802, Nov. 2020, doi: 10.1007/s00101-020-00829-z.

[5] K. Denecke, 'Automatic Analysis of Critical Incident Reports: Requirements and Use Cases', *Stud Health Technol Inform*, vol. 223, pp. 85–92, 2016.

[6] L. Tetzlaff, A. S. Heinrich, R. Schadewitz, C. Thomeczek, and T. Schrader, '[The analysis of CIRSmedical.de using Natural Language Processing]', *Z Evid Fortbild Qual Gesundhwes*, vol. 169, pp. 1–11, Apr. 2022, doi: 10.1016/j.zefq.2021.12.002.

[7] Y. Wang, E. Coiera, and F. Magrabi, 'Can Unified Medical Language System–based semantic representation improve automated identification of patient safety incident reports by type and severity?', *Journal of the American Medical Informatics Association*, vol. 27, no. 10, pp. 1502–1509, Oct. 2020, doi: 10.1093/jamia/ocaa082.

[8] Y. Wang, E. Coiera, and F. Magrabi, 'Using convolutional neural networks to identify patient safety incident reports by type and severity', *Journal of the American Medical Informatics Association*, vol. 26, no. 12, pp. 1600–1608, Dec. 2019, doi: 10.1093/jamia/ocz146.

[9] M. Grootendorst, 'BERTopic: Neural topic modeling with a class-based TF-IDF procedure', *arXiv preprint arXiv:2203.05794*, 2022.

[10] A. Hodkinson *et al.*, 'Preventable medication harm across health care settings: a systematic review and meta-analysis', *BMC Med*, vol. 18, no. 1, p. 313, Dec. 2020, doi: 10.1186/s12916-020-01774-9.

[11] M. Panagioti *et al.*, 'Prevalence, severity, and nature of preventable patient harm across medical care settings: systematic review and meta-analysis', *BMJ*, p. l4185, Jul. 2019, doi: 10.1136/bmj.l4185.

[12] S. Raoofi *et al.*, 'Global prevalence of nosocomial infection: A systematic review and meta-analysis', *PLoS One*, vol. 18, no. 1, p. e0274248, 2023, doi: 10.1371/journal.pone.0274248.

[13] J. H. LeLaurin and R. I. Shorr, 'Preventing Falls in Hospitalized Patients: State of the Science', *Clin Geriatr Med*, vol. 35, no. 2, pp. 273–283, May 2019, doi: 10.1016/j.cger.2019.01.007.

[14] G. E. Raskob *et al.*, 'Thrombosis: A Major Contributor to Global Disease Burden', *ATVB*, vol. 34, no. 11, pp. 2363–2371, Nov. 2014, doi: 10.1161/ATVBAHA.114.304488.

dHealth 2024
D. Hayn et al. (Eds.)

doi:10.3233/SHTI240003

Vivifying Knee Rehabilitation: SquatEmUp, a Serious Game Utilizing the Wii Fit Balance Board

René BARANYI[a,1], Marko ZIVOJINOVIC[a], Dominik HOELBLING[a], Christoph AIGNER[a], Werner HOERNER[b] and Thomas GRECHENIG[a]

[a] TU Wien, Institute of Information Systems Engineering, Research Group for Industrial Software (INSO) TU Wien
[b] Human&Digital (H&D)
Vienna, Austria

Abstract. Knee injuries are a common concern in orthopedic and sports medicine, often requiring extensive rehabilitation to restore function and alleviate pain. The rehabilitation process can be long and challenging, necessitating innovative approaches to engage and motivate patients effectively. Serious games have emerged as a promising tool in rehabilitation, offering an interactive and enjoyable way to perform therapeutic exercises. In this context, a new serious game that leverages the Wii Fit Balance Board as an input device supported by a keyboard to aid knee rehabilitation was developed. To tailor the game to the specific needs of knee rehabilitation, qualitative content analysis and requirement extraction based on three interviews with therapists were conducted. These insights were then iteratively integrated into the game's development, ensuring that the final game was both clinically relevant and engaging. A therapist subsequently re-evaluated the completed game, confirming its potential to enhance the rehabilitation process. It was also shown that further research is needed to detect squat movements on the Balance Board. But this initial approach, combining the Wii Fit Balance Board with squat exercises, uniquely addresses the challenges of knee rehabilitation, offering a novel video and game solution.

Keywords. Serious Game, Knee, Rehabilitation, Wii Fit Balance Board

1. Introduction

Knee joint diseases and injuries represent some of the most frequent issues related to the human postural and locomotive systems, comprising approximately thirteen percent of all ailments within the human postural and musculoskeletal systems. [1] Starting rehabilitation early is crucial; the longer the affected body part, like the knee, remains inactive without muscle building, the more challenging the healing process becomes, therefore, Physiotherapists must tailor exercises to the injury's severity, beginning with

[1] Corresponding Author: René Baranyi, TU Wien Institute of Information Systems Engineering, Research Group for Industrial Software (INSO), Vienna, Austria, E-Mail: rene.baranyi@inso.tuwien.ac.at

small movements and gradually progressing to more demanding exercises. [2] Due to the monotonous nature of rehabilitation exercises, motivation often decreases, which can be avoided using serious games. [3]

2. Methods

The methodology consisted of 5 steps. In the first step, a comprehensive narrative literature research was conducted, followed by semi-structured individual interviews (knee-injury, process, serious gaming, etc.) based on an interview guideline with three therapists and a requirements catalog. Afterwards, mockups and gaming ideas were drawn based on these results and adapted based on internal feedback loops. Then, a gaming concept was implemented with Unity and the Wii Fit Balance Board, trying to capture the movements to control this serious game. Finally, the game was evaluated by one therapist in the form of a semi-structured interview and demo of the serious game.

3. Results and Discussion

The key exercises emphasized by participating physiotherapists for rehabilitation and everyday strength are squats, lunges, and balancing exercises. Given the unanimous recommendation of squats by all interviewed physiotherapists, the primary focus was to incorporate squats into a mini-game format. Six requirements were identified. Continuous visual feedback for users during/after exercises for movement awareness and performance review (R01). Customizable game settings to accommodate the user's healing process and prevent injury (R02). Not too easy or too hard levels (R03). The game should be usable for all age groups (R04). Instructional videos for safe/correct exercise execution (R05). Add a high score system for user motivation (R06). The game itself was called SquatEmUp, where players shoot projectiles at flying enemies. The user performs a squat, descending to the lowest point they can and then holding that position. The squat duration influences the power and speed of the projectile; a longer squat awards more points if it hits an enemy. The challenge lies in timing the shot to hit enemies that move randomly across the screen. The user releases the shot by completing the squat with an upward movement. It was shown that the game mechanics might support the execution of these exercises, but further development needs to be done to check and capture correct positioning/identification when squats are done. This is currently supported with a keyboard (button pressing). Within further research, AI or integration of different sensors might help this initial approach.

References

[1] Schneider, O., Scharf, H. P., Stein, T., Knapstein, S., Hermann, C., & Flechtenmacher, J. (2016). Inzidenz von Kniegelenkverletzungen. Der Orthopäde, 45(12), 1015-1026.

[2] Terry Malone, Turner A Blackburn, and Lynn A Wallace. “Knee rehabilitation”. In: Physical therapy 60.12 (1980), pp. 1602–1610

[3] P. Rego, P. M. Moreira, and L. P. Reis, “Serious games for rehabilitation: A survey and a classification towards a taxonomy,” in 5th Iberian Conference on Information Systems and Technologies, 2010, pp. 1–6.

dHealth 2024
D. Hayn et al. (Eds.)

doi:10.3233/SHTI240004

Enhancing Dementia Risk Mitigation: Standardized Digital Monitoring in Clinical Trials and Platform Design for At-Risk Elderly Individuals

Hannes HILBERGER[a,b], Bianca BUCHGRABER-SCHNALZER[a,1], Simone HUBER[a], Theresa WEITLANER[a], Helena UNTERSTEINER[c,d], Carina SEERAINER[a] and Sten HANKE[a]

[a] *eHealth Institute, FH JOANNEUM University of Applied Sciences, Graz, Austria*
[b] *GSRC, Division of Medical Physics and Biophysics, Medical University of Graz, Graz, Austria*
[c] *Department of Neurology, Medical University of Vienna, Vienna, Austria*
[c] *Comprehensive Center for Clinical Neurosciences and Mental Health, Medical University of Vienna, Vienna, Austria*

Abstract. Background: Dementia is becoming a significant public health concern, affecting approximately 130,000 individuals in Austria, whereby nearly 40% of the cases are attributed to modifiable risk factors. Multidomain lifestyle interventions have thereby demonstrated significant effects in reducing the risk of dementia. Objectives: The goal was to define an interoperability framework to conduct standardized monitoring in clinical trials for enhancing dementia risk mitigation. In addition, the identified standards should be integrated into the components of the project. Methods: A step-by-step approach was used, where initially data collection, aggregation and harmonization was carried out with retrospective data from various clinical centers. Afterwards, the interoperability framework was defined including the prospective data that is gathered during a clinical trial. Results: A guideline for integrating healthcare standards was developed and incorporated into the technical components for the clinical trial. Conclusion: The interoperability framework was designed in a scalable way and will be regularly updated for future needs.

Keywords. Mobile Applications [L01.224.900.685], Clinical Trial [V03.175.250], Health Information Interoperability [L01.313.500.750.280.555, L01.470.813], Medical Informatics [L01.313.500], Health Information Systems [L01.313.500.750.300.361]

1. Introduction

As the global population ages, dementia is increasingly becoming a significant public health concern, affecting 50 million people worldwide. Projections indicate that this number will triple by 2050 [1]. The prevalence of dementia varies across regions [2], and in Austria, approximately 130,000 individuals are reported to be living with some form

[1] Corresponding Author: Bianca Buchgraber-Schnalzer, eHealth Institute, FH JOANNEUM University of Applied Sciences, Graz, Austria, E-Mail: bianca.buchgraber-schnalzer@fh-joanneum.at

of dementia [3]. Dementia is the loss of cognitive functioning and encompasses difficulties in memory, language as well as behavior, leading to a decline in the activities of daily living [1][4]. Approximately 40% of all dementia cases are attributed to 12 modifiable risk factors, which can be addressed at different stages of life: early (<45 years), mid (45-65 years), or late (>65 years) [2]. Previous multidomain lifestyle interventions have demonstrated significant effects in reducing the risk of dementia [5].

The European LETHE project wants to target those modifiable risk factors and aims to build upon the FINGER randomized controlled trial (RCT), a two-year multidomain lifestyle intervention that positively impacted cognition [6]. LETHE extends the existing FINGER protocol by incorporating personalized digital intervention methods for at-risk elderly individuals in a clinical trial. The primary objectives of the two-year LETHE RCT (ClinicalTrials.gov NCT05565170) include the assessment of participant retention, adherence, engagement in the intervention as well as the investigation of changes in dementia risk [7]. In total, four clinical sites are involved, with 160 participants evenly distributed across the sites. All participants are randomly allocated 1:1 to one of the study groups (intervention and control group).

The LETHE project introduces two key components for the digital intervention: the LETHE App, available in five languages (English, German, Finnish, Swedish and Italian) designed for study participants, reflecting personalized interventions based on the FINGER lifestyle intervention program, and the LETHE Clinical Trial Management System (CTMS), designed for clinical professionals. The CTMS allows clinical professionals to document, manage the content of the LETHE App and monitor participant progress. To ensure interoperability between current and future systems and to overcome language barriers, a guideline with all variables used in the LETHE project was developed.

This paper explores the creation of such an interoperability framework and presents the results of integrating various standards within the LETHE project.

2. Methods

The LETHE IT infrastructure was planned to deal with various data from certain data sources [7], which include sensor data, cognitive test and training data, clinical data collected by the LETHE CTMS and digital biomarkers collected via the LETHE App. To deal with this issue and to maximize interoperability and standardization, a step-by-step approach was used as visualized in Figure 1.

2.1. Data collection, aggregation, and harmonization

Data aggregation was divided into two phases. Phase I data included pre-existing features from the clinical sites (feature extraction - without patient data) such as variables from MRI data, laboratory data and general demographic data. During the pre-phase of the trial a harmonization process took place, where feature extraction from phase I data from the four different sites was carried out and resulted in a master dataset.

Phase I data features were then used as a basis for phase II data, where additional features were identified, i.e., clinical data like blood test results, data from sensors, data from the LETHE App and data from individual cognitive tests and trainings. Both feature sets were harmonized as far as the study setting allowed in study protocol meetings under considerations of technicians and clinical professionals.

2.2. Defining the interoperability framework

As a next step the identified features were defined in a variable set. To identify the most suitable standardization for interoperability, research was conducted. The research included the scanning for well-known and established standards within the healthcare sector such as standards and terminologies developed by HL7 [8], SNOMED [9] or by the Regenstrief Institute [10] (LOINC [11], UCUM [12]).

Finally, the research culminated in drafting a guideline of an interoperability framework with the aim to map all LETHE variables/features to potential standards in the healthcare sector. Following the completion of the interoperability framework for phase II data, several standards were integrated into the aforementioned components of the LETHE infrastructure.

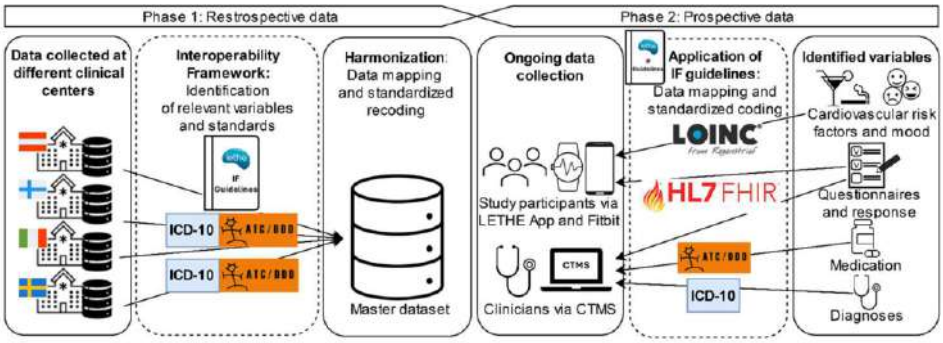


Figure 1. Development and use of the Interoperability Framework (IF) guideline which was conducted during phase I and phase II feature extraction.

3. Results

The LETHE Interoperability Framework [13] provides a guideline for the technical integration of potential standards used for phase I and phase II data, along with the description of data collection and data flow between various modules in the LETHE architecture.

During the initial implementation of the LETHE App and the LETHE CTMS recommended practices, such as integrating the FHIR questionnaire into the LETHE App were incorporated. ICD-10 [14] was used for recording the previous diagnoses of study participants in the electronic case report form (eCRF) of the LETHE CTMS. Originally planned as a FHIR resource, adjustments were made for tracking mood and cardiovascular risk factors, which are currently implemented using LOINC for identification. Similarly, the ATC-Codes [17] for documenting participants' medications were initially considered as a FHIR resource.

A more detailed explanation of the specific modules and how these standards are deployed is provided in the subsequent sections. An overview of integrated standards for the different modules is shown in Figure 1 and, for the LETHE App, in Figure 2.

3.1. Questionnaires

To collect information related to various topics (e.g., health status, sleep quality, nutrition and physical activities), the participants are asked to fill in questionnaires on a regular basis via the LETHE App. To transmit the questionnaires to the LETHE App, the Questionnaire resource [18] and QuestionnaireResponse resource [19] of HL7 standard FHIR R4 were used.

The use of FHIR is intended to enable easy integration of multi-language validated questionnaires from other resources (e.g., FHIR server) into the LETHE App, or to allow questionnaires created specifically for LETHE to be extracted in a standardized way for subsequent reuse and incorporation into other projects.

3.2. Documentation of medical conditions or medical events in the eCRF

Any diagnosed condition may pose a general risk of (developing) cognitive impairment, therefore it is important to frequently document all current and past medical conditions and medical events occurring during the LETHE study in a standardized manner. Professionals can enter diagnoses in an ICD-10-based dynamic list implemented in the CTMS eCRF for visits during the LETHE study. A description of the diagnosis with first occurring year and whether the disease is currently being treated should be entered.

3.3 Documentation of medications in the eCRF

Professionals can enter all medications, both prescribed and over-the-counter drugs as well as any dietary supplements in the CTMS eCRF for visits during the LETHE study.

Each entry for documenting all medications includes two text fields: one for entering the name of the medication and another for entering the ATC code. Additionally, there is a question asking whether the participant takes the medication regularly or as needed and a text field for entering the prescribed daily dose of the medication in milligrams.

Each entry for documenting any other dietary supplements, herbal products, vitamins etc. includes the name of the supplement product and a question about the frequency of intake, whether it is regular or occasional.

3.4 Diary of cardiovascular risk factors and emotions

To provide a medical diary for study participants, sections for alcohol intake, cigarette smoking and self-reported blood pressure were implemented in the activity and lifestyle section within the LETHE App, which would represent the domain vascular/metabolic risk management.

Initial considerations within the interoperability framework regarding to emotions was to use SNOMED-CT (emotion/mood), but an answer list provided by LOINC promised to achieve semantic standardization by using LOINC codes extracted from the already existing feelings and emotions answer list. Professionals received the whole value set and were asked to select a harmonized amount of 7 to 9 common feelings over all sites. The amount of selection was explicitly kept small due to usability reasons within the LETHE App, as our target group mainly consists of elderly individuals.

The final selection of moods included: sad, excited, ashamed, happy, calm, active, proud, afraid and lonely. As certain emotions can be indicators of modifiable risk factors (depression/social isolation), tracking emotions of participants is planned to be integrated into mood analysis at the end of the project.

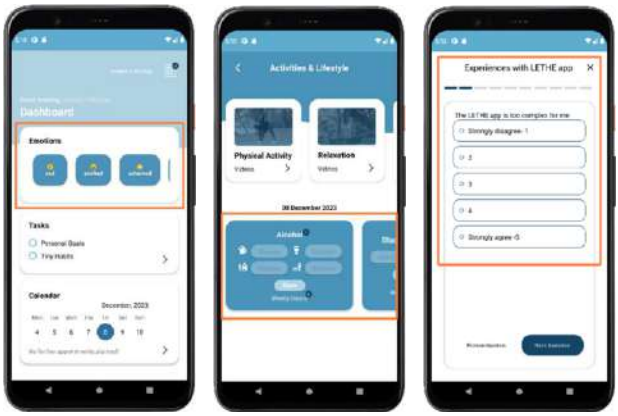


Figure 2. Visualization of the LETHE App including standardized elements such as LOINC for emotions (left screen), measures for cardiovascular risk (middle screen) and questionnaires based on FHIR (right screen).

4. Discussion

In this paper, a guideline for integrating potential healthcare standards, particularly for elderly individuals at risk of dementia was introduced. The development process of modules for two distinct components, in accordance with the provided recommendations, was detailed. To establish a foundation for the future, the interoperability framework was designed in a scalable manner and will be regularly updated to accommodate currently used variables, as some were not available at the time of guideline publication (e.g., ICD-10 for recording diagnoses in phase I data). Furthermore, certain variables were adjusted, such as the use of ATC for medication tracking. These alterations stem from the distributed environment created by diverse data sources from various entities.

Despite these modifications, adherence to recommendations, such as the incorporation of FHIR questionnaires in the LETHE App, was prioritized. The questionnaire section was designed as a reusable module, which can be used in other applications as well. Most FHIR questionnaires, available in multiple languages and based on validated instruments, align with European Interoperability Framework recommendations [20]. The LETHE App's questionnaire module currently and prospectively meets several key interoperability criteria, including openness (utilizing open-source software like HAPI FHIR for Android [21]), reusability (providing FHIR questionnaire structure and the FHIR Android module post-trial validation), data portability (transferable across systems and applications), accessibility (tailored for older users) and multilingualism.

Future enhancements of the LETHE CTMS may involve the integration of standards, such as UCUM (e.g., incorporation of a conversion tool to address variations in units used across different countries) or SNOMED-CT. The interoperability framework will continuously be updated to include the latest variables and consideration will be given to redesign modules for further interoperability improvements.

Declarations

Funding: The LETHE-Project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no 101017405 (https://ec.europa.eu/digital-single-market/en/policies/ehealth).

Acknowledgments: This research is supported by a larger consortium including Alzheimer Europe, Idryma Technologias kai Erevnas, Universiteit Maastricht, Kaasa Solution GmbH, i2Grow, Stichting EGI, Extra Red SRL, Infotrend Innovations Company Limited, Combinostics Oy, The Lisbon Council for Economic Competitiveness ASB.

References

[1] World Health Organization, Risk reduction of cognitive decline and dementia: WHO guidelines, Geneva, (2019). Accessed: Jan. 09, 2024. [Online]. Available: https://www.who.int/publications/i/item/9789241550543

[2] G. Livingston et al., Dementia prevention, intervention, and care: 2020 report of the Lancet Commission, *The Lancet* **396**(10248) (2020), 413–446

[3] S. Höfler et al., Österreichischer Demenzbericht 2014, Wien, (2015).

[4] L. Robinson et al., Dementia: timely diagnosis and early intervention, *BMJ* **350**(jun15 14) (2015), h3029–h3029

[5] X. Meng et al., Multidomain lifestyle interventions for cognition and the risk of dementia: A systematic review and meta-analysis, *International Journal of Nursing Studies* **130** (2022), 104236

[6] T. Ngandu et al., A 2 year multidomain intervention of diet, exercise, cognitive training, and vascular risk monitoring versus control to prevent cognitive decline in at-risk elderly people (FINGER): a randomised controlled trial, *The Lancet* **385**(9984) (2015), 2255–2263

[7] S. Hanke et al., AI-Based Predictive Modelling of the Onset and Progression of Dementia, *Smart Cities* **5**(2) (2022), 700–714

[8] Health Level Seven International - Homepage | HL7 International. https://www.hl7.org/ (accessed Dec. 22, 2023).

[9] Home | SNOMED International. https://www.snomed.org/ (accessed Dec. 22, 2023).

[10] Regenstrief Institute - Connecting and Innovating for Better Health. https://www.regenstrief.org/ (accessed Dec. 22, 2023).

[11] What LOINC is – LOINC. https://loinc.org/get-started/what-loinc-is/ (accessed Dec. 22, 2023).

[12] UCUM / About UCUM. https://ucum.org/about (accessed Dec. 22, 2023).

[13] LETHE Interoperability Framework. https://www.lethe-project.eu/wp-content/uploads/2022/01/LETHE-D2.5-LETHE-Interoperability-Framework.pdf (accessed Dec. 22, 2023).

[14] World Health Organization, The ICD-10 Classification of Mental and Behavioural Disorders, Geneva, (1993).

[15] WHO Collaborating Centre for Drug Statistics Methodology, Guidelines for ATC classification and DDD assignment 2023, Oslo, (2022).

[16] Index - FHIR v5.0.0. https://www.hl7.org/fhir/index.html (accessed Jan. 12, 2024).

[17] WHO Collaborating Centre for Drug Statistics Methodology, Guidelines for ATC classification and DDD assignment 2023, Oslo, (2022).

[18] FHIR R4 Questionnaire Resource. http://hl7.org/fhir/R4/questionnaire.html (accessed Jan. 12, 2024).

[19] FHIR R4 QuestionnaireResponse Resource. https://fhir-ru.github.io/questionnaireresponse.html (accessed Jan. 12, 2024).

[20] E. Commission and D.-G. for Digital Services, *New European interoperability framework – Promoting seamless services and data flows for European public administrations*. Publications Office, (2017).

[21] HAPI FHIR. https://hapifhir.io/hapi-fhir/ (accessed Jan. 15, 2024).

dHealth 2024
D. Hayn et al. (Eds.)

doi:10.3233/SHTI240005

Acceptance of Telepresence Robotics, Telecare and Teletherapy Among Stroke Patients, Relatives and Therapy Staff

Edda CURRLE[a,1] and Sonja HAUG[a]
[a] *Institute for Social Research and Technology Assessment (IST), Ostbayerische Technische Hochschule Regensburg (OTH), Germany*

Abstract. Background: Stroke as a cause of disability in adulthood causes an increasing demand for therapy and care services, including telecare and teletherapy. Objectives: Aim of the study is to analyse the acceptance of telepresence robotics and digital therapy applications. Methods: Longitudinal study with a before and after survey of patients, relatives and care and therapy staff. Results: Acceptance of the technology analysed is high in all three groups. Although acceptance among patients declined in parts of the cases in the second survey after having used telerobotics, all in all approval ratings remained high. With regard to patients no significant correlation was found between the general technology acceptance and the acceptance of use of telerobotics. Conclusion: Accepted new telecare and teletherapies can be offered with the help of telepresence robotics. This requires knowledge of and experience with the technology.

Keywords. Telehealth, stroke, acceptability of healthcare, telecare, evaluation study

1. Introduction

Stroke is one of the most common causes of death worldwide and is the third most common cause of permanent disability [1]. The risk of stroke increases with age, meaning that older people are more often affected. In Germany, in the year 2015 stroke was still the third most common cause of death [2]. In recent years, advances in acute treatment and therapy options as well as improved care structures, for example by increased numbers of stroke units [3, 4], have reduced the mortality rate of stroke patients [5]. The falling incidence rate, which is determined for Germany based on two population-based stroke registers [5], is in turn due to improved prevention measures [6]. Further advances in treatment and therapy options will presumably and hopefully further improve the survival rate. However, demographic change with the increasing ageing of Western European societies means that an increase in the number of new cases (incidence) and the number of people affected (prevalence) can be expected in the coming decades [7].

In the coming decades stroke as a cause of disability in adulthood will result in increased pressure on rehabilitation services and a significantly growing need for care and therapy for those who benefit from the described progress. This demand will come

[1] Corresponding Author: Edda Currle, Institute for Social Research and Technology Assessment, Regensburg, Germany, edda.currle@oth-regensburg.de

up against the ever-increasing challenges in the care and therapy sector because of the growing lack of skilled labour [8]. Particularly in rural regions patients find it difficult to receive adequate follow-up care for their rehabilitation. In this context, hopes are pinned on new digital applications, although these are not yet widespread. Consequently, a comparative study on digitalisation in the healthcare sector found that Germany ranked 16th out of 17 countries [9].

For these patients, telepresence robotics can offer an opportunity to supplement existing outpatient services with digital offerings. Our interdisciplinary field study on the use of telepresence robotics by stroke patients was based on this objective [10]. The research project has been issued a positive vote by the Joint Ethics Commission of the Bavarian Universities of Applied Sciences (GEHBa-202007-V-004-R). The additional provision of care and therapy via telepresence was intended to enable the study participants to remain in their own home environment enabling them to lead a self-determined life while at the same time participating in society. The accompanying social science study presented here analysed the acceptance and willingness of the participants to use the system as well as the ethical, legal, and social aspects of its use.

2. Methods

2.1. Telepresence robotics in the research project

Telepresence robots (TPR) are initially defined as relatively simple devices that enable communication over distance. They can be controlled via the Internet and can move autonomously within a fixed environment [11]. In the field study of our research project, two variants of TPR were used in the home environment of stroke patients. Study participants were assigned one of two telepresence robot systems in an alternating procedure for a test period of 24 weeks after enrolment in the study.

The TPRs were equipped with applications for independent training aimed at speech therapy and physiotherapy (of the upper extremities) as well as (psychosocial) care after a stroke. Both had a screen with a touchpad. While the autonomous mobile version included a navigation system with voice control, the second, so-called "DIY version" (Do-It-Yourself) had a larger screen. It was designed specifically for the project. The DIY was able to fill the gap in the market created by the lack of data protection compliance of other systems [12].

2.2. Research method of the acceptance surveys

The standardised surveys on the acceptance and willingness to use telepresence robotics after a stroke were conducted as part of the accompanying social science study (see figure 1). They included the test subjects (patients), their relatives and care and therapy staff in the environment of the patients [13]. Questionnaires of all surveys were developed based on tested items and scales and a technology acceptance model specifically adapted for the study and the respective target group [14]. The model named “TePUS-TAM” is in turn based on the technology acceptance model developed by Davis [15, 16, 17]. Here, acceptance is defined as a process in the context of an individual decision to use a technology. Attitude is a precondition for behaviour.

Study participants and relatives were surveyed in panel with two waves once before and once after the intervention using standardised, largely identical questionnaires.

Nursing and therapy staff were surveyed in a cross-sectional study. Even though the accompanying study itself did not include an intervention, the panel design can be described as a pretest-posttest design, as the intervention as well as the intervention evaluation took place between the first wave t1 and the second t2 [18]. The collected interview data (face-to-face, written form or via video communication using TPR) were transferred to an online tool. The data was analyed using IBM Statistics SPSS 27. The additional verbal statements from patients and relatives collected and written down during face-to-face interviews were coded and analysed qualitatively with MAXQDA. In addition, a qualitative study on acceptance and willingness to use was conducted with various stakeholders being not part of the field study [19].

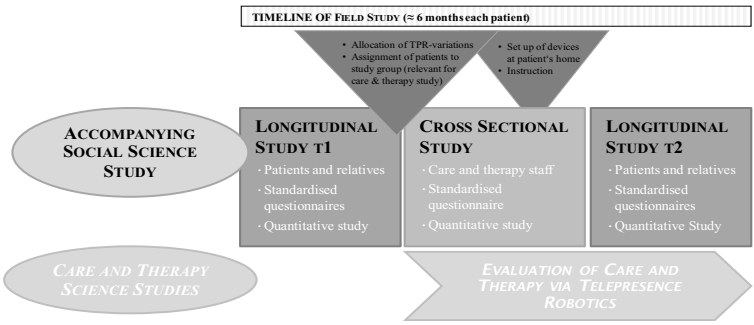


Figure 1. Research methods of the acceptance surveys

3. Results

3.1. How telepresence robotics are accepted

The study participants were asked about their attitudes towards using the technology both before the start of use (first wave) and at the end of the test phase (second wave). This is one of the core questions and the relevant target variable in our model. At the beginning of the study, patients' attitudes towards using the technology were at a very high level. Almost 91% stated that they could well or very well imagine using a TPR (n=44). This value was significantly lower after the end of the test phase and was still present in more than sixty per cent (64.9%, n=37). On the other hand, about 30% were negative and could not or could not at all imagine using it. The attitude towards both their own and the patient's behaviour was also positive at a high level among the family carers. It decreased slightly after the field test. In the first wave, n=51 family carers stated that they could well or very well imagine the test subjects using a TPR. This figure fell to 81.3% in the second wave (n=48). The nursing and therapy staff were only interviewed once after the test phase. At 93.5% (n=31), their (positive) attitude towards use by patients was at a very high level (see Figure 2).

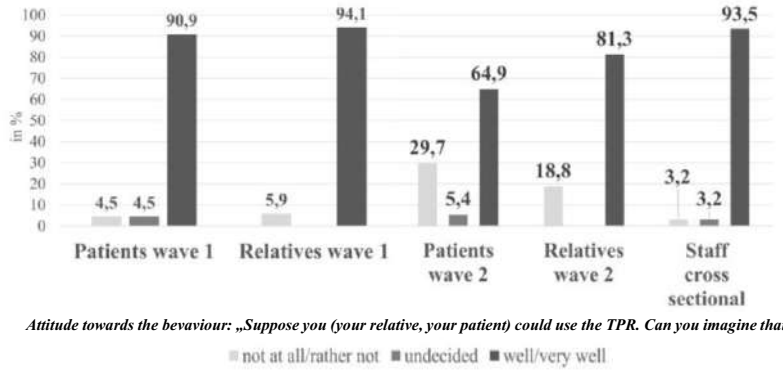


Figure 2. Attitude towards own behaviour (patients) and patient behaviour (relatives and staff) (in%) – first wave on the left, second wave and cross-sectional study of staff on the right.

The statistical correlation between the attitude values of the first and second wave is weakly positive for the patients (Rho = 0.261). Most patients (47%) remained stable in their attitudes (n=34). Few (8.8%) had more positive attitudes after the test phase than before. In each case, around 12% moved down one respectively two levels on the five-point scale. The attitudes of around 20% fell by three or more levels on the scale.

3.2. Technology acceptance

In our model, we assume that the general acceptance of technology plays a decisive role in whether there is a positive attitude towards the technology used and thus acceptance of use [20, 21]. The higher the level of technology acceptance, the more positive the attitude towards the use of telepresence robotics is expected to be. Technology acceptance was measured using an index made up of the constructs of technology use, technology access, technology affinity and technology competence. Validly tested items from various studies were used to formulate the questions for the construct of the TA Index [22-26]. The TA index can have values between 1 and 5, with 1 indicating a low level of acceptance and 5 indicating a high level. The index was collected in the first wave of patients and relatives as well as in the cross-sectional survey of therapy and care professionals. Although the three study groups differed in terms of age and gender, there were no significant differences in technology acceptance (table 1). A bivariate analysis showed the expected statistically significant age effect on the TA index for patients (n=44, Pearsons's r = -0,496, p<0,001) and relatives (n=48, ρ = -0,491, p<0,001), but not for nursing and therapy staff.

Table 1. Technology acceptance index of the study groups

	Mean	n	SD	Min	Max
Stroke patients	3,6	44	0,9	1,0	4,8
Relatives	3,7	50	0,8	1,2	5,0
Staff	3,9	31	0,7	2,2	4,8

3.3. Technology acceptance and attitude towards the use of telepresence robotics

The expected correlation between the TA index and attitude was found both in the target groups of relatives and staff in relation to their own use of telepresence robotics. The TA index and the attitude towards use, the acceptance of use, are statistically significantly related. However, such a correlation does not apply across the board for patients. Although the attitude towards behaviour correlates with the TA index in the first wave (Rho = 0.43, $p < 0.01$), this is not the case in the post-survey. Among relatives, the general technology acceptance index in the survey after the device test did not correlate with the attitude towards use by the patients being cared for. There was also no significant correlation between the general technology acceptance index and the attitude towards use by patients among the staff (Table 2).

Table 2. Acceptance of technology (TA Index) and attitude towards use of technology (acceptance of use)

	Patients	**Relatives**		**Nursing & therapy staff**	
Attitude to …	Own behaviour	Own behaviour	Patient behaviour	Own behaviour	Patient behaviour
First wave	0,43**	0,30*	0,35*		
Second wave	0,05	0,47**	0,26	0,38*	0,28

Spearman's Rho, * p< 0,05, ** p < 0,01

4. Discussion

As expected, in the bivariate analysis, age initially determined the acceptance of use of the TPR tested in each case. Overall, however, some of the expected correlations of the model regarding technology acceptance could not be confirmed. This applies in particular to the expected influence of general technology acceptance on the attitude to use *after* the device test: General technology acceptance still plays a significant role as an influencing factor for relatives before use in relation to patient behaviour and for nursing and therapy staff in relation to their own behaviour. After the device test, however, it was shown that technology use, technology access, technology affinity, technology competence and general technology acceptance among patients did not determine the acceptance of telepresence robotics. Even those who had not previously a pronounced affinity for technology were able to accept the technology at the end of the test phase. This also applies to nursing and therapy staff in terms of their attitude towards patients. The results of the study thus point in a direction that takes up the findings and demands of other studies. Age has not proven to be a determining factor for technology acceptance and therefore, like other studies imply, other variables should be included in the analysis [27-29].

Experience and knowledge of technology can increase acceptance. The level of knowledge on the subject of "teletherapy for stroke" surveyed in the first wave tended towards zero among both patients and their relatives. The variable could therefore not be used as an influencing factor to determine whether or not to use the technology. This is precisely where we need to start in the future, because, if necessary, patients will only consider what is already known [30]. Furthermore, our field study has shown that familiarisation, knowledge, experience, and handling of a technology can also lead to positive attitudes and acceptance of use, since even people who have been no "friends" of technology could imagine to use TPR. Possible reservations due to uncertainty in

dealing with the technology, which become barriers to acceptance, are reduced by concrete knowledge about the use and operation of the devices [19]. Relevant information combined with experience and familiarity are therefore ultimately one of the basic prerequisites for the sustainable use of technology in care and therapy.

References

[1] V. L. Feigin et al., Global, regional, and national burden of stroke and its risk factors, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019, *The Lancet Neurology* **20**(10) (2021), 795–820. doi: 10.1016/S1474-4422(21)00252-0.

[2] P. Kolominsky-Rabas, M. Weingärtner, H. Rosenthal, M. Hess C. Sedlak, Das Erlanger Schlaganfallregister – ein Modell für umfassende und nachhaltige Versorgungsforschung der Volkskrankheit Schlaganfall, *Public Health Forum* **25**(2) (2017), 128–130. doi: 10.1515/pubhef-2016-2174.

[3] V. Rücker, S. Wiedmann, M. O'Flaherty, M. A. Busch, P. U. Heuschmann, Decline in Regional Trends in Mortality of Stroke Subtypes in Germany from 1998 to 2015, *Stroke* **49**(11) (2018), 2577–2583. doi: 10.1161/STROKEAHA.118.023193.

[4] V. Rücker et al., Twenty-Year Time Trends in Long-Term Case-Fatality and Recurrence Rates After Ischemic Stroke Stratified by Etiology, *Stroke* **51**(9) (2020), 2778–2785. doi: 10.1161/STROKEAHA.120.029972.

[5] Robert Koch-Institut (ed.), *Gesundheit in Deutschland. Gesundheitsberichterstattung des Bundes: Gemeinsam getragen von RKI und Destatis,* RKI, Berlin, 2015.

[6] M. A. Busch, A. Schienkiewitz, E. Nowossadeck, A. Gößwald, Prävalenz des Schlaganfalls bei Erwachsenen im Alter von 40 bis 79 Jahren in Deutschland: Ergebnisse der Studie zur Gesundheit Erwachsener in Deutschland (DEGS1), *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz* **56**(5-6) (2013), 656–660. doi: 10.1007/s00103-012-1659-0.

[7] M. A. Busch, R. Kuhnert, 12-Monats-Prävalenz von Schlaganfall oder chronischen Beschwerden infolge eines Schlaganfalls in Deutschland, *Journal of Health Monitoring* **2**(1) (2017), 70–76. doi: 10.17886/RKI-GBE-2017-010.

[8] Bundesagentur für Arbeit - Statistik/Arbeitsmarktberichterstattung (ed.), *Arbeitsmarktsituation im Pflegebereich* (Berichte: Blickpunkt Arbeitsmarkt Mai 2023). Nürnberg, 2023. https://statistik.arbeitsagentur.de/DE/Statischer-Content/Statistiken/Themen-im-Fokus/Berufe/Generische-Publikationen/Altenpflege.html?__blob=publicationFile%C2%A0.

[9] R. Thiel et al., #SmartHealthSystems: Digitalisierungsstrategien im internationalen Vergleich, Bertelsmann Stiftung, Gütersloh, 2018. https://www.bertelsmann-stiftung.de/fileadmin/files/Projekte/Der_digitale_Patient/VV_SHS-Gesamtstudie_dt.pdf.

[10] D. Frommeld, S. Haug, E. Currle, K. Weber, Telepräsenzroboter in der Schlaganfallrehabilitation: Empirische Studie zur Unterstützung von Schlaganfallpatient*innen, *Pflege Zeitschrift* **75**(5) (2022), 52–55. doi: 10.1007/s41906-022-1251-7.

[11] H. Becker, Robotik in der Gesundheitsversorgung: *Hoffnungen, Befürchtungen und Akzeptanz aus Sicht der Nutzerinnen und Nutzer*, in: Pflegeroboter. O. Bendel (ed.). Springer Gabler, Wiesbaden, 2018. pp. 229–248.

[12] L. Middel, C. Popp, G. Raptis, T. Sutter, M. Gutbrod, *Konzeption und Aufbau einer technischen Telepräsenzrobotik-Plattform für die Unterstützung von Schlaganfallpatient*innen in der Pflege, Logopädie und Physiotherapie*, in: Digitale Technik für ambulante Pflege und Therapie: Herausforderungen, Lösungen, Anwendungen und Forschungsperspektiven. K. Weber, S. Haug, N. Lauer, A. Meussling-Sentpali, C. Mohr, A. Pfingsten, G. Raptis, G. Bahr (eds.), transcript, Bielefeld, 2022. pp. 19–58. doi: 10.1515/9783839462355-002.

[13] S. Haug, E. Currle, D. Frommeld, K. Weber, *Telepräsenzroboter für die Pflege und Unterstützung von Schlaganfallpatientinnen und -patienten: Das Forschungsdesign für die sozialwissenschaftliche Begleitforschung*, in: Digitale Technik für ambulante Pflege und Therapie: Herausforderungen, Lösungen, Anwendungen und Forschungsperspektiven. K. Weber, S. Haug, N. Lauer, A. Meussling-Sentpali, C. Mohr, A. Pfingsten, G. Raptis, G. Bahr (eds.), transcript, Bielefeld, 2022. pp. 175–194. doi: 10.14361/9783839462355-009.

[14] E. Currle, S. Haug, D. Frommeld, K. Weber, *TePUS-TAM: Entwicklung und Anwendung eines Technologieakzeptanzmodells für die Gesundheits- und Alternsforschung*, in: Digitale Technik für ambulante Pflege und Therapie: Herausforderungen, Lösungen, Anwendungen und Forschungsperspektiven. K. Weber, S. Haug, N. Lauer, A. Meussling-Sentpali, C. Mohr, A. Pfingsten,

G. Raptis, G. Bahr (eds.), transcript, Bielefeld, 2022. pp. 195–218. doi: 10.14361/9783839462355-010.

[15] F. D. Davis, *A technology acceptance model for empirically testing new end-user information systems: theory and results*, Dissertation, Massachusetts Institute of Technology, 1986. https://dspace.mit.edu/handle/1721.1/15192.

[16] F. D. Davis, Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology, *MIS Quarterly* **13(**3) (1989), 319, doi: 10.2307/249008.

[17] F. D. Davis, User acceptance of information technology: system characteristics, user perceptions and behavioral impacts, *International Journal of Man-Machine Studies* **38**(3) (1993), 475–487, doi: 10.1006/imms.1993.1022.

[18] H. Mayer, *Pflegeforschung anwenden: Elemente und Basiswissen für Studium und Weiterbildung*, 5th ed., Facultas, Wien, 2019.

[19] D. Frommeld, K. Weber, *Telepräsenzroboter zur Unterstützung von Pflege und Therapie: Eine qualitative Interviewstudie zur Nutzung und Ablehnung*, in: Digitale Technik für ambulante Pflege und Therapie: Nutzung, Akzeptanz, Wirkung und Lebensqualität. K. Weber, S. Haug, N. Lauer, C. Mohr, A. Pfingsten, G. Raptis, G. Bahr (eds.), transcript, Bielefeld, forthcoming 2024.

[20] D. Ferizaj, L. Perotti, R. Dahms, A. Heimann-Steinert, Technologienutzung im Alter: Zusammenhänge zwischen Akzeptanz, Kompetenz, Kontrolle, Interesse und sozialen Indikatoren bei Personen über 60 Jahre, *Z Gerontol Geriat*, (2023), doi: 10.1007/s00391-023-02225-9.

[21] S. R. Güsken, K. Frings, F. Zafar, T. Saltan, P. Fuchs-Frohnhofen, J. Bitter-Krahe, Einflussfaktoren auf die Nutzungsintention von Pflegekräften zur Verwendung digitaler Technologien in der ambulanten Pflege – Fallstudie zur Einführung eines Sensortextils, *Z. Arb. Wiss.*, **75** (2021), pp. 470-490, doi: 10.1007/s41449-021-00277-4.

[22] Institut für Arbeit und Technik (IAT), *Mobility Motivator – User Consultation Protocol and Tools*. 2014, http://www.aal-europe.eu/wp-content/uploads/2020/01/MoMo_D2.1_final_2-ch.pdf

[23] Generali Deutschland AG (ed.), *Generali Altersstudie 2017: Wie ältere Menschen in Deutschland denken und leben: repräsentative Studie des Instituts für Demoskopie Allensbach mit Kommentaren des wissenschaftlichen Beirats der Generali Altersstudie 2017,* Springer, Berlin, 2017.

[24] K. Karrer, C. Glaser, C. Clemens, C. Bruder, *Technikaffinität erfassen - der Fragebogen TA-EG*, in: Der Mensch im Mittelpunkt technischer Systeme: 8. Berliner Werkstatt Mensch-Maschine-Systeme, 7. bis 9. Oktober 2009, A. Lichtenstein, C. Stößel, and C. Clemens, (eds.), VDI-Verl., Düsseldorf, 2009, pp. 194–199.

[25] F. J. Neyer, J. Felber, and C. Gebhardt, Entwicklung und Validierung einer Kurzskala zur Erfassung von Technikbereitschaft, *Diagnostica* **58**(2) (2012), pp. 87–99. doi: 10.1026/0012-1924/a000067.

[26] A. Seifert, H. R. Schelling, *Digitale Senioren. Nutzung von Informations- und Kommunikationstechnologien (IKT) durch Menschen ab 65 Jahren in der Schweiz im Jahr 2015*. ProSenectute, Zürich, 2015.

[27] C. Paul, L. Spiru, *From Age to Age: Key Gerontographics Contributions to Technology Adoption by Older Adults*, in: ICT4AWE 2021: Proceedings of the 7th International Conference on Information and Communication Technologies for Ageing Well and E-Health. M.C. Ziefle, M. Mulvenna, L. Maciaszek (eds). SciRePress, Sétubal, 2021. pp. 121–129, doi: 10.5220/0010395701210129.

[28] H. Künemund, C. Vogel, *Alter und Altern – Kritik der Messung und Auswertung am Beispiel des Wohnens*, in: Wohnen und Gesundheit im Alter, A. Teti, E. Nowossadeck, J. Fuchs, H. Künemund (eds.). Springer VS, Wiesbaden, 2022. pp. 277–285.

[29] H. Künemund, *Wovon hängt die Nutzung technischer Assistenzsysteme ab? Expertise zum Siebten Altenbericht der Bundesregierung*, Deutsches Zentrum für Altersfragen, Berlin, 2016. https://nbn-resolving.org/urn:nbn:de:0168-ssoar-49994-1.

[30] B. Kramer, *Die Akzeptanz neuer Technologien bei pflegenden Angehörigen von Menschen mit Demenz*, Ruprecht-Karls-Universität, Heidelberg, 2016. https://archiv.ub.uni-heidelberg.de/volltextserver/20856/.

dHealth 2024
D. Hayn et al. (Eds.)

doi:10.3233/SHTI240006

Large Language Model-Based Evaluation of Medical Question Answering Systems: Algorithm Development and Case Study

Daniel REICHENPFADER[a,1], Philipp RÖSSLHUEMER[b] and Kerstin DENECKE[a]
[a] *Bern University of Applied Sciences, Biel/Bienne, Switzerland*
[b] *Department of Diagnostic, Interventional and Pediatric Radiology, Bern University Hospital, University of Bern, Bern, Switzerland*

Abstract. Background: Healthcare systems are increasingly resource constrained, leaving less time for important patient-provider interactions. Conversational agents (CAs) could be used to support the provision of information and to answer patients' questions. However, information must be accessible to a variety of patient populations, which requires understanding questions expressed at different language levels. Methods: This study describes the use of Large Language Models (LLMs) to evaluate predefined medical content in CAs across patient populations. These simulated populations are characterized by a range of health literacy. The evaluation framework includes both fully automated and semi-automated procedures to assess the performance of a CA. Results: A case study in the domain of mammography shows that LLMs can simulate questions from different patient populations. However, the accuracy of the answers provided varies depending on the level of health literacy. Conclusions: Our scalable evaluation framework enables the simulation of patient populations with different health literacy levels and helps to evaluate domain specific CAs, thus promoting their integration into clinical practice. Future research aims to extend the framework to CAs without predefined content and to apply LLMs to adapt medical information to the specific (health) literacy level of the user.

Keywords. Natural Language Processing, Consumer Health Information, Algorithms, Conversational Agents, Large Language Model

1. Introduction

In healthcare systems facing escalating resource constraints, conversational agents (CAs) represent a significant opportunity to improve patient-provider interactions and streamline information exchange. CAs are “computer programs designed to engage in human-like conversations with users” [1]. They can be grouped according to the direction of information flow: Information can flow either from the patient to the provider, e.g. facilitating history taking, symptom checking and triage, or from the provider to the patient, e.g. facilitating response to patient queries. In addition, CAs can support bi-directional information exchange. CAs designed to answer user questions are related to question-answering (QA) systems. These can be further categorized based on their

[1] Corresponding Author: Daniel Reichenpfader, Institute for Patient-centered Digital Health, Bern University of Applied Sciences, Biel/Bienne, Switzerland, E-Mail: daniel.reichenpfader@bfh.ch

technical implementation: On the one hand, patient questions can be mapped to the most similar example of pre-defined Question-Answer-Pairs (QAPs), either based on intent recognition or similarity-based techniques. With pre-defined QAPs, clinicians have absolute control over the provided content and answer generation is transparent, reducing the risk of providing wrong answers. On the other hand, Large Language Models (LLMs) might be used, including retrieval-augmented generation (RAG) [2] or immediate generation of an answer by the model itself [3]. Recent research shows that LLMs might provide comparably accurate and less biased answers to patients' medical questions compared to human experts, although still having limitations, e.g. interpretability of answers [3]. Therefore, many existing health CAs are using the approach with predefined QAPs.

In a setting, where a CA is supposed to answer patient queries, it is essential that it can handle a diverse set of user input. In healthcare, this means that there are users with a high health literacy and knowledge of medical terms, and those with low health literacy. Health literacy is defined as "the ability of an individual to obtain and translate knowledge and information in order to maintain and improve health in a way that is appropriate to the individual and system contexts" [4]. Pre-defined QAPs can never cover all possible wordings of user inputs to health CAs. The question arises how health CAs can be tested for their ability to understand user input formulated with different levels of health literacy. The aim of this paper is to present an approach that allows to simulate user input with different levels of health literacy and to investigate the accuracy of a health CA when confronted with this input. Specifically, we aim at answering the following research question: *How can LLMs be used to efficiently assess health CAs' ability to deal with input from a diverse set of patient populations?*

We will propose an evaluation framework that uses LLMs to emulate patient populations with varying levels of health literacy through in-context learning [5], with the aim of assessing health CAs. LLMs, defined as "deep learning models with a huge number of parameters trained in an unsupervised way on large volumes of text" [6], have demonstrated human-level performance in various academic and professional exams and benchmarks [7]. We therefore believe in their potential to effectively simulate user input.

2. Methods

2.1. Evaluation framework

Our evaluation framework consists of a fully automated procedure and a semi-automated procedure that requires the manual assessment of results by clinicians. Both procedures require the definition of QAPs. In the first approach, alternative questions are generated based on predefined questions, a task prompt and patient vignettes. The task prompt instructs the LLM what to do, i.e., to produce ten alternative questions for an existing question, based on a patient vignette. Patient vignettes are prompts that instruct an LLM to act as certain type of patient (e.g. having low (health) literacy). The generated alternative questions are sent to the CA with its answers being evaluated. Specifically, its answers are compared with the ground truth pre-defined answer. Pseudo-code for the automated evaluation is shown in Figure 1.

The total number of correctly and incorrectly classified questions is stored and used to calculate the accuracy averaged over each patient vignette and averaged over all questions, see Eq. (1). A question is considered correct if the answer returned by the

system matches the original predefined question. Questions for which an incorrect or no answer is returned are treated as misclassified.

Algorithm Automated evaluation of Question Variations

```
Input: dict predefinedContent, list patientVignettes
for each question-answer pair (question, groundTruthAnswer) in predefinedContent do
    for each vignette patientVignette in patientVignettes do
        generatedQuestionVariations ← SendToLLM(question, vignette)
        Save (generatedQuestionVariations, vignette, question, groundTruthAnswer) object
    end for
end for
for each object (generatedQuestionVariations, vignette, question, groundTruthAnswer) in the saved objects do
    for each question variation generatedQuestion in generatedQuestionVariations do
        generatedAnswer ← SendToQAmodule(generatedQuestion)
        Save (generatedQuestion, generatedAnswer) tuple to object
    end for
end for
Compute count of question-answer pairs: totalPairs ← CountAllPairs(saved objects)
Compute count of correct answers: correctCount ← CountCorrectAnswers(saved objects)
Compute percentage of correct answers: accuracy ← (correctCount / totalPairs) × 100 %
Output: accuracy
```

Figure 1. Pseudo-code of automated evaluation

$$Total\ Accuracy = \frac{Correct\ answers}{QAPs * No.\ of\ patient\ vignettes * No.\ of\ variations} \quad (1)$$

For the second, semi-automated approach, possible patient questions are generated for the domain covered by the CA. This procedure is based on a different task prompt that instructs the LLM to generate questions, impersonating each defined patient vignette. This approach additionally tests the CA's ability to handle unexpected user input. The generated questions are sent to the CA and stored together with the generated answer. To assess the quality of the CA, the generated answers are independently assessed and labelled as incorrect, partially correct, or correct by two authors (KD, DR). Inconsistencies are resolved by discussion. The evaluation score is the percentage of correct answers.

2.2. Case study

In order to test the proposed evaluation framework, we are using an existing health CA that is a rule-based, FHIR-conformant system for collecting the medical history of patients [8]. We added a QA module containing pre-defined QAPs. This module will be tested using the framework. The QA module is based on similarity matching: Let E = $\{e_1, e_2, \ldots, e_n\}$ be the set of n embedded predefined questions, where each e_i represents an embedded string. The embedded patient question to be answered is denoted as e_{query}. The cosine similarity between two embedded strings e_i and e_j is denoted as cosine similarity (e_i, e_j). To identify the most similar string, we formulate the objective as finding the index i that minimizes the cosine similarity, see Eq. (2).

$$Matching\ question = argmin_i \left(\text{cosine similarity}\left(e_{\text{query}}, e_i\right)\right) \quad (2)$$

The QA system does not provide an answer if the similarity score is too low or if the patient question is less than three words long. The minimum length of three words for questions ensures that enough information is available to be compared with the pre-defined content. Questions that cannot be answered are stored for continuous improvement and addition of QAPs. Furthermore, the system enables patients to flag wrong or complicated answers. Text embeddings are created with a pre-trained model [9] based on the sentence-transformers library [10] and stored in Chroma, an open-source vector database [11]. The QA module is currently implemented for the field of mammography. Two radiologists developed a set of 33 German QAPs for the domain of mammography, comprising popular questions asked by patients before an intervention (e.g. regarding costs, preparation, pain, expectations, duration etc.). These QAPs are integrated in the QA system.

3. Results

A total of three patient vignettes were experimentally developed, representing patients with high health literacy, low health literacy, and low literacy of German. The following text shows the patient vignette (translated from German) for low health literacy: *"You are a patient with low health literacy, and you are not well informed about the healthcare system. You use simple language and do not comprehend complicated medical terms. You will soon undergo a mammography."* For the automated evaluation, the following translated prompt was iteratively and experimentally developed to obtain ten question variations per QAP based on each patient vignette: *"Define, based on the following question, ten different variations of this question. Adapt your choice of words to your health literacy and literacy level. Return only the question, separated by a line break (\n)."* As LLM, OpenAI GPT-4 was used [7]. The results of the automated evaluation process are described in Table 1. In total, 990 alternative questions were generated by the LLM (33 QAPS * ten alternatives * three patient vignettes). Eleven questions (0.01 %) were not answered by the QA module as they contained fewer than three words, counting towards the number of errors.

Table 1. Results of case study (automated evaluation) tested on 990 alternative questions

Patient vignette	Errors (n)	Accuracy
High health literacy	101	0.69
Low health literacy	75	0.77
Low German literacy	69	0.79
Total	245	0.75

The results of the semi-automated evaluation involving manual judgement are described in Table 2. The following task prompt was used: *"Define, based on the following description [i.e. the patient vignette], ten different questions to ask your doctor. Adapt your choice of words to your health literacy and literacy level. Return only the question, separated by a line break (\n)."* In total, 30 new questions for the domain of mammography were generated (ten alternatives * three patient vignettes). Four questions were not answered by the QA module as they contained fewer than three words and no

answer was returned for three questions due to no pre-defined question being similar enough. Both categories are counting towards incorrect answers.

Table 2. Results of case study (semi-automated evaluation) tested on 30 automatically generated questions

Patient vignette	Incorrect (n)	Partial (n)	Correct (n)
High health literacy	5	5	0
Low health literacy	8	0	2
Low German literacy	5	0	5
Total	18	5	7

Obtaining the question variations and new questions resulted in total costs of $3.43 and therefore $0.10 per QAP using three patient vignettes. We make the source code of the evaluation framework including the QAPs for mammography as well as all developed prompts publicly available via Zenodo (10.5281/zenodo.10782323).

4. Discussion

Regarding automated evaluation, the results show that the 'high health literacy' patient vignette had the lowest accuracy. This may be because the LLM, as designed, generated relatively long alternative questions that tended to use more sophisticated wording, complicating the matching with possible responses. Conversely, the proportion of correctly answered questions derived from the 'low health literacy' vignette was ten percentage points higher. This might be explained by simple and short generated question variants. However, it can be assumed that patients with such a low level of German language literacy would not comprehend the pre-defined provided answers since their understanding requires a high health literacy and high readability skills.

As far as the semi-automated evaluation is concerned, none of the high literacy questions could be fully answered by the QA system. 50% of the questions were partially answered. This could be due to relatively long and complicated question variations, similar to the automated scoring. Only for the low health literacy vignette did the QA system report that it could not find a suitable answer in three cases. Although this response is technically wrong, it is a more desirable behavior than providing a completely wrong answer. In total, only seven newly generated questions were answered correctly, showing that the evaluated QA system does not yet generalize well to new data. The development of patient vignettes proved to be challenging: On the one hand, each patient vignette should lead to a unique formulation. On the other hand, the generated questions still need to be realistic. We recognized that for patient vignettes with high health literacy, the LLM tends to exaggerate the use of complex terms and phrases and domain-specific terms. Nevertheless, the QA system was able to answer 69 % of these questions correctly in the automated evaluation. See Table 3 for an example of an original question and LLM-generated question using the vignette-based alternatives.

Table 3. Comparison of original and LLM-generated questions

Group	Variant (translated from German)
Original question	How long does a mammogram take?
High health literacy	Assuming I undergo a mammogram, how long would I be expected to spend in the radiology department?
Low health literacy	How long will I sit there when they do this mammogram?
Low German Literacy	Mammogram, much time needed?

To improve performance of our QA module, the following improvement strategies can be applied: If two questions with different answers are very similar to each other, the model may constantly misclassify one of them. To improve this, these questions could either be merged into one QAP, or each question could be rephrased. Next, the distance threshold defines the minimum level of similarity between a predefined question and a patient question for the system to return an answer. As every question exceeded this threshold in the automated test, a higher threshold could lead to higher precision, albeit at the expense of accuracy. Finally, better models are continually being published. The model used to implement the evaluated QA module is easily replaceable due to its compatibility with the sentence-transformers framework.

The limitations of our study are as follows: Only one commercially available LLM was tested. In future, the performance of other LLMs, including open-source models, should be compared. Moreover, our case study only focused on a single clinical domain with a relatively small register of QAPs (n = 33) and newly generated questions (n = 30).

In this paper, we present a scalable, semi-automated evaluation framework to evaluate domain-specific QA systems. The developed task prompts and patient vignettes can be easily used with other models. Using LLMs and patient vignettes, QA inputs from different patient populations can be easily simulated. Based on this evaluation framework, QA systems can be evaluated with low effort cost-effective, fostering their application in clinical practice. Future research directions include adapting this evaluation framework to QA systems without pre-defined content and elaborating on the development of more detailed prompts to simulate patient populations. Furthermore, we recommend investigating whether LLM-based adaptation of responses based on the patient's (health) literacy level improves understandability and patient satisfaction.

References

[1] K. Denecke, 'Potential and pitfalls of conversational agents in health care', Nat Rev Dis Primers, vol. 9, no. 1, Art. no. 1, Nov. 2023, doi: 10.1038/s41572-023-00482-x.

[2] S. Siriwardhana, R. Weerasekera, E. Wen, T. Kaluarachchi, R. Rana, and S. Nanayakkara, 'Improving the Domain Adaptation of Retrieval Augmented Generation (RAG) Models for Open Domain Question Answering', Transactions of the Association for Computational Linguistics, vol. 11, pp. 1–17, Jan. 2023, doi: 10.1162/tacl_a_00530.

[3] K. Singhal et al., 'Large language models encode clinical knowledge', Nature, vol. 620, no. 7972, pp. 172–180, Aug. 2023, doi: 10.1038/s41586-023-06291-2.

[4] C. Liu et al., 'What is the meaning of health literacy? A systematic review and qualitative synthesis', Fam Med Community Health, vol. 8, no. 2, p. e000351, May 2020, doi: 10.1136/fmch-2020-000351.

[5] T. Brown et al., 'Language Models are Few-Shot Learners', in Advances in Neural Information Processing Systems, Curran Associates, Inc., 2020, pp. 1877–1901.

[6] A. Birhane, A. Kasirzadeh, D. Leslie, and S. Wachter, 'Science in the age of large language models', Nat Rev Phys, vol. 5, no. 5, Art. no. 5, May 2023, doi: 10.1038/s42254-023-00581-4.

[7] OpenAI et al., 'GPT-4 Technical Report'. arXiv, Dec. 18, 2023. doi: 10.48550/arXiv.2303.08774.

[8] K. Denecke, N. Cihoric, and D. Reichenpfader, 'Designing a Digital Medical Interview Assistant for Radiology', Stud Health Technol Inform, vol. 301, pp. 60–66, May 2023, doi: 10.3233/SHTI230012.

[9] P. May and deepset GmbH, 'German BERT large paraphrase cosine'. Deutsche Telekom, Dec. 17, 2023. Accessed: Jan. 05, 2024. [Online]. Available: https://huggingface.co/deutsche-telekom/gbert-large-paraphrase-cosine

[10] N. Reimers and I. Gurevych, 'Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks'. arXiv, Aug. 27, 2019.

[11] Chroma, 'Chroma'. Chroma, Jan. 05, 2024. Accessed: Jan. 05, 2024. [Online]. Available: https://github.com/chroma-core/chroma

dHealth 2024
D. Hayn et al. (Eds.)

doi:10.3233/SHTI240007

Mapping the Bulgarian Diabetes Register to OMOP CDM: Application Results

Evgeniy KRASTEV[a], Emanuil MARKOV[b], Simeon ABANOS[a], Ralitsa KRASTEVA[c] and Dimitar TCHARAKTCHIEV[d,1]

[a] *Sofia University St. Kliment Ohridski, Bulgaria (EFMI institutional member)*
[b] *Technical University, Sofia, Bulgaria*
[c] *Specialized Hospital for Active Treatment of Children's Diseases, Sofia, Bulgaria*
[d] *Medical University, Sofia, Bulgaria (EFMI institutional member)*

Abstract. *Background*: The Bulgaria Diabetes Register (BDR) contains more than 380 millions of pseudonymized outpatient records with proprietary data structures and format. *Objectives:* This paper presents the application results and experience acquired during the process of mapping such observational health data to OMOP CDM with the objective of publishing it in the European Health Data and Evidence Network (EHDEN) Portal. *Methods:* The data mapping follows the activities of the well-structured Extract-Transform-Load process. Unlike other publications, we focus on the need for preprocessing the data structures of raw data, cleaning data and procedures for assuring quality of data. *Results:* This paper provides quantitative and statistical measures for the records in the CDM database as published in the EHDEN Portal. *Conclusion:* The mapping of data from the BDR to OMOP CDM provides the EHDEN community with opportunities for including these data in large-scale project for evidence generation by applying standard analytical tools.

Keywords. Medical Informatics Applications, Diabetes Mellitus, Registries, OMOP CDM, Electronic Health Records

1. Introduction

Diabetes mellitus is a socially significant illness that rapidly increases its prevalence over the world. National diabetes registries (NDR) like the Bulgarian Diabetes Register (BDR) provide opportunities to analyze huge amounts of clinical data and develop new approaches for timely prevention of this disease by focusing on risk factors with age-specific quantitative effects on diabetes [1] [2].

The BDR manages pseudonymized electronic health records of the patients with diabetes (Type 1 and Type 2) since 2013, where the latest dataset is from 2018 [3] [4]. Over 380 million outpatient records were collected during this time period by the National Health Insurance Fund (NHIF) from all General Practitioners and Health Professionals in specialized healthcare nationwide for every visit of a patient suffering from diabetes. The data structure of the outpatient records follows a proprietary XML schema introduced by the NHIF for its specific data processing purposes. Similarly to

[1] Corresponding Author: Dimitar Tcharaktchiev, Department of Medical Informatics, Medical University-Sofia, University Hospital of Endocrinology, 2 Zdrave street, Sofia 1431, Bulgaria; E-mail: dimitardt@gmail.com, ORCID: https://orcid.org/0000-0001-5765-840X

other NDR, this proprietary structure of observational health data strongly constrains the exchange and integration of data as well as the interoperability of the tools for data analysis and evidence generation [5].

In this paper we describe the implementation of the Extract, Transform and Load (ETL) process for mapping the dataset with outpatient records in the BDR from year 2018 to the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) in the Observational Health Data Sciences and Informatics (OHDSI) community [6] [7]. The OMOP CDM offers a feasible solution to interoperability challenges in health data processing by transforming data into a common format and standardized vocabulary [8] [9]. It is the CDM employed to build the European Health Data and Evidence Network (EHDEN), a growing network of 187 Data Partners across 29 countries, with access to more than 850 million anonymized health records [10]. We focus on overcoming some of the major difficulties in the implementation of the ETL process that emerge at the stage of preprocessing the source data, vocabulary mapping and overall quality assurance of data workflow [11]. Although each step of the ETL process is supported by OHDSI open-source software tools [12], the nature of these problems makes the ETL process a challenging experience. It is required to overcome a lot of technical difficulties, knowledge management problems and quality concerns in order to make the OMOP CDM part of EHDEN [13] [14]. Thus, the OMOP CDM of the BDR with data for more than 500K distinct patients was published in the EHDEN portal [15] [16]. Section 2 describes the methods to execute the ETL process, while Section 3 presents the OMOP CDM database in terms of quantitative measures and results from data quality tests. Section 4 summarizes the findings in this paper, compares the results with existing literature sources and outlines future research plans.

2. Methods

The source dataset comprises all the outpatient records of 501,065 patients of all ages with diabetes (42,452 Type 1; 458,613 Type 2) collected in 2018 by the NHIF. These records were provided for research purposes to the National eHealth Scientific Program after applying pseudonymization procedures [4].

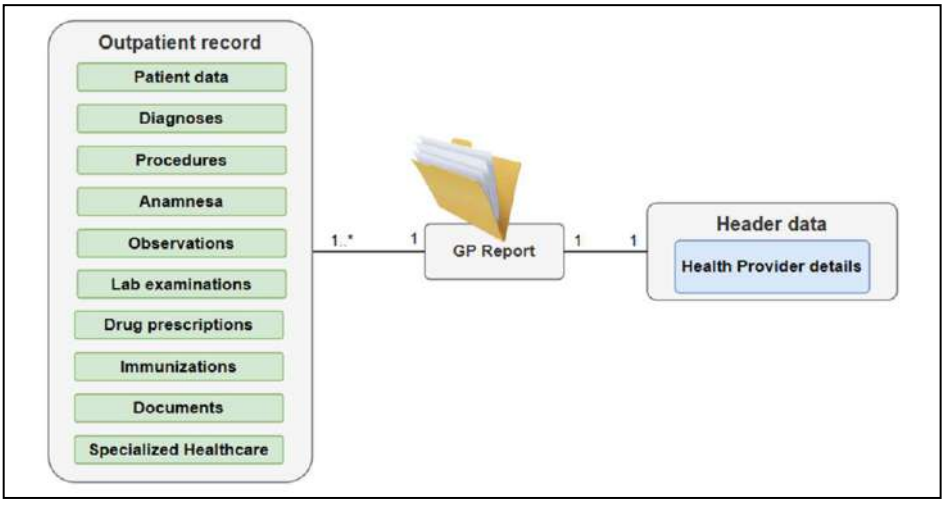


Figure 1. Patient-centric structure of the source dataset with outpatient records.

Mapping these nationally representative observational health data to OMOP CDM allows to extend the scope of shared data at international level and generate evidence using standardized analytics tools. Besides, the ETL process of mapping is enhanced by the patient-centric structure of both the source dataset (Figure 1) and the OMOP CDM. It is noteworthy, that in our case we were dealing with a large number (6,887,876) of original clinical documents in XML format whose XML structure appeared to be damaged by the pseudonymization procedure, for instance, by inserting the '&' character or various control symbols between opening and closing XML tags. Therefore, a stage of converting the pseudonymized records into valid XML documents and transforming XML data into relational database tables preceded the well-structured ETL process.

Java applications were developed for this purpose and it facilitated significantly the execution of the Extraction and Transform stages in the ETL process. For example, these applications created a relational model of the source database that retains the patient-centric structure of the outpatient records and resembles the semantics of the tables in the OMOP CDM (v. 5.3.0). On the other side, the extraction of observational health data by processing original clinical documents is a problem that requires non-traditional approach for finding a solution. It is common to discover such data recorded in raw text in sections Anamneses, Observations or Lab examinations of outpatient records. In our case, this text was provided in unstructured native (Cyrillic) language, where one and the same clinical concept appeared with different abbreviations in huge number of records and besides, sometimes with wrongly specified measurement values or units. Python scripts using regular expressions for extraction of observational health data (blood pressure, blood sugar etc.) proved to be more effective than traditional Natural Language Processing (NLP) tools in resolving extraction problems with such complexities [17].

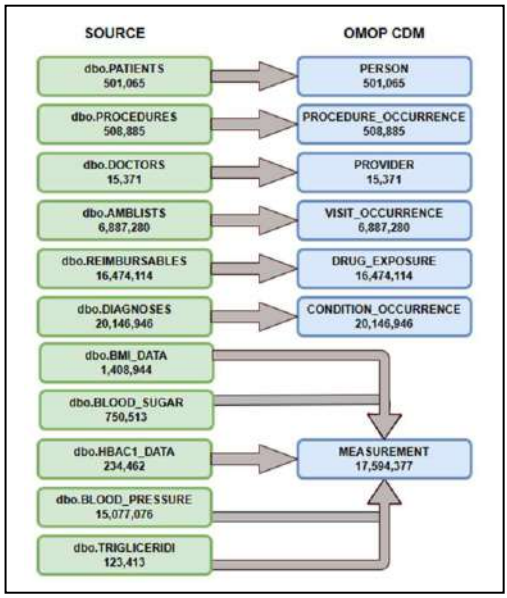


Figure 2. Data table mapping and number of rows per table after the ETL process.

Dealing with raw data required to complement the extraction of observational health data with procedures for cleaning data and ensuring EHDEN data quality criteria

(conformance, completeness and plausibility) are satisfied [11] [13] [18]. For example, blood pressure measurements might appear with a different separator ('-', ',', '/') or without any separator between the systolic and diastolic values located inside the unstructured text in elements Anamneses or Observations of the XML documents. Moreover, the systolic and diastolic values might appear exchanged. It is just an example, that illustrates the complexity of the procedures for measurement extraction.

At the end of the preprocessing stage, the source database was loaded with all the data from the original dataset and the ETL process followed the development process suggested by the OHDSI community [7]. It entails the usage of specific OHDSI tools like White Rabbit, Rabbit-in-a-Hat, Achilles, Data Quality Dashboard at each step of the process [12]. For example, the White Rabbit report stated that there were no missing values in the source database. The standardization of the medical terminologies in the source usually requires significant efforts in this process. In this use case it helped that the NHIF applied the ATC classifier for drug encoding. Similarly, the source encoding for diagnoses and procedures followed ICD10 and ICD9Proc, respectively. Data from the national specialty classifier is mapped manually to the Medicare Specialty dictionary. It significantly facilitated the translations of these common source terminologies to standard concepts using the OMOP Vocabularies [19]. Once the structural and semantic standardization design was finalized, a complete set of instructions were produced to map the source database tables to the OMOP CDM tables (Figure 2).

3. Results

The preprocessing of the original dataset described in the previous section contributed to the successful completion of the ETL process. For better efficiency, both the source database and the OMOP CDM database were hosted on the same instance of Microsoft SQL Server. It enabled direct data extraction and transformation via SQL scripts making the procedure to load the OMOP CDM database stepwise and fully reproducible.

Table 1. Distribution of records per person with diabetes Type 1 (T1) and Type 2 (T2) in major data domains.

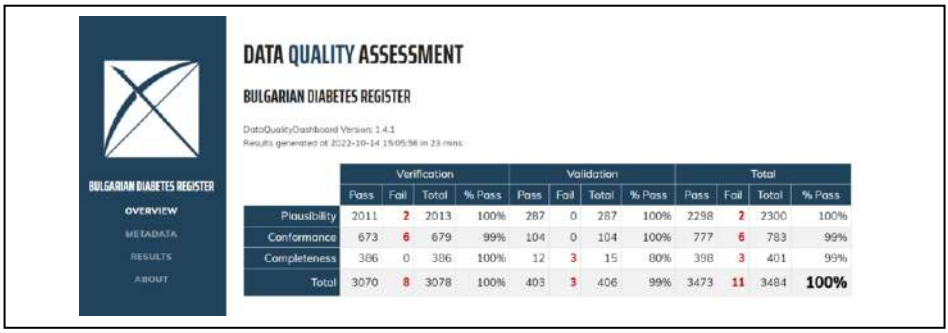
DESCRIPTION	VISIT OCCURENCE		CONDITION OCCURRENCE		DRUG EXPOSURE		MEASUREMENT	
	T2	T1	T2	T1	T2	T1	T2	T1
Min	1	1	1	1	0	0	0	0
Percentile 10%	5	7	10	14	0	8	5	8
Percentile 25%	7	11	18	25	11	18	16	18
Median	12	17	33	45	30	32	30	32
Percentile 75%	18	23	54	72	49	56	50	56
Percentile 90%	24	29	77	98	67	76	68	76
Max	142	125	417	319	467	390	369	217

This way, all the clinical documents (100%) from the source dataset for 501,065 patients with diabetes were mapped to the OMOP CDM. Accordingly, data from the source dataset were loaded in tables of a Microsoft SQL Server database with relational model matching the OMOP CDM (Figure 2). Table 1 and Table 2 provide insight about the burden caused by the diabetes illness on the healthcare system during 2018. Although detailed analysis of the numerical data in these tables is outside the scope of this paper, they are good example for the potential benefits of using the OMOP CDM in research.

Table 2. Top 10 of the mapped drugs.

Row No.	Concept name	Number of records	Subjects
1	metformin; oral	1,676,600	224,700
2	bisoprolol; oral	1,115,400	143,300
3	gliclazide; oral	768,400	104,700
4	rosuvastatin; oral	522,600	71,700
5	amlodipine; oral	462,900	66,200
6	glimepiride; oral	406,600	51,700
7	nebivolol; oral	397,700	54,400
8	metoprolol; systemic	364,500	47,100
9	lercanidipine; oral	309,900	45,400
10	atorvastatin; oral	306,400	42,200

At the end of the ETL process we ran the Achilles and the Data Quality Dashboard tools provided by the OHDSI community for data quality assessment. Figure 3 shows the results of the quality checks executed by the Data Quality Dashboard. These results prove (1) conformance with OMOP CDM standards and format; (2) plausibility i.e. data values are valid with respect to integrity rules for domains, concept classes and vocabulary IDs and (3) completeness, meaning that data values are present.



	Verification				Validation				Total			
	Pass	Fail	Total	% Pass	Pass	Fail	Total	% Pass	Pass	Fail	Total	% Pass
Plausibility	2011	2	2013	100%	287	0	287	100%	2298	2	2300	100%
Conformance	673	6	679	99%	104	0	104	100%	777	6	783	99%
Completeness	386	0	386	100%	12	3	15	80%	398	3	401	99%
Total	3070	8	3078	100%	403	3	406	99%	3473	11	3484	100%

Figure 3. The Total summary result of executing the DataQualityDashboard v1.4.1 tool is 100% pass.

The quality assessment procedures confirm the successful completion of the ETL process for mapping the BDR in OMOP CDM. It is among the 153 OMOP CDM from 28 countries that are currently published in the EHDEN Portal [16].

4. Discussion

Mapping of the BDR to the OMOP CDM resolves interoperability issues across similar data sources by standardizing management of observational health data in terms of common structure (data model) and terminology (vocabulary). Unlike other papers [5] [9], this study addresses the usage of data quality assessment procedures supporting high-quality in evidence generation. Preprocessing of the source dataset and the extraction of measurements of observational health data were the major challenges to overcome. The mapping of a nationally representative sample of outpatient records of patients with diabetes to OMOP CDM allows us to join the EHDEN community and apply standard analytical tools in the large-scale EHDEN MegaStudy project entitled “Studying Drug shortages in Europe: A Multinational, Multidatabase Network Study”.

Acknowledgement

This research is supported by Grant 80-10-8/11.04.2023 of Sofia University, Bulgaria.

References

[1] Bak JCG, Serné EH, Kramer MHH, Nieuwdorp M, Verheugt CL. National diabetes registries: do they make a difference? Acta diabetologica. 2021; 58(3): 267–278.
[2] Boytcheva S, Angelova G, Angelov Z, Tcharaktchiev D. Data Mining and Analytics for Exploring Bulgarian Diabetic Register. In Data Analytics and Management in Data Intensive Domains.: Springer International Publishing; 2018.
[3] University Specialized Hospital for Active Treatment in Endocrinology. Diabetes Register. [Online].; 2024 [cited 2024 January 11. Available from: https://usbale.org/bg/registar-zaharen-diabet/.
[4] National Scientific Program eHealth. National Scientific Program "Electronic Healthcare in Bulgaria" (e-health). [Online].; 2024 [cited 2024 January 10. Available from: https://ehealth.fmi.uni-sofia.bg/.
[5] Korntheuer RL, Katsch F, Duftschmid G. Transforming Documents of the Austrian Nationwide EHR System into the OMOP CDM. In Pfeifer B, et. al.. Volume 301: dHealth 2023, Studies in health technology and informatics (dHealth 2023).: IOS Press; 2023. p. 54–59.
[6] Observational Health Data Sciences and Informatics. OMOP Common Data Model. [Online].; 2024 [cited 2024 January 10. Available from: https://ohdsi.github.io/CommonDataModel/.
[7] Observational Health Data Sciences and Informatics. The Book of OHDSI: https://ohdsi.github.io/TheBookOfOhdsi/; 2021.
[8] Kent S, Burn E, Dawoud D, Jonsson P, Østby JT, Hughes N, et al. Common Problems, Common Data Model Solutions: Evidence Generation for Health Technology Assessment. Pharmacoeconomics. 2021; 39(3): 275-285.
[9] Reinecke I, Zoch M, Reich C, Sedlmayr M, Bathelt F. The Usage of OHDSI OMOP - A Scoping Review. In German Medical Data Sciences 2021: Digital Medicine: Recognize – Understand – Heal, Studies in health technology and informatics.: IOS Press; 2021. p. 95–103.
[10] EHDEN. A federated network of Data Partners. [Online].; 2022 [cited 2024. Available from: https://www.ehden.eu/datapartners/.
[11] Blacketer C, Voss EA, DeFalco F, Hughes N, Schuemie MJ, Moinat M, et al. Using the Data Quality Dashboard to Improve the EHDEN Network. Applied Sciences. 2021; 11(24): 11920.
[12] Observational Health Data Sciences and Informatics. Software Tools. [Online].; 2024 [cited 2024 January 10. Available from: https://www.ohdsi.org/software-tools/.
[13] Voss EA, Blacketer C, van Sandijk S, Moinat M, Kallfelz M, van Speybroeck M, et al. European Health Data & Evidence Network-learnings from building out a standardized international health data network. Journal of the American Medical Informatics Association : JAMIA. 2024; 31(1): 209–219.
[14] Biedermann P, Ong R, Davydov A, Orlova A, Solovyev P, Sun H, et al. Standardizing registry data to the OMOP Common Data Model: experience from three pulmonary hypertension databases. BMC medical research methodology. 2021; 21(1): 238.
[15] Krastev E, Markov E. EHDEN Community Calls. [Online].; 2023 [cited 2024 January 12. Available from: https://www.ehden.eu/bulgarian-diabetes-register-mapping-onto-omop-cdm-cc10/.
[16] Krastev E, Tcharaktchiev D, Abanos S. Application of OMOP Common Data Model for Data Integration: The Bulgarian Diabetes Register. In Volume 309: Telehealth Ecosystems in Practice, Studies in Health Technology and Informatics. Torino, Italy: IOS Press; 2023. p. 141 - 142.
[17] Wiesmüller F, Hayn D, Kreiner K, Pfeifer B, Pölzl G, Kastner P, et al. Natural Language Processing for Free-Text Classification in Telehealth Services: Differences Between Diabetes and Heart Failure Applications. In Volume 279: Navigating Healthcare Through Challenging Times, Studies in Health Technology and Informatics (dHealth 2021).: IOS Press; 2021. p. 157 - 164.
[18] Brownlee J. Data Preparation for Machine Learning. Data Cleaning, Feature Selection, and Data Transforms in Python: Machine Learning Mastery; 2020.
[19] Reich C, Ostropolets A, Ryan P, Rijnbeek P, Schuemie M, Davydov A, et al. OHDSI Standardized Vocabularies-a large-scale centralized reference ontology for international data harmonization. Journal of the American Medical Informatics Association : JAMIA. 2024; 31(3): ocad247.

dHealth 2024
D. Hayn et al. (Eds.)

doi:10.3233/SHTI240008

Supporting Well-Being: Exploring the Value of a Digital Coach for Older Adults in the Transition from Work to Retirement

Katja NEUREITER[a], Leen BROECKX[b], Markus GARSCHALL[a,1], Kim HELSEN[b] and Johannes KROPF[c]

[a] *AIT Austrian Institute of Technology GmbH, Center for Technology Experience, Vienna, Austria*
[b] *Thomas More University of Applied Sciences, Centre of Expertise Care and Well-being, Geel, Belgium*
[c] *Salumentis OG, Vienna, Austria*

Abstract. Supporting older adults' health and well-being in the transition from work to retirement requires a holistic perspective and needs to address physical, mental, and social aspects of life. In a field study, applying a mixed-methods approach, we investigated to what extent the prototype of a digital coach can support older adults in this sensitive phase. We aim at answering the central research question: How can a digital coach support older adults in the transition from work to retirement to establish and maintain a healthy lifestyle? Overall, 32 participants from Austria and Belgium took part in an eight-week trial. App-based interventions in different domains (physical, mental, social) were provided and aimed at motivating the target group to become and/or stay active. The study shows that the digital coach has potential to support health and well-being on various levels. In particular, the mental activities proofed valuable and supported older adults' well-being.

Keywords. Well-being, digital coach, older adults, retirement

1. Introduction

For several years Europe has been confronted with the effects of demographic ageing, i.e., the number of older people is increasing relative to the number of people of working age. By 2050, the proportion of older adults (aged 65 years and older) is expected to increase from 90,5 million (2019) to 129,8 million, having implications on an individual and societal level, e.g., rising costs for health and long-term care [1].

A particularly sensitive and critical phase in older age is the transition from work to retirement, when older adults do not only gain freedom but face losses, for example, on a social level, losing a variety of social contacts [2]. Time and relationships need to be redefined [3] and negative impacts might be decreasing income, a loss of social capital, and decrease of physical activity [4]. Older adults who retire need time to restructure

[1] Corresponding Author: Markus Garschall, AIT Austrian Institute of Technology, Center for Technology Experience, Vienna, Austria, E-Mail: markus.garschall@ait.ac.at

their daily activities and routines and establish new social contacts to stay healthy and maintain their own well-being.

According to the WHO constitution, health and well-being does not only encompass physical aspects (absence of illness) but needs to be understood as a holistic concept including measures to promote health (e.g., physical activities to stay active and remain independent) but also measures that address mental and social aspects of life (e.g., meaningful activities, social activities).

There are various approaches to support older adults' health and well-being, ranging from smart watches to context aware lifestyle interventions, [5], or robotic systems [6, 7]. Whereas smart watches, such as Fitbit, basically allow to monitor physical activities, health records, or sleeping habits [8], other systems address different dimensions of health and well-being e.g., social, physical, mental, emotional, and environmental factors [5]. Moreover, there is an increasing awareness of the potential of robotic systems to support older adults to stay independent and increase well-being in older age [9]. Although there is a general interest in improving health and well-being for older adults, only a few digital health interventions exist that target the specific phase of retirement [10] and studies on this topic are still scarce [11].

2. Study design

In this study the prototype of the digital coach ProSelf was tested in a field trial over a period of eight-weeks. We aim at answering the following research question: How can a digital coach support older adults in the transition from work to retirement to establish and maintain a healthy lifestyle? The prototype provides for various activities in the areas physical well-being, mental well-being, social well-being, and retirement. These activities are tailored to personal objectives that can be set within the app (see Figure 1). The target group consisted of healthy individuals aged 55 years or older, either recently retired or in the final three years before retirement to whom a healthy lifestyle is important and who would like to change some habits in their lives. They were invited to participate in the study through e-mail, newsletters, and social media posts. The initial sample consisted of 21 Austrian and 19 Flemish participants of whom 32 completed all research activities[2] Data was collected between October 2022 and February 2023. All participants provided informed consent.

2.1. Coaching Approach

The ProSelf app aims to support users to practice a healthy lifestyle with personalized activities tailored to their individual objectives. The app is available on iOS, Android and as a web application (https://proself.org). The program starts with a query of the user's current level of well-being addressing three different areas: physical, social, and mental well-being. Based on the results of the query and personal goals, individualized activities addressing all three areas and the additional topic retirement, are suggested (for an

[2] Drop-outs were due to not meeting expectations (1), dissatisfaction with the overall idea of the app and programming (2), loss of interest in the study (4), health problems (2), and too high effort (1).

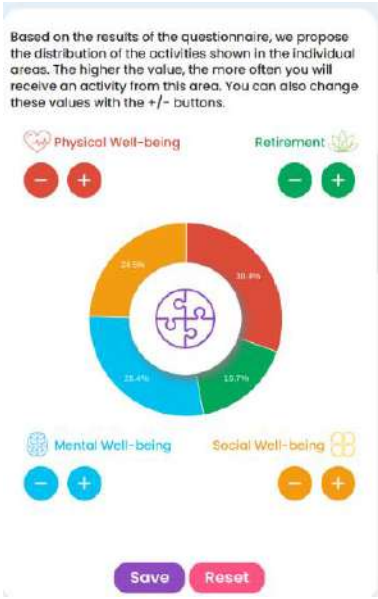


Figure 1: Screenshot of the app allowing users to set their goals in the four different areas.

example activity see Figure 2). The exercises intend to motivate participants to acquire positive habits, e.g., by doing more physical or social activities, or deal with the topic of retirement. Currently, the prototype offers 55 distinct activities. Participants have the possibility to adjust the distribution of activities within the different areas (see Figure 1). Additionally, the app allows you to connect with other users who are members of the community and to exchange personal messages. It is possible to see users who live in the immediate vicinity in order to promote physical contact with these users. Moreover, users can set up and post events that other users can join, e.g., a visit to the theatre or a joint walk. They can also connect with others based on similar interests.

Figure 2: Example activity in the area mental well-being

To support the target group to adopt a healthy lifestyle frequently employed persuasive strategies used in mobile health applications were applied, such as reminders (e.g., to finish activities that have not been accomplished yet), or self-monitoring (e.g., to allow users to keep track of their progress).

2.2. Data collection and analysis

Participants were actively involved in several research activities using a mixed methods approach. To assess participants' well-being standardized questionnaires were applied, i.e., the SF-36v2 (personal well-being), the Lubben Scale [13] (social life) and the GSE-6 [14], (adaptation to stress and chronic illness). The UEQ [15] and the SUS [16] were used to assess user experience and usability. The questionnaires were created online using LimeSurvey and Qualtics. Qualitative methods encompassed co-creation sessions and interviews in the beginning of the study (T0), telephone interviews after four weeks (T1) and qualitative interviews and co-creation at the end of the study (T2). These methods allowed us to gain background information into how and through which activities participants' well-being changed (or not) and why.

The co-creation session in Flanders and interviews in Austria were conducted to obtain insights into the challenges and experiences of the new life phase of retirement. Participants received their dummy profile for using the digital coach and filled out the standardized questionnaires. They were asked to use the app for eight weeks. After four weeks, participants were contacted for an intermediate evaluation (T1), after the test period (T2), they completed the online questionnaires for a third time to assess potential changes in wellbeing, social life, and adaptation to stress and chronic illness. Finally, interviews were conducted to discuss the extent to which the digital coach supported participants in leading a healthy lifestyle, e.g. what kind of activities they found valuable (or not), whether they were able to adopt new healthy habits or socialize. Barriers, motivators and improvements to the design were discussed in co-creation sessions.

For the qualitative data analysis all interviews, notes from the telephone interviews and content of the co-creation sessions were transcribed and a thematic analysis was conducted. As two organizations from two countries were involved in conducting the study and data analysis, a joint Miro Board was set up to work together in an iterative analysis process. Firstly, key themes were identified and discussed. The transcripts and notes were then reviewed again by both organizations and categories were reorganized and refined in sub-themes. For the quantitative analysis the online questionnaires were analyzed using SPSS. Descriptive statistical analyses were applied. In this article, we focus on presenting the qualitative findings and discussing the most important results.

3. Findings

The data analysis revealed four key themes, that already reflect the potential of the mobile coaching app and its main pitfall: reflection, action, confirmation, and transition.

3.1. Reflection.

Besides physical activities, the app provides a couple of mental activities and aims at encouraging participants to become aware of different aspects in life, e.g., being thankful,

seeing things positive, or reflecting upon memories. Positive feedback could be identified towards mental activities, raising awareness for the good things in life. During the interviews, users were enthusiastic about the activities that were suggested. It made them, for example, stop and become *"more conscious to live a healthy lifestyle"* (A12). Some participants became aware that they have *"never reflected upon themselves"* (B5) and that the app supported them to *"stop and reflect and see what I have accomplished so far"* (A2). A lot of the participants appreciated and valued those mental activities. One participant, for example pointed out *"The emotional aspect is a big plus."* (B14), illustrating his/her appreciation for these activities. Only one participant indicated that the exercise *"reflection upon memories"* (A15) made him/her sad because s/he realized that s/he was not that fit anymore.

3.2. Action.

The theme action encompasses a variety of aspects that supported users to become active on different levels (e.g., physically, socially). Sub-topics identified in this context were, e.g., stimulation, engagement, motivation, curiosity, easiness, appreciation, or consciousness of benefits. The daily activities suggested by the app were perceived as appealing, stimulating, engaging, and motivating. At the beginning of the study, participants indicated that they were curious and that they liked the activities and tips. One participant commented as follows: *"My first activity was rope jumping. I have been doing that every day since then and it is energizing."* (B14) Unfortunately, this curiosity and motivation waned over time, mainly due to the limited number of activities currently available in both languages. Participants appreciated the easiness of the activities. Most of the suggestions were "*doable for everyone"* (A15) (e.g., wall sit ups) and were "*easy to integrate in daily life"* (A12). This ease proved to be a pitfall during the course of the study for those participants who were already quite active and did a lot of sport and physical exercise. One of them, for example, was disappointed that s/he could not adapt the activities to his/her personal fitness level.

Users had a rather ambivalent impression of the social activities. On the one hand, the opportunity to socialize with others was perceived positive, e.g., *"It is a great opportunity to be in contact ... when I was ill and at home, I hardly had any contacts."* (A2). On the other hand, the interviews also revealed that not all participants experienced a need to connect with others via the app: *"For other social contacts I do not have any time."* (A15) *"The threshold is too high to connect."* (B16), "*I was not interested in connecting with others - that was too effortful for me."* (A11). Moreover, some participants pointed out that they did not like to connect with strangers and prefer face-to-face contacts: *"I do not like to write with others I do not know."* (A14), *"I do not like the contact via technology."* (B14).

3.3. Confirmation.

Another interesting finding was that the app did not only encourage participants to take action (becoming active) but also confirmed that they were already making a good contribution to their personal health and well-being. Some participants were motivated because they became aware of all the positive things, they were already doing which was expressed in the following quotes: *"The app confirmed that I am on track to a healthy*

lifestyle." (A20), *"It is somehow a confirmation that I am on a good way because of good nutrition and daily exercises."* (A15).

3.4. Transition.

Even though the app encouraged participants to take action (e.g., to perform physical activities) and has potential to raise curiosity among participants and increase awareness on mental well-being (e.g., being more conscious to live a healthy lifestyle or reflecting upon important milestones that have already been reached), one important finding is that they lost interest throughout the course of the study. This can be traced back to the fact that the prototype only provided a limited number of activities, a lack of personalization and visible evolution mainly regarding physical activities. For example, a lot of participants raised the need to further personalize the physical activities to specific needs. As one user argues that "The a*ctivities do not fit my personal situation as I am not yet retired.*" (A6). One user who regularly did physical training raised the need for more advanced activities to raise his/her fitness level and points out that *"Physical activity did not increase in terms of level of difficulty.*" (A15). Others reported specific requirements due to health issues (e.g., age-related or accident-caused limitations) and experienced restrictions when performing certain physical activities. These results indicate a need to further elaborate on the persuasive strategies that are currently employed.

4. Discussion

In this study we investigated how a mobile application can support older adults' well-being in the transition from work to retirement. The results illustrate that the holistic approach can make a valuable contribution to the well-being of the target group, i.e., supporting them in various areas. In particular, the mental activities proofed to be valuable and supported older adults beyond training and physical activities, which was appreciated by all participants who completed the trial. Hence, this is an area that will be further developed and refined. The results also show that the digital coach has potential to encourage older adults to become and/or stay active. Participants felt engaged, motivated, and were curious, especially in the beginning of the study. In this context, personalization (e.g., adapting the level of difficulty of activities) plays a vital role and will be further developed in the future. While personalization plays a significant role in increasing one's fitness level and maintaining challenges, the results show that confirmation could play a significant role to motivate users, who are not so active now but have the goal of becoming active. Hence, activities which can be easily accomplished or that participants are already doing might be as important as challenging ones.

The present work needs to be considered in the light of variety of limitations. First, the number of activities (55) was quite small, considering the period of eight weeks. Due to limited resources for the translation of activities, it was not possible to provide additional tasks. Participants, who were quite active at the beginning gradually lost interest, indicating that they missed new and challenging activities. The prototype did allow for personalization in the different areas of well-being, i.e., participants could choose in which areas they would like to receive activities and could set personal goals (e.g., social, mental). However, the prototype did not allow to adjust intensity levels of the activities based on one's personal fitness level.

As a next step, it is planned to increase the number of activities, available for potential users and increase the personalization options to better address the personal needs of potential end users in the future and support sustainable behavioural change.

5. Acknowledgements

We would like to thank Salumentis OG (https://www.salumentis.at), who provided the app for the study and gave technical support during the course of the field trial. This research was done in the context of the VITALSE project (https://vitalise-project.eu). The financial support from European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 101007990 is gratefully acknowledged.

References

[1] Eurostat 2020, Aeging Europe – statistics on population developments. , *https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Ageing_Europe_-_statistics_on_population_developments*.

[2] R. Pocinho et al., Study of the perception of active workers in preparation for retirement, *Proceedings of the Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality* (2016), 405–409.

[3] T. Knearem et al., Constructing Well-being Together: Older Adults Engagement in Coproduction Through Volunteering, *C&T '21: Proceedings of the 10th International Conference on Communities & Technologies - Wicked Problems in the Age of Tech* (2021), 169–178.

[4] P. Eibich, Understanding the effect of retirement on health: Mechanisms and heterogeneity, *Journal of Health Economics* **43** (2015), 1–12.

[5] A. Seiderer et al., Exploring Digital Image Frames for Lifestyle Intervention to Improve Well-being of Older Adults, *Proceedings of the 5th International Conference on Digital Health 2015* (2015), 71–78.

[6] V. N. Antony et al., *Co-designing with older adults, for older adults: Robots to promote physical activity*, **1**(1). Association for Computing Machinery, (2023).

[7] S. A. McGlynn et al., PARO's Stress-Reduction Potential for Older Adults, *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* **60**(1) (2016), 1799–1803.

[8] S. Farivar et al., Wearable device adoption among older adults: A mixed-methods study, *International Journal of Information Management* **55** (2020), 102209.

[9] I. Dratsiou et al., Assistive Technologies for Supporting Wellbeing of Older Adults, *Proceedings of the 14th PErvasive Technologies Related to Assistive Environments Conference* (2021), 338–341

[10] V. Stara et al., Digital Health Coaching Programs Among Older Employees in Transition to Retirement: Systematic Literature Review, *Journal of Medical Internet Research* **22**(9) (2020), e17809.

[11] S. Santini et al., A Digital Coach Promoting Healthy Aging among Older Adults in Transition to Retirement: Results from a Qualitative Study in Italy, *Sustainability* **12**(18) (2020), 7400.

[12] M. Morfeld and M. Bullinger, Der SF-36 Health Survey zur Erhebung und Dokumentation gesundheitsbezogener Lebensqualität, *Physikalische Medizin, Rehabilitationsmedizin, Kurortmedizin* **18**(05) (2008), 250–255.

[13] J. Lubben and M. Gironda, Measuring social networks and assessing their benefits, *Social networks and social exclusion: Sociological and policy perspectives* **20**(04) (2004), 14–20.

[14] M. Romppel et al., A short form of the General Self-Efficacy Scale (GSE-6): Development, psychometric properties and validity in an intercultural non-clinical sample and a sample of patients at risk for heart failure, *GMS Psycho-Social-Medicine* **10** (2013).

[15] B. Laugwitz et al., Construction and evaluation of a user experience questionnaire, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **5298 LNCS** (2008), 63–76.

[16] J. Brooke, Sus: a "quick and dirty" usability, *Usability evaluation in industry* **189**(3) (1996), 189–194.

dHealth 2024
D. Hayn et al. (Eds.)

doi:10.3233/SHTI240009

Symbiosis of Technology and Ethics: Preliminary Results of an Inquiry into the Moral Dimensions in the Use of Robotic Systems in Patient Care

Carolin MIRBETH[a,1], Christoph OHNEBERG[a] and Inge EBERL[a]
[a] *Catholic University of Eichstätt-Ingolstadt, Germany*

Abstract. The present study aims to describe ethical and social requirements for technical and robotic systems for caregiving from the perspective of users. Users are interviewed in the ReduSys project during the development phase (prospective viewpoint) and after technology testing in the clinical setting (retrospective viewpoint). The preliminary results presented here refer to the prospective viewpoint.

Keywords. Assistive Robotics, Ethical Issues, Nursing Research, Qualitative Research

1. Introduction

Technical systems for vulnerable groups must be oriented towards the needs of the users. In this context, ethical and social requirements from the perspective of users play a crucial role already during the technology development phase [1]. The ReduSys project addresses these challenges and aims to relieve nursing professionals and improve patient care through a multimodal technical system.

2. Methods

The qualitative longitudinal design integrates the Care-Centered Framework and the methodology of Care Centered Value-Sensitive Design (CCVSD). The five dimensions of CCVSD (Context, Practice, Actors involved, Type of robot, and Manifestation of moral elements) were incorporated into the episodic interview guide [2, 3]. The first phase of the study is part of the technology development phase (prospective viewpoint). A total of 23 individuals were interviewed: nursing professionals (n=7), health care professionals (n=6), patients (n=9), and family members (n=1). The data were analyzed using the Thematic Analysis approach by Braun and Clarke [4]. This procedure will be repeated in a second part of the study to elucidate comparisons between the different phases where the retrospective viewpoint is captured.

[1] Corresponding Author: Carolin Mirbeth, Catholic University of Eichstätt-Ingolstadt, Kapuzinergasse 2, 85072 Eichstätt-Ingolstadt, Germany; E-mail: Carolin.Mirbeth@ku.de

3. Results

The preliminary results suggest that participants express concerns regarding a potential loss of interpersonal interaction and empathy in the context of the application of robotic systems, as well as fears of social isolation: "[…] Nursing staff might only come in personally to see the patient twice per shift, and the rest is somehow controlled through the robot. I view this critically if the patient no longer encounters any human presence" (IP6, Pos. 36). Moreover, the participants express the concern "[…] that people could become anxious if a robot suddenly rolls into the room […]" (IP7, Pos. 10). Patient-related factors such as age, illness, and cognitive abilities play a crucial role in this regard. The participants emphasize the importance of considering these factors in the development and application of technical systems: "I work with very severely affected patients, and they need to be awakened partially at first. […] I also don't know if they are just waking up or in a transitional stage, whether they might get scared if a robot speaks to them because patients in rehabilitation clinics are often older and not so tech-savvy" (IP5, Pos. 12). Participants are furthermore of the opinion that robots should only be used as a supplement to nursing staff and must not replace them: "I have to emphasize that it is not a replacement for me but rather serves as a supportive process" (IP10, Pos. 18).

4. Discussion

Ethical and social requirements can vary depending on the technical system and its potential applications. The preliminary results align with normative and societal positions. It remains to be seen how users evaluate the technology after testing in the clinic and what similarities and differences emerge in the assessment. Furthermore, the study provides an opportunity to evaluate and further develop the CCVSD.

Acknowledgements

The project ReduSys (Reduced-contact Care in the Clinical Environment through Multimodal Systems and Robotics.") is funded by the Federal Ministry of Education and Research (identification number 16SV8831).

References

[1] Deutscher Ethikrat, Robotik für gute Pflege: Stellungnahme, https://www.ethikrat.org/publikationen/publikationsdetail/?tx_wwt3shop_detail%5Bproduct%5D=130&tx_wwt3shop_detail%5Baction%5D=index&tx_wwt3shop_detail%5Bcontroller%5D=Products&cHash=61efeb07abf2347f3834e309c5df15b3, last access: 17.01.2024

[2] U. Flick, *Qualitative Sozialforschung: Eine Einführung*. Rohwolt-Taschenbuch-Verlag, Reinbek bei Hamburg, 2017.

[3] A. van Wynsberghe, Designing Robots for Care: Care Centered Value-Sensitive Design. *Sci Eng Ethics*, **19**(2) (2013), 407–433. https://doi.org/10.1007/s11948-011-9343-6

[4] V. Braun, V. Clarke, *Thematic Analysis*: *A practical guide*, SAGE, London, 2022.

dHealth 2024
D. Hayn et al. (Eds.)

doi:10.3233/SHTI240010

Evaluation of a Digital Dementia Registry's IT Architecture After a Three-Year Period in Practice: digiDEM Bayern

Michael ZEILER[a,1], Nikolas DIETZEL[b], Peter L. KOLOMINSKY-RABAS[b], Elmar GRAESSEL[c] and Hans-Ulrich PROKOSCH[a]

[a] *Friedrich-Alexander-Universität Erlangen-Nürnberg, Department of Medical Informatics, Erlangen, Germany*
[b] *Friedrich-Alexander-Universität Erlangen-Nürnberg, Interdisciplinary Center for Health Technology Assessment (HTA) and Public Health (IZPH), Erlangen, Germany*
[c] *University Hospital Erlangen, Friedrich-Alexander-Universität Erlangen-Nürnberg, Center for Health Services Research in Medicine, Department of Psychiatry and Psychotherapy, Erlangen, Germany*

Abstract. Introduction: The project "digiDEM Bayern" aims to set up a registry with long-term follow-up data on people with dementia and their family caregivers. For that purpose an Electronic Data Capture (EDC) system linked with a Participant Management (PM) system has been established. This study evaluates the acceptance and usability of the IT tools supporting all data management processes in order to further improve the system and associated processes. Methods: For this purpose we collected the key numbers of the registry, and used the System Usability Scale (SUS) to evaluate the interactions of the data management systems in a wide area. Results: Thirty-six research partners (RP) and six study team (ST) members completed the anonymous online survey. The EDC system overall reached an average SUS score of 73.42 and the PM system of 77.92. Discussion: The two systems fulfil their required task and, therefore, simplify the work of the RP in the data collection process and of the ST during the data quality checks. Conclusion: Integrating the used systems is therefore recommended for registry studies in other medical areas.

Keywords. Evaluation, IT Architecture, Dementia, Registry, Online Survey

1. Introduction

1.1. Background

Dementia is a widespread disease that currently affects over 55 million people worldwide, with annually almost 10 million new cases. Diagnosis and treatment for people with dementia is going to be one of the biggest challenges for healthcare systems worldwide [1]. Digitalisation offers additional opportunities to improve dementia care and health outcomes research to enhance national health planning [2,3]. Registries are a useful

[1] Corresponding Author: Michael ZEILER, Friedrich-Alexander-Universität Erlangen-Nürnberg, Department of Medical Informatics, Wetterkreuz 15, 91058 Erlangen, Germany; E-mail: mze.m.zeiler@fau.de.

research tool for collecting long-term data. In addition, the registries provide insight into disease-related services [4].

1.2. Data Collection Process

The "Digital Dementia Registry Bavaria – digiDEM Bayern" has the goal of recording the healthcare situation around dementia diseases and the long-term progression of dementia-related illnesses to improve the care situation of people with dementia and their family caregivers. Participants are people with mild cognitive impairment and mild to moderate dementia living at their own homes. Experts from various specialist areas of dementia care have defined the data set, which contains among other things data on sociodemographic, diagnoses, activities of daily living, use of resources, the burden on caregivers, needs of people with dementia. The survey instruments focus on the home environment and include caregivers' questions. Research partners (RP) are dementia care professionals from multiple related institutes, such as care facilities and clinics for people with dementia. A broad network of RPs is necessary to ensure geographical coverage throughout Bavaria. The participants are then questioned in a baseline interview (t0) followed by four follow-up surveys 6, 12, 24, and 36 months later [5]. The RPs collect the necessary data through online or face-to-face interviews. Data can be entered using stationary and mobile devices via the web browser. A permanent internet connection is required for data collection. In digiDEM Bayern, 117 active facilities currently participate as RPs (as of 30.09.2023). The RPs have included 1.068 people with cognitive impairments in the study. The geographical evaluation shows that most of the study area is covered. Each survey is subject to data quality assurance supported by the Electronic Data Capture software. Through timely feedback loops, missing values are added, and implausible values are corrected. The study team (ST) consists of six research fellows responsible for the quality control of the entered data and the administration of the follow-up surveys.

1.3. IT infrastructure

The IT infrastructure provides standardized workflows to support an electronic collection of registry data and monitor every participant's study progress. To effectively provide these services, the IT architecture is divided into two components:

(1) Electronic Data Capture (EDC) Software: digiDEM Bayern uses the EDC system Research Electronic Data Capture (REDCap), a secure Web-based software platform designed to support EDC for studies to collect data.

(2) Participant Management (PM) Software: The registry uses a digital and automated monitoring process with escalation levels for the patient's progress. Therefore, webMODYS (web-based modular control and documentation system) was selected to manage the decentralized RP and participants' identity data and monitor the follow-up surveys' times.

The IT architecture of digiDEM Bayern was developed based on an iterative, stakeholder-oriented process that focuses on joint development regarding requirements and architecture [6]. The success of a registry study depends on two key factors: Recruiting the required number of participants [7] and their subsequent participation in the study [8]. The expected decrease in cognitive abilities and general health of dementia patients throughout this study may lead to additional challenges regarding the questioning format [9]. The RPs use the EDC system to enter the collected data digitally

and the PM system to receive the visit dates of the follow-up surveys and the participants' master data. The RPs once transmit the patient master data electronically. The physical separation of the EDC system (survey data) and the PM system (master data) ensures higher security of the patient master data. The PM system is localised in the hospital's internal network and cannot be accessed externally. The RPs have no access to this system. Automatic reminder e-mails for the follow-up interviews and letters with the participants' master data are sent out. The ST uses the EDC system to check the quality of the data entered and the PM system to administer the follow-up surveys and master data of the study participants. This study aims to evaluate the IT architecture of digiDEM Bayern after a three-year period in practice. The well-known instrument, the System Usability Scale (SUS), was used to evaluate the interactions of the data management systems for the software architecture of a digital dementia register in a wide area.

2. Methods

For this work, corresponding indicators and usage data of the registry's IT architecture were analysed. We also conducted an anonymous online survey across RPs in Bavaria/Germany and the ST. The survey was carried out during a two-month period from the beginning of September to the end of October 2023. Invitations to participate in the survey were sent via e-mail; it included a cover letter describing the study's aim and providing the survey link. The content of the online questionnaire was developed based on scientific literature [10, 11]. It was developed using the online SoSci Survey program (www.soscisurvey.de) and made available online. Participants were asked to provide information about their professional background and use of technology. They gave their subjective assessment (on a scale of 1 to 10) and complete the SUS for the systems. The scale can take values between 0 and 100; the higher the value, the higher the user-friendliness is categorised [10]. Furthermore, the RPs and the ST were asked to name the most significant benefit and the most considerable problem associated with the systems they used.

3. Results

3.1. Registry key numbers

Due to the integration of the EDC system and the PM system, tracking the dates of the follow-up surveys is possible. The PM system leads to reduced administration and is less time-consuming. Suppose an RP is unable to complete a survey. In that case, another RP can be granted access to the contact's specific master data and the corresponding data set in order to complete the survey. This procedure has been accomplished in 59 cases so far. Monitoring and ensuring timely follow-up surveys in the PM system leads to the high data quality of the registry data. Since the start of the project, only 27 of the 779 (3.47%) follow-up surveys have taken place outside the defined periods.

3.2. Feedback from the research partners

146 people were asked to take part in the online survey by email. Thirty-six RPs completed the survey, resulting in a response rate of 25.53%. They were, on average, 51 years old, with an age range of 21 to 66, and predominantly female (91.67%). 16 people (44.44%) are professionally involved in counselling those affected (people with dementia and caregivers), 16 participants (44.44%) work in medical or nursing care, while 4 volunteers in the field of dementia (11.11%) also took part in the survey. They most frequently use modern technology (4.29) and rated their competence in dealing with modern technology as average (3.54), while fear of failure did not play a significant role (2.29). The Data entry via the web interface of the REDCap system achieved a score of 7.88 (out of 10) and an average SUS score of 71.77. The software for managing the follow-up surveys reached a score of 8.50 and an SUS score of 77.28.

The feedback from the research partners who collected the survey data using the software components described was predominantly positive. The most frequently mentioned benefit (16 responses) was the clarity of data entry, directly followed by the immediate transfer of data to the ST and the fact that no paper is required for data collection, with six responses each. Four people mentioned other program features as a benefit. To the open question about the most significant benefit, an RP answered, "Available everywhere and quickly, completeness of data entry is displayed, partial survey with later continuation possible." Another RP mentioned that the real-time input and the simultaneous access by several employees of an institution helped them a lot in the survey process. Some RPs (12 responses) were bothered by the effort required to familiarise themselves with the software or the sometimes cumbersome electronic case report forms. The problem is exacerbated if there is a significant time gap between individual surveys. Nine other RPs reported other technical or structural problems. Four other respondents were bothered by the fact that the device actively required internet access to enter data.

3.3. Feedback from the study team

The ST consists of 6 people. They are, on average, 34 years old, with a wide age range between 27 and 55. Two of the scientists are male, while four are female. They frequently use modern technology (4.67), rated their competence in using modern technology as very high (4.67) and have no fear of failure (2.00). Their subjective assessment of the data entry via the web interface of the REDCap system was 8.83 (out of 10) and an average SUS score of 83.33. The software for managing the follow-up surveys reached a score of 8.50 and an SUS score of 81.76.

The feedback from the ST was unanimous. The PM system can monitor compliance with the deadlines for the follow-up survey in a transparent manner. One scientist noted that eliminating double documentation (first manually on paper, then digitally) means an enormous time advantage for them. The functions built into the EDC system also allow initial data evaluations. For example, this leads to time savings when checking the number of cases for individual scientific evaluations. The EDC system accelerates data quality control significantly. The display of missing or implausible values, such as a daily caregiving time of more than 24 hours. One research fellow named this the most significant benefit: "Checking the software for non-plausible data and inconsistencies as well as the possibility of checking missing values." However, it was also criticized for the system's inability to perform a general and fully automatic data quality check. The

meaningfulness of the data still has to be checked manually by the researcher. Correcting some values is also sometimes time-consuming, as the built-in filters mean that some questions only appear when specific criteria are met.

4. Discussion

The SUS is often used to evaluate the usability of a stand-alone system [12]. In this study, it was used to evaluate the usability of the IT architecture as well as the complete data collection and data management process. As the IT architecture of the digital dementia register consists of several individual systems, the SUS was also used in this study to test their interaction in terms of user-friendliness. For this purpose, both the ST of the university and the RPs who carried out the data entry were surveyed. Bangor et al. describe that products with an SUS score of 90 points and above were rated as exceptional, products rated 80 points as good, and products rated 70 points as acceptable. Anything below 70 points had usability issues that were a cause for concern [14]. The statistically average SUS score (at the 50th percentile) is 68 [10]. That means all SUS assessments are above the average and at least rated as acceptable. The ST rating of the used systems was slightly better and therefore rated as good (EDC-system: 83.33, PM-system: 81.76). However, the RP rating was just above the acceptable limit (EDC-system: 71.77, PM-system: 77.28). In addition to that, the subjective assessment of the EDC system (RP: 7.88, ST: 8.83) and the PM system (RP: 8.5, ST: 8.5) was quite positive. These numbers suggest that the project-used systems were generally perceived positively by the two target groups in their application. However, the RPs report some difficulties. The missing availability of internet connection in most of the participants' homes leads to the problem that permanent internet access is required for data entry. According to the university's data protection regulations, no study data can be stored locally on a private device. This would have led to constantly checking a device for potentially high-risk applications, such as social media platforms. Checking all devices is time-consuming and vulnerable to mistakes; offline data collection and entry on private devices is not sustainable. One potential solution would be equipping researchers with mobile devices limited to preinstalled applications. The need for manual checking of data by the ST poses another problem. For some data entries, automatic filters can be defined that check the entries for plausibility. However, this is only possible in rare cases, as contextual information is necessary for an automatic plausibility check by the system. Another approach is increasing the input quality. For this purpose, the RPs are briefed on frequent entry errors and system-based features in binding regular training and follow-up training courses. Furthermore, the ST offers a monthly meeting to answer questions.

5. Conclusion

This evaluation showed that the feedback was largely positive. The two systems fulfil their required task and, therefore, simplify the work of the research partners in the data collection process and of the study team during the data quality checks. The partially automated process of informing research partners about upcoming follow-up interviews with participants and monitoring their timely implementation leads to accumulating high-quality data. Nevertheless, the software can only solve some problems in study registry data collection processes; process reorganization is sometimes needed.

Integrating an electronic data capture system and a participant management system is therefore recommended for registry studies in other medical areas.

Declarations

Authors' contribution: Conceptualization: MZ, ND, HUP; Formal analysis: MZ; Methodology: MZ, ND, HUP; Project administration: PKR, EG, HUP; Supervision: MZ, ND, HUP; Writing - original draft: MZ; Writing - review and editing: MZ, NDI, HUP.

Acknowledgements: The project is funded by the Bavarian State Ministry of Health, Care and Prevention (funding code: G42d-G8300–2017/1606–83). The present work was performed by Michael Zeiler in (partial) fulfilment of the requirements for obtaining the degree "Dr. rer. biol. hum." at the Friedrich-Alexander-Universität Erlangen-Nürnberg.

Conflict of Interest: All authors state that they have no conflict of interests.

Ethics vote: Ethical approval was not required.

References

[1] World Health Organization. Global Status Report on the Public Health Response to Dementia. 2021. Accessed March 10, 2024 at: https://www.who.int/publications/i/item/9789240033245.
[2] Labrique A, Vasudevan L, Mehl G, et al. Digital health and health systems of the future. Glob Health Sci Pract 2018;6:S1–4.
[3] Ramsey M, Lee EE. Digital Dementia Care for the Future: Opportunities and Challenges. Am J Geriatr Psychiatry. 2022 Jan;30(1):12-14.
[4] Dreyer NA, Garner S. Registries for robust evidence. JAMA 2009; 302(07):790–791
[5] Dietzel N, Kürten L, Karrer L, et al. Digital Dementia Registry Bavaria-digiDEM Bayern: study protocol for a multicentre, prospective, longitudinal register study. BMJ Open 2021;11(02):e043473
[6] Reichold M, Dietzel N, Chmelirsch C, Kolominsky-Rabas PL, Grässel E, Prokosch HU: Designing and implementing an IT architecture for a digital multicenter dementia registry: digiDEM Bayern. Applied clinical informatics 12. 2021: 551–563.
[7] Lacey JV Jr, Savage KE. 50 % Response rates: half-empty, or halffull? Cancer Causes Control 2016;27(06):805–808
[8] Fogel DB. Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: a review. Contemp Clin Trials Commun 2018;11:156–164
[9] Connell CM, Shaw BA, Holmes SB, Foster NL. Caregivers' attitudes toward their family members' participation in Alzheimer disease research: implications for recruitment and retention. Alzheimer Dis Assoc Disord 2001;15(03):137–145
[10] Brooke, J. SUS -- a quick and dirty usability scale, 1996.
[11] Neyer, F. J., Felber, J., & Gebhardt, C. 2016. Kurzskala Technikbereitschaft (TB, technology commitment). Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS). https://doi.org/10.6102/zis244.
[12] Hajesmaeel-Gohari S, Bahaadinbeigy K. The most used questionnaires for evaluating telemedicine services. BMC Med Inform Decis Mak. 2021 Feb 2;21(1):36.
[13] Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. J Usability Stud 2009; 4: 114-123

dHealth 2024
D. Hayn et al. (Eds.)

doi:10.3233/SHTI240011

Dual Implementation Guides in FHIR and CDA

Nikola TANJGA[a,1] and Oliver KRAUSS[b]

[a] *ELGA GmbH, Vienna, Austria*
[b] *Advanced Information Systems and Technology, University of Applied Sciences Upper Austria, Hagenberg, Austria*

Abstract. Background: The Fast Healthcare Interoperability Resources (FHIR) and Clinical Document Architecture (CDA) are standards for the healthcare industry, designed to improve the exchange of health data by interoperability. Both standards are constrained through what are known as Implementation Guides (IG) for specific use. Objectives: Both of these two standards are widely in use and play an important role in the Austrian healthcare system. Concepts existing in CDA and FHIR must be aligned between both standards. Methods: Many existing approaches are presented and discussed, none are fully suited to the needs in Austria. Results: The IG Publisher has already been used for CDA IGs, beside of its intended FHIR support, but never for both in one IG. Even the International Patient Summary (IPS), existing as CDA and FHIR specification, does not solve the needed comparability between these two. Conclusion: As the IG Publisher is widely used and supports CDA, it should be used for Dual Implementation Guides. Further work and extension of IG Publisher is necessary to enhance the readability of the resulting IGs.

Keywords. Dual Implementation Guides, FHIR, CDA

1. Introduction

The current transition between two active health data standards is forcing standard communities such as HL7 Austria to ask themselves how Dual IGs should be built, including the transformation in-between and the balloting of the whole package. We compare approaches and recommend extending the IG Publisher to solve this challenge. This work is based on the results of a workshop held by the ELGA GmbH and Members of HL7 Austria's FHIR Technical Committee.

The Fast Healthcare Interoperability Resources (FHIR) and Clinical Document Architecture (CDA) are standards in the healthcare industry designed to improve the interoperability and exchange of health data. Both standards are constrained through what are known as Implementation Guides (IG) for specific use. An IG is a document, or in the case of FHIR, a website generated from a FHIR resource, that provides guidelines and instructions for the proper implementation and utilization of these health standards. These guides serve as a kind of manual, aiding developers, healthcare institutions and other stakeholders in effectively integrating and using the standards in their systems, with the goal of achieving interoperability between different vendors and domains.

[1] Corresponding Author: Nikola Tanjga, ELGA GmbH, Vienna, Austria, E-Mail: research@tanjga.eu

In the context of the Austrian healthcare system, the use of Dual IGs for CDA and FHIR can offer several important advantages over the next few years. Currently, the central hub of the Austrian healthcare system is the Electronic Health Record (Elektronische Gesundheitsakte - ELGA), which is based on CDA and implemented, maintained and further developed by ELGA GmbH. In recent years, ELGA GmbH has also started to provide FHIR interfaces, such as their self-made open source package TerminoloGit [1] for hosting and maintaining healthcare terminologies [2].

Since both of these two standards are widely in use, the challenge arises that concepts existing in the Austrian healthcare system must be correctly implemented and aligned in both standards. To enhance the acceptance of IGs by Austrian healthcare service providers (Gesundheitsdiensteanbieter - GDA) and other stakeholders, alignment between the two standards is necessary. Currently, this alignment is achieved by HL7 Austria [3] closely adhering to the existing CDA guidelines when creating new FHIR IGs [4].

At least for the next few years, during which both standards are actively being used in Austria, there is a need for so-called Dual Implementation Guides (IGs). The goal is to closely align the definitions in the standard and offer parallel interfaces providing healthcare data exchange via CDA or FHIR. For example, such IGs would enable the description of a guide for a specific healthcare use case, such as laboratory results or medication data, simultaneously for both standards.

As the active development of IGs in Austria requires a Ballot Cycle, another necessary consideration is how comments and issues in the IGs can be effectively reported and discussed. Currently, this is being done via shared documents for which comments are captured in Excel sheets, creating a significant workload of manual triage, duplicate alignment and discussion.

Particular considerations must be taken into account for Dual Implementation Guides (IGs), especially regarding the following questions:

- Which tools can be used to create Dual IGs?
- How can FHIR and CDA be presented alongside each other in the IG?
- How can mappings between the CDA and FHIR standard be defined?
- How can the content of the IG be balloted?

2. Methods

This section gives an overview of currently available methods, as a basis for possible answers to the questions and challenges regarding Dual IGs.

2.1. Which tools are being used to create (Dual) IGs?

There are closed source tools such as Simplifier and open source tools like ART-DECOR and IG Publisher, that can be used to create IGs in general.

Simplifier [5] is a product by Firely, with additional online hosting and a GUI in the form of Forge to create, validate and document FHIR IGs, but do not support CDAs. Its primary advantage over competitors is the cloud hosting that is available for all profiles per default.

ART-DECOR [6] is used to create and maintain HL7 templates, value sets, scenarios and data sets. It supports the creation of implementation guides that adhere to the core principles of using CDA R2. These principles include understanding the use of data types

in HL7 V3, the use of XML, the structure of a CDA document and more. It currently has a limited support of FHIR resources [7], possibly enabling Dual IG work.

Examples of outputs of such tools are the Implementation Guides by the Kassenärztliche Bundesvereinigung (KBV) [8], using Simplifier and a extra website and the ELGA CDA implementation guides [9], using ART-DECOR and a Wiki. This highlights a glaring issue in the current processes for implementation guide definition. The CDA standards suite has no sensible way of creating IGs for the standard and even tools such as Art Decor must be supplemented with other options, such as the Wiki HL7 Austria is using to create the Narrative for their IGs [10]. FHIR itself seems to have issues too, as KBV has also not chosen to use the FHIR IG resource, which was created to address this is- sue in the standard. While they have defined FHIR profiles in Simplifier, the narrative is hosted on a website [8] and most pages have an option to add comments for issues in the IG. Additionally, an overview page allows viewing all comments together.

IG Publisher [11] is the official tool by HL7 to create FHIR IGs and the FHIR standard itself is also created and published via the tool. In addition, SUSHI is a simple language to create FHIR Profiles, Extensions and Examples for IGs and directly integrates in IG Publisher. It also offers support for CDA, [12] with C-CDA being published as a full IG [13].

None of the aforementioned tools that currently exist support the creation of Dual IGs. Currently, Lantana is working on Dual IGs [14], which considers a separate process for the FHIR and CDA standards and creating mappings and transformations between them. The result would be two separate packages.

The Dental Data Exchange IG is another example of a currently existing IG which considers CDA and FHIR and goes the route of applying C-CDA on FHIR [15], e.g. using a Composition resource in FHIR to enable using FHIR in the context of documents. This option would not be considerable as a Dual IG but rather only uses the FHIR standard, to achieve its goals.

An example of an existing Dual IG is the International Patient Summary (IPS) [16]. The IPS is written in both CDA and FHIR and the two versions are closely aligned, however they are still two separate IGs which are written separate from each other in ART-DECOR and in IG Publisher.

2.2. How are FHIR and CDA presented alongside each other in the IG?

This is currently not done at all. Not even the IGs that exist in CDA and FHIR have any direct alignment whatsoever.

2.3. How are mappings between the CDA and FHIR standard currently defined?

There are two ways to create mappings. The simplest way of mapping from CDA to FHIR and vice versa, is in the implementation guide, in the Mappings section of an artifact (profile, extension, ...) [17]. This represents a simple mapping as free text, which is usually enough for implementers to understand the differences in the two standards.

The more complex mappings are possible via the *StructureMap* and *ConceptMap* resources, which can be used in the authoring friendly format FHIR Mapping Language (FML) to create unidirectional mappings. Using these resources in FHIR would allow automated translation between the CDA documents and FHIR resources. The advantage of such an approach would in addition to automated translation also be a definite-alignment between the two standards in the context of a Dual IG, since the considerations

must be made to enable automation. Executing a FML is possible by some open source tools written in Java, Pascal, JavaScript and C# [18]. Even the Java FHIR validator, which is well known in the FHIR community, is able to perform the transform operation. FML got a big upswing with the publication of the FHIR R5 specification, that implements a conversion for all resources to R4 as well as R4B and vice versa [19]. The disadvantage being that creating FML present a significant overhead in IG creation and that effort would double if a bidirectional mapping was required.

2.4. How is the content of the IG balloted?

Currently, balloting is done in one of three ways, balloting with web-support, balloting via issue tracking tools and balloting through Excel sheets.

Examples for balloting with web-support are the KBV FHIR profiles as mentioned above, as well as currently ongoing SNOMED-CT German translation project [20]. In the case of KBV, the comments can also be publicly viewed.

Issue tracking is the official process that HL7 International follows when developing their standards, such as FHIR and CDA. The advantage of this is, that there is less redundancy since ballot participants can actively search for issue duplicates and issues are public and can be commented by other participants to enable discussion. The disadvantage of this approach is that it may be less accessible to non-technical users.

The final method, currently used by HL7 Austria, to enable participation from non-technical users is Microsoft Excel. A ballot is announced via email, to which participants can respond with Excel sheets, filled with their comments or concerns. This approach can be used for any type of ballot and is a very common misuse of Excel for business cases it was not designed for. The primary disadvantages of this approach are that they prevent discussion and pose a severe overhead of sifting through and aligning comments.

3. Results

The following are recommendation based on the need of a Dual IG that supports both CDA and FHIR and the available state of the art.

3.1. Which tools can be used to create Dual IGs?

The best way to go with a Dual IG is most likely IG Publisher. It has a large base of developers and implementers that are actually using the toolchain, including HL7 Austria. ART-DECOR is open source and could be used in a similar manner, however it does not support complete IG create for CDAs or FHIR as IG Publisher already does.

IG Publisher has base templates for both CDA and FHIR. A necessary tooling step to enable Dual IGs would firstly be to combine both templates to support both the CDA-XML based generation process and the FHIR generation process based on *StructureDefinitions* (machine readable definitions of FHIR resource types), enabling the IG Publisher toolchain to build artifacts for both standards at the same time.

Some approaches have been made in a working group from the FHIR technical committee HL7 Austria, by testing the usage of the FHIR logical model. By adapting an existing FHIR logical model for C-CDA, a logical model analog to the Austrian extended Schema could be created and used as a basis for FML transformation on CDA. As a next test step, SUSHI will be applied for profile creation, to fit specific Austrian use cases.

3.2. How can FHIR and CDA be presented alongside each other in the IG?

There is no current solution for presenting FHIR and CDA alongside each other. When following the IG Publisher approach, this would require more severe changes to the publishing process, as IG Publisher was designed to present artifacts individually. As the narrative between overlapping FHIR and CDA artifacts is the same and only the structural views are duplicate, this would require modifying IG Publisher to allow presenting two artifacts on the same page. The presentation itself via a HTML view is no issue.

3.3. How can mappings between the CDA and FHIR standard be defined?

Mappings between CDA and FHIR could be similar to the current way of the mapping table in FHIR. The presentation could be improved to not only use the mapping tab, but rather extend the visualization to draw lines between the two artifacts. However, this would require following a syntax instead of free text and possibly even require the full use of the *StructureMap* resource for mappings.

Similar to the FHIR R5 IG an own conversion tab per resource would give commenters a full insight and show developers how Dual IGs could be implemented.

3.4. How can the content of the IG be balloted?

The process by HL7 International to use issue tracking has clear advantages, which also align with HL7s values of open development and discourse.

The disadvantage for non-technical users exists and could be remedied similar to the proprietary solutions that exist. Comment-plugins in the IG website could be used to automatically create issues in an issue tracker and maybe even show already existing issues. This enables discourse and eases the balloting process as the comments are made at the point of issue. An open challenge with this balloting process is, that the underlying technical resources are separate from the rendered HTML, thus preventing directly referencing the underlying sources that define the CDA or FHIR profiles or extensions.

4. Discussion

The final recommendation of this work is to use IG Publisher to create Dual IGs based on the existing templates for both standards, which should be combined. This would be a relatively low-effort first step to create IGs which contain both standards.

Further topics such as mappings between the individual artifacts and side-by-side visualizations, as well as better balloting support via plugins to connect to issue trackers are all also recommended, but can be worked on in subsequent steps. These would not only provide advantages on Dual IGs for CDA and FHIR, but also open the possibility to create comparative views between different IG versions of the same topic or different FHIR major versions or even different standards themselves, other than CDA and FHIR.

Acknowledgements

The authors want to thank the international and Austrian HL7 community for sharing their wisdom with us and for their willingness to engage in various thought experiments, especially the other ELGA workshop participants Reinhard Egelkraut, Emmanuel Helm, Gabriel Kleinoscheg and Andreas Schuler.

References

[1] G. Kleinoscheg, N. Tanjga, N. Svec, S. Rainer-Sablatnig and S. Sabutsch, *TerminoloGit – An Open Source Terminology Server for Large Scale eHealth Environments*, 2022. doi:10.3233/SHTI220351.
[2] ELGA GmbH, ö. e-Health-Terminologien, last access 2024-01-18. https://termgit.elga.gv.at/.
[3] HL7 Austria, HL7® Austria FHIR® Core IG, last access 2024-01-18. https://fhir.hl7.at/.
[4] R. Egelkraut, B. Traxler, E. Helm, O. Krauss and A. Schuler, *Open Infrastructure for Standardization of HL7® FHIR® Implementation Guides in Austria*, 2022. doi:10.3233/SHTI220372.
[5] Firely, About Simplifier.net, last access 2024-03-14. https://simplifier.net/about.
[6] ART-DECOR Open Tools GmbH, ART-DECOR® – an open-source collaboration tool that supports healthcare interoperability, last access 2024-01-18. https://art-decor.org/.
[7] agile health data information systems, artdecor2ig - scripts to create a FHIR Implementation Guide for an ART-DECOR project, last access 2024-01-18. https://github.com/ahdis/artdecor2ig.
[8] Kassenärztlichen Bundesvereinigung (KBV), KBV-BASIS-PROFILE, last access 2024-01-18. https://mio.kbv.de/display/BASE1X0.
[9] HL7 Austria, Wiki-Portal von HL7 Austria, last access 2024-01-18. https://wiki.hl7.at/index.php?title=ILF:Ambulanzbefund.
[10] S. Sabutsch, M. Frohner, G. Kleinoscheg, A. Klostermann, N. Svec, S. Rainer-Sablatnig and N. Tanjga, *ELGA Outpatient Clinic Report and ELGA Telehealth Note: Two HL7-CDA®-Based Modular Electronic Documents*, 2022. doi:10.3233/SHTI220350.
[11] HL7 International, IG Publisher Documentation, last access 2024-03-14. https://confluence.hl7.org/display/FHIR/IG+Publisher+Documentation.
[12] HL7 International, Template used for most HL7-defined CDA Implementation Guide, last access 2024-01-18. https://github.com/HL7/ig-template-cda.
[13] HL7 International, Consolidated CDA Release 2.2, last access 2024-01-18. https://build.fhir.org/ig/HL7/cda-ccda-2.2/.
[14] Lantana Consulting Group, Dual FHIR/CDA IG Development, last access 2024-01-18. https://www.lantanagroup.com/wp-content/uploads/2019/12/Dual-FHIR-CDA-IG-Development.pdf.
[15] HL7 International and American Dental Association, Dental Data Exchange, last access 2024-01-18. https://build.fhir.org/ig/HL7/dental-data-exchange/index.html.
[16] Joint Initiative Council for Global Health Informatics Standardization, The International Patient Summary, last access 2024-01-18. https://international-patient-summary.net/.
[17] HL7 Austria, HL7® Austria FHIR® Core Austrian Patient Mappings, last access 2024-01-18. https://fhir.hl7.at/r4-core-ballot_2021_01/StructureDefinition-austrian-patient-mappings.html.
[18] HL7 International, Using the FHIR Mapping Language, last access 2024-01-18. https://confluence.hl7.org/pages/viewpage.action?pageId=76158820#UsingtheFHIRMappingLanguage-CodeLibraries.
[19] HL7 International, FHIR StructureMap Conversions, last access 2024-01-18. https://hl7.org/fhir/extensions/conversions-StructureMap.html.
[20] SNOMED-CT German Translation Group, Kommentierung von Übersetzungen, last access 2024-01-18. https://confluence.ihtsdotools.org/pages/viewpage.action?pageId=146867107

dHealth 2024
D. Hayn et al. (Eds.)

doi:10.3233/SHTI240012

DiGA in Primary Care and the Influence of Patient-Related Factors - Results of a Survey of Family Doctors

Anna-Lena BRECHER[a,1], Julia HERRGEN[a] and Tobias MUELLER[a]

[a] *University of Applied Sciences - Faculty of Health Sciences, Giessen, Germany*

Abstract. Background: In 2019, the Digital Healthcare Act created the legal basis for prescribable mobile health applications, referred to as DiGA (in German: Digitale Gesundheitsanwendungen), as a novel healthcare delivery option in Germany [1, 2]. Objectives: The aim of this study is to analyze the use of DiGA in primary care, focusing on the influence of socio-demographic characteristics of family doctors (FDs) and patient-related factors. Methods: Pen-and-paper survey among 97 FDs in the district of Giessen, Hesse, Germany. Results: 59.4% of surveyed FDs have already prescribed DiGA. The age and digital affinity of FDs as well as the location of their practice are significantly correlated with the level of information and willingness to use DiGA. Male und younger FDs rate their digital affinity higher. When deciding whether to prescribe DiGA, 72.9% of surveyed physicians take patient-related factors such as digital affinity, motivation, age and health literacy into consideration. Conclusion: Socio-demographic characteristics of FDs and patient-related factors have an influence on the use of DiGA.

Keywords. digital health, telemedicine, primary health care, general practitioners, health inequality, digital divide

1. Introduction

The healthcare system in Germany is facing major and complex challenges [3, 4]. Digital technologies, such as mobile health applications, are considered key to ensuring high-quality healthcare, reducing the prevalence of chronic diseases through preventive measures, and containing exploding healthcare costs [2, 5]. As part of the Digital Healthcare Act, the legal basis for prescribable mobile health applications as a new healthcare delivery option was established in 2019 [1, 6].

The current state of research shows that awareness and acceptance among physicians has increased significantly since the introduction of DiGA in 2020 [2, 7]. However, the willingness to use DiGA on a long-term basis has been rather low so far [2, 8, 9]. There is initial evidence that younger age and higher digital affinity of physicians as well as the location of their practice in an urban area are associated with a higher willingness to use DiGA [2, 8, 10]. In addition, DiGA usage metrics and current research indicate that patient-related factors such as age, gender, digital affinity, and health literacy correlate

[1] Corresponding Author: Anna-Lena Brecher, University of Applied Sciences – Faculty of Health Sciences, Giessen, Germany, E-Mail: anna-lena.brecher@ges.thm.de

with the extent to which patients use and benefit from mobile health applications [5, 10, 12, 13].

The aim of this study is to analyze the use of DiGA in the context of primary care in the district of Giessen. The focus is on gaining insights into the current awareness, acceptance and use of DiGA among FDs and the influence of socio-demographic factors of the physicians on these aspects. Furthermore, the influence of patient-related factors on the use of DiGA from the perspective of FDs is investigated.

2. Methods

The data collection was conducted through a pen-and-paper survey, to which all FDs in the district of Giessen, Hesse, Germany were invited (n=212) during August and September 2023 [14]. 97 participants responded (response rate=45.8%). The questionnaire used primarily Likert scales (questions 1-8) and contained the following 17 questions (translated from German) [2, 8]:

1. Short text with basic information about DiGA - How much do you agree with the following statement? “I am aware of this information.”
2. How much do you agree with the following statement? “I feel sufficiently informed about DiGA to prescribe them.”
3. How much do you agree with the following statement? “I am in favor of physicians and therapists being able to prescribe DiGA.”
4. How often do your patients you ask about DiGA in general or their prescription?
5. How often do you yourself mention DiGA in general or their prescription when talking to patients?
6. How often do you prescribe DiGA?
7. How likely is it that you will prescribe DiGA in the next twelve months?
8. How much do you agree with the following statement? “When deciding to prescribe DiGA to patients, I consider patient-related factors (age, gender, motivation, digital affinity, health literacy, etc.) in addition to the medical indication.”
9. How significant do you consider the influence of the following patient-related factors on the extent to which patients can benefit from using DiGA?
    a. Age
    b. Gender
    c. Migrant background
    d. Health literacy
    e. Digital affinity
    f. Health status
    g. Motivation/Interest
10. Apart from the medical indication, are there other patient-related factors that, in your opinion, influence the extent to which patients can benefit from the use of DiGA? (open-ended)

Socio-demographic information
1. Gender
2. Age
3. Digital affinity
4. Medical specialty

5. Practice location – city/town population
6. Type of medical practice – individual or group
7. Number of patients per quarter

The responses were statistically analyzed using IBM SPSS 29.0. The Spearman rank correlation (α=5%) was used for the determination of statistical correlations.

3. Results

		n	%
Gender	Male	48	50.5
	Female	46	48.4
	Diverse	1	1.0
Age	<35 years	2	2.1
	35-49 years	29	30.5
	50-64 years	52	53.6
	≥65 years	12	12.6
Specialty	General medicine	80	83.3
	Internal medicine	10	10.4
	Multiple specialties	6	6.3
Practice location	<5,000 inhabitants	7	7.5
	5,000-50,000 inhabitants	61	65.6
	50,001-100,000 inhabitants	25	26.9
Type of medical practice	Group practice	68	73.1
	Individual practice	25	26.9
Number of patients per quarter	<1,000	11	12.1
	1,000-2,000	40	44.0
	>2,000	40	44.0

Note: Not all participants answered all the questions, which explains missing values. Due to rounding, the percentages may not add up to 100%.

Table 1. Socio-demographic and practice characteristics of the surveyed FDs.

The socio-demographic and practice characteristics of the 97 FDs participating in the study are shown in Table 1.

On a scale of 1 to 10, FDs rated their digital affinity on average as 6.8 (median=7.5, SD=2.4). The subjective digital affinity is significantly correlated with both age (ρ=-.339, p<.001) and gender (ρ=-.241, p=.019) of the respondents, meaning younger and/or male FDs rate their digital affinity higher (see Figure 1 and 2).

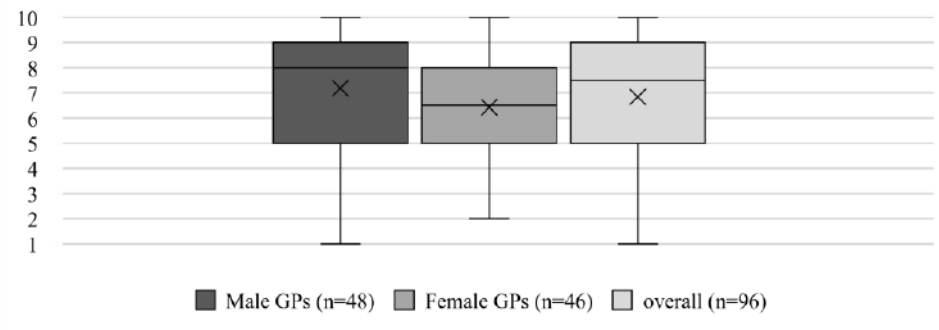


Figure 1. Self-rated digital affinity of FDs by gender

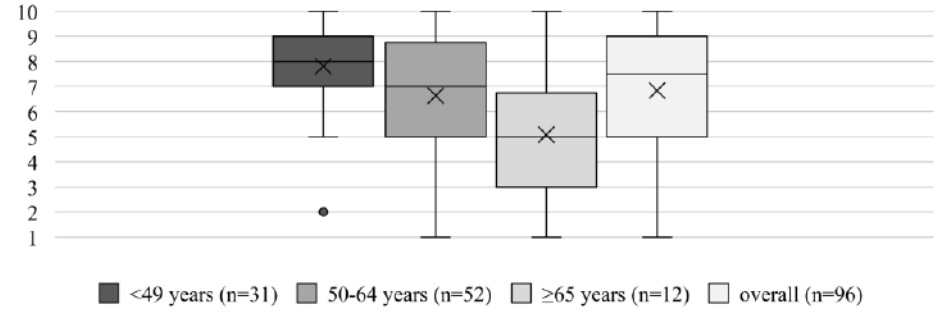
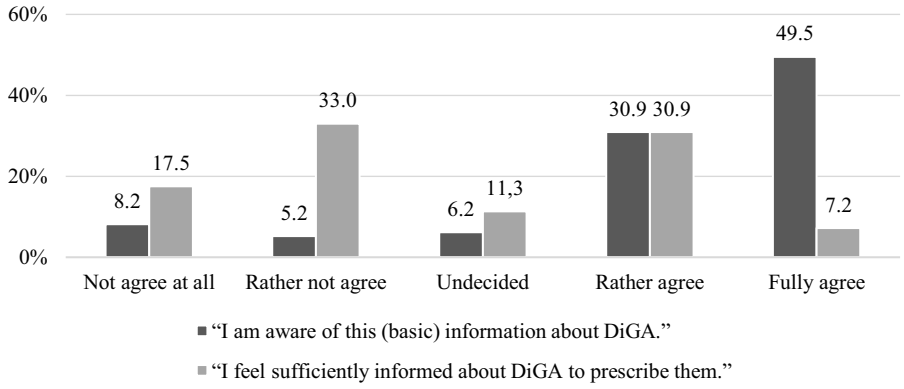


Figure 2. Self-rated digital affinity of FDs by age

Most FDs surveyed (80.4%, n=78) are aware of DiGA. However, only 46.7% (n=15) of respondents with a self-rated low digital affinity (rating 1-4) state that they are familiar with DiGA, while this proportion is 89.6% (n=48) among doctors with a high digital affinity (rating 8-10). Only 38.1% (n=37) of FDs feel sufficiently informed to prescribe them (see Figure 3). This level of information is significantly correlated with the willingness of FDs to prescribe DiGA (so far: ρ=.574, p<.001; future: ρ=.405, p<.001). About two thirds of the FDs (61.1%, n=59) in the district of Giessen have a positive attitude towards prescribable mobile health applications. 40.6% (n=39) of the FDs stated that they had never prescribed a DiGA. Those who have already used DiGA usually prescribe it less than once a month (49.0%, n=47). 38.5% (n=37) of the FDs in the district of Giessen consider it (very) likely that they will prescribe DiGA within the next twelve months.



Note: Due to rounding, the percentages may not add up to 100%.

Figure 3. Basic knowledge and sufficient information about DiGA.

There is a significant correlation between socio-demographic characteristics of the FDs and their awareness of DiGA, as well as their willingness to use them. Specifically, younger physicians tend to have a more positive attitude towards prescribable mobile health applications (ρ=-.240, p=.010), mention DiGA more frequently in conversations with patients (ρ=-.236, p=.011), and have a higher willingness to prescribe them both so

far (ρ=-.230, p=.013) and in the future (ρ=-.215, p=.019). A subjectively higher level of digital affinity is associated with better information (ρ=.212, p=.019) and a more positive attitude towards prescribable mobile health applications (ρ=.217, p=.018) but does not significantly affect the practical use of DiGA (so far: ρ=.156, p=.065; future: ρ=.146, p=.078). Furthermore, FDs working in a practice located in the city of Giessen are better informed about DiGA than their colleagues from the surrounding smaller towns (ρ=-.202, p=.026) and also mention them more frequently when advising patients (ρ=-.208, p=.047). However, their attitude towards prescribable mobile health applications (ρ=-.039, p=.358) is not significantly associated with the location of their practice.

About three quarters of the physicians surveyed (72.9%, n=70) consider sociodemographic and individual patient-related factors when deciding whether to prescribe a DiGA. According to the FDs, especially a patient's digital affinity, motivation or interest, age and health literacy significantly influence the extent to which they can benefit from the use of DiGA. Whereas, migration background, patients health status and gender have less influence (see Figure 4). The open-ended question (No. 10) identified compliance, language and cognitive skills, physical condition, and time and technology resources of patients as additional influencing factors.

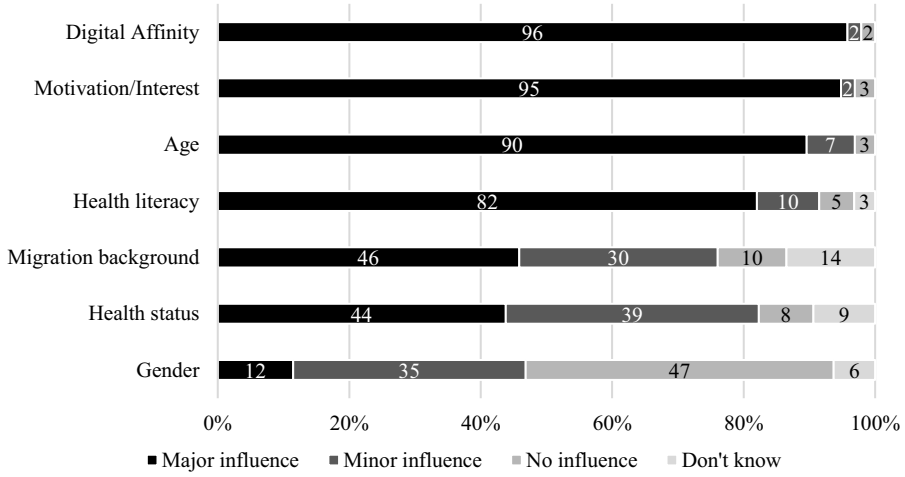


Figure 4. Influence of patient-related factors.

4. Discussion

Due to the regional limitation of the survey, the validity of these results for other German regions and Germany as a whole is limited. Nevertheless, they can provide initial insights and indicate future research needs.

Although many FDs in the district of Giessen are aware of and generally in favor of prescribable mobile health applications, the proportion of physicians who feel sufficiently informed to advise and prescribe DiGA to their patients is still low. The level of information and the willingness to use DiGA are correlated with the age and digital affinity of the physicians as well as the degree of urbanization of the practice environment. This is in line with current evidence [8]. To encourage the broader adoption of DiGA in primary care, it is recommended to provide FDs, especially those with a

lower digital affinity, with demand-oriented and reliable sources of information as well as measures to increase their digital affinity [2, 8]. Possible approaches are e.g. specific information on DiGA from medical societies, continuing education events and the provision of test accounts by the manufacturers [15].

Based on our survey, patient-related factors such as digital affinity, age and health literacy strongly influence FDs prescribing behavior related to DiGA. These factors are in line with those known from patient studies regarding their willingness to use DiGA [16]. However, a well-founded assessment of digital affinity and health literacy according to scientific criteria is time-consuming, and the level perceived by FDs during a brief consultation could differ from the actual level. It could also be assumed that age still plays a strong role as a predictor but also as a surrogate parameter for digital affinity. This should be investigated further. In practice differing health literacy and digital affinity among patients should be considered and promoted at all levels of implementing digital healthcare services [2, 5, 17, 18].

Fortunately, gender, migration background and health status are given little influence by many FDs. This means that there is possibly little social discrimination in terms of access to DiGA and ensures that the benefits of DiGA are accessible to all population groups and further promote equal health opportunities [2, 5].

References

[1] Bundesministerium für Gesundheit, Ärzte sollen Apps verschreiben können, https://www.bundesgesundheitsministerium.de/digitale-versorgung-gesetz.html, last access: 21.1.2024.

[2] F. Dahlhausen, M. Zinner, L. Bieske, J.P. Ehlers, P. Boehme, L. Fehring, Physicians' Attitudes Toward Prescribable mHealth Apps and Implications for Adoption in Germany: Mixed Methods Study, *JMIR Mhealth Uhealth* **9** (2021), e33012.

[3] T. Gerlinger, Baustelle Gesundheitssystem: Aktuelle Herausforderungen in der Gesundheitspolitik, https://www.bpb.de/shop/zeitschriften/apuz/270312/baustelle-gesundheitssystem/, last access: 21.01.2023

[4] A.C. Saß, T. Lampert, F. Prütz, S. Seeling, A. Starker, L.E. Kroll, A. Rommel, L. Ryl, T. Ziese, *Gesundheit in Deutschland*, in: Gesundheitsberichterstattung des Bundes. Robert-Koch-Institut, Berlin, 2015.

[5] A.C. Müller, B. Wachtler, T. Lampert, Digital Divide – Soziale Unterschiede in der Nutzung digitaler Gesundheitsangebote, *Bundesgesundheitsbl* **63**(2) (2020), 185–191.

[6] Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM). *Das Fast-Track-Verfahren für digitale Gesundheitsanwendungen (DiGA) nach § 139e SGB V: Ein Leitfaden für Hersteller, Leistungserbringer und Anwender*, Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM), Bonn & Köln, 2023.

[7] C. Jacob, A. Sanchez-Vazquez, C. Ivory, Social, Organizational, and Technological Factors Impacting Clinicians' Adoption of Mobile Health Tools: Systematic Literature Review, *JMIR Mhealth Uhealth* **8**(2) (2020), e15935.

[8] J. Wangler, M. Jansky, Welche Potenziale und Mehrwerte bieten DiGA für die hausärztliche Versorgung? – Ergebnisse einer Befragung von Hausärzt*innen in Deutschland. *Bundesgesundheitsbl* **65**(12) (2022), 1334–43.

[9] F. Dahlhausen, M. Zinner, L. Bieske, J.P. Ehlers, P. Boehme, L. Fehring, There's an app for that, but nobody's using it: Insights on improving patient access and adherence to digital therapeutics in Germany, *Digit Health* **8** (2022), 1–12.

[10] T. Lampert, M. Niels, S. Müters, B. Wachtler, J. Hoebel, Robert Koch-Institut, *Gesundheitliche Ungleichheit* in: Datenreport 2021: Ein Sozialbericht für die Bundesrepublik Deutschland. Bundeszentrale für politische Bildung, Bonn, 2021. pp. 334–345.

[11] GKV-Spitzenverband, *Bericht des GKV-Spitzenverbandes über die Inanspruchnahme und Entwicklung der Versorgung mit digitalen Gesundheitsanwendungen (DiGA-Bericht) gemäß § 33a Absatz 6 SGB V*, GKV-Spitzenverband, Berlin, 2024.

[12] C. Friesendorf, S. Lüttschwager, *Digitale Gesundheitsanwendungen: Assessment der Ärzteschaft zu Apps auf Rezept*, Springer, Wiesbaden, 2021.

[13] D. Schaeffer, E-M. Berens, D. Vogt, Health Literacy in the German Population, *Dtsch Arztebl Int* **114**(4) (2017), 53–60.
[14] Kassenärztliche Vereinigung Hessen, Arzt- und Psychotherapeutensuche Hessen, https://arztsuchehessen.de/, last access: 27.11.2023.
[15] J.G. Richter, G. Chehab, P. Stachwitz, J. Hagen, D. Larsen, J. Knitza, M. Schneider, A. Voormann, C. Specker, One year of digital health applications (DiGA) in Germany – Rheumatologists' perspectives, *Front Med* **9** (2022).
[16] M. Uncovska, B. Freitag, S. Meister, L. Fehring, Patient Acceptance of Prescribed and Fully Reimbursed mHealth Apps in Germany: An UTAUT2-based Online Survey Study, *J Med Syst* **47** (2023), 14.
[17] T. Sawert, J. Tuppat, Social inequality in the digital transformation: Risks and potentials of mobile health technologies for social inequalities in health, *SOEPpapers* No. 1079 (2020).
[18] W. Greiner, D. Gensorowsky, J. Witte, M. Batram, *DiGA-Report 2022*, Techniker Krankenkasse, Hamburg, 2022.

dHealth 2024
D. Hayn et al. (Eds.)

doi:10.3233/SHTI240013

Unpacking Sociotechnical Discourses on Telehealth Use and Data Protection: A Path Towards Digital Health Value Creation

Javad POOL[a,1], Farhad FATEHI[a], Salma SHARIFI[a], Morteza NAMVAR[a] and Saeed AKHLAGHPOUR[a]

[a] *The University of Queensland, Brisbane, Australia*

Abstract. Background: Telehealth uptake will remain sub-optimal without consumer trust. Safeguarding the security and privacy of health information plays an important role in building trust and acceptance of telehealth. Objectives: This study seeks to unpack the sociotechnical discourses on the use of telehealth with a focus on privacy and security in the context of United States health services. Methods: A search of the media outlets facilitated via the Factiva database was conducted. Using a qualitative method, thematic analysis was performed on the news texts to identify the key themes and provide contextual explanations. Results: The analysis led to the identification of three key themes: 'data protection practice', 'clinical resilience', and 'digital health business value' perspectives. These themes focus on various concepts of telehealth use including data privacy, security, public health emergency, compliance activities in the use of telehealth, meeting stakeholders' needs, reducing costs of service delivery, the potential of telehealth for informed action, and improving users' experience. Among these themes, 'data protection practice' was directly associated with privacy compliance and telehealth use. Other thematic discourses have provided an indirect reflection on the role of privacy compliance, with a greater emphasis placed on health service delivery and market dynamics rather than compliance in practice. Conclusion: Our study revealed the importance of the COVID-19 pandemic in telehealth use, highlighting the move towards 'good faith' and responsible use of telehealth.

Keywords. Telehealth, digital health, data protection, privacy, security

1. Introduction

Telehealth is considered a key solution for improving access to healthcare services in public health emergencies and beyond [1]. The COVID-19 pandemic has catalyzed the use of telehealth and the actualization of its benefits for informed health delivery in a time of crisis [2]. This pandemic crisis, however, presented unique challenges and tensions between public health interests and privacy protection against improper health data use. While telehealth applications have the potential to generate beneficial outcomes for public health, there is also the risk of their misuse in bad faith, leading to privacy violations. Ensuring telehealth data protection is challenging due to the escalating threat of cybercriminals exploiting vulnerabilities in health systems, aiming to gain

[1] Corresponding Author: Javad Pool, The University of Queensland, Brisbane, Australia, E-Mail: j.pool@uq.net.au

unauthorized access to valuable health data. These data protection concerns have adversely affected telehealth use among clinicians and patients. Furthermore, a survey on telehealth use found that health professionals frequently expressed concerns regarding privacy [3]. In achieving the goal of health service delivery and effective use of telehealth, compliance with health data protection laws and regulations is an essential requirement. This study places its emphasis on the US Health Insurance Portability and Accountability Act (HIPAA), serving as a prominent example of data protection regulations in the health services context. HIPAA and its privacy and security rules provide a rich context for our study of telehealth use. To contribute to telehealth literature, our paper takes a holistic and multidisciplinary approach by analyzing multiple aspects of telehealth use and data protection practices. This research also aims to inform public health policy debates and health informatics research, enabling a deeper understanding of the complexity and logic of telehealth data protection, responsible use, and HIPAA compliance.

2. Methods

This research applied an inductive qualitative methodology to offer insights into the regulatory landscape encompassing telehealth use. This particular methodology holds considerable prominence in the contexts of digital health, telehealth, and health services. We used the Factiva database to identify pertinent sociotechnical discourses on HIPAA and the use of telehealth. Factiva is a global news database with a broad collection of sources that reach across disciplines from business, health, communications, and technology to political science. The search was conducted between January 1990 and June 2023 , using search terms related to telehealth and HIPAA (see Appendix, Tabel A1 for search strategy). We identified 105 records by employing the search strategy. After removing duplicates (29 records) and excluding irrelevant news content (14 records) through screening and eligibility steps, 62 news articles were included for qualitative analysis.

The qualitative analysis was conducted employing an inductive approach for contextual investigation and conceptualization. The analysis was performed by the guidance offered by the "Gioia Methodology" for inductive research [4]. We followed the steps of this methodology for constructing a data structure that is composed of '*first-order concepts'*, '*second-order themes'*, and '*aggregate dimensions'*. First-order concepts emerge from the news text. To begin, the first author open-coded the news text to identify perspectives on the use of telehealth. We then categorized the emerging concepts into second-order themes and a theoretical dimension.

3. Results

The analysis led to the identification of three major discursive themes: data protection practice, clinical resilience, and digital health business value. These themes provide a holistic perspective on the use of telehealth as a digital health artifact. Figure 1, informed by data structure (see Appendix, Figure A1) and practice lens [5], illustrates a model of telehealth data protection as practice. The following subsections explain the three perspectives of data protection, clinical resilience, and digital health business value.

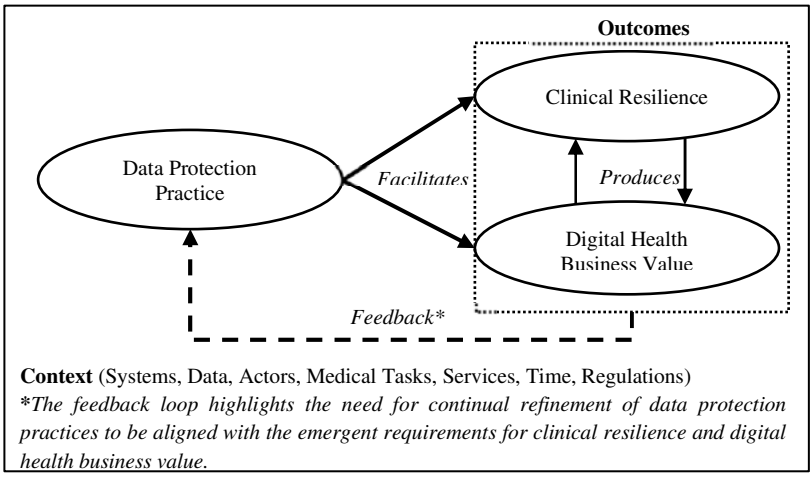


Figure 1. A model of Telehealth Data Protection as Practice (TDPP).

3.1. Data Protection Practice Perspective

The first sociotechnical theme that emerged from our analysis is *Data Protection Practice Perspective*. It consists of concepts around the protection of patient data privacy and security, and information assets. Practices related to the protection of patient data privacy and ensuring telehealth security are boldly represented in media discourses. This dominant concept reflects compliance with HIPAA rules in telehealth use, including video-based and audio-only modalities. Using telehealth technology with full HIPAA compliance in mind can help to achieve data protection [6]. For example, to avoid "bad faith" in telehealth provision and protect patient data, health professionals should not use public-facing video communication platforms, even in public health emergencies (e.g., Facebook Live) [7]. Bad faith in telehealth provision extends beyond malicious intent, covering acts like criminal behavior, privacy breaches, violations of licensing laws, and use of inappropriate communication platforms. Privacy practices such as gaining informed consent for use of telehealth are also important. Providers should inform health service customers that telehealth applications potentially introduce privacy breach risks. Beyond the informed consent, telehealth providers should enable security and privacy features of telehealth (e.g., encryption) when using third-party telehealth [8]. The success of telehealth use is associated with the security of patient data and digital assets. Data privacy and security concerns can lead to refusal of telehealth use by customers. As the number of telehealth users increases, telehealth providers must prioritize the management of emerging cyber risks associated with telehealth use [6]. They should consider effective cybersecurity practices during the high demand for telehealth. Extending encryption technology to mobile health, designing specific training programs for secure use of telehealth are examples of such cybersecurity and data protection practices [9].

3.2. Clinical Resilience Perspective

The second theme in the sociotechnical discourse on HIPAA is *Clinical Resilience Perspective*. It comprises three concepts: public health emergency as a trigger of

response and system use, providing healthcare service, and meeting stakeholders' needs. Clinical resilience perspective can be seen as beliefs and health practices that provides meaning and guides for telehealth actions in responding to public health crises. We found this salient theme on HIPAA and telehealth use are constructed during the COVID-19. While complying with HIPAA indicated in the background, our observation is that the perspective was explicit in the acknowledgment of the role of telehealth services for public health. For instance, the term "good faith" was introduced to the HIPAA discourse during COVID-19. In particular, the US Department of Health and Human Services' (HHS) Office of Civil Rights (OCR) provided guidance on telehealth use that highlighted the good faith. Good faith implies that "*a covered entity or workforce member would exercise a degree of discretion appropriate for its role when deciding to use or disclose [Personal Health Information] PHI*" [10]. As a response to the crisis, this notification facilitated and expanded the use of telehealth platforms which were not considered fully HIPAA compliant before. During the COVID-19, OCR did not penalize service providers for using less secure telehealth platforms to provide timely and accessible care while maintaining good faith provision [8]. But to meet good faith in telehealth provision, providers were prohibited from improper data processing and misuse [11]. From a clinical resilience perspective, relaxing HIPAA restrictions derive additional benefits. This includes increased access to healthcare services via telehealth (especially for rural and underserved patients), reducing the risk of in-person visits for patients and clinicians, and helping to limit the virus spread [12].

3.3. Digital Health Business Value Perspective

The final theme that emerged from our analysis is *digital health business value*. In our contextualization, three concepts represent digital health business value: reducing costs of service delivery, the affordance of IT platform for informed action, and solution view on access to care and improving users' experience. The generalized assumption about IT business value is that digital health can be used in a way that enhances organizational relationships, service operations, efficiency, and coordination [13]. In our analysis, the digital health business value for telehealth emerges as a regulatory compliant tool. From this perspective, business value is generated through various means, including cost reduction, service delivery enhancement, improved telehealth experience, and informed medical actions. However, value creation from telehealth does not take place in isolation. It occurs from a synergistic combination of organizational context, practices, policies, and business processes [14]. This theme reports that telehealth businesses promote their platforms as "HIPAA compliant" to the market, indicating the value creation from data protection. Similarly, HIPAA compliant platforms, from a telehealth designer's perspective, are intended to help providers keep technology operations costs low and in turn, through cost savings generate value for the clients [15]. Digital health business value also can be generated from actualized potentials of HIPAA compliant telehealth for health professionals and end-users. Telehealth creates value by making appointments and diagnoses of patients easier and safer for healthcare professionals, especially during the pandemics and other crises [12, 16]. A high-quality telehealth platform generates value via reducing barriers to optimal care and enabling enhanced feedback. It ultimately improves users' telehealth experiences [17]. In support of our analysis, Table A2 (see Appendix) presents inductive concepts accompanied by illustrative quotes. Overall, our theoretical perspectives on telehealth data protection, which provide research avenues for future theory elaboration and theory testing, are summarized in Table 1.

Table 1. Theoretical overview of Telehealth Data Protection as Practice (TDPP).

Proposed theoretical overview	
We proposed a theoretical model of telehealth data protection (see Figure 1) and provided a sociotechnical perspective which made contribution to digital health data protection and digital health business value literature.	
Type of theoretical perspective	Contextual explanations on Telehealth Data Protection as Practice (TDPP)
Primary constructs	*Data Protection Practice*: actions to protect patient data privacy and security involve adhering to data protection regulations, gaining informed consent, and utilizing encryption and cybersecurity measures during telehealth interactions and use. Additionally, data protection practices encompass safeguarding against malicious attacks, including those orchestrated by hackers, by implementing robust cybersecurity and continuously monitoring for potential security breaches. *Clinical Resilience*: ability to adapt and respond effectively to shocks, and public health emergencies, such as the COVID-19 crisis, by employing telehealth services in compliance with data protection regulations and maintaining good faith provision. *Digital Health Business Value*: creating digital health benefit through use of safe and trustworthy telehealth platforms, offering health service cost reduction, improved service delivery, informed medical actions, and a better telehealth user experience.
Level	Telehealth data protection practice operates on multiple levels, encompassing individual and organizational levels. It extends beyond singular efforts to collective practices, acknowledging collaborative responsibility for safeguarding health data.
Context	The context of telehealth data protection encompasses various factors and actors influencing the facilitation or limitation of data protection. It involves telehealth systems, personal health data, actors (health professionals, managers, admins, IT department, patients, compliance enforcer and data protector (e.g. OCR, privacy officers), regulators, medical tasks (e.g. diagnoses, consultations), time (e.g. public health emergency), law and regulations (e.g. HIPAA)
Overarching proposition	The extent to which telehealth users effectively practice data protection can facilitate clinical resilience and generate digital health business value.
Theoretical advancements	Providing sociotechnical perspectives and contextual explanations which contribute the theories of data protection in digital health such as “Mobile Health Data Protection (MHDP)” [18] and “Context-Privacy Concerns-Practice-Outcomes (CPCPO)” in telehealth [19].

4. Conclusions

This study has unpacked sociotechnical discourses on telehealth use, data security, and privacy. Among the emerging themes, data protection practices were directly associated with HIPAA. Other concepts such as ‘public health emergency’, and ‘reducing costs of service delivery’ offered an indirect privacy perspective on the use of telehealth. In other words, HIPAA was used as background information in the discourse, but the emphasis was elsewhere such as introducing healthcare technologies. Our analyses also revealed that most of the texts on HIPAA and telehealth were produced during the COVID-19 pandemic, highlighting the impact and the momentum created by a public health emergency and the consequent move towards ‘good faith’ in using telehealth. As healthcare providers around the world are following the path towards the post-pandemic, compliance with data protection regulations is essential. Non-compliant behavior can adversely impact the effective use of telehealth systems for providing health services.

Also, security measures in telehealth can unintentionally deter engagement due to added complexity. This burden disproportionately affects those with low digital literacy. Stakeholders like healthcare providers and telehealth designers must prioritize user-friendly data protection solutions and offer training to alleviate this barrier. Nevertheless, this study has boundary conditions that can limit the generalizability of the findings. This research relied on the study of one major data privacy regulation in the context of the US health service. Future research can examine discourses on other regulations and privacy acts such as the EU's General Data Protection Regulation (GDPR).

Supplementary Material

Supplementary material (Appendix) is available at: https://tinyurl.com/nh9rnss7

References

[1] Hollander JE, Carr BG. Virtually perfect? Telemedicine for COVID-19. New England Journal of Medicine. 2020;382(18):1679-81.
[2] Gaitán JA, Ramírez-Correa PE. COVID-19 and telemedicine: A netnography approach. Technological Forecasting and Social Change. 2023;190:122420.
[3] Montoya MI, Kogan CS, Rebello TJ, Sadowska K, Garcia-Pacheco JA, Khoury B, et al. An international survey examining the impact of the COVID-19 pandemic on telehealth use among mental health professionals. Journal of Psychiatric Research. 2022;148:188-96.
[4] Gioia DA, Corley KG, Hamilton AL. Seeking qualitative rigor in inductive research: Notes on the Gioia methodology. Organizational research methods. 2013;16(1):15-31.
[5] Jarzabkowski P, Kaplan S, Seidl D, Whittington R. On the risk of studying practices in isolation: Linking what, who, and how in strategy research. Strategic Organization. 2016;14(3):248-59.
[6] Worth T. HIPAA Compliance: Prepare for the End of Pandemic Telehealth Waivers. Renal & Urology News. 2020.
[7] Castles C. COVID-19 Emergency Efforts: HHS To Allow Non- HIPAA Compliant Telehealth Remote Communications. Mondaq Business Briefing. 2020.
[8] Lacktman N. FAQs On Telemedicine And HIPAA During The Public Health Emergency. Mondaq Business Briefing. 2021.
[9] Hoar S. OCR Announces HIPAA Telehealth Security Waiver In Response To COVID-19 Pandemic. Mondaq Business Briefing. 2020.
[10] US Department of Health and Human Services. Proposed Modifications to the HIPAA Privacy Rule To Support, and Remove Barriers to, Coordinated Care and Individual Engagement 2021 [8.03.2022]. Available from: https://www.regulations.gov/document/HHS-OCR-2021-0006-0001.
[11] Feldman J. COVID-19: OCR HIPAA Enforcement Discretion For Telehealth Mondaq Business Briefing. 2020.
[12] Pfeffer H. HIPAA Restrictions Exist to Protect Patient Privacy. Can We Have Telehealth More Broadly and Still Comply? Risk & Insurance. 2021.
[13] Davidson E. Technology frames and framing: A socio-cognitive investigation of requirements determination. MIS Quarterly. 2002;24(6):329-58.
[14] Kohli R, Grover V. Business value of IT: An essay on expanding research directions to keep up with the times. Journal of the association for information systems. 2008;9(1):23-39.
[15] PR Newswire. DocVilla- HIPAA Compliant Telemedicine Platform to Connect Doctors and Patients through Smartphone. PR Newswire Association 2018.
[16] Nichols J. Revolve Robotics and swyMed Team to Deliver Secure, HIPAA-Compliant Robotic Telepresence for Telemedicine PRWeb. 2015.
[17] Dow Jones Institutional News. Medallia LivingLens Video Feedback Now HIPAA Compliant to Improve Telehealth Experience. 2021.
[18] Pool J, Akhlaghpour S, Fatehi F. Towards a contextual theory of Mobile Health Data Protection (MHDP): A realist perspective. International Journal of Medical Informatics. 2020;141:104229.
[19] Pool J, Akhlaghpour S, Fatehi F, Gray LC. Data privacy concerns and use of telehealth in the aged care context: an integrative review and research agenda. International Journal of Medical Informatics. 2022;160:104707.

dHealth 2024
D. Hayn et al. (Eds.)

doi:10.3233/SHTI240014

A Sustainable Approach to Telerehabilitation in Europe: Patients Are Ready, but Caregivers Are Essential

Andrea CATTANEO[a,1], Andrea VITALI[a], Daniele REGAZZONI[a] and Caterina RIZZI[a]
[a] *University of Bergamo, Bergamo, Italy*

Abstract. *Background*: Several studies have demonstrated the effectiveness of telerehabilitation. However, it remains unclear what proportion of people in need of rehabilitation can confidently use telecommunications networks and related devices. *Objectives*: The aim of this study is to estimate the proportion of patients who possess either the requisite digital literacy to perform telerehabilitation independently or have a family caregiver capable of providing effective support. *Methods*: Synthetic populations with a realistic kinship network (i.e. family trees) representative of European countries are built. Age, sex, and location-specific prevalence rates of rehabilitation needs and digital skills are combined to estimate the percentage of digitally literate patients and patients with digitally literate relatives. *Results*: In Europe, 86% of people in need of rehabilitation are potentially eligible for telerehabilitation. However, in four out of five cases, eligible patients over the age of 65 require caregiver support. *Conclusion*: Telerehabilitation has the potential to spread in Europe. Caregivers have an essential social role in ensuring sustainable access to telerehabilitation.

Keywords. Telemedicine, Telerehabilitation, Caregivers, Kinship Care, Sustainable Development.

1. Introduction

Although the concept of remote healthcare has been well-known for decades, the global impact of the COVID-19 pandemic accelerated the delivery of rehabilitation services at a distance. As we move past the crisis, it is critical to consider the future role of telerehabilitation in the broader healthcare landscape [1], especially in light of its contribution to the achievement of the Sustainable Development Goals (SDGs) [2].

Rehabilitation services encompass all interventions necessary when an individual faces limitations in physical, mental, or social functioning due to aging or a health condition. Telerehabilitation, the delivery of rehabilitative care via telecommunication or the Internet [3], has proven effective in numerous clinical trials [4]. However, a key question has yet to be addressed: how many people are truly ready to use this form of rehabilitation? Some patients may face accessibility challenges due to the severity of their conditions, but a broader barrier is a lack of familiarity with digital technologies and access to a reliable Internet connection [5]. This is particularly relevant given that a significant proportion of those in need of rehabilitation services are in the older

[1] Corresponding Author: Andrea Cattaneo, University of Bergamo, Bergamo, Italy, E-Mail: andrea.cattaneo@unibg.it

demographic [6]. However, patients' lack of digital knowledge is often compensated for by caregivers who help them set up and conduct telerehabilitation sessions. Although patients can receive support from a variety of professionals, caregiving is often provided by family members [7]. Younger individuals typically have greater technological proficiency [8], thus sons and grandsons can often effectively assist patients in accessing telerehabilitation. In addition, the "demographic winter" experienced by most European countries results in families composed of fewer young people and more elderly individuals [9].

In this context, this work aims to estimate the percentage of people in Europe who require rehabilitation and are suitable candidates for telerehabilitation. This includes people who possess the digital skills necessary to access telerehabilitation services, as well as those who have relatives capable of effectively assisting them.

2. Methods

We developed a methodology based on a computational model to estimate the percentage of the population with rehabilitation needs who are eligible for telerehabilitation services (Figure 1). First, using microsimulation, we constructed virtual populations and kinship networks (i.e., family trees) reflecting European nations in 2023. Then, we randomly marked each person as either digitally skilled or not according to age-sex-location-specific rates from open data. Similarly, we marked the people in need of rehabilitation. Lastly, we calculated the proportion of those in need who are suitable candidates for telerehabilitation. We assumed that individuals are eligible if (a) they are digitally literate, or (b) they have a digitally literate relative. The R programming language was used to construct kinship networks, while Python was utilized for data preprocessing and analysis.

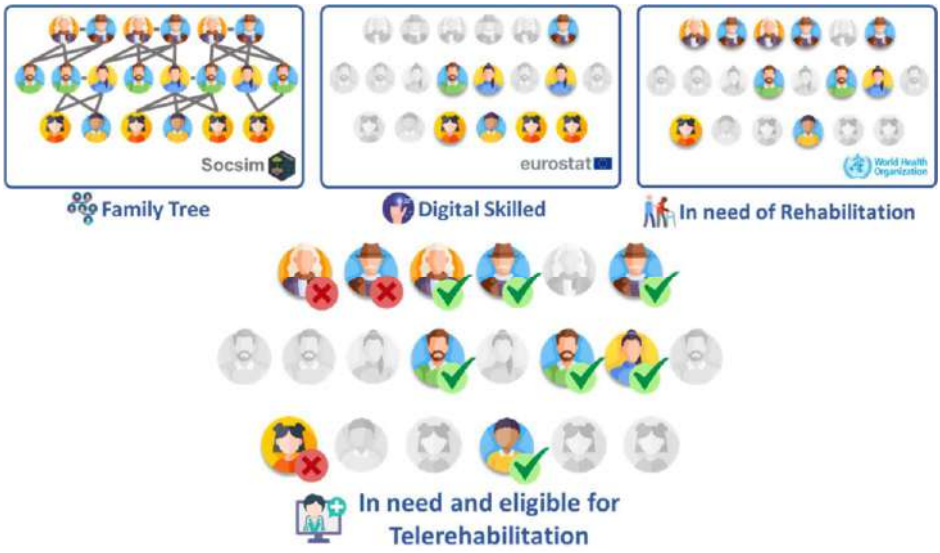


Figure 1. Conceptual framework of the proposed methodology.

2.1. Kinship network

We generated realistic family kinship networks using Socsim, a demographic microsimulation model developed at the University of California, Berkeley [10]. It generates a kinship network by iteratively simulating births and deaths in a synthetic population according to given age- and sex-specific rates. Model output can be combined with disease prevalence rates [11] to study care time demands [12]. Simulations began with a population of 20,000 individuals, initialized by simulating 100 years under the assumption of stable growth. Spanning from 1950 to 2023, the simulation incorporated historical age-specific and sex-specific life tables, fertility rates, and sex ratios at birth from the UN World Population Prospects [13]. To ensure consistency with Socsim's monthly time resolution, we converted the annual rates using Wachter's method [14]. The simulation considered the formation of new couples concurrent with women giving birth to their first child. For partner matching, we employed Socsim's "one-queue" system, which fits a target distribution of age differences between couple members. The mean age difference of couples was sourced from UNECE [15], while the standard deviation was derived from Insee [16]. In defining relatives, we considered the following: romantic partners, parents, offspring, siblings, grandparents, and grandchildren.

2.2. Prevalence of rehabilitation needs and digital skills

The prevalence rates of rehabilitation needs were determined using age, sex, and location-specific data from the World Health Organization (WHO) [6]. These rates are based on a set of diseases and conditions identified by WHO experts as causes of rehabilitation needs. To evaluate the digital literacy and Internet access of the population, we collected from Eurostat age- and sex-specific indicators of activities related to Internet and software use in European countries [8]. We assumed that individuals with at least a basic or higher level of digital literacy are capable of accessing telerehabilitation services. The raw data on digital literacy are available in 5-year age groups, which we interpolated into single-year age groups using second-degree splines smoothed with a first-order Savitzky-Golay filter. Since people in their early years are not expected to engage in self-reliant telerehabilitation, it is assumed that telerehabilitation skills do not materialize until the age of 14.

2.3. Estimation of telerehabilitation eligibility shares

Starting with the synthetic populations of countries as of 2023, we randomly assigned each individual two Boolean attributes representing (a) the need for rehabilitation and (b) digital literacy. The assignment relied on the demographic characteristics of the individuals. To achieve this, we drew from Bernoulli distributions parametrized with the rates previously described. Subsequently, for each nation, the percentage of individuals requiring rehabilitation with digital literacy skills was calculated, marking them as eligible for "self-reliant" telerehabilitation. In cases where individuals had a digitally literate caregiver, they were labeled as eligible "with family caregiving". The union of these two groups constituted the overall number of suitable candidates for telerehabilitation. This calculation was independently performed for all the nations included in the study. Ultimately, the average European proportion was determined by weighting each nation's share based on its population size as of 2023.

3. Results

On average, 86.2% of people in need of rehabilitation in Europe are potentially good candidates for telerehabilitation (Figure 2). Northern and Western countries lead the ranking (92.3% and 89.3%, respectively), while Southern and Eastern countries are below the European average (83.2% and 81.9%, respectively). The country with the highest potential access to telerehabilitation is the Netherlands, with nearly all patients eligible (97.5%). The Netherlands is also the country with the highest proportion of people eligible for "self-reliant" telerehabilitation (67.3%), while the European average is 38.2%. Self-sufficiency varies greatly by age. The percentage of independent patients is 58.0% in the 20-64 age group, 22.8% in the 65-80 age group, and 5.7% in the 80+ age group.

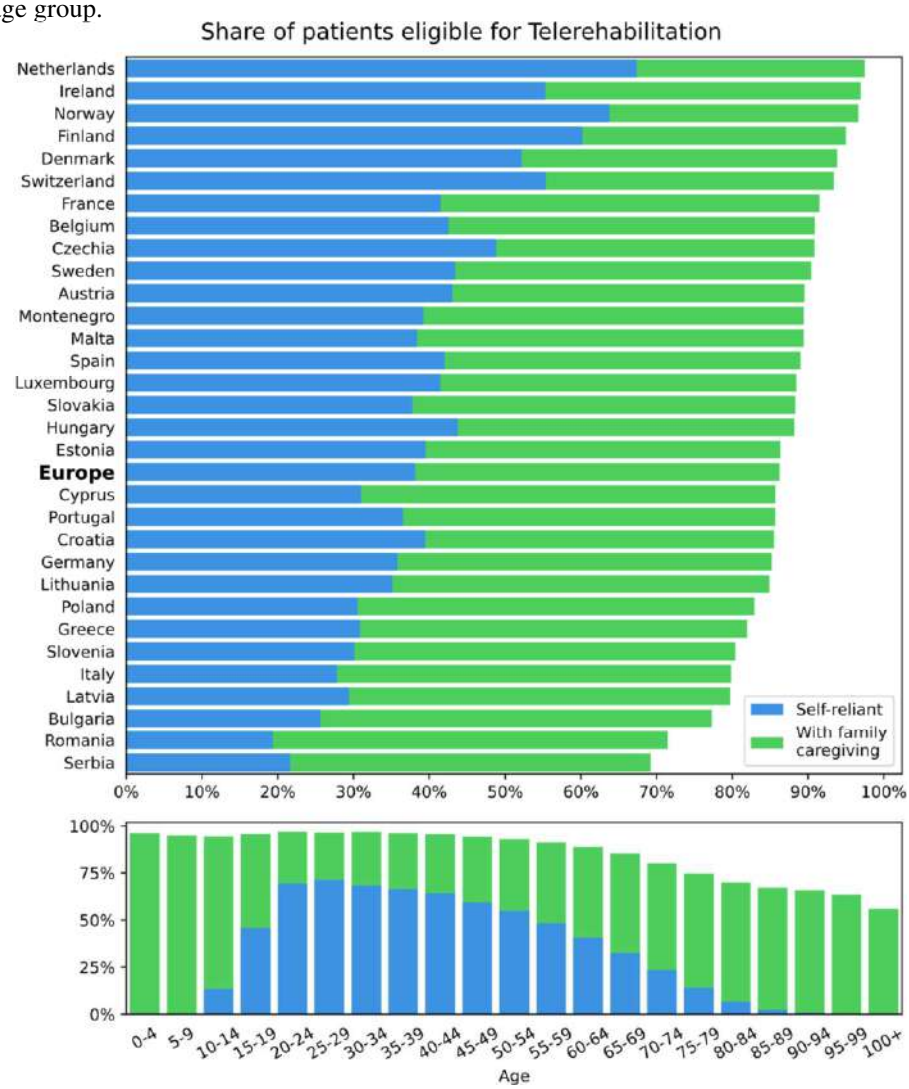


Figure 2. Percentage of people with rehabilitation needs who are potentially suitable for telerehabilitation in Europe. People with basic or above digital literacy are considered eligible for "self-reliant" telerehabilitation, whereas those with at least one digitally literate relative are eligible "with family caregiving".

4. Discussion

The results show that telerehabilitation could be a viable option for the vast majority of Europeans in need of rehabilitation, ranging from 71% to 98% depending on the country. This finding indicates that remote rehabilitative treatments have the potential to become widespread. Although 56% of EU citizens possess basic digital skills [8], we discovered that only an average of 38% of individuals in need of rehabilitation can independently perform telerehabilitation. This is not surprising since the majority of rehabilitation needs come from the elderly population, who also tend to have lower levels of digital literacy. However, it is important to note that the percentage of self-reliant telerehabilitation patients varies significantly among countries, ranging from 19% to 67%.

For patients over the age of 65, we found that nearly four out of five are eligible for telerehabilitation only with caregiver support. Previous studies have emphasized the valuable contribution of caregivers to the sustainability of the traditional health system [17]. Our results suggest that caregivers also play a crucial role in the sustainability of remote healthcare. However, to achieve a truly sustainable system, it is essential to consider that caregiving is associated with intense physical and psychological strain [18], which can lead to reduced care provision, lower quality of life, and physical and mental health deterioration [19]. Therefore, efforts must be made to reduce caregiver burden and improve health outcomes for both caregivers and patients [20]. Telerehabilitation can effectively reduce caregiver burden while improving their knowledge and competence [21], as well as providing significant financial and time savings [22].

The percentage of people eligible for telerehabilitation is influenced by two major factors. The first factor is the number of people who possess the digital literacy and connectivity needed to effectively use telecommunications networks and devices. The second factor is related to family composition, specifically the number of relatives each individual has. The likelihood of having a caregiver with digital skills increases with the size of the family. However, the relationship between the two factors is not linear. Here we present a methodology that takes both factors into account. Our results suggest that the ranking of countries by percentage of people potentially eligible for telerehabilitation is very similar to the ranking by digital skills, highlighting the importance of digital literacy for telehealth.

This work has limitations. Firstly, we used Socsim's kinship network to determine if patients had relatives who could assist with telerehabilitation, but we were unable to ascertain whether the need for care translated into actual caregiving efforts. Additionally, we did not explore the role of caregiver distance from the patient. Secondly, as with any simulation, some assumptions had to be made, resulting in a certain degree of approximation. We did not include marriage or divorce rates, and we assumed no cross-border migrations.

Our findings have implications for policymakers and healthcare providers. The increasing demand for long-term care, driven by population aging, poses a major challenge to the sustainability of public health services. The prevailing trend in Europe is to limit eligibility for institutional care, accompanied by an increasing reliance on family caregiving [23]. Policymakers should recognize telerehabilitation as a viable option for sustainable delivery of rehabilitation services, recognizing that a significant portion of the population can be effectively engaged. At the same time, the critical role of caregivers should be recognized and supported.

Acknowledgments

Cattaneo A. is funded by PON R&I 2014-2020 (DM 1061/2021). Figure 1 has been designed using images from Flaticon.com. This work is partially funded by the National Plan for NRRP Complementary Investments - project n. PNC0000003 - AdvaNced Technologies for Human-centrEd Medicine (ANTHEM). This work reflects only the authors' views and opinions, neither the Ministry for University and Research nor the European Commission can be considered responsible for them.

References

[1] M. Scherrenberg et al., The future is now: a call for action for cardiac telerehabilitation in the COVID-19 pandemic from the secondary prevention and rehabilitation section of the European Association of Preventive Cardiology, European Journal of Preventive Cardiology 28(5) (2021), 524–540.
[2] World Health Organization, Rehabilitation: fact sheet on Sustainable Development Goals (SDGs), (2019),. https://www.who.int/europe/publications/i/item/WHO-EURO-2019-2384-42139-58051
[3] M. Alexander, Telerehabilitation: Principles and practice, Telerehabilitation: Principles and Practice (2022), 1–431.
[4] L. Suso-Martí et al., Effectiveness of Telerehabilitation in Physical Therapist Practice: An Umbrella and Mapping Review With Meta–Meta-Analysis, Physical Therapy 101(5) (2021), 1–9.
[5] T. V. Le et al., The impact of digital competence on telehealth utilization, Health Policy and Technology 12(1) (2023), 100724.
[6] A. Cieza et al., Global estimates of the need for rehabilitation based on the Global Burden of Disease study 2019: a systematic analysis for the Global Burden of Disease Study 2019, The Lancet 396(10267) (2020), 2006–2017.
[7] E. K. Eifert et al., Family Caregiver Identity: A Literature Review, American Journal of Health Education 46(6) (2015), 357–367.
[8] Eurostat, Individuals who have basic or above basic overall digital skills by sex, 2023. https://ec.europa.eu/eurostat/databrowser/view/tepsr_sp410/default/table (accessed Jan. 11, 2024).
[9] The Lancet Regional Health – Europe, Securing the future of Europe's ageing population by 2050, The Lancet Regional Health - Europe 35 (2023), 100807.
[10] MPIDR/rsocsim: Socsim Microsimulation R Package, https://github.com/MPIDR/rsocsim
[11] C. M. Rutter et al., A Simple Method for Simulating Dementia Onset and Death within an Existing Demographic Model, Medical Decision Making 42(1) (2022), 43–50.
[12] D. Alburez-Gutierrez et al., The "Sandwich Generation" Revisited: Global Demographic Drivers of Care Time Demands, Population and Development Review 47(4) (2021), 997–1023.
[13] United Nations, World Population Prospects 2022,. https://population.un.org/wpp/
[14] K. W. Wachter, Essential demographic methods, Harvard University Press (2014).
[15] UNECE Statistical Database, Mean Age at First Marriage by Sex, Country and Year, https://w3.unece.org/PXWeb2015/pxweb/en/STAT/STAT__30-GE__02-Families_households/052_en_GEFHAge1stMarige_r.px/ (accessed Jan. 12, 2024).
[16] Insee, En 2017, dans deux tiers des mariages entre personnes de sexe différent, la femme est plus jeune que son mari , (2019),. https://www.insee.fr/fr/statistiques/3715700 (accessed Jan. 12, 2024).
[17] K. M. Kokorelias et al., "Caregiving is a full-time job" impacting stroke caregivers' health and well-being: A qualitative meta-synthesis, Health & Social Care in the Community 28(2) (2020), 325–340.
[18] H. Y. Chang et al., Impact of mental health and caregiver burden on family caregivers' physical health, Archives of Gerontology and Geriatrics 50(3) (2010), 267–271.
[19] Z. Liu et al., Caregiver burden: A concept analysis, International Journal of Nursing Sciences 7(4) (2020), 438–445.
[20] K. E. Uhm et al., Influence of preparedness on caregiver burden, depression, and quality of life in caregivers of people with disabilities, Frontiers in Public Health 11 (2023), 1153588.
[21] W. J. Sun et al., Telerehabilitation for Family Caregivers of Stroke Survivors: A Systematic Review and Meta-Analysis, Journal of Nursing Management (2023).
[22] E. M. G. de S. e. Silva et al., Impact of physiotherapy with telerehabilitation on caregivers of patients with neurological disorders: A systematic review protocol, Frontiers in Aging Neuroscience 14 (2022).
[23] S. Elayan et al., The Economic Costs of Informal Care: Estimates from a National Cross-Sectional Survey in The Netherlands, European Journal of Health Economics (2024).

dHealth 2024
D. Hayn et al. (Eds.)

doi:10.3233/SHTI240015

Generating Actionable Insights from Patient Medical Records and Structured Clinical Knowledge

Natasha TRAJKOVSKA [a], Michael ROISS [b], Sophie BAUERNFEIND [a], Mohammad ALNAJDAWI [b], Simone SANDLER [a], Daniel HERZMANEK [b], Matthias WINKLER [b], Michael HAIDER [b], and Oliver KRAUSS [a,1]
[a] *University of Applied Sciences Upper Austria*
[b] *Treetop Medical*
ORCiD ID: Natasha Trajkovska https://orcid.org/0000-0001-8265-9041, Sophie Bauernfeind https://orcid.org/0009-0005-5290-3900, Simone Sandler https://orcid.org/0000-0001-8919-7804, Oliver Krauss https://orcid.org/0000-0002-8136-2606

Abstract. While adherence to clinical guidelines improves the quality and consistency of care, personalized healthcare also requires a deep understanding of individual dis- ease models and treatment plans. The structured preparation of medical routine data in a certain clinical context, e.g. a treatment pathway outlined in a medical guideline, is currently a challenging task. Medical data is often stored in diverse formats and systems, and the relevant clinical knowledge defining the context is not available in machine-readable formats. We present an approach to extract informa- tion from medical free text documentation by using structured clinical knowledge to guide information extraction into a structured and encoded format, overcoming the known challenges for natural language processing algorithms. Preliminary re- sults have been encouraging, as one of our methods managed to extract 100% of all data-points with 85% accuracy in details. These advancements show the potential of our approach to effectively use unstructured clinical data to elevate the quality of patient care and reduce the workload of medical personnel.

Keywords. Electronic Health Records, Treatment Adherence and Compliance, Data Science

1. Introduction

We present an approach to extract clinical data from existing medical documentation and reconstruct the course of treatment in a certain clinical context e.g. a treatment pathway, for example treatment of a lung carcinoma, outlined in a medical guideline or a local standard operating procedure.

Understanding a patients past treatment course is crucial for personalized medicine [1]. Individuals respond differently to certain treatments, and clinicians need to take

[1]Corresponding Author: Oliver Krauss, oliver.krauss@fh-hagenberg.at

known clinical findings and patient characteristics into account when they tailor upcoming treatments [2]. Cross-referencing a patient's past treatment course with predefined and evidence based clinical pathways provides a powerful tool for clinical decision support. It helps healthcare professionals to consistently deliver the best possible quality of care, ultimately leading to better patient outcomes. The availability of accurate and detailed knowledge on their patient's past treatment course is also crucial for patient safety. It helps to prevent severe errors that occur due to a lack of awareness of past clinical findings but also helps to reduce duplicate tests or procedures. This, in turn, reduces the likelihood of adverse events and enhances overall patient safety [3,4,5].

The structured preparation of medical routine data in a certain clinical context is a challenging task, as medical data is stored in diverse formats and systems [6]. Integrating information from different sources can be complex. We solve this issue by using the existing medical documentation in a free text format, but are well aware that extracting treatment information from unstructured medical text is a non-trivial task [7,8].

2. Methods

We use patient-specific medical records and laboratory reports for reconstruction of the treatment course. Our methodology adopts an iterative process to transform unstructured data into a cascade of progressively refined stages: structured data, encoded data, interpreted data, and ultimately, actionable knowledge.

Initially the transition from unstructured data to structured data, documents are analyzed, and relevant information is extracted. During this phase, machine-readable treatment pathways, and disease models based on Treetop's framework are used to infuse the various extraction models with in-depth-knowledge about the presumed treatment and disease of the patient. Next, the structured data is encoded utilizing different coding systems such as SNOMED CT. In the following step, process mining techniques are employed to construct a chronological treatment timeline of a patient, thereby providing detailed insights of the medical events. The final stage again leverages Treetop treatment pathways and disease models to put the data into a certain clinical context and gain actionable knowledge on the past course of the treatment and prospective steps in the treatment of the patient.

2.1. Unstructured to Structured Data

To obtain relevant information about the patient's course of treatment, the PDF patient letters are converted to plain text. Each document is classified into *patient letter*, *laboratory report*, which can currently be processed, or an *unknown* type of document. Each document is then divided into different sections, such as diagnosis or medication. With a large language model (LLM), specifically the Llama-2-70b-orca-200k model [9] which was not trained on medical data, each section is processed and data points are extracted in JSON format. Two different approaches are used for prompting the LLM (see Figure 1. In the *LLM* approach, the section text is used in its entirety. The *GuidedLLM* approach, uses medical knowledge infused LLM prompting. We apply the Treetop Medical treatment pathways and disease model information to reduce the section text to relevant sentences before using it in the prompting step. For extraction, we use the LLAMA-

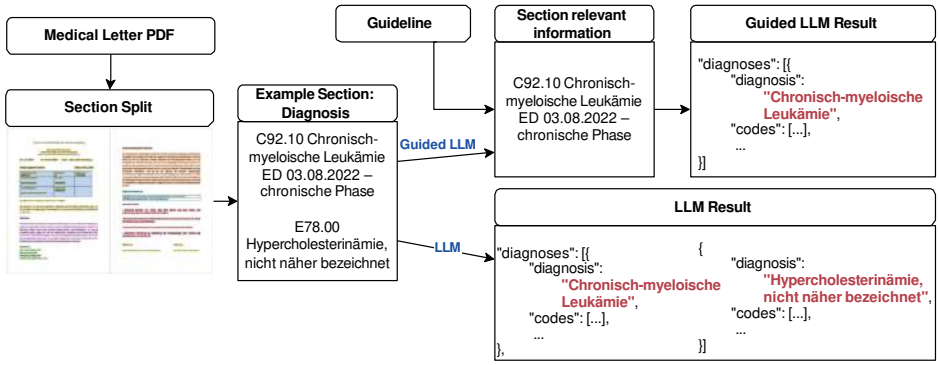


Figure 1. Patient letter example with both approaches

ORCA model [10]. The extracted information is checked to prevent hallucinations from the LLM.If a hallucination is found, the results are flagged and the hallucinated values are discarded for the further process. The validated and extracted information from the different sections are collected across all documents and combined in one file for one patient.

2.2. Data Encoding

The extracted data is mapped to clinical ontology codes. This enhances data interpretation, and provides a systematic code that identifies clinical terms. We employ a hybrid approach involving lexical and semantic mapping techniques. Initially, candidate matches are identified by comparing extracted terms with a translated source of SNOMED CT, utilizing Levenshtein distance for lexical comparison. Top candidates are determined, For the semantic approach, we computed the similarity between the target clinical term and SNOMED CT concepts. This involved converting and representing both entities in a high-dimensional vector space. Subsequently, we generated the top semantic candidates based on their similarity using the cosine similarity function. Subsequently, a reranking process is executed for the top candidates from both approaches based on matching scores, selecting the highest score as the final match. This integrated methodology ensures a comprehensive and accurate mapping of extracted data to clinical ontologies.

2.3. Constructing a Treatment Timeline

We use process mining to construct a treatment timeline for the patient from the extracted data points. To use these events for process mining, we need each event to contain the following three pieces of information: ”**When** did it happen?”, ”Who was the involved **patient**?” and ”**What** did happen?”. The advantage of using the encoded data is that the treatment timeline resulting from these events can then be automatically compared to the treatment plan that should be followed, or to the treatment timeline of other patients.

2.4. Create Actionable Knowledge

Finally, we match the certain treatment pathway and disease model with the treatment timeline resulting from the process mining for comparison. This final step helps to iden-

tify relevant findings in the treatment course that might be relevant for upcoming treatments or procedures. Furthermore, the identification of deviations between predefined treatment pathways and actual treatment courses e.g. too late follow-up checks helps to detect potential weak spots in clinical processes, or bottlenecks such as availability of resources.

3. Results

In our comparative analysis of two approaches *GuidedLLM* and *LLM* for extracting information from medical documents, the emphasis is placed on the evaluation of information that is relevant for patient's diagnoses and medications. The evaluative scope is defined by four medical letters, each originating from patients diagnosed with Chronic Myeloid Leukemia (CML). Two distinct ground truth documents were created manually by medical experts: the first one contextualizes the assessment and focuses solely on the target diagnosis (CML) present in all 4 patients and the medications relevant to it (total 4 diagnoses and 3 medications), while the second one encompasses all diagnoses and medications that are listed in the letters, totaling 11 conditions and 13 medications across the 4 patients. The instances used in the ground truth documents encapsulate relevant details by being organized in JSON object lists.

Each diagnosis contains: diagnosis title, codes (comprising a list of objects denoting code and the code system), notes, clinical status, start date, end date, and verification status.

A medication contains the following attributes: drug (represented by commercial name and substance), route, dosage (with value and unit), dose form, interval, frequency, notes, start date, end date, and duration.

3.1. Contextualized Ground Truth

The contextualized ground truth (4 diagnoses and 3 medications) is used to make a targeted evaluation of the performance of our two approaches. We define a diagnosis match as the accurate detection of the diagnosis title. A medication match is recorded when either the commercial name or the substance of the extracted and expected values match. The textual comparison involves the removal of punctuation and white spaces, with capitalization disregarded. A match is marked when the expected and extracted values are equivalent, or when one encompasses the other. The results showed that the medical knowledge infused prompting (*GuidedLLM*) demonstrates a perfect detection rate, identifying all 4 diagnoses and the 3 medications. The *LLM* approach detected 3 target diagnoses and 2 of the target medications, as shown in Table 1. It failed to get a response for one of the patients.

Table 1. Sensitivity Results (Contextualized Ground Truth)

	Sensitivity		
Approach	Diagnoses	Medications	Combined
GuidedLLM	4/4	3/3	7/7
LLM	3/4	2/3	5/7

Subsequently, we evaluate the proficiency of our methods in extracting the details of the diagnoses and medications. As a metric, we used Jaccard similarity coefficient, which considers the similarity between two JSON objects. Each JSON object was treated as a set of its key-value pairs (including the nested ones). Then we calculated the ratio of the number of common key-value pairs to the total number of unique key-value pairs across both objects. In this comparison, we used the diagnoses and medications that belong to the intersection of the instances detected by both methods and the target instances in the contextualized ground truth (3 diagnoses and 2 medications, see Table 2).

Table 2. Jaccard Coefficient Results (Contextualized Ground Truth)

	Jaccard coefficient		
Approach	Diagnoses	Medications	Combined
GuidedLLM	0.739	0.958	0.851
LLM	0.783	0.958	0.872

The lower Jaccard coefficient for the diagnoses stems mostly from the extraction tendency to identify the diagnosis code system, despite not being explicitly stated in the document. Additionally, in most of the cases, the extraction failed to detect notes. The small difference in the coefficient between the two approaches is due to the difference in start date extraction.

3.2. Generalized Ground Truth

We further employed a generalized ground truth, intended to serve as an indicator of potentially extractable additional medical information regarding the diagnoses and medications of the four patients. The *LLM* approach demonstrated a detection capability for 8 out of the 13 diseases, while *GuidedLLM* identified 4 of the ground truth diagnoses (and detected one additional, not listed in the ground truth diagnoses). The detection rate for medications is 8 for *LLM* and 6 for *GuidedLLM* out of 13, as seen in Table 3.

Table 3. Sensitivity Results (Generalized Ground Truth)

	Sensitivity		
Approach	Diagnoses	Medications	Combined
GuidedLLM	4/11	6/13	10/24
LLM	7/11	8/13	15/24

The results in Table 4 illustrate the Jaccard Coefficients. We again used only the target instances that both approaches achieved to detect. In this case, the diagnoses' evaluation remained the same as in the contextualized comparison. On the other hand, the two approaches have one additional medication in common, which induces change in the coefficients.

These findings show that under specific clinical scenario, the *GuidedLLM* approach provided better results. The Jaccard coefficients demonstrate a reasonably good overlap, but also provided a direction for further prompt improvements.

Table 4. Jaccard Coefficient Results (Generalized Ground Truth)

	Generalized Ground Truth -Jaccard coefficient		
Approach	Diagnoses	Medications	Combined
GuidedLLM	0.739	0.944	0.864
LLM	0.783	0.944	0.881

4. Discussion

We introduce an approach aimed at extracting disease-relevant clinical information from unstructured patient records and, subsequently, cross-referencing extracted data with disease models and treatment pathways. Our study demonstrates that leveraging structured medical knowledge can enhance the accuracy and effectiveness of various sub-tasks of the pipeline. This improvement was demonstrated in an empirical manner through comparative analysis, with the *GuidedLLM* approach finding 100% of the relevant items discovered and 85.1% of details extracted, primarily missing physician notes. In an evaluation of the performance of an *LLM*-only method in contrast to the *GuidedLLM*, we observed a better detection rate of medical entities with the latter.

There is a wide array of LLM applications in medicine [11], albeit with mixed results. We show that including expert knowledge can improve this aspect.

It is important to acknowledge that this assessment was conducted using a limited sample of four patient cases. To validate these findings further, an investigation with a larger dataset is necessary.

While our study demonstrates the advantages of integrating structured medical knowledge into data extraction processes utilizing LLMs, there are some challenges that need to be addressed in future research. In the short term, we plan to expand the scope of entity extraction within medical documents while focusing on more complex aspects, such as evaluating the verification status of statements. Additionally, we aim to address the identification of missing information that should ideally be present in accordance with outlined treatment pathways.

Our approach holds potential for enhancing the accuracy and efficiency of clinical information extraction from unstructured patient records. Leveraging the extensive knowledge encapsulated within digital treatment pathways and disease models, our research has showcased the potential to elevate the precision of clinical decision-making.

Acknowledgements

We thank the Austrian Research Promotion Agency FFG, for funding the project *Treetop Medical* (program number: 900850) in the General Programme, with research budget provided by the Federal Republic of Austria. We also thank our project partners from the Medical University of Vienna.

References

[1] M. Cellina, M. Ce`, M. Al`ı, G. Irmici, S. Ibba, E. Caloro, D. Fazzini, G. Oliva and S. Papa, Digital Twins: The New Frontier for Personalized Medicine?
13(13), 7940. doi:10.3390/app13137940. https://www.mdpi.com/2076-3417/13/13/7940.

[2] F. Sugandh, M. Chandio, F. Raveena, L. Kumar, F. Karishma, S. Khuwaja, U.A. Memon, K. Bai, M. Kashif, G. Varrassi, M. Khatri and S. Kumar,
Advances in the Management of Diabetes Mellitus: A Focus on Personalized Medicine. doi:10.7759/cureus.43697.

[3] M. Packer and M. Metra, Guideline-directed Medical Therapy for Heart Failure Does Not Exist: A Non-judgmental Framework for Describing the Level of Adherence to Evidence-based Drug Treatments for Patients with a Reduced Ejection Fraction
22(10), 1759–1767. doi:10.1002/ejhf.1857.

[4] M. Jarjour, C. Henri, S. De Denus, A. Fortier, N. Bouabdallaoui, A. Nigam, E. O'Meara, C. Ahnadi, M. White, P. Garceau, N. Racine, M.-C. Parent, M. Liszkowski, G. Giraldeau, J.-L. Rouleau and A. Ducharme, Care Gaps in Adherence to Heart Failure Guidelines
8(9), 725–738. doi:10.1016/j.jchf.2020.04.019. https://linkinghub.elsevier.com/retrieve/pii/S2213177920303334.

[5] M. Vaismoradi, S. Tella, P. A. Logan, J. Khakurel and F. Vizcaya-Moreno, Nurses' Adherence to Patient Safety Principles: A Systematic Review
17(6), 2028. doi:10.3390/ijerph17062028. https://www.mdpi.com/1660-4601/17/6/2028.

[6] E.L. Covington, X. Chen, K.C. Younge, C. Lee, M.M. Matuszak, M.L. Kessler, W. Keranen, E. Acosta, A.M. Dougherty, S.E. Filpansick and J.M. Moran, Improving Treatment Plan Evaluation with Automation
17(6), 16–31. doi:10.1120/jacmp.v17i6.6322.

[7] T.A. Koleck, C. Dreisbach, P.E. Bourne and S. Bakken, Natural Language Processing of Symptoms Documented in Free-Text Narratives of Electronic Health Records: A Systematic Review
26(4), 364–379. doi:10.1093/jamia/ocy173.

[8] S. Sheikhalishahi, R. Miotto, J.T. Dudley, A. Lavelli, F. Rinaldi and V. Osmani, Natural Language Processing of Clinical Notes on Chronic Diseases: Systematic Review
7(2), e12239. doi:10.2196/12239. http://medinform.jmir.org/2019/2/e12239/.

[9] MHK, Llama-2-70b-orca-200k (Revision 1ab69d4), Hugging Face, 2024. doi:10.57967/hf/1687. https://huggingface.co/ddobokki/Llama-2-70b-orca-200k.

[10] ddobokki/Llama-2-70b-orca-200k · Hugging Face, 2024, [Online; accessed 20. Jan. 2024]. https://huggingface.co/ddobokki/Llama-2-70b-orca-200k.

[11] A.J. Thirunavukarasu, D.S.J. Ting, K. Elangovan, L. Gutierrez, T.F. Tan and D.S.W. Ting, Large language models in medicine, *Nature Medicine* **29**(8) (2023), 1930–1940–. doi:10.1038/s41591-023-02448-8.

dHealth 2024
D. Hayn et al. (Eds.)

doi:10.3233/SHTI240016

Exploring the Environmental Impact of Telemedicine: A Life Cycle Assessment

Anna SAVOLDELLI[a,1], Daniele LANDI[a] and Caterina RIZZI[a]

[a] *Department of Management, Information and Production Engineering, University of Bergamo, via Pasubio, Dalmine (Bergamo), Italy*

Abstract. Background: Telemedicine has emerged as a potential solution to mitigate the significant greenhouse gas emissions of the healthcare sector. A comprehensive evaluation is required to quantify the environmental benefits of its implementation. Objectives: The study aims to compare the environmental sustainability of in-person and virtual examinations for heart failure patients. Methods: A standard life cycle assessment has been applied to quantify the equivalent CO_2 of direct and indirect activities required to release a medical examination (virtual or physical) for a patient in an Italian hospital. Inputs of the analysis include electronic devices of hospital and patients, energy consumption, wastes, internet usage and patient travel. Depending on the type of visit (virtual or physical), inputs have been processed differently, considering actual consumption and utilization. Results: Televisit reduces emissions from 9.77 $kgCO_2e$ to 0.41 $kgCO_2e$. Transport and internet data use are key inputs for in-person (i.e., 98%) and telemedicine visits (i.e., 72%), respectively. Discussion: Given the frequent car travels, telemedicine emerges as a tool to improve environmental benefits and reduce time for patients and caregivers.

Keywords. telemedicine, life cycle, carbon footprint, environmental impact, sustainable development

1. Introduction

The increasing negative impacts of climate change on human health requires more efforts to mitigate emissions and alleviate the overall environmental burden [1]. The healthcare sector, while crucial for addressing patient needs. It participates significantly to environmental impact, becoming one of the largest contributions to greenhouse gas (GHG) emissions [2]. Chronic patients care is responsible of a significant amount of healthcare carbon footprint, due to the release of long-term assistance, repetitive follow-up examination and monitoring devices [3]. Specifically, cardiovascular and heart failure (HF) chronic diseases affect millions of people worldwide [4], representing a significant aspect of this burden.

Several research studies have highlighted telemedicine as a potentially effective solution to mitigate the environmental impact of healthcare, aiming to balance emissions without compromising the quality of care [5][6]. Telemedicine techniques have found extensive application in chronic patients' management, allowing also the reduction of the travel to hospitals. Following the increase adoption of these promising technologies,

[1] Corresponding Author: Anna Savoldelli, Department of Management, Information and Production Engineering, University of Bergamo, Dalmine (Bergamo), Italy, E-Mail: anna.savoldelli@unibg.it

it is crucial to quantify overall environmental impacts generated by their use. Hence, standardized, and quantitative methods should be developed and applied to evaluate impacts across the life cycle [7]. Several studies have investigated travel reductions, focusing on its associated aspects [8] (e.g., distance avoided, transportation mode, time, and cost savings, ...). However, these analyses are often confined to direct environmental impacts of telemedicine processes. Recently, a few articles have applied life cycle assessment (LCA) to analyze environmental sustainability of remote care. LCA has been developed in the industrial field for the evaluation of all life cycle phases of product and processes. It includes materials extraction, production of disposable and reusable equipment, transportation, utilization of products and materials, energy consumption, and disposal [9]. Previous evaluations have considered only activities performed during the time of a televisit [10]. A comprehensive analysis should include also indirect activities, required to perform the medical examination.

Hence, the present study aims to conduct a comprehensive LCA of virtual and in-person visits, encompassing the entire phases of the care process. The methodology is applied to a real case, involving chronic patients from the HF clinic of an Italian hospital. This care provider has begun trials of televist to manage the critical phase of the Covid-19. Remote and face-to-face processes are compared basing on GHGs emissions, for quantifying the environmental sustainability of remote care.

2. Methods

A standard life cycle assessment has been applied to evaluate the sustainability of a care process for chronic HF patients, comparing traditional in-person follow-up examinations to virtual examinations. Steps have been derived from ISO 14040/14044 (ISO (International Organization for Standardization), 2006a/2006b)) [11][12]. The approach has been applied to evaluate follow-ups appointments, that do not require the employment of additional specific medical devices.

2.1. Goal and scope definition

The analysis aims to quantify the amount of equivalent CO_2 (CO_2e) associated to the activities required to deliver virtual and in-person follow-up examination. The functional unit has been defined as a complete process to perform a medical examination (virtual or physical) for a single HF patient. As both care processes serve the same function (i.e., refer patient's clinical conditions), a comparative analysis between the two modalities can be conducted. Information has been acquired during previous research [13] and include both the activities carried out during the medical examination and the ones required to engage patients and assess the quality of service. Figure 1 shows inputs included in system boundaries and their contribution for the phases of the care process. The impact of devices, wastes and transportation and Italian electricity mix came from primary data. Secondary data have been used to get energy consumption [14] and network data usage. For the latter, a minimum network bandwidth for various operations has been approximated, considering a data rate of 0.015 kWh/GB. Each visit lasted an average of 15 minutes. For patients connected via televisit, only laptop use has been included. Conversely, for patients attending in-person visits, an average round trip travel of 30 km has been considered. Given the mountainous region and the limited public

transportation service, only travel by private car has been calculated. For these types of visits, protective personal equipment has not been used. Heating and lighting in rooms have omitted in the analysis, as they have been deemed non-discriminatory factors.

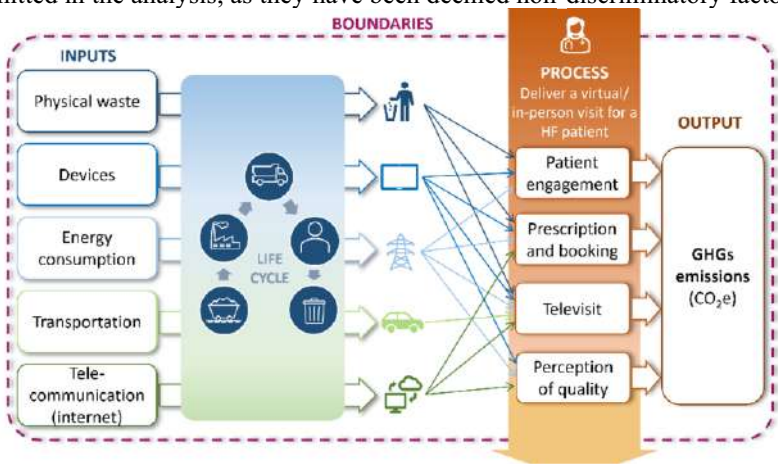


Figure 1. System boundaries and inputs flows for calculating GHGs outputs.

2.2. Life cycle inventory

Inventory has been defined acquiring real data, thanks to the collaboration between the hospital and the university. When information was not available, secondary data has been extrapolated from scientific literature [14]. Inventory has been imported in FootprintCalc [15], a free footprint calculator, used to evaluate the environmental impact during the function life cycle. EcoInvent V3.8 database [16] has been used for transportation assessment. In addition, impacts of internet network usage has been manually included.

2.3. Calculation of emissions

The amount of CO_2e has been computing for the functional unit by summing the impacts of inputs included in Table 1. The same inputs, necessary to release the patient medical examination, have been considered in both cases, to carry out the comparison. Depending on the mode of service release (in-person or virtual), the inputs have been processed differently, considering actual consumption and utilization.

The calculation of device usage has been performed using the following formula:

$$\frac{CF \times time\ of\ use}{life\ span} \tag{1}$$

In formula (1), CF identifies the kilograms of CO_2e of the entire life cycle, calculated by means of the FootprintCalc database [15]. The amount of paper sheets and hand sanitizer used have been extrapolated from the process. The energy consumption of devices has been determined based on the Italian electricity mix. Using emission factor electricity, kilograms of CO2e per kWh were derived. Transportation emissions have been calculated considering an average Euro 5 car in EcoInvent V3.8 [16]. Finally, network data usage has been quantified as defined by Sillcox et al. [9], summing internet usage and server utilization, according to the following formulas:

$$Interet\ usage = E \times I \times D \times t \times \frac{3600}{800} \quad (2)$$

$$Server\ usage = E \times W \times t \quad (3)$$

In formula (2) and (3), E represents the amount of GHGs per kWh (0.31 $kgCO_2e$/kWh), calculated using the Italian electricity mix. I is the energy density of the internet (0.015 kWh/GB), D is the internet data transfer rate (Mbps), and W is the wattage of the server (0.6 kWh/server [9]).

Table 1. Life cycle inventory related to the defined functional unit (F.u.).

Inputs		F.u. (Televisit process)	F.u. (In-person process)
Devices	1 desktop (hospital)	42 min	20 min
	1 laptop (patient)	27 min	5 min
	1 webcam/microphone/ sound system	15 min	0 min
	2 local network area	69 min	25 min
	2 telephones	0 min	16 min
Physical wastes	Paper sheets, FSC 80 gr/m2	0.75 m2	0.88 m2
	Hand sanitizer (benzyl alcohol)	0 kg	0.003 kg
Transportation	Car travel	0 km	30km
Energy consumption	Desktop	0.105 kWh	0.05
	Laptop	0.018 kWh	0.003 kWh
	LAN	0.023 kWh	0.008 kWh
	Webcam	0.0024 kWh	0 kWh
	Sound system	0.001 kWh	0 kWh
	Microphone	0.0006 kWh	0 kWh
	Telephone	0 kWh	0.0011 kWh
	Printing process	0.75 m^2	0.88 m^2
Network data usage	Webpage/applications	35 min (5.5 Mbps)	17 min (5.5 Mbps)
	Upload/download	430 KB	90 KB
	Email/fax sending	4 units	2 units
	Videocall (2 participants)	15 min (13.6 Mbps)	0 min

3. Results

The quantity of CO_2e generated during the telemedicine process has been resulted significantly lower compared to that produced in a traditional in-person visit scenario, decreasing from 9.77 $kgCO_2e$ to only 0.41 $kgCO_2e$ (Figure 2a).

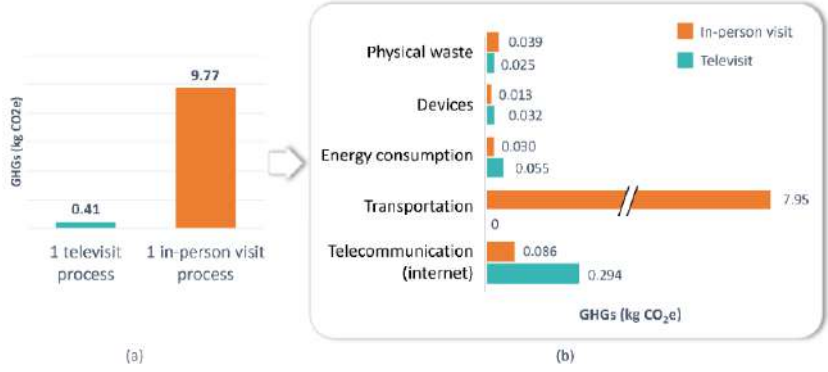


Figure 2. Total GHGs emissions related to the life cycle of a process to deliver in-person and virtual visit.

The primary sources of GHGs emissions of televisit process (Figure 2b) have been identified in the use of internet data (i.e., 72%) and the consumed energy (i.e., 14%). Minor contributions have involved the use of paper materials (i.e., 6%) and the life cycle of the devices (i.e., 8%). During the in-person visit process, emissions have been mainly associated with travel of patients to the hospital (i.e., 98%). A modest contribution came from the use of internet data (i.e., 1%), paper waste and hand sanitizing gel (i.e., 0.4%), and the energy consumption due to device usage (i.e., 0.4%). The impact of the life cycle of the devices can be considered negligible, given the short duration of their use.

4. Discussion

Comparing a single visit for HF patients, telemedicine results in a drastic reduction of over 95% in CO_2e emissions compared to traditional in-person appointments. The findings highlight a significant environmental benefit resulting from the reduction of car travels. In the literature, few studies have adopted an LCA approach, allowing for a standardized and comprehensive assessment. Among these, transportation has emerged as the most critical aspect [17][7]. In the considered province of Bergamo, patient transportation emerges as a critical issue. The hospital's location outside the city extends public travel times, often necessitating the change of different public transportations. For this reason, patients often choose to use private cars, increasing related emissions. A deeper evaluation, including hybrid/electric vehicles, could offer more insights. Despite the analyzed limited distance, telemedicine shows benefits not only in terms of environmental sustainability but also in saving time for both patients and caregivers. Indeed, considering the traffic conditions in the area, Google Maps estimates approximately 50 minutes by car to cover that distance. Additionally, patients must spend time for searching parking and arriving ahead of the appointment. Thus, the adoption of telemedicine can also be seen as an improvement in social impact. Reached benefits could be especially relevant for patients with chronic conditions, requiring long-term assistance and periodic follow-up analyses.

Despite often being overlooked, telecommunication internet usage emerged as the most impactful component within the telemedicine process (i.e., 72%), due to both televisit delivery and management activities. GHGs missions of internet usage almost doubled from in-person (i.e., 0.098 kgCO_2e) to virtual examination (i.e., 0.295 kgCO_2e). While the impact generated by internet use is considerably lower than that emitted by patient transportation, an efficient management of digital activities can further contribute to emissions reduction. In real practices, the elimination of valueless information can minimize environmental costs. Conversely, the transitioning to energy-efficient devices presents a promising prospect for significantly enhancing overall impact [18]. However, despite the reachable benefits, it is essential to define the appropriateness of televisit, which lacks the specific instrumentation available in a hospital setting [19].

In conclusion, given the considerable adoption of televisit and remote care activities, the ability to objectively quantify impacts becomes fundamental to move beyond theoretical discussions and provide actionable insights. In contrast with the most employed approaches, this research considers all stages of inputs life cycle, from material extraction to waste management. Furthermore, a novel perspective has been defined by analyzing both direct and supporting activities, required for the visit execution. Sensitivity and data quality analyses will be investigated at later and more articulated stage, overcoming the limitations of the present study. Future work will extend the

evaluation to other environmental indicators and incorporate the social aspect, as essential issued to move towards a sustainable development framework.

Acknowledgements

The authors would like to thank the Hospital ASST Bergamo Est for the contribution to data acquisition. The author Savoldelli A. is funded by PON R&I 2014-2020 (DM 1061/2021). This work is partially funded by the National Plan for NRRP Complementary Investments - project n. PNC0000003 - AdvaNced Technologies for Human-centrEd Medicine (ANTHEM). This work reflects only the authors' views and opinions, neither the Ministry for University and Research nor the European Commission can be considered responsible for them.

References

[1] A. Luyten et al., Health impact studies of climate change adaptation and mitigation measures – A scoping review, *The Journal of Climate Change and Health* **9** (2023), 100186

[2] P.-P. Pichler et al., International comparison of health care carbon footprints, *Environmental Research Letters* **14**(6) (2019), 064004

[3] T. J. Lewis et al., Reduction in Chronic Disease Risk and Burden in a 70-Individual Cohort Through Modification of Health Behaviors, *Cureus* (2020),

[4] N. R. Jones et al., Survival of patients with chronic heart failure in the community: a systematic review and meta-analysis, *European Journal of Heart Failure* **21**(11) (2019), 1306–1325

[5] C. Morcillo Serra et al., Impact on the reduction of CO2 emissions due to the use of telemedicine, *Scientific Reports* **12**(1) (2022), 12507

[6] R. X. Qin et al., Building sustainable and resilient surgical systems: A narrative review of opportunities to integrate climate change into national surgical planning in the Western Pacific region, *The Lancet Regional Health - Western Pacific* **22** (2022), 100407

[7] Z. Lokmic-Tomkins et al., Assessing the carbon footprint of digital health interventions: a scoping review, *Journal of the American Medical Informatics Association* **29**(12) (2022), 2128–2139

[8] A. Savoldelli et al., Sustainability in Healthcare: Methods and Tools for the Assessment, (2023),.

[9] R. Sillcox et al., The environmental impact of surgical telemedicine: life cycle assessment of virtual vs. in-person preoperative evaluations for benign foregut disease, *Surgical Endoscopy* **37**(7) (2023), 5696–5702

[10] C. L. Thiel et al., Telemedicine and the environment: life cycle environmental emissions from in-person and virtual clinic visits, *npj Digital Medicine* **6**(1) (2023), 87

[11] ISO (International Organization for Standardization), 2006a, Environmental management - life cycle assessment - principles and framework (ISO 14040:2006). [online]. Available from https://www.iso.org/obp/ui/#iso:std:iso:14040:ed-2:v1:en (2006),.

[12] ISO (International Organization for Standardization), 2006b, Environmental management-Life cycle assessment-Requirements and guidelines (ISO 14044:2006). [online]. Available from https://www.iso.org/obp/ui#iso:std:iso:14044:ed-1:v1:en (2006),.

[13] A. Savoldelli et al., Environmental Sustainability of a Televisit Process: Definition of Parameters and Preliminary Results, (2024), 250–258.

[14] Å. Holmner et al., Carbon Footprint of Telemedicine Solutions - Unexplored Opportunity for Reducing Carbon Emissions in the Health Sector, *PLoS ONE* **9**(9) (2014), e105040

[15] https://www.footprintcalc.org/ ,.

[16] ecoinvent v3.0, https://ecoinvent.org/the-ecoinvent-database/data-releases/ecoinvent-3-0/,.

[17] R. Ravindrane and J. Patel, The environmental impacts of telemedicine in place of face-to-face patient care: a systematic review, *Future Healthcare Journal* **9**(1) (2022), 28–33

[18] N. Papadakis and D. A. Katsaprakakis, A Review of Energy Efficiency Interventions in Public Buildings, *Energies* **16**(17) (2023), 6329

[19] M. E. Reed et al., Patient Characteristics Associated With Choosing a Telemedicine Visit vs Office Visit With the Same Primary Care Clinicians, *JAMA Network Open* **3**(6) (2020), e205873

dHealth 2024
D. Hayn et al. (Eds.)

doi:10.3233/SHTI240017

User Personas for a "Better Design" of Nation-Wide EHRs Based on Thorough Expert Evaluation and Field Analysis: Modeling Users as Individuals Plus Family Members for an Enhanced Mapping of Healthcare Situations

Daniela GLATZ[a,1], Alexander ZEIRINGER[a], Johannes HARMS[b,1], René BARANYI[b], Karin KAPPEL[b] and Thomas GRECHENIG[b]
[a] *Research Industrial Systems Engineering (RISE), Vienna, Austria*
[b] *TU Wien, Institute of Information Systems Engineering, Research Group for Industrial Software (INSO), Vienna, Austria*

Abstract. Electronic Health Records (EHRs) are pivotal in prevention, therapy, and care. Their design necessitates the representation of users, activities, context, and technology. Among various participative and ethnographic design methods, user personas are an effective tool for encapsulating users in the design process. **Goals and methods:** This research focused on the creation of user personas for the design of a nation-wide EHR, specifically the German "elektronische Patientenakte" (ePA). We employed qualitative methods, such as field analysis and expert workshops, to generate, assess, and refine a set of user personas that can cover the complexity of real-life familial care environments. We used an innovative bottom-up approach applying a whole new process for persona generation especially in the context of family management. **Results:** The research yielded an initial set of five personas that accurately represent fictional user types. Importantly, at least two of these personas encapsulate the unique challenges inherent in family care work. These results provide a foundation for future work, which can utilize these personas for EHR design, as well as for further evaluation and refinement.

Keywords. Electronic Health Records (EHRs), User-centered Design (UCD), User Personas

1. Introduction

Taking care for sick relatives is an important familial task. In aging societies, this is not just a marginal phenomenon but an essential part of everyday life. For example, in 2021, over 4 million people (i.e. 84% of all those requiring care) in Germany received care at home, with more than 3 million of them being cared for by family members [1]. A still increasing amount of female part-time work indicates that this care work is predominantly provided by women [2].

[1] Corresponding Authors: Johannes Harms, TU Wien Institute of Information Systems Engineering, Research Group for Industrial Software (INSO), Vienna, Austria, E-Mail: johannes.harms@inso.tuwien.ac.at. Daniela Glatz, Research Industrial Systems Engineering (RISE), Vienna, Austria, E-mail: daniela.glatz@rise-world.com

According to the World Health Organization (WHO), more than half of their member states are implementing a national EHR system [3] with the main purpose of ensuring the communication of patient-centered medical data between patients and healthcare institutions [4]. The most evident advantages of EHRs, like an increase in the efficiency in healthcare delivery, reduction of paper-based errors in medical documents, encouragement of patients in health self-monitoring, and overall, in improving the quality of diagnosis and treatments seem promising to also ease the lives of those taking non-institutional care of others by allowing them fast and easy access to the caretakers' relevant data in case they are not yet or no more able to speak for themselves [5-7].

Germany launched the "elektronische Patientenakte" (ePA), a national opt-in EHR system, in 2021, which is offered as mobile applications by health insurers. To use it, the users must request access, allowing them to control which medical documents are shared and for how long [8]. Two years post-launch, only 595,000 enrollees used ePA [9]. actors influencing this low rate include trust issues, the opt-in model, access difficulties due to security measures, and varying attitudes towards data security and digital/health literacy [10] [11]. A user-friendly interface based on user-centered design (UCD) is also expected to be key to increasing the overall usage [10] [12]. UCD, central to this work, emphasizes end-user needs and behaviors in digital solution design and implementation. [12] Various methods are used for user inclusion, like for example personas [13] [14], participatory design [15] and ethnographic methods [16] and user journeys [17].

This study focuses on insured individuals and their health-related activities, including care for dependents. In this work user personas were developed based on established best guidelines (see [18]) in a bottom-up approach, assessing its applicability in resource-limited design and development processes. A comprehensive set of attributes and attitudes has been developed, explored and enriched in a field study, filling a gap in existing literature, where the focus lies on a comparably smaller sets of user characteristics. [19–21]. With Germany's shift to an opt-out ePA by 2025 [22] and the constantly increasing part of people in need for health care, understanding care-providing families through persona creation becomes also crucial.

2. Methods

The used methodology consisted of four big parts, which are outlined below:

(1) State-of-the-art Analysis and development of templates: Based on a thorough research of state of the art, an extensive set of attributes and attitudes to establish a foundation for insightful, relatable, and relevant personas representing the users within the ePA design process was defined. These attributes served as major input for a interview guideline as well as a template, based on which the personas development was set to be founded on.

(2) Field Analysis: 30 qualitative interviews have been conducted in a 2-step-approach: (a) in the first part, participants who were chosen based on a quota sampling representing the main sociodemographic characteristics of German residents with health insurance, were asked in an explorative open interview to tell their story of personal experiences with the health care system and the social environment involved. The goal of this first narrative part of the interview was to delve into the lives of potential users of the ePA to understand social dependences, relations, role-models, and environments. (b) In the second part of the interviews, standardized and structured questions delivered input to attributes and attitudes in case it had not been given by respondents already in

the first part of the interviews. The goal of the second part was to gather specific information around the relevant characteristics as a foundation to build personas in the next step.

(3) Content Analysis and Creation of Personas: (a) Based on the interview transcripts content was analyzed and interpreted qualitatively with the aim to identify further attributes and attitudes that might be important to build deeper and more complex personas. Based on the interview-content, personas have been created. (b) Parallel to this analysis, content of the interview has been used to generate AI-prompts for an alternative user personas creation based on a large conversational language model. The goals of this second set of personas was not only to validate the expert findings, but also to the ability of AI to generate visualization for the personas.

(4) Expert Evaluation of Personas in 2 steps: (a) In (3a) generated personas have been compared, adapted, and validated by experts in a workshop to capture any possible bias. Included experts represented the domains of user experience design, product management, requirements engineering and e-health. (b) After the evaluation of the personas based on qualitative analysis the AI-generated personas have been analyzed as well. Both were compared and evaluated and led to the final set of user personas.

3. Results

Based on the described methodology five personas have been deducted and validated. These personas (see Figure 1) include specific images, as well as systematically chosen psychographic, socio-demographic and medical attributes and attitudes. They also provide insights into the role families play in matters of non-institutional, informal health care, as at least two of them (Persona 2 and Persona 4) represent typical familial care constellations: parents taking care of their children (Persona 2), grown-up family members taking care of other grown-up, mostly older, family members (Persona 4). In addition, another persona (Persona 1) contains a transition into a period of life with increasing medical issues and hence need for support or even care by others. Involved experts have found the AI-based visualizations to be adequate to represent the described user types.

Also, attitudes across all personas deliver insights that need to be considered for the further design and development process of ePA, as 3 out of 5 personas are rather skeptical towards digitalization in general. It will be important to involve representatives of these user groups in the process and explore their reasons and how a user-centered designed ePA might be able to add value to their lives and overcome their skepticism. A hint may be found in Persona 4 who is not very interested in digitalization but finds the idea of ePA useful when it comes to organize her mother's health life and/or serve as her representative for certain tasks. In this regard, also the social environment of users will need to be explored deeper in the future as very often skepticism towards new tools, services or systems can be overcome by observing positive role models in the peer group. Methodologically this could mean that it might be reasonable not only to include individuals in the evaluation and optimization of artefacts but to work in groups to get more useful insights in social dynamics that influence the users' experience.

In total, the following characteristics were included for persona creation: Basic demographic attributes such as age, name, and gender as well as an image to improve the designers' empathy and imagination of the fictional users. The socioeconomic background of the personas was described through the income level, education, living

situation, and place of residency. Psychographic attributes like needs and expectations and factors that drive motivation and goals regarding ePA were included. A set of attributes described medical characteristics regarding possible medical challenges personas might face and how they interact with the healthcare system in general. Persona attributes additionally included aspects of the users' lifestyle, interests, and dietary habits. Attitudes of users towards, digitalization, data privacy, the healthcare system and ePA itself revealed additional insights into possible acceptance patterns. A description of the social environment was also essential to understand the potential impact family and friends have on the personas' behavior. Further attributes were health and digital literacy that reveal relevant information regarding the interaction with the ePA system. Additionally included was an attribute describing the devices and software that are used daily. Furthermore, aspects of accessibility, like limitations the personas might perceive while using technology were considered. This collection of attributes has been perceived as relevant to the design process according to the expert validation.

ID-Name-Age	Persona 1 – Gertrud – 73 years	Persona 2 – Max – 32 years	Persona 3 – Klaus – 57 years	Persona 4 – Andrea – 45 years	Persona 5 – Ayesha – 27 years
Image					
Background information	• Widowed, 2 adult children • Lives alone • Owns a house in the Bavarian countryside	• In a relationship, 1 child (3 years) • Rental flat in Frankfurt am Main • Works in marketing • Stressful lifestyle	• Married • Owns a house in a small town • Works in metal production • Unhealthy habits (alcohol, exercises rarely)	• Married, 1 child (17 years) • Rental house in the suburbs • Works in retail (part-time) • Takes care of her mother (dementia)	• Single • Shared apartment in Berlin • Studies and has a part-time job • In Germany for 1 year already, is learning German
Medical characteristics	• Problems with the thyroid gland, osteoporosis • Regular medical check-ups • Often forgets to take her prescribed thyroid medication • Low health literacy, high trust in her long-term GP	• A herniated disc, contracted while playing football • Regular physiotherapy • Highly motivated to implement the therapy recommendations • Low health literacy, relies mostly on expert opinions	• Regular severe headaches • Visits doctor rarely, afraid of diagnosis • Tries to keep a headache diary, often forgets to keep track • Low health literacy, relies mainly on his wife	• Diabetes mellitus type 2 • Regular medical check-ups • Takes her medication regularly • Joins her mother for doctor visits • Has acquired medical knowledge	• Suspicion of endometriosis, no diagnosis yet • Sees a doctor once a month • High treatment adherence, wants clarity about her condition • Low health literacy
Attitudes	• Sceptical toward digitalisation • Positive towards the healthcare system, aware of its advantages • Positive interest in ePA, supported by her children and her GP	• High interest in digital trends • Neutral towards the healthcare system • Sees a great benefit in the ePA, especially for the management of his child's medical records	• Sees digitalisation as a constraint, is afraid of AI • Sceptical about the healthcare system • Unsure if he benefits from the ePA	• Follows digital trends only casually • Often finds the healthcare system tedious due to regular interactions • Interested in using the ePA to manage her mother's healthcare	• Appreciate the convenience caused by digitalisation • Lacks understanding of the healthcare system (language barrier) • Sees the benefit of the ePA
Digital characteristics	• Low digital literacy • Uses her smartphone only for messaging and calls • Supported in digital matters by her children	• High digital literacy • Uses several digital tools both professionally and privately • Uses health apps regularly	• Average digital literacy, basic technical understanding • Uses mainly social media and messaging on his smartphone	• Average digital literacy, confident in using technology • Uses mainly social media and messaging on her smartphone • Her son supports her	• High digital literacy, confident in using new unfamiliar technology • Uses several different devices and software • Uses a period-tracking app
Social environment	• Her friends are in her age group and share her views on digitalisation	• His girlfriend has a stressful lifestyle as well • Friends and Family have similar attitudes	• His wife trusts in alternative treatments and is sceptical about the healthcare system	• Her husband and sister support her in caring for her mother • She has a close circle of friends who support her morally	• Her friends are mainly international students from various countries
Motivation and Goals	• Easier access to medical records for doctor visits • Overview of her medication	• Wants to implement therapy recommendations successfully • Taking care of his child's medical affairs	• Keep a consistent headache diary • Overcome fear of going to doctors • Actively work on his health	• She wants to manage her diabetes effectively • Organise her mother's healthcare effectively	• She wants clarity about her health conditions • She wants a better understanding of the healthcare system

Figure 1. Overview of the five ePA user personas.

For the further development of ePA the 5 personas will be applied to the practical work in a variety of ways, like for example as guidelines for specific design decisions, or to serve as a feature-prioritizing guideline.

4. Discussion

This work set out to create user personas that are ready to be used in the design process for the German „Elektronische Patientenakte" (ePA) - a nationwide EHR system. In addition, we expected these personas to provide valuable insights into non-institutional and informal care-work environments that mostly are and increasingly will be found in

familial settings, as western societies - such as the German - are constantly aging. An „ePA for all" will have to provide solutions for every kind of health problem and personal social setting, which makes a user-centered approach to tackling specific patient tasks even more urgent. A secondary goal of this research was to apply a specific bottom-up process for persona creation, assessing its applicability in resource-limited design and development processes and exploring the possibilities of a thoughtful use of generative AI. As a result, this work presents and contributes a set of five user personas that can be used for designing nationwide EHRs on a regular and continuous base.

When choosing suitable methods for creating the user personas, we followed established best guidelines (see [18]). Qualitative evaluation with design experts showed that our approach worked very well. The careful use of AI by selectively picking basic information about personas from literature in combination with expert evaluation limited the potential bias. Also, thorough qualitative research, systematically deducted sets of attributes and attitudes as well as persona templates proved to be crucial for the creation of comprehensive and coherent persona descriptions. However, during this research, specific limitations and need for future work have been identified:

- Collaboration on the same health record: It is very likely that one and the same health record will be used by more than one user (outside the medical or insurance systems, like picking up prescriptive medicine. In this regard, future research will have to provide more detailed insights into inner-familial care settings and enrich and very likely extend the initial set of five personas.
- Extreme user personas: The set of five personas will also have to stand the test of patients with very frequent and constant touchpoints with the medical system, like for example people with severe and/or chronic illnesses, pregnancy, etc.
- Prospective and Anti Personas: It will also be necessary to learn more about potential users who have not yet shown interest or the need to engage with ePA. Also, the creation of an „anti-persona" could be helpful in the process of prioritizing improvements and marketing communication.
- Practicability: The found personas will be applied in real-world design processes, as planned by the authors. This will enable further qualitative evaluation on a continuous base.
- Quantification: Based on the qualitative research, a quantitative study will have to follow, that includes a representative sample of the German population as a further way to validate and enrich our set of personas. That way, segments of society that might be under-represented in the quota sampling or the areas of expertise and experience of the experts of this study can be included in the set of personas.
- Methodological plurality: With an "ePA for all" on the horizon, the inclusion of users in the design process should be established as a matter of course. Besides continuous forms of user testing it will also have to include participatory and ethnographic methods as well as methods that are able to tackle social dynamics that are relevant for the acceptance of ePA as a national EHR system.

In summary, we conclude that user personas can be deducted based on a thorough qualitative research setting that is also the fundament for logic and consistent persona descriptions and images delivered by AI-tools. These personas also deliver valuable insights in real-life care environments within the family without having to specifically address or deduct these aspects in the interviews or AI prompts.

References

[1] Statistisches Bundesamt. Demografischer Wandel: Mehr Pflegebedürftige. Accessed: Jan 17, 2024. [Online]. Available: https://www.destatis.de/DE/Themen/Querschnitt/Demografischer-Wandel/Hintergruende-Auswirkungen/demografie-pflege.html

[2] C. Albrecht and B. Rude, 'Wo steht Deutschland 2022 bei der Gleichstellung der Geschlechter?', ifo Schnelldienst, vol. 22, pp. 01-11, 2022

[3] World Health Organization. Regional Office for Europe, From innovation to implementation: eHealth in the WHO European Region. World Health Organization. Regional Office for Europe, 2016. Accessed: Aug. 23, 2023. [Online]. Available: https://apps.who.int/iris/handle/10665/326317

[4] W. Currie and D. Finnegan, Integrating Healthcare with Information and Communications Technology. CRC Press, 2018.

[5] R. Henkenjohann, 'Role of Individual Motivations and Privacy Concerns in the Adoption of German Electronic Patient Record Apps—A Mixed-Methods Study', Int. J. Environ. Res. Public. Health, vol. 18, no. 18, p. 9553, Sep. 2021, doi: 10.3390/ijerph18189553.

[6] C. L. Anderson and R. Agarwal, 'The Digitization of Healthcare: Boundary Risks, Emotion, and Consumer Willingness to Disclose Personal Health Information', Inf. Syst. Res., vol. 22, no. 3, pp. 469–490, 2011.

[7] J. H. Weber-Jahnke and C. Obry, 'Protecting privacy during peer-to-peer exchange of medical documents', Inf. Syst. Front., vol. 14, no. 1, pp. 87–104, Mar. 2012, doi: 10.1007/s10796- 011-9304-2.

[8] 'Elektronische Patientenakte (ePA) | Datenschutz 2023', Datenschutz. Accessed: Sep. 16, 2023. [Online]. Available: https://www.datenschutz.org/elektronische-patientenakte/

[9] V. S. Biesdorf and M. Redlich, 'E-Rezept und ePA – die Schlüssel zur Digitalisierung des deutschen Gesundheitswesens?'. Accesssed: Jan 17, 2024. [Online]. Available: https://www.mckinsey.de/publikationen/e-rezept-und-epa-sind-schluessel-zur-digitalisierung-des-deutschen-gesundheitswesens

[10] T. Schmitt, 'Implementing Electronic Health Records in Germany: Lessons (Yet to Be) Learned', vol. 23, no. 1, Art. no. 1, Mar. 2023, doi: 10.5334/ijic.6578.

[11] O. Fennelly et al., 'Successfully implementing a national electronic health record: a rapid umbrella review', Int. J. Med. Inf., vol. 144, p. 104281, Dec. 2020, doi: 10.1016/j.ijmedinf.2020.104281.

[12] D. A. Norman, Ed., User Centered System Design: New Perspectives on Human-computer Interaction. Boca Raton: CRC Press, 1986. doi: 10.1201/9780367807320.

[13] A. Cooper, 'The Inmates are Running the Asylum', in Software-Ergonomie '99: Design von Informationswelten, U. Arend, E. Eberleh, and K. Pitschke, Eds., in Berichte des German Chapter of the ACM. , Wiesbaden: Vieweg+Teubner Verlag, 1999, pp. 17–17. doi: 10.1007/978-3-322- 99786-9_1.D.

[14] D. A. Norman, The design of everyday things, Revised and Expanded edition. New York, New York: Basic Books, 2013.

[15] D. Schuler and A. Namioka, Participatory Design: Principles and Practices, 1st ed. CRC Press, 2017.

[16] M. Hammersley and P. Atkinson, Ethnography: principles in practice, 3. ed., Reprinted. in An Informa business. London: Routledge, 2010.

[17] J. Kalbach and J. Kalbach, Mapping Experiences: A Guide to Creating Value Through Journeys, Blueprints, and Diagrams. O'Reilly, 2016.

[18] S. Mulder and Z. Yaar, The User is Always Right: A Practical Guide to Creating and Using Personas for the Web. in Voices That Matter. Pearson Education, 2006.

[19] C. LeRouge, J. Ma, S. Sneha, and K. Tolle, 'User profiles and personas in the design and development of consumer health technologies', Int. J. Med. Inf., vol. 82, no. 11, pp. e251–e268, Nov. 2013, doi: 10.1016/j.ijmedinf.2011.03.006.

[20] K. Bredies, 'Using System Analysis and Personas for e-Health Interaction Design', presented at the Undisciplined! Design Research Society Conference 2008, Sheffield Hallam University, Sheffield, UK, Aug. 2009. Accessed: Oct. 16, 2023. [Online]. Available: https://shura.shu.ac.uk/464/?template=drs2008_item

[21] E. Stojmenova, D. Dinevski, J. Guna, and M. Pogacnik, 'Persona-Based Design for Personal Health Information Management Systems', Int. Ser. Inf. Syst. Manag. Creat. EMedia CreMedia, no. 2013/1, Art. no. 2013/1, Aug. 2013.

[22] "Digitalgesetze im Bundestag beschlossen." Accessed: Jan. 22, 2024. [Online]. Available: https://www.bundesgesundheitsministerium.de/presse/pressemitteilungen/bundestag-verabschiedet-digitalgesetze-pm-14-12-23.html

dHealth 2024
D. Hayn et al. (Eds.)

doi:10.3233/SHTI240018

Should We Better Stick to Pen and Paper? An Empirical Investigation on Functionality, Privacy and Data-Security of Physiotherapy Telehealth Applications

Lukas MAUL[a,1], Ines KRAMER[b], Lena RETTINGER[a] and Franz WERNER[a]

[a] *University of Applied Science Campus Vienna, Competence Center Digital Health and Care, Vienna, Austria*
[b] *University of Applied Science Campus Vienna, Competence Center for IT Security, Vienna, Austria*

Abstract. Background: Telehealth and mHealth apps become increasingly popular in health professions such as physiotherapy calling for increased awareness on functionality, privacy, and data security. Objectives: This work presents a functionality, privacy, and data-security evaluation of four telehealth services commonly used in physiotherapy. Methods: We examined functionality and features, data protection, privacy implementations and data-security with a questionnaire and performed an in-depth investigation of the services. Results: Privacy and security relevant findings such as use of outdated webservers, problems with certificate renewal as well as questionable GDPR compliance were reported. Conclusion: Due to the privacy and security relevant findings in this analysis it can be concluded that there is a need for improvement in design, development, operation as well as regulation of telehealth apps and services.

Keywords. Telehealth, Physiotherapy, Data Security, Mobile Health

1. Introduction

Telehealth and the use of mHealth apps are increasingly common in health professions such as physiotherapy and telehealth can be feasible and also effective in terms of rehabilitation [1]. Official institutions such as the Austrian government promote the digitalization of health care services [2].

Thus, digitalization can also be critical in terms of privacy and data-security. Several studies report privacy and data-security issues in mHealth applications [3–8] and that privacy and data-security concerns are a relevant barrier for the use of telehealth or mHealth apps in allied health professions [9, 10].

This work presents a functionality, privacy, and data-security evaluation of telehealth applications in physiotherapy following a questionnaire developed within this evaluation. Four physiotherapy telehealth services based on similar functionality, comparable features, availability in German language and most popular in the field of

[1] Corresponding Author: Lukas Maul, University of Applied Science Campus Vienna, 1100 Vienna, Austria, E-Mail: lukas.maul@fh-campuswien.ac.at

physiotherapy in Austria were chosen for further analysis. Physiotherapy telehealth services enable remote communication between practitioner and patient to suggest training exercises, they offer feedback for patients and some of them also provide synchronous functionalities such as video chatting. They process sensitive health data such as pain level and patient feedback; the suggested training exercises can reveal a physical impairment or diagnosis of the patient. Our analysis should highlight whether privacy and data security are implemented in a best-practice approach by the corresponding service providers (SP).

2. Methods

A questionnaire consisting of 70 questions regarding functionality and features, availability, operating conditions and legal frameworks, such as licensing and billing, data protection and privacy implementations as well as data-security was created. The questions are based on findings of the 24h-QuaAALity research project conducted at University of Applied Science Campus Vienna and examining the OWASP Top 10 Mobile Security Risks version 2016 [11] as well as the OWASP Mobile Application Security Verification Standard MASVS v1.5.0 [12]. The authors supplemented physiotherapy telehealth specific questions based on a team discussion. This questionnaire was sent by email to four telehealth SP who were invited to co-operate in answering the questionnaire. Unfortunately, co-operations couldn´t be established, therefore an external empirical investigation was conducted by the researchers. This approach is comparable with other studies in this field of research [3, 8, 13, 14]. Thus, without co-operation, the questionnaire~~s~~ could not be answered to full extent. The following results present the physiotherapy specific functionality, data protection and privacy implementations as well as data-security parts of the questionnaire which could partly be answered by the research team between April and June 2023. The full questionnaire is not yet published but the authors are pleased to provide it on e-mail request.

2.1. Physiotherapy specific functionality analysis

For the functionality analysis, fictional therapist's as well as patient's accounts were created in each of the four investigated telehealth services. Data of a fictional rehabilitation scenario was entered over the course of five days by two researchers in April 2023 to test the full functionality of the services and to be able to answer all eleven physiotherapy specific functionality questions (Q1 to Q11) listed in Table 2.

2.2. Data protection and privacy implementations

To analyze data protection and privacy compliance with the General Data Protection Regulation (GDPR) [15], a two-step method was chosen. First, the SP homepages as well as the apps were screened for privacy policy and GDPR related documents. Secondly, after entering the fictional rehabilitation scenario for five days, a request for access and copy of entered personal data following article 15 GDPR [15] was sent to the SPs via email in May 2023. Additionally, the questionnaire encompassed four data protection

questions. As we could not find exhaustive answers to those questions without support of the SPs, only one question (Q12) is partly discussed in chapter 3.2.

2.3. Data-security

Due to the lack of cooperation, active tests against the productive systems of the application SP were not performed in this evaluation. Therefore, non-invasive tests have been selected that don´t have any impact on the availability of services and remain within normal user behavior. Data-security testing followed a similar approach as Zhao et al. examining privacy and security of mHealth applications via a static and dynamic analysis using MobSF, app store information and privacy policies [13]. MobSF (v3.6 beta) is an open-source mobile security testing framework used for pen-testing, malware analysis and security assessment [16]. All Android apps were directly installed from Google Play Store on the phone. Android Debug Bridge (1.0.39) [17] was used to pull the apps from the phone. Four of five investigated apps are bundled apps, which means they consist of several split APK files. MobSF cannot manage split APKs in a consistent analysis, therefore each of these APK files needed to be analyzed separately. The dynamic testing capabilities with MobSF were limited, because four of five investigated apps are ARMv8 compiled binaries, which are not supported by emulators. For testing network traffic encryption these apps have been installed and examined on a rooted Android phone using Magisk [18] and FRIDA [19]. In addition, we conducted security analysis of corresponding web applications to test TLS configuration, authentication, database endpoints, DNS records and running services as well as third party services for video chatting. 13 questions on data security were stated in the questionnaire of which eight could partly be answered with this approach. Those questions (Q13 to Q20) are discussed in chapter 3.3.

3. Results

The four investigated services all provide apps for practitioners, in the following context called “therapist”, and other users, named “patient”. Table 1 shows the available app interfaces for the roles therapist and patient. Results for Web apps and Android apps are presented in this study and were pseudonymized due to security relevant findings.

Table 1. App interfaces for roles therapist and patient

App	Role	Android	Browser / Web app
A	Patient	X	X
	Therapist	-	X
B1	Patient	X	X
B2	Therapist	X	X
C	Patient	X	X
	Therapist	-	X
D	Patient	X	-
	Therapist	X	-

In apps A and C, the patients are administered by the therapist over a Web app and patients use either Web or Android apps. SP B offers role defined apps and websites, which are named B1 for patients and B2 for therapists. App D can be used by both

therapists and patients. Apps A to C are paid licensed services for therapists, but without fee for patients. App D supports mobile apps only and is a free service for both therapist and patient. Thus, we examined five apps of four SP and four web interfaces of three SP.

3.1. Physiotherapy specific functionality analysis results

The functional analysis highlighted differences in the applications' content. Apps A, B1/B2 and C provide extensive exercise libraries, whereas app D solely has a functionality for therapists to create their own exercise instructions within the application. In addition, both apps A and B1/B2 offer a video chat function for synchronous telehealth service. App A as well as app D are CE certified medical products within risk class 1 [20]. The results of the functionality analysis questionnaire are summarized in Table 2. Note that the results in Table 2 suite as qualitative information on the apps functionality and do not serve a rating purpose.

Table 2. Questions and answers to physiotherapy specific functionality analysis of applications A to D

#	Question	Answer App A	Answer App B1/B2	Answer App C	Answer App D
Q1	Which exercises (exercise categories) does the tool contain?	8 main categories	11 main categories	11 main categories	No categories
Q2	How many different exercises can the therapist choose from?	450	7,154	789	19
Q3	Are there exercises for the right and left side (if applicable)?	Yes	No	No	No
Q4	How does the therapist find exercises that suit the patient?	Search terms / Categories	Search terms / Categories	Search terms / Categories	Selection menu
Q5	Can own exercise videos and/or descriptions be created?	Personalized exercise description, no video / pictures	Personalized exercise video / pictures	Personalized exercise video / pictures	Personalized exercise video / pictures
Q6	Can selected exercises be modified? If yes, how? Which parameters can be set?	Modification possible on different levels	Modification possible on different levels	Modification possible on different levels	All parameters can be modified
Q7	In what form do patients receive their exercises?	Patient App	Patient App	Patient App	Patient App
Q8	Can the exercises be monitored by the therapist? If yes, what information is provided?	Direct response possible via (video) chat function	Direct response possible via (video) chat function	No direct response - only post exercise feedback	No direct response – only post exercise feedback
Q9	Is feedback given to the patient? If yes, how?	Direct response possible via (video) chat function	Direct response possible via (video) chat function	No	No
Q10	What happens if the patient experiences pain while exercising?	In exercise safety feedback + post exercise feedback	Post exercise feedback	Post exercise feedback	Post exercise feedback

Q11	Are there any other additional functions that have not yet been mentioned?	Safety questions before exercise program	Clinical questionnaires and reports	Reports + Therapist search function	-

3.2. Data protection and privacy implementations results

All four SP answered the request for access and copy of entered personal data following article 15 GDPR [15] within four weeks. Only one SP partially fulfilled the request and sent a copy of stored personal data in the form of a user account database extraction but no information on other user entered data for example patient feedback to the therapist was delivered. Two SP answered with a general information concerning their GDPR compliance. One SP argued that following GDPR, they only act as data processor, managing data on behalf of the data controller. Further they stated, this data controller would be one's general practitioner who would be the appropriate contact for inquiries regarding personal data. Notably, this standardized answer was sent to both fictional patient and therapist accounts.

Question Q12: Are cloud services or third-party services used? If yes, which? During installation of the apps all users were requested to confirm to the data protection policy and terms of use of the app. One app did not provide explicit information which third party services will process user data with what purpose. All tools use third party services, e.g., all of them use a Google Firebase data storage. The tracker analysis using MobSF highlighted that all five apps use Google Firebase Analytics and four of five apps use Google CrashLytics. Malware Analysis revealed that none of the apps contained any known malware.

3.3. Data-security results

Question Q13 Is the product open source? None of the apps is open source, app A to C are paid licensed services for therapists. All apps are free to use and install for patients and can be found in the Google app store. Three of five apps use code obfuscation methods and anti-debug code. Four of five apps implement detection of a rooted phone to abort installation.

Question Q14 How are therapists and clients authenticated? Authentication methods in the apps and web interfaces are the same. Three SP use different authentication methods for patients using an app code and/or birthyear/birthday of the patient. The main authentication method is email and password authentication, only one SP offers two factor authentication via mobile phone for therapists. One SP implemented a password policy with very low complexity, even 'aaaaaa' evaluates as a valid password.

Question Q15 How can forgotten credentials be reset? Apps which support email and password authentication provide a mechanism for changing the password via email ownership verification.

Question Q16 Are there precautions against brute force attacks on the access data? This was not tested intensively, but there was no limit in login attempts mentioned and found.

Question Q17 What permissions does the client-side application (APK) need? MobSF permission analysis showed some extensive use of permissions for two of five apps. Investigated apps use two to nine so called dangerous permissions, which can access resources on the phone and require user consent during installation. Most had a

legit use case in the apps, such as CAMERA and AUDIO access for video chatting or recording own videos, or a GPS location permission for a service to search a therapist's location. For one app we could not find a related use case for following permissions READ_PHONE_STATE, READ_PHONE_NUMBERS and GET_ACCOUNTS.

Question Q18 Is data encrypted and authenticated between client and server, which TLS version is used? The APK static analysis showed issues with a high-risk score in four of five apps, such as enabled clear text traffic, permitted clear text traffic to specific domains or accepted self-signed certificates. Further examination showed these are issues related to SP development and testing environments, which is a bad practice to include in the app. All apps use certificate pinning for SP certificates. The recorded Transport Layer Security (TLS) handshakes in the dynamic app testing showed that three of five apps use Version 1.2 and two of five apps 1.3. Testing web server TLS configurations with Qualys SSLabs [21] highlighted that two SP support older TLS versions 1.0 or 1.1 and only one SP supports TLS 1.3. In none of the tested web interfaces any attempted TLS attack succeeded.

Question Q19 Is the communication between patient and therapist encrypted end to end? As far as we have seen the asynchronous communication works over firebase database service or third-party nodes with transport layer encryption. Two SP offer video chatting function of third-party services, which also support end-to-end media encryption. During our tests we did not verify the used encryption methods.

Question Q20 Are there update mechanisms in the software, how often are updates and patches applied? Public app store information reveals that two of five apps did not get any updates or patches for more than a year. One app still supported Android version 4.1, showing several security implications in the scoring of MobSF static analysis, such as the Janus vulnerability for APK v1 signatures. We did not find internal update mechanisms of the apps.

Further security related findings: During our testing phase one SP had maintenance issues, twice expired web server certificates were discovered and reported to this corresponding SP. Another SP was using an outdated webserver first released in 2014 with several security issues. One app offers a therapists localization service for patients to find a nearby therapist and sends all public reachable therapist user accounts with unfiltered fields to the patient app. These contain not only consented information such as email and address of the therapists, in some cases even the activation token or access token were displayed, including information about account creation, last login or the purchased license of the therapists.

Important questions, whether external security audits were performed and if those are publicly available could not be answered due to missing SP co-operation, making it hardly possible for non-professionals to quickly assess whether an application is secure or not.

4. Discussion

The present study provides findings from a functionality, privacy, and data-security evaluation of four telehealth services commonly used in physiotherapy. In summary, the evaluation shows that two of the four evaluated services are CE certified medical products in risk class 1. All services offer asynchronous functionalities such as exercise prescription and two services additionally offer synchronous elements such as video chatting. Only one SP answered the GDPR [15] request with delivering information on

stored personal data. All SPs implemented basic security measures such as authenticated and encrypted network traffic at transport level (TLS) and certificate pinning. None of the applications contained malware and only one or two trackers were found in the apps. One app uses dangerous permissions which can hardly be legitimated with the use case of the service. Privacy and security-relevant findings such as the use of outdated webservers, problems with certificate renewal as well as GDPR compliance were reported.

Limitations of the study are that no co-operations could be established with the SP to fully answer the questionnaire. If the SP would have co-operated, further and more extensive security tests could have been carried out. Further only a small number of four services could be tested due to the still limited distribution of telehealth services in physiotherapy. Therefore, this study can only suite as an example and general statements are not possible. Nevertheless, Iwaya et al. also report unnecessary permissions in mental health apps [5] and Alfawzan et al. report poor data privacy and security standards in their study on women´s mHealth apps [6]. Therefore, findings of this study are in line with reports in other works [3–8], highlighting possible need for better development and regulation.

Answering the title question, we should not stick to pen and paper, as the investigated applications provide various functionalities which can promote digitalization and further increase efficiency as well as quality within the physiotherapeutic treatment. Thus, due to the privacy and security-relevant findings in the analysis of all four services, it can be concluded that there is a need for improvement in design, development, and operation. Privacy and data security are not yet implemented in a best-practice approach and do not follow privacy by design principles. It is also notable that – given the complex and time-consuming research needed – it is currently not possible to quickly assess the functionality, security, and safety of telehealth applications by non-experts without publicly available and easily comprehensible information on for example security audits. Hence a need for better regulation of this sector to allow health professionals to quickly identify usable and secure applications and services can be expressed.

Funding: This work is produced as part of the research project “Telehealth Blocks” funded by the City of Vienna, Magistratsabteilung 23, under grant number MA23-338474-2021-2.

Conflicts of Interest: The authors declare no conflict of interest.

References

[1] M. T. Muñoz-Tomás et al., Telerehabilitation as a Therapeutic Exercise Tool versus Face-to-Face Physiotherapy: A Systematic Review, *International Journal of Environmental Research and Public Health* **20**(5) (2023), 4358

[2] Bundeskanzleramt; Bundesministerium Kunst, Kultur, öffentlicher Dienst und Sport; Bundesministerium Finanzen, Digital Austria Act Für mehr Wohlstand, Sicherheit und neue Chancen durch Innovation - Beilage zum Ministerratsvortrag zum Digital Austria Act,. (2023),. Accessed: Jul. 20, 2023. [Online]. Available: https://www.digitalaustria.gv.at/dam/jcr:fd6e0768-2917-48c2-83df-456d3e75a837/MRV%20Digital%20Austria%20Act%20DAA-61_10_bei%20BF.pdf

[3] D. He et al., Security Concerns in Android mHealth Apps, *AMIA ... Annual Symposium proceedings. AMIA Symposium* **2014** (2014), 645–54

[4] L. H. Iwaya et al., Security and Privacy for mHealth and uHealth Systems: A Systematic Mapping Study, *IEEE Access* **8** (2020), 150081–150112

[5] L. H. Iwaya et al., On the privacy of mental health apps, *Empirical Software Engineering* **28** (2022),
[6] N. Alfawzan et al., Privacy, Data Sharing, and Data Security Policies of Women's mHealth Apps: Scoping Review and Content Analysis, *JMIR mHealth and uHealth* **10** (2021),
[7] K. Knorr and D. Aspinall, Security testing for Android mHealth apps, *2015 IEEE Eighth International Conference on Software Testing, Verification and Validation Workshops (ICSTW)* (2015), 1–8
[8] J. Müthing et al., Server-Focused Security Assessment of Mobile Health Apps for Popular Mobile Platforms., *Journal of medical Internet research* **21**(1) (2019), e9818
[9] L. Rettinger and S. Kuhn, Barriers to Video Call–Based Telehealth in Allied Health Professions and Nursing: Scoping Review and Mapping Process, *Journal of Medical Internet Research* **25** (2023), e46715
[10] G. D. Giebel et al., Problems and Barriers Related to the Use of Digital Health Applications: Scoping Review, *Journal of Medical Internet Research* **25** (2023), e43808
[11] C. Holguera et al., *OWASP Mobile Application Security Testing Guide*, **1.5.0**. (2022),. Accessed: May 12, 2023. [Online]. Available: https://github.com/OWASP/owasp-mastg/
[12] S. Schleier et al., *OWASP Mobile Application Security Verification Standard*, V2.0.0. OWASP Foundation, (2023),.
[13] W. Zhao et al., Security and Privacy Analysis of mhealth Application: A Case Study, *2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)* (2020), 1882–1887
[14] D. Schultchen et al., Stay Present with Your Phone: A Systematic Review and Standardized Rating of Mindfulness Apps in European App Stores, *International Journal of Behavioral Medicine* **28**(5) (2021), 552–560
[15] EUROPEAN PARLIAMENT AND THE COUNCIL OF THE EUROPEAN UNION, VERORDNUNG (EU) 2016/ 679 DES EUROPÄISCHEN PARLAMENTS UND DES RATES - vom 27. April 2016 - zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/ 46/ EG (Datenschutz-Grundverordnung), *Amtsblatt der Europäischen Union* (2016),
[16] A. Abraham and 85 more, Mobile Security Framework (MobSF),. (2022),. Accessed: Jun. 06, 2023. [Online]. Available: https://github.com/MobSF/Mobile-Security-Framework-MobSF
[17] G. The Android Open Source Project., Android Debug Bridge (ADB),. Accessed: Apr. 10, 2023. [Online]. Available: https://developer.android.com/tools/adb
[18] Magisk Manager, Magisk Beta, (2020),. https://magiskmanager.com/magisk-beta/ (accessed Jan. 17, 2024).
[19] Frida Instrumentations, Frida A world-class dynamic instrumentation toolkit,. https://frida.re/ (accessed Jan. 17, 2024).
[20] EUROPEAN PARLIAMENT AND THE COUNCIL OF THE EUROPEAN UNION, *Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC (Text with EEA relevance)Text with EEA relevance*. (2023),. Accessed: Jan. 23, 2024. [Online]. Available: http://data.europa.eu/eli/reg/2017/745/2023-03-20/eng
[21] I. Ristić, Qualys SSL Labs, (2024),. https://www.ssllabs.com/ (accessed Jan. 21, 2024).

dHealth 2024
D. Hayn et al. (Eds.)

doi:10.3233/SHTI240019

Editing Physicians' Responses Using GPT-4 for Academic Research

Magdalena T. WEBER[a,1], Jannik SCHAAF[a], Holger STORF[a], Thomas O.F. WAGNER[b], Alexandra BERGER[c] and Richard NOLL[a]
[a] *Institute of Medical Informatics, Goethe University Frankfurt, University Hospital Frankfurt, Frankfurt, Germany*
[b] *European Reference Network for Rare Respiratory Diseases (ERN-LUNG), University Hospital Frankfurt, Frankfurt, Germany*
[c] *Reference Center for Rare Diseases (FRZSE), University Hospital Frankfurt, Frankfurt, Germany*

Abstract. The integration of Artificial Intelligence (AI) into digital healthcare, particularly in the anonymisation and processing of health information, holds considerable potential. **Objectives:** To develop a methodology using Generative Pre-trained Transformer (GPT) models to preserve the essence of medical advice in doctors' responses, while editing them for use in scientific studies. **Methods**: German and English responses from EXABO, a rare respiratory disease platform, were processed using iterative refinement and other prompt engineering techniques, with a focus on removing identifiable and irrelevant content. **Results**: Of 40 responses tested, 31 were accurately modified according to the developed guidelines. Challenges included misclassification and incomplete removal, with incremental prompting proving more accurate than combined prompting. **Conclusion**: GPT-4 models show promise in medical response editing, but face challenges in accuracy and consistency. Precision in prompt engineering is essential in medical contexts to minimise bias and retain relevant information.

Keywords. Artificial Intelligence, Data Anonymisation, Natural Language Processing, Medical Informatics

1. Introduction

In digital health communication, EXABO is emerging as an integral platform [1]. Central to the European Reference Network for Rare Respiratory Diseases (ERN-LUNG), EXABO provides expertise in the management of rare respiratory diseases, addressing challenges such as delayed diagnosis and lack of expertise. With the online expert advice system, patients, relatives, and physicians can ask specific questions about rare respiratory diseases to experts who are specialised and experienced in this field [1, 2].

In digital healthcare, the focus on the use of Artificial Intelligence (AI) to improve the data processing of physician-patient interactions offers considerable potential. Our research focuses on editing physician responses, with the aim of improving research utility while protecting Protected Health Information (PHI). We propose the use of Generative Pre-trained Transformer (GPT) models for PHI protection and response

[1] Corresponding Author: Magdalena T. Weber, Institute of Medical Informatics, Goethe University Frankfurt, University Hospital Frankfurt, Frankfurt, Germany, E-Mail: ma.weber@med.uni-frankfurt.de

refinement. The process of tailoring and adapting information in our study is closely related to the concept of prompt engineering, which is central to guiding GPT to produce desired results. Prompt engineering involves the strategic design and refinement of input prompts to effectively communicate the task or information need to the model [3]. In this study GPT-4, a Large Language Model (LLM) developed by OpenAI, was chosen for its good performance and accessibility [4].

In our specific use case, physician responses on the EXABO platform are not only populated with PHI, but also with other information irrelevant to further clinical research and analysis, such as references to technical problems or apologies for late responses. While there have been studies on anonymising medical free text using GPT models [5], there is a notable gap in research that specifically focuses on the targeted adaptation of information for study-specific purposes. Our study seeks to fill this research gap.

2. Methods

A comprehensive analysis was performed on the EXABO dataset, consisting of 93 question-answer pairs in German and 49 pairs in English, relating to a variety of rare respiratory diseases. Ten English responses were selected to optimise the prompt. For the analysis, 20 question-answer pairs were selected from each language and edited once with the final prompt. The questions used for the analysis do not contain examples that were used for optimisation. A German prompt was designed for the German responses and an English prompt for the English responses. The translation was done with DeepL [6].

Prompt engineering involves various techniques [3, 7, 8] to optimise the input prompts to obtain the desired output. The following techniques were used in our study.

(1) Conditional Prompting: The use of conditional statements in prompts can guide the model to produce outputs based on certain conditions or criteria, thereby increasing the relevance of the response to the given scenario.
(2) Negative Prompting: Specify not only what you want the model to do, but also what you don't want it to do. This can help to avoid unwanted outputs or biases.
(3) Delimiters, represented by unique sequences of special characters, serve as markers that enhance the clarity of the prompt and help the model to focus on specific aspects during processing. They are crucial for structuring prompts in complex tasks.
(4) Providing context and objective serves displaying the purpose of the task and guides the model towards the desired result.

The framework we have constructed contains five sections. In the first section we provide the context and in the second section we describe the objective, giving a basic understanding of the task. The instructions and guidelines needed to complete the task follow in the third and fourth sections. The fifth section prompts the model to start the task after receiving the instructions.

The initial prompts were tested and iteratively refined with ChatGPT 4 using data from the original EXABO dataset. The model was asked to edit its own prompts based on given human feedback (meta-prompting). We identified the following context cues in the physicians' responses of the EXABO dataset that should be removed: (1) Identifications: name, title, and organisation of the author; (2) Introductions: off-topic introductions, such as excuses for delays; (3) Quotations: to focus on the medical content, we aim to seamlessly integrate quotations.

We explored two scenarios: An incremental prompting approach, where tasks or subtasks are presented to the model message by message, providing step-by-step guidance, and presenting them all at once in a combined prompt.

When analysing the edited responses, we categorised errors into eight groups, based on their deviation from the guidelines and general instructions defined in the prompt. Issues were identified by screening and comparing the edited answers with the original answers and counting the occurrence of different errors. Furthermore, the count of edited answers adhering completely to the instructions was taken.

3. Results

Presented in Fig. 1 is the English version of our prompt to edit the answers of EXABO:

```
#CONTEXT#
This task is part of a preparatory phase for a study aimed at assessing a ChatBot's proficiency
in responding to patient inquiries. To facilitate a comparative analysis, both the ChatBot's
responses and those provided by a medical doctor will be evaluated by independent raters.

#OBJECTIVE#
Editing the physician's responses to ensure anonymity, preventing immediate recognition of the
author by the raters. Crucially, all edits must preserve the original medical content and tone,
while effectively depersonalizing the authorship.

#GENERAL INSTRUCTIONS#

Do not change the original text:
Retain the original wording  without making any changes or adjustments except for the changes
required in the guidelines. This also applies to text in brackets.

Retain relevant third party information and external sources:
Retain third-party information that provides valuable context or background information, but
does not contribute to identification.
Maintain links to relevant medical or informational resources.

#GUIDELINES#
Provided guidelines are to be applied to the physician's answer sequentially, here is the first
guideline:

1. Remove identifying information:
     Remove all names, titles and organizations that could identify the sender or persons
          mentioned, including indirect references.
     Mark the removal with "[REMOVED: Identification]".

#START TASK#
If you understand, ask for the physician's answer.

----------------------------------------next message----------------------------------------

2. Removal of introductions and excuses:
     Remove introductory sentences with excuses for delays or technical problems.
     Mark these with "[REMOVED: Introduction]".
     Keep introductory sentences that acknowledge the question as relevant to maintain tone and
          context

----------------------------------------next message----------------------------------------

3. Adjust the presentation of quotations:
     Remove quotation marks around quotes for seamless integration into the text
     Remove introductions from quotes, such as "... responded to your question:", and mark them
          with "[REMOVED: quote]".
```

Figure 1. English prompt with the incremental prompting approach.

Delimiters such as "# ... #" were used strategically to provide structural cues for interpreting different sections. Task initiation is indicated in the #START TASK# section, which serves to engage the model with the task at hand and establishes the beginning of the incremental prompting and editing.

An analysis of the most common problems, categorised to visualise their frequency and extract valuable insights, is presented in Table 1.

Table 1. Categorisation of edited responses after applying the guidelines via the GPT prompt.

	German	English	Total
Complete Implementation of Instructions	18	13	31
Incorrect classification of Identification (Guideline 1)	1	2	3
Incomplete removal of Identification (Guideline 1)	0	2	2
Incorrect classification of Introduction (Guideline 2)	1	1	2
Incomplete removal of Introduction (Guideline 2)	0	0	0
Incorrect classification of Quotation (Guideline 3)	0	4	4
Incomplete removal of Quotation (Guideline 3)	1	1	2
Removal of relevant Information (General Instructions)	1	1	2
Change of Wording (General Instructions)	0	1	1

A review of the edited responses shows a positive trend, with 31 out of 40 responses accurately edited according to the guidelines provided.

The comparison between incremental prompting and combined prompting is summarised in Table 2. The results indicate a higher accuracy rate when the incremental prompting method was used.

Table 2. Comparison of incremental and combined prompting.

	Combined Prompt	Incremental Prompt	Difference
Complete Implementation of Instructions	7	15	+8
Incorrect classification of Identification (Guideline 1)	6	1	-5
Incomplete removal of Identification (Guideline 1)	2	1	-1
Incorrect classification of Introduction (Guideline 2)	0	0	0
Incomplete removal of Introduction (Guideline 2)	0	0	0
Incorrect classification of Quotation (Guideline 3)	2	2	0
Incomplete removal of Quotation (Guideline 3)	4	1	-3
Removal of relevant Information (General Instructions)	1	2	+1
Change of wording (General Instructions)	0	0	0

Table 3 presents and analyses examples of edited physician responses. In the first example, the model failed to integrate the quotation. As per guideline sentences like '... responded to your question' should be removed and marked. It also incorrectly removed an introduction that should have been retained as it did not meet the criteria for removal set out in the guidelines. It is an excuse, but there is no connection to a delay or technical problem. The second example presented further challenges as the model did not mark the removal of the sentence introducing the quotation, which is against the guidelines for clear identification of edited content. Removal of identification was also incomplete.

Table 3. Two examples of edited responses resulting from the incremental prompt.

Question	Answer	Edited Answer
Hello, Our son, born 2018 prematurely 29+1 weighing 700 g, has developed very severe BPD. He is now almost 18 months old and his condition has not improved a slightest bit. He is on mechanical ventilation in NICU with high settings on ventilator and FiO2 100%. He has never been off mechanical ventilation and his oxygen requirement have since 3rd day of life always been very high (70-100%). He has been previously on high frequency ocillator and NO. He has had very many courses of steroids: dexamethasone, methyl-predinosolone, and is currently on hydrocorti-sone. He is sedated with dormicum and dexdor for almost 1 year now. His other main daily medications are: atrovent, pulmicort, ACC, Sildenafil, Bosentan, Verospiron. His feeding is partly through nasogastric tube, but most part is parenteral with Numetah. Our doctors in Estonia have no further ideas how to improve his condition. We would very much appreciate any advice or guidelines how to proceed and help him or recommendations about options where to turn to for help. We thank you in advance fro any help you can give. Best regards, (parents)	Dear parents, Prof. Sch…, head of our neonatology Dept. has given the following answer: "Dear family, I am sorry that I also have few ideas about how to help your son Albert really. In my opinion, the colleagues have done all options in this severe case of BPD. The medication leaves hardly any effective changes open. I don't know if your son is still intubated or has a tracheostoma. This may help to make ventilation easier. Of course, I can't say more without X-rays. Best regards Rolf Schlößer Prof. Dr. med. Rolf Schlößer Universitätsklinik Frankfurt" Best wishes Thomas Wagner, M.D., Coordinator of ERN-LUNG	Dear parents, [REMOVED: Identification] has given the following answer: [REMOVED: Introduction] In my opinion, the colleagues have done all options in this severe case of BPD. The medication leaves hardly any effective changes open. I don't know if your son is still intubated or has a tracheostoma. This may help to make ventilation easier. Of course, I can't say more without X-rays. Best regards, [REMOVED: Identification] Best wishes, [REMOVED: Identification]
My question is : which treatment or follow up for a fibrotic pulmonary sarcoidosis patient with recent inflammatory activity (uveitis/pericarditis) and hugh degree fatigue, muscles weakness and dyspnea with dramatic air privation episodes (spring / summer) obstructive asthma also and immune condition T / B partial deficiency ?	Hello, Prof. F. Bon…, Essen. Germany, wrote: "For the sarcoidosis manifestations, a treatment with Prednisolon and Methotrexate should be taken into consideration, but after a careful immunological evaluation of the T/B partial deficiency (do you mean a t- or b-cell dysfunction?) in a specialized center." Best regards Dr TOF Wagner, Coordinator ERN-LUNG	For the sarcoidosis manifestations, a treatment with Prednisolon and Methotrexate should be taken into consideration, but after a careful immunological evaluation of the T/B partial deficiency (do you mean a t- or b-cell dysfunction?) in a specialized center." Best regards, [REMOVED: Identification] Coordinator, [REMOVED: Identification]

4. Discussion

This study examined the effectiveness of GPT-4 in editing responses, assessing its ability to maintain the clinical relevance and specificity required for medical advice, while removing identifiable markers of responding physicians and pre-identified contextual cues.

While prompt editing is a critical factor in determining the quality of output, the inherent unpredictability of GPT-generated results must be considered, particularly in

professional or academic contexts where reproducibility is a key aspect of reliability and validity. Bias minimisation is particularly important in medical settings [9].

Our prompt provides clear instructions for retaining the essential text in responses, which is essential for our further research. In the absence of these precise instructions, the model tends to reformulate, which can be beneficial by correcting errors and improving the structure of the text. However, this rewording carries the risk of including incorrect information or omitting important details.

In our exploration of GPT-4's capabilities, it was noted that despite significant investment in development, its consistency in providing accurate responses falls short, with approximately one-fourth of responses deviating from the intended outcome. While these results suffice for our needs, there is clear room for improvement.

One potential avenue for enhancing GPT-4's consistency lies in refining the prompt. The Chain of Thought (CoT) approach stands out as a promising method, encouraging the model to reason and articulate its answers step by step [10]. However, for our specific application, as we are planning to interact with the model through an API, this approach proves impractical. In our scenario, a fixed prompt is required for each query, precluding real-time model guidance.

This predicament underscores the importance of striking a balance between specificity and generality in prompt engineering. While a narrowly tailored prompt increases the likelihood of a correct response, it may not cover all scenarios.
Problems occurred more frequently when the task was presented in a combined prompt than when it was divided into subtasks. With incremental prompting, the model showed a more consistent ability to understand and follow individual instructions.

While existing research has yet to explore the approach of incremental prompting, into API-driven interactions with GPT-4, our findings suggest promising avenues for its integration. Moving forward, we are committed to investigating and implementing incremental prompting to enhance model consistency and performance within our API-driven framework.

References

[1] Walther D, Steinmann O, Wagner TOF, Storf H. Implementation of EXABO - an expert advisory board for the european reference network for rare respiratory diseases. Vol. 270. Studies in health technology and informatics; 2020. p. 986–990. doi:10.3233/SHTI200309
[2] EXABO. [accessed 2024 Jan 19]. https://exabo-lung.mig-frankfurt.de/
[3] Prompt engineering. [accessed 2024 Jan 19]. https://platform.openai.com/docs/guides/prompt-engineering
[4] OpenAI. ChatGPT (Large language model). [accessed 2024 Jan 19] . https://chat.openai.com
[5] Liu Z, Huang Y, Yu X, Zhang L, Wu Z, Cao C, Dai H, Zhao L, Li Y, Shu P, et al. Deid-gpt: zero-shot medical text de-identification by gpt-4. arXiv; 2023. doi:10.48550/arXiv.2303.11032
[6] DeepL Translate: The world's most accurate translator. [accessed 2024 Jan 19]. https://www.deepl.com/translator
[7] White J, Fu Q, Hay S, Sandborn M, Olea C, Gilbert H, Elnashar A, Spencer-Smith J, Schmidt DC. A prompt pattern catalog to enhance prompt engineering with chatgpt. ArXiv; 2023. doi:10.48550/arXiv.2302.11382
[8] Meskó B. Prompt engineering as an important emerging skill for medical professionals: tutorial. Journal of Medical Internet Research. 2023. doi:10.2196/50638
[9] Siala H, Wang Y. SHIFTing artificial intelligence to be responsible in healthcare: A systematic review. Social Science & Medicine. 2022 doi:10.1016/j.socscimed.2022.114782
[10] Kojima T, Shixiang SG, Reid M, Matsuo Y, Iwasawa Y. Large Language Models are Zero-Shot Reasoners. arXiv; 2022. doi:10.48550/arXiv.2205.11916

dHealth 2024
D. Hayn et al. (Eds.)

doi:10.3233/SHTI240020

Towards an Electronic Health Prevention Record Based on HL7 FHIR and the OMOP Common Data Model

Markus FALGENHAUER[a,1], Aaron LAUSCHENSKY[a], Karl KREINER[a], Stefan BEYER[a], Kristina REITER[b], Andreas ZIEGL[a,b], Robert MODRE-OSPRIAN[b], Bernhard PFEIFER[c,d], Sabrina NEURURER[c,d], Susanne KRESTAN[c], Hanna WAGNER[c], Andreas HUBER[c], Sandra PLAIKNER[e], Sarah KUPPELWIESER[e,f], Martin WIDSCHWENDTER[e,f] and Günter SCHREIER[a]

[a] *AIT Austrian Institute of Technology GmbH, Graz, Austria*
[b] *telbiomed Medizintechnik und IT Service GmbH, Graz, Austria*
[c] *Tyrolean Federal Institute for Integrated Care, Tirol Kliniken, Innsbruck, Austria*
[d] *Division for Digital Health and Telemedicine, UMIT TIROL - Private University for Health Sciences and Health Technology, Hall (Tyrol), Austria*
[e] *EUTOPS Institute, University of Innsbruck, Innsbruck, Austria*
[f] *Tirol Kliniken GmbH, Innsbruck, Austria*

Abstract. Background: Approximately 40% of all recorded deaths in Austria are due to behavioral risks. These risks could be avoided with appropriate measures. Objectives: Extension of the concept of EHR and EMR to an electronic prevention record, focusing on primary and secondary prevention. Methods: The concept of a structured prevention pathway, based on the principles of P4 Medicine, was developed for a multidisciplinary prevention network. An IT infrastructure based on HL7 FHIR and the OHDSI OMOP common data model was designed. Results: An IT solution supporting a structured and modular prevention pathway was conceptualized. It contained a personalized management of prevention, risk assessment, diagnostic and preventive measures supported by a modular, interoperable IT infrastructure including a health app, prevention record web-service, decision support modules and a smart prevention registry, separating primary and secondary use of data. Conclusion: A concept was created on how an electronic health prevention record based on HL7 FHIR and the OMOP common data model can be implemented.

Keywords. Electronic Health Records, Primary Prevention, Secondary Prevention, HL7 FHIR, OMOP CDM

1. Introduction

The life expectancy of Austria's population shows a rising trend. In 2021 it was at 83.7 years for females and 78.8 years for males, which is slightly above the EU's average (2021, female: 82.9, male: 77.2) [1]. A well established parameter for the quality of the aging society is the so-called "healthy life years" (HLY). The HLY is a measure for the expected number of years that a person – of a defined age group – can live in a healthy condition and without limitations and disabilities. In contrast to the average life

[1] Corresponding Author: Markus Falgenhauer, AIT Austrian Institute of Technology GmbH, Graz, Austria, markus.falgenhauer@ait.ac.at

expectancy of Austria mentioned above, the HLY are – compared to the EU – below average. In 2019, the difference of HLY at birth were 2.9 years for females (AT: 61.3, EU: 64.2) and 2.4 years for males (AT: 61.5, EU: 63.1) [2]. Impacting factors for estimating the HLY are chronic diseases, frailty, disabilities and mortality. According to the Austria Country Health Profile, approximately 40% of the recorded deaths in 2019 were due to behavioural risk factors. These include (tobacco) smoking, low physical activity, dietary risks, and alcohol consumption [3].

Prevention describes the process of avoiding or delaying the onset (primary prevention) and early detection (secondary prevention) of an illness or an adverse medical event. Preventive health interventions offered to Austrian citizens include early detection of common cancers – such as cervical, breast or colon cancer – as well as an annual preventive medical examination. The latter is comprised of early detection of risk factors for cardiovascular diseases and metabolic diseases, prevention of addictive disorders, periodontal diseases and age-related diseases [4].

The use of electronic health records (EHR) and electronic medical records (EMR) with respect to preventive health services has been widely studied in. In the area of primary intervention studies have shown that automatic reminders built on top of EMRs for vaccination and/or antibody screening [5][6][7] can have a positive health outcome. In the area of secondary prevention, predictive models have been built using existing EHR/EMR in areas including dementia [8], early detection of pancreatic cancer or gestational diabetes [9]. Wang et al. [10] investigated the use of deep learning models to detect cognitive decline from clinical notes in EMRs. Jauk et al. [11] showed how delirium in hospitalized patients can be predicted through data stored in EMRs.

Patel et al. [12] address gaps in the accuracy of EMRs regarding smoking history of patients which has an impact on potential lung cancer screenings. This is in line with the observation of Kruse et. al [13] as missing and/or incorrect data in EHR/EMRs being a main barrier in the adoption of the technology.

Our work contributes to the field of primary and secondary prevention by extending the concept of EHRs to incorporate prevention and addressing the gap of missing and/or inaccurate data on disease prevention. The concept of an EHR for prevention includes the usage of Health Level 7 Fast Healthcare Interoperability Resources (HL7 FHIR) to express various elements of the record and the Observational Medical Outcomes Partnership (OMOP) Common Data Models (CDM) to enable secondary use of data for research purposes.

2. Methods

2.1. Identifying a structured prevention pathway

A structured approach to care in a multi-professional and multi-disciplinary prevention network was chosen, based on the principles of P4 medicine [14]:

- Predictive using a risk-stratified approach
- Personalised with the individualisation of measures
- Preventive in the sense of primary and secondary prevention
- Participatory by an active involvement of the user

Following the principles of P4 Medicine, a pathway with a modular structure was designed. Each module, defining an individual action within the pathway, could be operated in person or remote using telemedical solutions. We organized workshops with stakeholders from different medical fields and analysed existing prevention programs, such as the preventive medical examination program of Austria.

2.2. Architecture of an IT infrastructure

In order to comply with and extend the standard for Austrian EHRs (ELGA), interoperability with them had to be achieved. A concept for further development of ELGA to an eHealth architecture, designates HL7 FHIR for the mapping of medical and administrative information in a finely granular and standardized form using resources, and the OMOP CDM as standardized starting point for a possible persistence scheme with the focus on processing medical data for research purposes [15].

HL7 FHIR standardized data exchange between software systems in healthcare sector using resources. There were a variety of resources, the most relevant for the IT infrastructure were:

- Patient: contains demographic and administrative information about an individual receiving care or other services which are health related.
- PlanDefinition: aims to standardize clinical decision support, guidelines and care plans in healthcare.
- CarePlan: enables a structured planning, documentation and coordination of care of patients and can be designed to be patient-centred by considering the patient's individual needs and preferences. CarePlans can be linked to other FHIR resources. This allows seamless integration with clinical data, which in turn contributes to holistic and coordinated patient care.
- Questionnaire: focuses on the standardized collection of data in the healthcare sector. These questionnaires provide a structured framework for collecting patient information, clinical data and other relevant information.
- Goal: describes intended objectives e.g. weight loss.
- DiagnosticReport: represents the set of information that is typically provided by a diagnostic service when investigations are complete.

Possible different sources, types and formats of data, related to prevention were identified. Suitable OMOP CDMs were determined for harmonisation. A data-transfer-specification was created that defines the process of extracting data from the source system, transforming it into a harmonised form and loading it in a target system (ETL-process).

3. Results

3.1. Modular Prevention Pathway

Management of prevention was identified as the core module of the prevention pathway (**Figure 1**), as it enabled a personalised view on each participant. Information exchanged via computer interfaces, e.g. from preventive medical examinations and the collection of findings from EHR/EMR, determined the further direction along the pathway. This could be enabled by a decision support module or by the assistance from a prevention nurse -

identified as a necessary role - to aid participants with advice regarding prevention e.g. to find the most appropriate intervention considering the personal environment.

Continuing from the prevention management, three branches of the pathway were identified:

- Risk assessment carried out by a physician to determine health potentials and risks.
- Diagnostic measures for early detection of diseases, attributable to secondary prevention, like different screening programs or laboratory.
- Interventions as preventive measures like physical activity, smoking cessation, nutrition, addiction management, resilience strategies, or vaccinations, attributable to primary prevention.

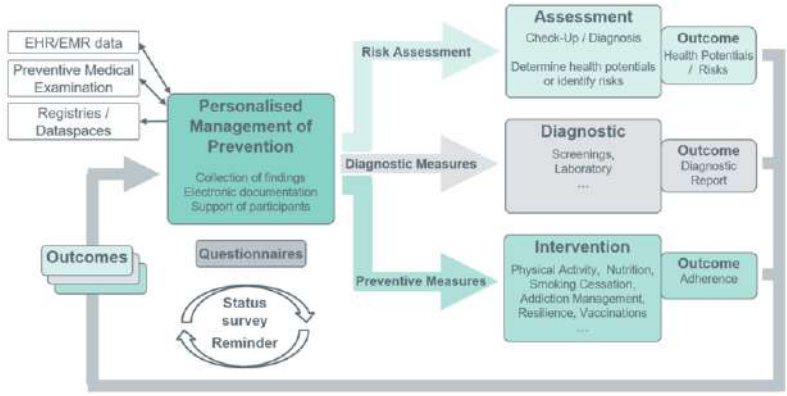


Figure 1. Overview of the prevention pathway. A personalized management of prevention module directs the way to risk assessment, diagnostic measures or preventive measures and enables questionnaires for a status survey and reminders.

For each of these branches, a specific outcome was identified to ensure measurability of the effectiveness of the action. Outcomes of previous routes were used again at the prevention management to allow even more personalised decisions for a further cycle within the prevention pathway. Possible outcomes for the aforementioned branches were a list of health potentials and risks, diagnostic reports or information about a participant's adherence with an agreed intervention (e.g. adhering to a program or vaccinations).

3.2. Architecture of IT infrastructure

The IT infrastructure for an electronic health prevention record (EHPR) has been designed as a web-based IT service that could basically be divided into the areas of primary and secondary use of data (Figure 2). The technological foundation of these services was the HL7 FHIR standard to ensure interoperability with other systems (e.g. ELGA 2.0). An HL7 FHIR server was used to store the data and offers REST interfaces to process and exchange data.

In the primary use, a web service was provided for the record itself, in which healthcare professionals could manage prevention and the associated measures for participants. This web service allowed the definition of configurable FHIR Questionnaires that could be used e.g. for anamnesis. A web-based health app was conceptualised to stay in contact with participants. Thus, the progress and results of the

prevention measures could be monitored. The IT service allowed personalized suggestions with recommendations for action to be stored for every participant. A decision support module could be utilized for automatically generated suggestions based on available data. Architecturally, these modules were separated by clear FHIR interfaces, as they are potential medical products relevant to undergo a certification process.

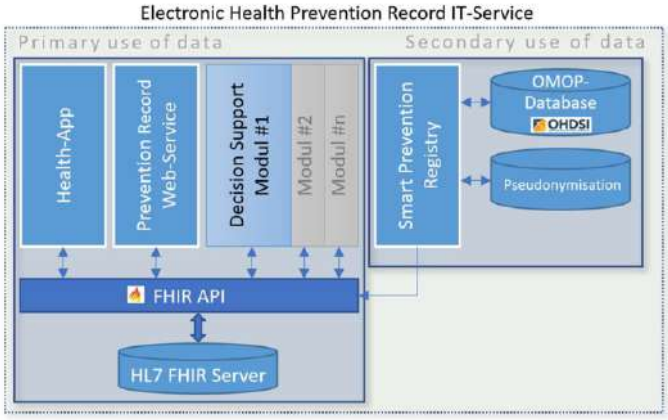


Figure 2. Architecture of the EHPR IT-Service. The primary use of data shows a HL7 FHIR server which can exchange data with e.g health-apps or prevention record web services via a dedicated API. The secondary use of data depicts a smart prevention registry, which uses a OMOP database that contains pseudonymized data, and also connects to the FHIR server via the API.

The core element of the EHPR were personalised FHIR CarePlans that were transformed from generic FHIR PlanDefinitions, containing all relevant information to support the participant in terms of prevention referencing other FHIR Resources like Questionnaires, ServiceRequests, DiagnosticReports, Tasks or Goals.

Besides the primary use, the data collected as part of prevention programs was also relevant for research questions, e.g. in the field of population health. For this purpose, the architecture provided a separate area for secondary use in the form of a Smart Prevention Registry (SPR) or dataspaces. The SPR was automatically filled with data from the record, whereby the person's identification data was pseudonymized. In the registry itself the data was mapped to according OMOP CDMs. The data of the prevention record was transformed in a relational, analysis-friendly format. Furthermore, this harmonized format allows for linkage with other OMOP-compatible databases to provide even more insight in the participants medical histories.

4. Discussion

In this work the concept of a modular prevention pathway based on the principles of P4 medicine was created and used to design the architecture for an electronic health prevention record based on HL7 FHIR enabling interoperability and the OMOP CDM harmonising data and providing a structure to analyse data in an efficient way. To the best of our knowledge, this is a novel approach in the field of electronic health prevention. The architecture of an IT infrastructure provided a separation into primary and secondary use of data, where the primary data structured in FHIR could be transformed to the secondary data in the OMOP CDM.

Furthermore, to enable a personalised approach supported by software, it was considered to separate the decision support modules, because they needed to be considered as software as medical device, covered by the medical device regulation (MDR - (EU) 2017/745). This aspect has not been addressed in this publication. The intended purpose of such a module - with application to humans for the diagnosis, prevention, monitoring, prediction, prognosis, treatment or alleviation of disease - could be considered a Class IIa or higher medical device according to "Annex VIII: Classification rules" Rule 11. To implement such a device, clear guidelines for prevention following rules would be needed, which could also be used for a software.

The presented results provide a concept how a prevention pathway based on P4 medicine can be implemented with digital support. The next step could be an implementation in a pilot study in a real-world setting. Such a study would help with the validation of the concept, as well as with identifying possible obstacles within the pathways (e.g. prevention of multiple identical records).

Acknowledgements

The research leading to these results was co-financed by the Standortagentur Tirol.

References

[1] eurostat, Mortality and life expectancy statistics, *https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Mortality_and_life_expectancy_statistics*, (2023),.

[2] eurostat, Glossary: Healthy life years (HLY) , *https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Healthy_life_years_(HLY)*, (2023),.

[3] OECD/European Observatory on Health Systems and Policies, Austria: Country Health Profile 2021, State of Health in the EU, Paris, (2021),.

[4] R. Griebler et al., Österreichischer Gesundheitsbericht 2022, Wien, (2023),.

[5] J. G. Wijesundara et al., Electronic Health Record Portal Messages and Interactive Voice Response Calls to Improve Rates of Early Season Influenza Vaccination: Randomized Controlled Trial, *Journal of Medical Internet Research, 22(9), e16373* (2020),

[6] E. Mulhem et al., Electronic Health Record Reminder Effect on Hepatitis C Antibody Screening, *The Journal of the American Board of Family Medicine* **33**(6) (2020), 1016–1019

[7] R. C. Hechter et al., Impact of an electronic medical record reminder on hepatitis B vaccine initiation and completion rates among insured adults with diabetes mellitus, *Vaccine* **37**(1) (2019), 195–201

[8] Z. Ben Miled et al., Predicting dementia with routine care EMR data, *Artificial Intelligence in Medicine* **102** (2020), 101771

[9] N. S. Artzi et al., Prediction of gestational diabetes based on nationwide electronic health records, *Nature medicine* **26**(1) (2020), 71—76

[10] L. Wang et al., Development and Validation of a Deep Learning Model for Earlier Detection of Cognitive Decline From Clinical Notes in Electronic Health Records, *JAMA Network Open* **4** (2021), 2135174

[11] S. Jauk et al., Risk prediction of delirium in hospitalized patients using machine learning: An implementation and prospective evaluation study, *Journal of the American Medical Informatics Association : JAMIA* **27** (2020),

[12] N. Patel et al., A Comparison of Smoking History in the Electronic Health Record With Self-Report, *American Journal of Preventive Medicine* **58**(4) (2020), 591–595

[13] C. Kruse et al., The use of Electronic Health Records to Support Population Health: A Systematic Review of the Literature, *Journal of Medical Systems* **42** (2018),

[14] L. Hood and S. H. Friend, Predictive, personalized, preventive, participatory (P4) cancer medicine, *Nature Reviews Clinical Oncology* **8**(3) (2011), 184–187

[15] ELGA GmbH, ELGA 2.0" aka elga2 - Konzept zur Weiterentwicklung der ELGA- und eHealth-Zielarchitektur, Version 1.0 vom 21.3.2023, Vienna, (2023),.

dHealth 2024
D. Hayn et al. (Eds.)

doi:10.3233/SHTI240021

KIJANI: Designing a Physical Activity Promoting Collaborative Augmented Reality Game

Florian SCHWEIZER[a,1], Laura WILLINGER[b], Renate OBERHOFFER-FRITZ[b], Jan MÜLLER[b], Stephan JONAS[c] and Lara Marie REIMER[a,c]

[a] *School of Computation, Information and Technology, Technical University of Munich, Munich, Germany*
[b] *Chair of Preventive Pediatrics, Technical University of Munich, Munich, Germany*
[c] *Institute for Digital Medicine, University Hospital Bonn, Bonn, Germany*

Abstract. Background: There is an increased need for physical activity among children and adolescents. KIJANI, a mobile augmented reality game, is designed to increase physical activity through gamified exercises. Objectives: The primary aim of this study is to get feedback on the design and implementation of potentially physical activity-increasing features in KIJANI. Methods: A mixed-method study (n=13) evaluates newly implemented game design features quantitatively through measuring physical activity and qualitatively through participant feedback. Results: Preliminary results are limited and need further studies. Participants' feedback shows a positive trend and highlights the game's potential effectiveness. Conclusion: KIJANI shows potential for increasing physical activity among children and adolescents through gamified exercise. Future work will refine the game based on user feedback and findings presented in related work. The game's long-term impact is to be explored.

Keywords. Physical Activity, Augmented Reality, Mobile Applications, Video Games, Child

1. Introduction

The World Health Organization (WHO) recommends that children and adolescents (aged 5 to 17) do at least an average of 60 minutes of moderate-intensity physical activity (PA) per day. Additionally, they recommend vigorous activity on three days per week. Time spent being sedentary, particularly recreational screen time, should be limited [1]. Research has shown that activity levels during childhood and adolescence correlate with sedentary behavior, activity, and the rate of illnesses and diseases as adults, making it important to start being active as early as possible [2, 3]. It is recommended that children move at least 60 minutes per day [1, 4], and a 2008 review suggests that setting and monitoring PA goals can positively impact children's activity [5]. The review suggests that parents should do this, but games could also take that control role. PA can be done in many ways, with a common concept being to increase the daily step count [6]. Tabata

[1] Corresponding Author: Florian Schweizer, School of Computation, Information and Technology, Technical University of Munich, Munich, Germany, E-Mail: florian.schweizer@tum.de

[7] classifies walking as a non-exercise PA. They differentiate between regular and "brisk walking". They also suggest increasing walking or bicycling time as a good start to increasing PA. A common way of measuring PA is by counting steps. Doing so manually isn't feasible, which is why step-counting devices have become common. In addition to standalone step counters, modern smartphones and smartwatches also support counting steps during everyday activities.

Counting steps is commonly considered part of the "quantified self" movement [8, 9], with the idea of tracking metrics of one's life to quantify different aspects of everyday life. As such, step counts are used to measure PA. Given that research suggests that wearable step trackers increase the steps walked by individuals [10], this offers an opportunity to gamify the experience to increase PA.

According to the WHO's 2020 catalog of PA recommendations, sedentary behavior in children can be associated with various health issues: "increased adiposity; poorer cardiometabolic health, fitness, behavioural conduct/pro-social behaviour; and reduced sleep duration" [4]. Usually, smartphone usage is classified as sedentary behavior [10, 11], and children are suggested to limit their screen time. Location-based games that promote outdoor activities (like KIJANI, the game proposed in this research) could change that narrative.

Through a thorough investigation, Arsenault analyzed two lists of video game genres and discussed their rigidness and history [12]. As an ever-expanding market, it is important to note that such lists are often incomplete. An often overlooked genre of video games is location-based (LB) games. LB games are not limited to being digital, with an example of a non-video LB game being *geocaching*. In *geocaching*, players roam the physical world to find hidden caches. This concept is fundamental for most LB games, especially those that have arisen in the last few years, such as Pokémon Go. A 2012 review of LB educational games categorizes them into three categories: ludic, pedagogic, and hybrid approaches to learning [13].

Most video games are based on extrinsic motivation [14], meaning that players aren't motivated by themselves (intrinsic motivation) but rather through other people or external factors like gameplay-related goals. When discussing goals, a duration-based distinction must be made between short-and long-term goals. In video games, short-term goals should be achievable in a few days. Consequently, long-term goals refer to several days to week-long goals. Research suggests that "achievable short- and long-term goals" are relevant for successful activity-based video games [15]. According to a 2014 study by Staewen et al. [16], different player types get different motivation levels depending on the goal duration. People who play to avoid boredom prefer quick and easily achievable goals, while those players who enjoy difficulty are less likely to pursue short-term goals. True long-term goals in video games often include streaks (repeating a behavior for several consecutive days) and immersive experiences that allow players to shape the game through their actions. Immersive experiences include AR, which directly allows manipulation of the perceived reality.

Research also highly suggests that social features are a significant driver of motivation to play a specific video game. Research by Shameli et al. and Zielinski et al. suggests that in-game competitions can increase PA in players [17, 18]. Shameli et al. found that players participating in a competition increase their daily step count by 23% [17]. One of the simplest forms of competition in video games (aside from player vs. player game modes) is ranking. Such a leaderboard also provides another way of instant feedback for competitive players. As existing research suggests, only some enjoy having that public comparison enabled through a leaderboard [19].

It is suggested that video game players often prefer gaming over socializing with friends in real life and might even sacrifice real-life social interactions to be able to play video games [20, 21].

This research describes KIJANI, a location-based mobile augmented reality game that aims to promote PA in children and adolescents through gamification. The game promotes moderate PA by incentivizing players to walk outside to unlock in-game content through external motivational factors hidden behind a game. Step counts are used as a measure of PA. KIJANI can also be categorized as a hybrid educational game, as increasing PA is merely a side effect of completing in-game challenges. In KIJANI, players can create block-based buildings in an AR environment (Fig. 1). Blocks can be unlocked with coins earned through real-life fitness challenges.

This paper explains KIJANI's strategy to increase PA, game design, and user experience decisions and presents preliminary results from a chair-internal study.

2. Methods

2.1. KIJANI's Goals

KIJANI has a high-level goal of making the player go outside, move around, and have fun. The game tries to motivate its players to increase their PA while enjoying time in nature, a common concept in the literature around active games [22, 23]. KIJANI aims to increase step counts in children and adolescents, as that metric is used to approximate player PA in the context of this research.

2.2. Requirements

In addition to common video game requirements regarding usability, user experience, and ease of use, specific requirements for the KIJANI game and application were defined:

R1: KIJANI should be played at specific real-world locations suitable for children and adolescents.
R2: KIJANI should reward exploration and movement.
R3: KIJANI should increase players' motivation to interact with the game through long- and short-term goals.
R4: KIJANI should promote social interactions and communication in the physical world.

2.3. KIJANI Overview

KIJANI is a collaborative AR game built for iOS devices. It combines location-based gameplay with a focus on the gamification of PA. In KIJANI, players place virtual blocks to create buildings in augmented reality (Fig. 1). This can happen solo or in a group of players. More blocks can be unlocked with coins. Fitness challenges reward the player with KIJANI coins upon completion of PA tasks. Currently, walking a certain number of steps is used as a PA measure for fitness challenges. KIJANI aims to increase PA both while playing and outside of gameplay.

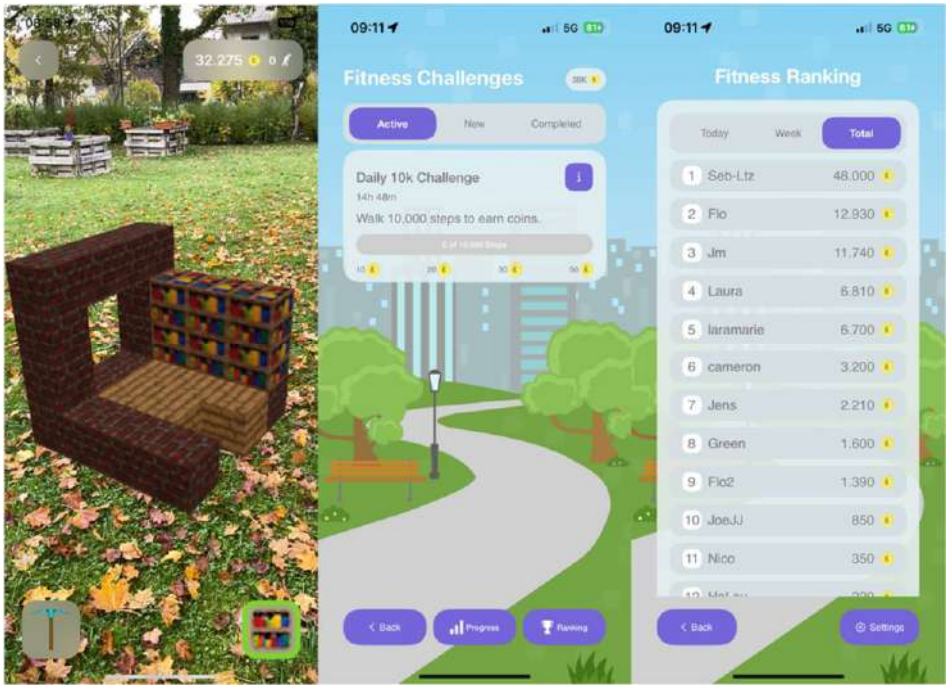


Figure 1. AR Building; Fitness Challenge; Fitness Ranking

2.4. KIJANI Features

2.4.1. Locations

KIJANI is designed to be unusable at home and in most public places except for specific outdoor locations. In their 2017 review, Licoppe found that the LB game "Mogi" sometimes made players go to "strange, unanticipated, and sometimes risky places" [24]. KIJANI doesn't promote visiting such places in the context of the game, as the set of available locations is finite and pre-selected. They were chosen based on several criteria and mostly represent public parks or playgrounds in the greater Munich metropolitan area (and other selected areas in Germany) with sufficient space and safety to allow children to play. These physical locations are limited in space. Because of that, they are eventually filled with blocks, disallowing players to continue building. To circumvent this issue, the concept of servers was introduced. Like other games' concepts of servers, KIJANI servers offer the option to play in the same physical location while accessing a different digital world. As such, players can construct multiple space-filling buildings at the same location, making it feasible to play KIJANI at the same place over a prolonged time. Servers can be configured to be private, and players can set a list of visitors, builders, and admins. This eliminates a set of potential misuse issues regarding inappropriate behavior among players. This setup allows for a distinction between private and group building without changing physical location, which might be infeasible in certain areas (due to lack of parks and playgrounds).

2.4.2. Fitness Challenges

Fitness Challenges are KIJANI's main tool to increase players' motivation to play and, thus, their PA. Fitness Challenges are of varying duration and intensity, ensuring a mixture of short-term and long-term goals. There are three kinds of fitness challenges: the daily 10k challenge and other non-recurring, manually created, challenges that might last longer than a day (e.g., weekend or week-long) are visible to the user and can be completed through PA. The third challenge type is a hidden background challenge that automatically tracks the players' PA while running studies. This system is a backup to the manual data collection described in the study overview. The daily 10k challenge automatically gets activated when players open KIJANI for the first time on any day. On the technical side, fitness challenges are implemented as a modular system, allowing for full customization of rewards, sub-goals, and metadata such as the title and description. They are fetched from a remote server and managed through a separate app, allowing the KIJANI team to update and change challenges without the user needing to update their app. An in-app notification system was added to KIJANI to keep players updated about their challenge progress without opening the respective screen (Fig. 1). The fitness challenges collect progress even while the app is in the background, using Apple's HealthKit as the data source.

2.4.3. Friends

KIJANI players can add each other as in-game friends by sending and accepting a friend request. Doing so enables them to play on the same server and inspect each other's fitness challenge progress on a leaderboard. KIJANI servers allow players to specify who can explicitly visit, build, and administrate a server. This setup ensures that no uninvited players destroy buildings or interfere with the fun of others.

A leaderboard (Fig. 1) was added to improve motivation toward the fitness challenge feature described in the next section. The leaderboard has two participation options: global and friends. This means that both players who want to compare their progress with everyone and those who only want to compare themselves to their friends can do so by changing the setting. Given that no significant correlation between access to the leaderboard and fitness challenge progress could be found in the thesis study, a more prominent placement for the leaderboard should be considered.

2.5. Study Overview

To get the first results regarding the effectiveness of both KIJANI and the impact of social features in the game and an overview of the game's core concept, an exploratory study was conducted as part of a master's thesis at our chair. The study was based on a randomized controlled design (albeit noted that familiar participants could have interacted), with the participants split into two groups. The first group had a basic, stripped-down version of the game without social features (friends, leaderboard, and multiplayer servers), while the second group had those social features enabled. Two questionnaires were included: At the beginning and at the end of the study duration. The questionnaires contained free-text feedback questions and recorded the participants' PA, based on HealthKit metrics, during the study duration. Participants were instructed to record their PA in the first week while following their usual routine. In the second week, participants received access to KIJANI and were instructed to play for at least 30 minutes during at least two sessions to ensure at least one return to the game. The cohort contained

predominantly students and was completed by scientific staff. Both male (n=12) and female (n=1) players were included, with most people (n=11) 20 to 25 and a minority (n=2) 30 to 31 years of age.

3. Results

According to the first questionnaire, participants did sports an average of 3.43 times per week, which decreased to 3 times per week in the second questionnaire. The average amount of steps walked in the first week was 46556, which increased to 47269. 5 participants had an increased step count in the second week (Fig. 2). Before joining the study, participants averaged 4.35 hours of weekly game-playing time, including about 1.06 hours of mobile games. The average gameplay duration of KIJANI was 24.85 minutes (min: 12 min, max: 40 min). 5 participants played for >= 30 minutes, 9 for >= 20 minutes, and 4 for < 20 minutes.

During the second week of the study duration, two fitness challenges were available: A week-long challenge with the goal of 35000 steps and a reward of 1600 coins total, and a 3-day challenge with a goal of 10000 steps and a cumulative reward of 650 coins. The only study participant who completed a fitness challenge was part of the social feature group. A total of 4 (2 from each group) of the 13 study participants completed at least one sub-goal of a fitness challenge. 10 participants took on at least one fitness challenge. One of them took on both available challenges.

In their final questionnaire, 3 study participants mentioned liking KIJANI's approach to promoting a healthy lifestyle through walking challenges. Two study participants explicitly mentioned that they liked the multiplayer aspects of KIJANI's gameplay. Another participant mentioned that there wasn't enough to offer for solo players, which underlines the importance of social interactions. One of the non-social-feature participants also mentioned that they wanted more social aspects to the game. It was also mentioned that adding friends should be easier than entering an ID.

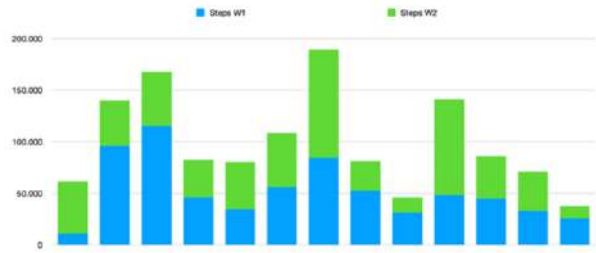


Figure 2. Step count results from the questionnaires

4. Discussion and Future Work

The location feature satisfies R1 partly and ensures KIJANI isn't considered sedentary behavior. It enforces gameplay at suitable real-world locations. R2 specified that KIJANI should reward exploration of the physical world. The current implementation has no incentive to play at unique locations, as staying at the same location is easiest for everyday gameplay. An exploration-focused route feature, rewarding players for walking

along a predefined route, would be a possible implementation that satisfies R2. To satisfy R3, the fitness challenge system was introduced. A positive effect on player motivation couldn't be recorded through the study. KIJANI promotes social interactions through the collaborative gameplay feature in the server system, as specified in R4. The positive impact on playing motivation that was expected wasn't found through the collected data.

Even though the data couldn't support the idea of social aspects increasing PA in KIJANI players, the collected feedback was still valuable, as several participants had positive feelings toward the game's social features. This supports our belief that an intervention like KIJANI can positively impact players' PA. HealthKit, the integration powering fitness challenges, is a secure health metric store that collects health data through phone movement, manual input, and peripherals like the Apple Watch [25]. An investigation into its data accuracy has shown that recorded step counts (using just an iPhone) are much more accurate than estimated distances walked [26]. An Apple Watch could be worn to further increase step-counting accuracy [27]. Since KIJANI directly reads health data using HealthKit, both approaches are suitable, and this could be explored further in the future.

We were able to implement a competitive social feature through the leaderboard. The preliminary study's sample size is limited and based on non-target-group participants, making it hard to interpret the findings w.r.t. the research's original goals.

More collaborative gameplay elements should be considered to improve the players' need to interact with each other further. Introducing collaboration and contest features, such as the possibility to attack other players' buildings, could allow for more direct real-world communication between players. Currently, KIJANI uses step counts as a proxy for PA. Other measurements, such as workout count, active minutes, heart rate, etc., could also be adapted. They could be added to the fitness challenge system, allowing users to narrow down on a specific area of activity they want to engage in. This would make the system more approachable and personalized. KIJANI's impact on PA should be evaluated in a study with a bigger sample size and a longer timeframe to reduce the impact of external factors on step counts.

Acknowledgments

This research was funded by an unrestricted grant from the Bavarian State Ministry for Health and Care (StMGP) as part of the "Gesund.Leben.Bayern" initiative. We want to express our gratitude to Birgit Böhm, the study participants, and the WalkiMUC team for their valuable contributions to this research.

Declaration of generative AI and AI-assisted technologies in the writing process During the preparation of this work, the authors used ChatGPT to proofread the text and eliminate typos and grammatical flaws. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] Chaput, Jean-Philippe, et al. "2020 WHO guidelines on physical activity and sedentary behaviour for children and adolescents aged 5–17 years: summary of the evidence." International Journal of Behavioral Nutrition and Physical Activity 17 (2020): 1-9.

[2] He, Qian, and Emmanuel O. Agu. "Smartphone usage contexts and sensable patterns as predictors of future sedentary behaviors." 2016 IEEE Healthcare Innovation Point-Of-Care Technologies Conference (HI-POCT). IEEE, 2016.
[3] Telama, Risto, et al. "Physical activity from childhood to adulthood: a 21-year tracking study." American journal of preventive medicine 28.3 (2005): 267-273.
[4] Nettle, Heather, and Elizabeth Sprogis. "Pediatric exercise: truth and/or consequences." Sports medicine and arthroscopy review 19.1 (2011): 75-80.
[5] Trost, Stewart G., and Paul D. Loprinzi. "Exercise—Promoting healthy lifestyles in children and adolescents." Journal of Clinical Lipidology 2.3 (2008): 162-168.
[6] Banach, Maciej, et al. "The association between daily step count and all-cause and cardiovascular mortality: a meta-analysis." European journal of preventive cardiology 30.18 (2023): 1975-1985.
[7] Tabata, Izumi. "Exercise and Physical Activity Guide for Health Promotion 2006."
[8] den Braber, Maarten. "The emergence of quantified self as a data-driven movement to promote health and wellness." Proceedings of the first Workshop on Lifelogging Tools and Applications. 2016.
[9] Appelboom, Geoff, et al. "The quantified patient: a patient participatory culture." Current medical research and opinion 30.12 (2014): 2585-2587.
[10] Chaudhry, Umar AR, et al. "The effects of step-count monitoring interventions on physical activity: systematic review and meta-analysis of community-based randomised controlled trials in adults." International Journal of Behavioral Nutrition and Physical Activity 17 (2020): 1-16.
[11] Hwang, In-Whi, et al. "Association between Physical Activity, Sedentary Behavior, Satisfaction with Sleep Fatigue Recovery and Smartphone Dependency among Korean Adolescents: An Age-and Gender-Matched Study." International Journal of Environmental Research and Public Health 19.23 (2022): 16034.
[12] Arsenault, Dominic. "Video game genre, evolution and innovation." Eludamos: Journal for computer game culture 3.2 (2009): 149-176.
[13] Avouris, Nikolaos M., and Nikoleta Yiannoutsou. "A review of mobile location-based games for learning across physical and virtual spaces." J. Univers. Comput. Sci. 18.15 (2012): 2120-2142.
[14] G. Reid. “Motivation in video games: a literature review.” In: The computer games journal 1.2 (2012), pp. 70–81.
[15] J. Yim and T. N. Graham. “Using games to increase exercise motivation.” In: Proceedings of the 2007 conference on Future Play. 2007, pp. 166–173.
[16] Staewen, Randal, Philip Trevino, and Chang Yun. Player characteristics and their relationship to goals and rewards in video games. IEEE, 2014.
[17] A. Shameli, T. Althoff, A. Saberi, and J. Leskovec. “How gamification affects phys- ical activity: Large-scale analysis of walking challenges in a mobile application.” In: Proceedings of the 26th international conference on world wide web companion. 2017, pp. 455–463.
[18] S. Zielinski, J. Emmerich, and M. Schellenbach. “Mobile App based Game Mechan- ics to Motivate Enhanced Physical Activity in Early Adolescense.” In: Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts. 2018, pp. 677–682.
[19] Lister, Meaghan. "Gamification: The effect on student motivation and performance at the post-secondary level." Issues and Trends in Educational Technology 3.2 (2015).
[20] B. Harrington and M. O’Connell. “Video games as virtual teachers: Prosocial video game use by children and adolescents from different socioeconomic groups is associated with increased empathy and prosocial behaviour.” In: Computers in Human Behavior 63 (2016), pp. 650–658.
[21] M. Griffiths, Z. Hussain, S. M. Grüsser, R. Thalemann, H. Cole, M. N. Davies, and D. Chappell. “Social interactions in online gaming.” In: International Journal of Game-Based Learning (IJGBL) 1.4 (2011), pp. 20–36.
[22] Foley, Louise, and Ralph Maddison. "Use of active video games to increase physical activity in children: a (virtual) reality?." Pediatric exercise science 22.1 (2010): 7-20.
[23] Barnett, Anthony, Ester Cerin, and Tom Baranowski. "Active video games for youth: a systematic review." Journal of Physical Activity and Health 8.5 (2011): 724-737.
[24] Licoppe, Christian. "From Mogi to Pokémon GO: Continuities and change in location-aware collection games." Mobile Media & Communication 5.1 (2017): 24-29.
[25] Inc., Apple, HealthKit, 19. January 2024, https://developer.apple.com/documentation/healthkit
[26] van Zandwijk, Jan Peter, and Abdul Boztas. "The iPhone Health App from a forensic perspective: can steps and distances registered during walking and running be used as digital evidence?." Digital investigation 28 (2019): S126-S133.
[27] Veerabhadrappa, Praveen, et al. "Tracking steps on apple watch at different walking speeds." Journal of general internal medicine 33 (2018): 795-796.

dHealth 2024
D. Hayn et al. (Eds.)

doi:10.3233/SHTI240022

Visualizing Study Profiles Using the GMDS Competency Catalog for Medical Informatics Programs

Elske AMMENWERTH[a,1], Peter HAAS[b] and Oliver J. BOTT[c]

[a] *UMIT TIROL – Private University for Health Sciences and Health Technology, Hall in Tirol, Austria*
[b] *University of Applied Sciences and Arts Dortmund, Germany*
[c] *University of Applied Sciences Hannover, Germany*

Abstract. Background: Medical informatics programs cover a variety of topics. Objectives: To test the utility of the GMDS medical informatics competency catalog in comparing programs by developing study profiles. Methods: Coverage of 234 competencies is recorded and visualized in a spider diagram. Results: Spider diagrams allow visualizing various study profiles. Conclusion: The GMDS catalog seems useful for comparing medical informatics study programs, e.g., for interested students, employers, or accreditation reviewers.

Keywords. Medical informatics, learning objectives, competencies

1. Introduction

Medical informatics programs encompass a variety of topics, posing challenges for interested students, employees, or accreditation reviewers in quickly identifying competence profiles and program differences. To address this, the German Society for Medical Informatics, Biometry and Epidemiology (GMDS) created a catalog outlining core competencies for bachelor programs in medical informatics [1], based on international recommendations [2]. The catalog comprises 234 competencies organized into 13 topics covering core competencies in medical informatics, medicine & health, computer science & mathematics, and personal competencies, and is freely available [1]. While the catalog aims to aid in program development and comparison, it effectiveness in comparing study profiles remains uncertain. Our objective is to test its utility in developing study profiles for medical informatics programs.

2. Methods

A quantitative analysis tool was devised for users to rate the fulfillment of the 234 competencies on a scale from 0 (not covered in the curriculum) to 2 (comprehensively covered), alongside ECTS credits allocation. Percentages per topic then generate the

[1] Corresponding Author: Elske Ammenwerth, UMIT TIROL, Hall i.T., elske.ammenwerth@umit-tirol.at

study profile. Spider diagrams visualize competency fulfillment and ECTS credits allocatation for comparision. ECTS credits are differentiated into compulsory and elective. Data from six medical informatics programs in Austria and Germany were collected via expert contacts.

3. Results

Data entry on the coverage of all 234 competencies takes < 2 hours for someone with excellent program knowledge. Figure 1 displays competence-related study profiles for two programs, demonstrating the catalogue's ability to distinguish programs profiles.

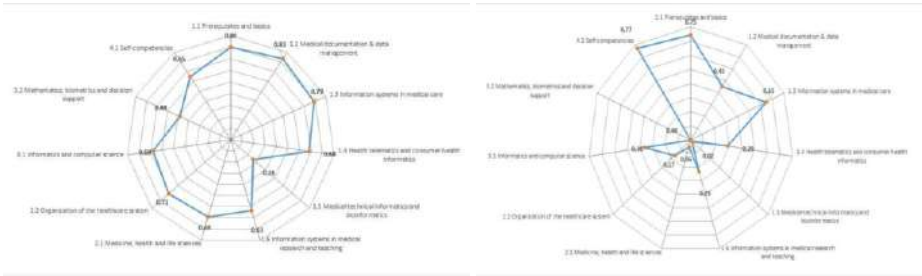


Figure 1. Two study profiles, derived from the GMDS competency catalog [1]. Left: Bachelor program. Right: Postgraduate program. Numbers indicate the percentage of covered competencies in each topic.

4. Discussion

Using spider diagrams for visualization of competency profiles has been used in other fields [3], but to our knowledge not for medical informatics programs. The first experience using the GMDS competency catalog to quantitatively and visually compare study programs in medical informatics is promising. The individual study profiles can be recorded with manageable effort and easily be visualized. The study profiles represent the quality aspect of the programs, and invested ETCS for competencies represent the quantity aspect of the programs.

The GMDS competency catalog covers competencies in bachelor's programs. The first experiences show that it can also be used for postgraduate programs. For full-time master programs, additional competencies would need to be added, and others can be removed. Heads of bachelor programs interested in using the GMDS competency catalog and the tool to visualize their study profile are invited to contact the authors.

References

[1] GMDS. Kompetenzkatalog für Bachelor-Studiengänge der (Bio-)Medizinischen Informatik und des Medizinischen Informationsmanagements der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie (gmds) (2021). Version 1.0. https://www.gmds.de/index.php?id=1142.

[2] J. Bichel-Findlay, S. Koch, J. Mantas, et al. (2023). Recommendations of the International Medical Informatics Association (IMIA) on Education in Biomedical and Health Informatics: Second Revision. *International Journal of Medical Informatics*, **170**, 104908. DOI: 10.1016/J.IJMEDINF.2022.104908.

[3] T.S. Gugapriya, S. Banarjee, M. Girish, S. Damke, S. Bhattad. Attainment of Indian Medical Graduate (IMG) Roles Through the Curriculum. National J Clin Anatomy 11(1):p 30-36, 2022.

dHealth 2024
D. Hayn et al. (Eds.)

doi:10.3233/SHTI240023

Comparison of WebSocket and Hypertext Transfer Protocol for Transfer of Electronic Health Records

Nikola KIRILOV[a,1]

[a] *Institute of Medical Informatics, Heidelberg University Hospital, Germany*

Abstract. Background: Electronic health records (EHR) emerged as a digital record of the data that is generated in the healthcare. Objectives: In this paper the transfer times of EHRs using the Hypertext Transfer Protocol and WebSocket in both local network and wide area network (WAN) are compared. Methods: A python web application to serve Fast Health Interoperability Resources (FHIR) records is created and the transfer times of the EHRs over both HTTP and WebSocket connection are measured. 45000 test Patient resources in 20, 50, 100 and 200 resources per Bundle transfers are used. Results: WebSocket showed much better transfer times of large amount of data. These were 18 s shorter in the local network and 342 s shorter in WAN for the 20 resource per Bundle transfer. Conclusion: RESTful APIs are a convenient way to implement EHR servers; on the other hand, HTTP becomes a bottleneck when transferring large amount of data. WebSocket shows better transfer times and thus its superiority in such situations. The problem can be addressed by developing a new communication protocol or by using network tunneling to handle large data transfer of EHRs.

Keywords. electronic health record, fhir, http, websocket, rest

1. Introduction

Electronic health records (EHR) emerged as a digital record of the data that is generated in the healthcare. They offer interoperability and immediate access to important information [1]. Nowadays EHRs are based on web servers and use databases, which makes transferring medical information from one facility to another easier and furthers the creation of health information exchange networks among medical organizations [2].

The recently developed Fast Health Interoperability Resources (FHIR) standard is gaining popularity and adoption by companies and institutions. The architecture of FHIR is organized by resources, which are predefined profiles of concepts in healthcare (e. g. Patient, Encounter, Condition, Procedure etc.) [3]. This makes the representational state transfer (REST) protocol suitable for the implementation of their application programming interfaces (APIs). The APIs are based on the Hypertext Transfer Protocol (HTTP), although in the Roy Fielding's dissertation HTTP was never mentioned as the only protocol for REST. He also mentioned the inefficiency of HTTP

[1] Corresponding Author: Nikola Kirilov, Institute of Medical Informatics, Heidelberg University Hospital, Germany, E-Mail: Nikola.Kirilov@med.uni-heidelberg.de

concerning its single request/response per connection behavior [4]. When requesting FHIR resources from the server these are usually provided in Bundles [5]. The number of resources per Bundle is limited (by default 20) by the server and in order to retrieve large number of resources, multiple requests need to be made. This makes such transactions time consuming.
WebSockets have been recently introduced to bring real-time capabilities to the web and to solve multiple problems [6]. The connection is established using HTTP GET request with an upgrade header and allows long-lived exchange of messages [7]. HTTP has already been identified as a bottleneck when transferring large data using REST [8].

2. Methods

To compare the transfer time of the EHRs from the server using HTTP requests and the transfer time of the same data using a WebSocket connection a python web application is created. The application was implemented using the Hypercorn library serving the resources from a HL7 application programming interface (HAPI) PostgreSQL database with the psycopg driver. The Hypercorn library offers a support for WebSocket connection over HTTP. The execution time of the database queries was similar for both HAPI FHIR and the test python implementation.
The test dataset consisted of 45000 test FHIR Patient resources. The 45000 resources were transfered using 20, 50, 100 and 200 resources per Bundle, which resulted in 2250, 900, 450 and 225 HTTP requests and WebSocket messages respectively. The test was carried out while the client was connected to the local network and then while the client was connected to the wide area network (WAN) (Figure 1).
Asynchronous JavaScript And XML was used for the execution of the HTTP requests and JavaScript for the WebSocket messages. Requests and messages were sent serially to avoid server performance issues which might alter the results. The total execution time was logged using the HTTP Archive files for each request.

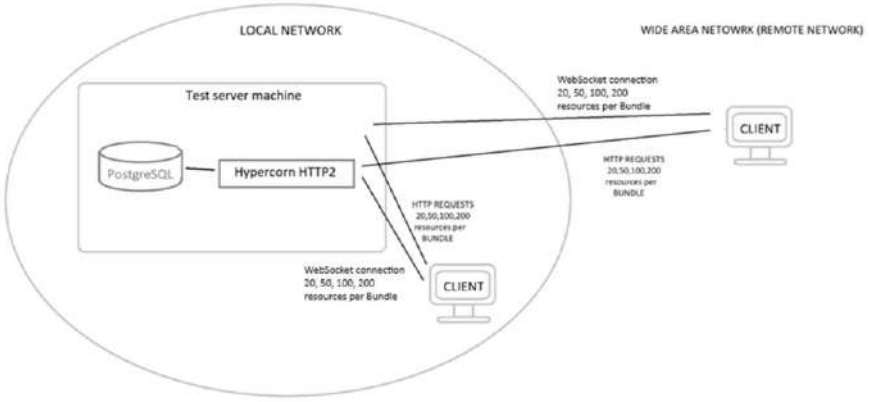


Figure 1. Graphical depiction of the test setup: server – client. Requests/Messages from WAN and in the local network.

3. Results

In the first test where the resources were transferred in the local network using HTTP requests there was a reduction of the transfer time as the number of resources per Bundle increased starting from 56.88 s for 20 resources per Bundle and dropping to 10.42 s for 200 resources per Bundle. Using WebSocket messages 20 resources per Bundle were transferred in 38.03 s and 200 resources per Bundle for 7.33 s. (Table 1, Figure 2).

Table 1. Transfer time of EHR resources using HTTP requests and WebSocket messages in the local network for 20, 50, 100 and 200 resources per bundle.

resources per Bundle	HTTP Requests	WebSocket Messages
20	56.78 s	38.03 s
50	37.95 s	17.89 s
100	18.24 s	10.80 s
200	10.42 s	7.33 s

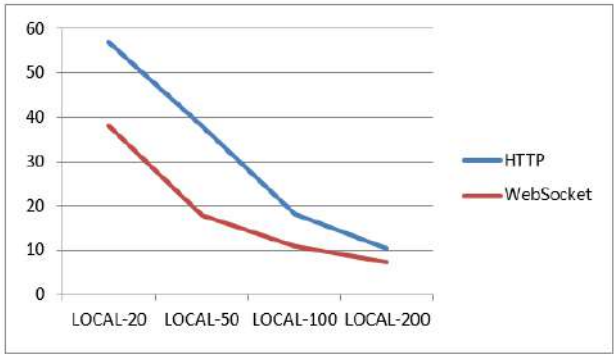


Figure 2. Transfer time of the EHR resources HTTP against WebsSocket in the local network.

The second test was carried out with a connection of the client from the WAN. The same reduction of transfer time was observed as in the local network starting from 509.51 s for 20 resources per Bundle and dropping to 60.22 s for 200 resources per Bundle using HTTP Requests. The observed transfer times using WebSocket messages started from 167.14 s for 20 resources per Bundle and fell to 24.22 s for 200 resources per Bundle. (Table 2, Figure 3).

Table 2. Transfer time of EHR resources using HTTP requests and WebSocket messages in the remote network for 20, 50, 100 and 200 resources per bundle.

resources per Bundle	HTTP Requests	WebSocket Messages
20	509.51 s	167.14 s
50	205.35 s	74.54 s
100	103.88 s	39.15 s
200	60.22 s	24.22 s

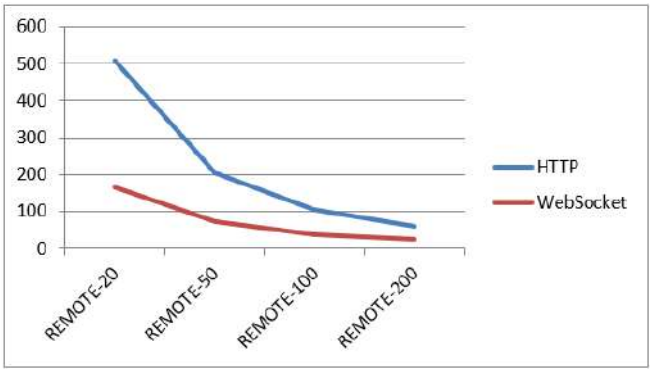


Figure 3. Transfer time of the EHR resources HTTP against WebSocket in the WAN.

An example related to the healthcare domain could be given when transferring data between two healthcare institutions. Each institution has a private network and the data needs to pass across the WAN. A possible solution to the problem where the better performance of WebSocket could be taken advantage of and the HTTP connection protocol could still be utilized is to use tunneling. Tunneling is the process of transmitting data across the network by encapsulating a packet and protocol into another packet with different protocol. A proxy server could be placed in the demilitarized zone (DMZ) of the private network of the healthcare institution where the FHIR server storing the data is located. This proxy can be used to tunnel the HTTP requests over WebSocket. A WebSocket client located in the second institution establishes a secure connection over the WAN and requests the data over this WebSocket connection tunneling HTTP. (Figure. 4)

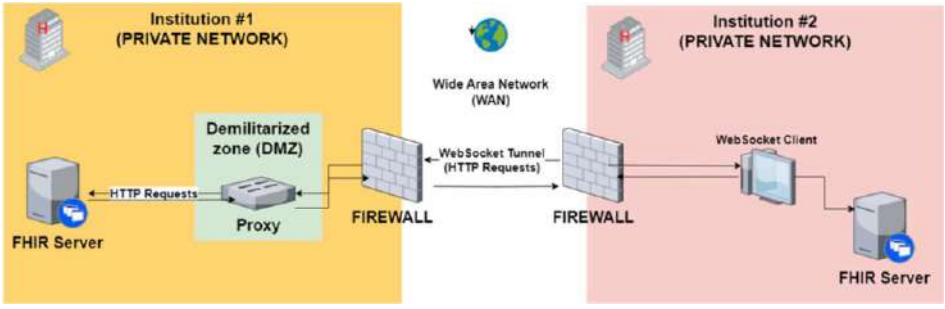


Figure 4. Clinical example: transfer of EHR resources using WebSocket tunnel.

4. Discussion

The results show that the most important factor for the transfer time of the data is the number of resources per Bundle. Increasing this number leads to reduction of the requests or messages needed. The requests needed to transfer 45000 Resources using 20 resources per Bundle are 2250 and for 200 resources per Bundle these are 225. With the HTTP connection the client makes a request and after the server response the

connection is closed. This process is repeated for each Bundle until all data is received. In the local network the transfer time decreases almost fivefold as the resources per Bundle increase to 200. In the remote network where each HTTP connection suffers from high-latency this decrease is over 9 times. The use of WebSocket connection improves the transfer times dramatically. The duration is 18 s shorter in the local network and the staggering 342 s shorter in the WAN for the 20 resources per Bundle transfer. The difference is becoming less prominent with the reduction in the number of requests. In our work we do not consider the time needed for establishing the WebSocket connection, which starts with an HTTP request with an upgrade header. This could be considered as a limitation of the work.

5. Conclusion

HTTP-based RESTful APIs are a convenient way to implement EHR servers; on the other hand, HTTP becomes a bottleneck when transferring large amount of data. WebSocket shows much better transfer times and thus its superiority in such situations. This problem can be addressed by developing a new communication protocol or by using network tunneling to handle large data transfer of EHRs.

References

[1] Cowie MR, Blomster JI, Curtis LH, Duclaux S, Ford I, Fritz F, Goldman S, Janmohamed S, Kreuzer J, Leenay M, Michel A, Ong S, Pell JP, Southworth MR, Stough WG, Thoenes M, Zannad F, Zalewski A. Electronic health records to facilitate clinical research. Clin Res Cardiol. 2017 Jan;106(1):1-9. doi: 10.1007/s00392-016-1025-6.

[2] Evans RS. Electronic Health Records: Then, Now, and in the Future. Yearb Med Inform. 2016 May 20;Suppl 1(Suppl 1):S48-61. doi: 10.15265/IYS-2016-s006. PMID: 27199197; PMCID: PMC5171496

[3] Ayaz M, Pasha MF, Alzahrani MY, Budiarto R, Stiawan D. The Fast Health Interoperability Resources (FHIR) Standard: Systematic Literature Review of Implementations, Applications, Challenges and Opportunities. JMIR Med Inform. 2021 Jul 30;9(7):e21929. doi: 10.2196/21929. Erratum in: JMIR Med Inform. 2021 Aug 17;9(8):e32869. PMID: 34328424; PMCID: PMC8367140.

[4] Fielding, Roy Thomas (2000). "Architectural Styles and the Design of Network-based Software Architectures". Dissertation. University of California, Irvine..

[5] https://build.fhir.org/bundle.html

[6] V. Wang, F. Salim, and P. Moskovits, The Definitive Guide to HTML5 Websocket. New York, NY: Apress, 2013.

[7] D. Skvorc, M. Horvat and S. Srbljic, "Performance evaluation of Websocket protocol for implementation of full-duplex web streams," 2014 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, Croatia, 2014, pp. 1003-1008, doi: 10.1109/MIPRO.2014.6859715.

[8] R. K. L. Ko, M. Kirchberg, B. S. Lee and E. Chew, "Overcoming Large Data Transfer Bottlenecks in RESTful Service Orchestrations," 2012 IEEE 19th International Conference on Web Services, Honolulu, HI, USA, 2012, pp. 654-656, doi: 10.1109/ICWS.2012.107.

dHealth 2024
D. Hayn et al. (Eds.)

doi:10.3233/SHTI240024

Relationship Between Dietary Intakes and Elevated Diastolic Blood Pressure Among Children: A Cross-Sectional Study Using Wearable Devices

Kukiat TUDPOR[a,b], Panthicha CHINNAKHOT[a]
Montira BOONSORN[a], Wiriyaporn LOMTHAISONG[a]
Worrameth KAREECHUM[c], Charnyuth SRIPHUWONG[c], Le Ke NGHIEP[d] and
Niruwan TURNBULL[a,b,1]

[a] *Faculty of Public Health, Mahasarakham University, Maha Sarakham, Thailand*
[b] *Public Health and Environmental Policy in Southeast Asia Research Cluster (PHEP SEA Thailand), Mahasarakham University, Maha Sarakham, Thailand*
[c] *Ban Yang Sub-district Health Promoting Hospital, Maha Sarakham, Thailand*
[d] *Vinh Long Department of Health, Vietnam*

Abstract. Background: Elevating systolic blood pressure (SBP) and diastolic blood pressure (DBP) independently influences clinical outcomes and adverse cardiovascular events. Blood pressure can be affected by modifiable (such as diets and physical activities) and non-modifiable factors (such as age and gender). Elevated blood pressure (EBP or formerly prehypertension) during childhood is associated with hypertension incidence in later adulthood. Objectives: This cross-sectional study investigated modifiable risk factors for blood pressure among children (aged 3-12). Methods: We employed wearable devices to monitor the blood pressure of 45 preschool and primary school children and analyze this data with secondary blood pressure data of their parents from electronic medical records. Results: EBP phenotypes in children (offspring) were not related to their parent's blood pressure phenotypes (P = 0.15 and 0.19 for SBP). Consumption of high saturated fat (P = 0.032), copper (P = 0.026), and vitamin B12 (P = 0.032) was associated with a significant increase in DBP. Daily sodium intakes between normal and DBP hypertensive groups were not significantly different (P = 0.75). Conclusion: This study indicates that dietary intakes of high saturated fat, copper, and vitamin B12, but not parental blood pressure statuses, determine high diastolic blood pressure among children regardless of daily sodium intake. Early dietary consumption behavioral adaptation should be considered to prevent further hypertension in adulthood.

Keywords. Elevated blood pressure, Pediatric hypertension, Dietary intake, Wearable device, Electronic medical records

[1] Corresponding Author: Niruwan Turnbull, Mahasarakham University, Maha Sarakham, Thailand, E-Mail: niruwan.o@msu.ac.th

1. Introduction

Both diastolic blood pressure (DBP) and systolic blood pressure (SBP) elevations have an independent effect on unfavorable cardiovascular events and clinical outcomes [5]. Both modifiable (diet and exercise) and non-modifiable (age and gender) factors can have an impact on blood pressure [10]. According to a meta-analysis, there is a substantial correlation between isolated diastolic hypertension (IDH), as defined by the 2018 European Society of Cardiology (ESC) definition (SBP of < 140 mmHg with DBP ≥ 90 mmHg), and a higher risk of composite cardiovascular diseases (CVD), cardiovascular mortality, and hemorrhagic stroke [8].

Three independent cohort studies showed a substantial correlation between a declined left ventricular systolic performance and elevated DBP, whereas elevated SBP was linked to decreased diastolic function [16]. In adults, total and low-density lipoprotein (LDL) cholesterol levels are highly associated with dietary saturated fatty acids and trans-fatty acids, recognized CVD indicators [19]. In contrast, according to evidence-based dietary guidelines, consuming a diet rich in fruits and vegetables, whole grains, and lean protein sources is beneficial to maintain cardiovascular health [13]. It is known that childhood elevated blood pressure (EBP, formerly known as prehypertension) is linked to the incidence of hypertension in later life [18]. Thus, pediatric blood pressure monitoring and treatment can prevent adult hypertension and CVD [2]. It is well established that children's consumption behaviors might be related to their parents. However, nowadays, most children live with their grandparents in Thailand, and it is challenging to collect their parent's health data directly. The health data of Thai people aged ≥ 30 are stored as electronic medical records (EMR) called the HOSxP system. This digital health database can be helpful for statistical analysis. However, there is no research on the relationship between dietary intake and blood pressure among schoolchildren and their parents in Thailand. The objective of the present study was to investigate the relationship between dietary intake and blood pressure among preschool and primary school children. Their SBP and DBP phenotypes were also compared with their parents.

2. Methods

2.1. Study design, participants, and research tools

This cross-sectional study included 45 school students (3-12 years old) and their parents (n = 90) from 3 schools in Kantarawichai district, Maha Sarakham, Thailand. The demographic data are depicted in Table 1. The children's blood pressure was collected using a wrist-worn monitoring device calibrated with the hospital's standard blood pressure measuring machine (HBP-9030, Omron, Japan) (Fig. 1) between 8:00-10:00 AM. Cutoffs for EBP and hypertension were based on a clinical practice guideline for screening and management of high blood pressure in children and adolescents based on their age, gender, and height [6]. On the other hand, the EBP criteria for their parents were 120-129 mmHg (SBP) or 80-89 mmHg (DBP), according to the JNC-7 guideline [4]. Physical activity levels were categorized as adequate and inadequate according to the youth guidelines – those between 6 and 18 should engage in moderate-to-vigorous physical activity for at least 60 min/day [14]. Three structured questionnaires were used to interview the students and their parents or guardians. Nutrient consumption amounts were obtained by nutritionists from 24-hour dietary records and the INMUCAL-Nutrients V.4.0 program. Most students lived with their grandparents; thus, demographic data of their parents were obtained from electronic medical records (HOSxP for primary healthcare, version 3).

Figure 1. A wearable device for blood pressure monitoring among primary school students

2.2. Statistical analyses

The relationships between categorical and continuous data were analyzed using χ^2 and Pearson's correlation tests, respectively. Means between two independent groups were compared using the independent t-test.

3. Results

3.1. Demographic data

Demographic data of the participants revealed that approximately 89% and 36% had elevated SBP and DBP, respectively, with a similar proportion of males and females. Age, height, weight, heart rate, sleep duration, and physical activity levels are detailed in Table 1.

Table 1. Demographic data of the preschool and primary school children (n = 45)

Variable	Normal SBP	EBP SBP	Normal DBP	EBP DBP
Gender (male/female)	2/3	21/19	14/15	9/7
Age (years)	10.6±0.7	8.6±0.5	9.1±0.5	8.4±0.9
Height (cm)	133.6±9.0	130.3±3.2	130.9±3.5	130.2±5.5
Weight (kg)	36.0±7.6	31.2±2.5	32.8±2.9	29.8±4.2
Heart rate (beat/min)	91.0±4.6	91.4±2.3	92.8±2.4	88.6±4.0
Sleep duration (hour/day)	8.4±0.2	8.3±0.2	8.3±0.2	8.4±0.2
Physical activity level (adequate/inadequate)	2/3	24/16	15/14	11/5

EBP, elevated blood pressure; SBP, systolic blood pressure; DBP, diastolic blood pressure.

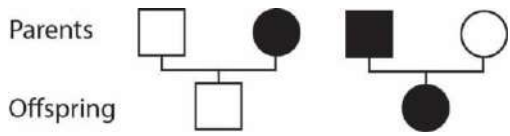


Figure 2. Representatives of blood pressure phenotypes of parents from electronic medical records and offsprings from wearable devices. White, normal phenotype; black, EBP phenotype; square, father; circle, mother.

3.2. Relationship between blood pressure phenotypes of parents and offspring

To determine the relationship between parental and offspring (preschool and primary school children) SBP and DBP phenotypes, the parental SBP and DBP were classified into normal, paternal EBP, maternal EBP, and parental EBP subgroups. Representatives of blood pressure phenotypes of parents from electronic medical records and offspring from wearable devices are shown in Fig. 2. The χ^2 test results showed that there were no significant relationships between parental and offspring SBP and DBP phenotypes (P-value = 0.15 and 0.19, respectively, Tables 2 and 3).

Table 2. Relationship between parental (n = 90) and offspring (n = 45) SBP phenotypes

Offspring SBP phenotype	Parental SBP phenotypes			
	Normal	**Paternal EBP**	**Maternal EBP**	**Familial EBP**
Normal	1 (5.3)	2 (33.3)	2 (18.2)	0 (0.0)
EBP	18 (94.7)	4 (66.67)	9 (81.8)	9 (100.0)
$\chi^2 = 5.34$	P-value = 0.15			

Note: Familial EBP**,** elevated blood pressure phenotypes in father and mother.

Table 3. Relationship between parental (n = 90) and offspring (n = 45) DBP phenotypes

Offspring DBP phenotype	Parental DBP phenotypes			
	Normal	**Paternal EBP**	**Maternal EBP**	**Familial EBP**
Normal	20 (64.5)	2 (33.3)	3 (100.0)	4 (80.0)
EBP	11 (35.5)	4 (66.67)	0 (0.0)	1 (20.0)
$\chi^2 = 4.72$	P-value = 0.19			

3.3. Relationship between dietary intakes and diastolic blood pressure among preschool and primary school children

Pearson's correlation analyses were performed to determine the relationship between preschool and primary school children's dietary intakes, SBP, and DBP. Significant correlations existed between saturated fat (P = 0.032), copper (P = 0.026), vitamin B12 (P =0.032) consumption, and DBP (Fig. 3). None of the diets were related to the SBP. In addition, sodium intakes were not significantly different between the normal and EBP DBP groups (Fig. 4).

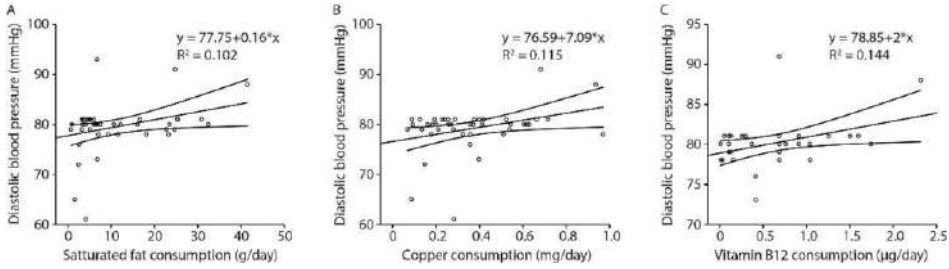


Figure 3. Pearson's correlation analyses of preschool and primary school children's nutrient consumption, SBP, and DBP.

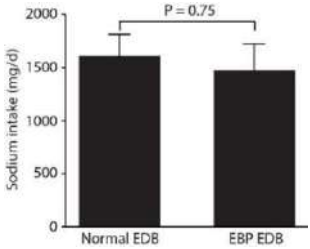


Figure 4. Sodium intakes in preschool and primary school children's normal DBP and EBP DBP.

4. Discussion

This cross-sectional study illustrates that the blood pressure phenotypes of the offspring were not related to the phenotypes of their parents. These results suggest that the genetic factors were not responsible for EBP in our children participants. Instead, the high dietary intakes of saturated fat, copper, and vitamin B12 were likely in charge of the phenotypes despite equal sodium intake. Based on a recent nationwide cross-sectional study, approximately 10% of the offspring (aged 10-18) inherited hypertension from their only fathers or mothers. The prevalence of hypertension doubled when the parents were hypertensive [9]. However, the EBP condition might not be susceptible enough to show the inheritance. Further study of the relationship of hypertension between parents and offspring is of fundamental importance.

Dietary intake is pivotal to blood pressure control. Previous studies showed that macronutrients (carbohydrates, proteins, and fats) leverage blood pressure [7]. Saturated fats raise blood pressure through oxidative stress activation [15]. Interestingly, it has been observed that children who consumed less saturated fat had significantly lower levels of LDL cholesterol, total cholesterol, and DBP, but not SBP [19]. These findings substantiate the cruciality of the saturated fat intake limit in children. Our findings were also in line with earlier studies showing that micronutrients such as minerals (potassium, magnesium, zinc, selenium, calcium, and copper) and vitamins (B6, B12, C, D, and E) potentially affect blood pressure [3]. However, most studies reported their beneficial impacts. The reasons for these discrepancies are unknown and need to be clarified using blood chemistry data.

According to recent epidemiological research, too high levels of micronutrients may be a significant factor in the onset of hypertension [12]. Although the recommended plasma vitamin B12 levels in children are approximately 200-1200 pmol/L [1], it has been found that > 700 pmol/L plasma vitamin B12 is associated with hypertension [20]. Similarly, plasma copper concentrations for EBP vs. healthy people were 125 and 112 μg/dL [12]. Therefore, micronutrient supplements should be cautiously used and personalized under healthcare professionals' supervision. Other confounders should also be considered. Children between 6 and 12 require between 9 and 12 hours of sleep daily [17]. A child's inadequate sleep is less than 9 hours [21]. In the present study, it was not clear whether physical activity affected blood pressure. However, it was previously shown that relatively low physical activity might be related to the EBP [11]. Future studies on the relationship between these factors are captivating.

In conclusion, this cross-sectional study demonstrates that the blood pressure phenotypes of the parents and their offspring did not correlate. These results suggest no genetic component to our child participants' EBP. Even though salt intake was the same, the phenotypes were likely brought on by increased dietary intakes of copper, vitamin B12, and saturated fat. Personalized diets and continuous monitoring of the EBP by healthcare professionals might be recommended.

Acknowledgment

This research project was financially supported by Mahasarakham University.

References

[1] A. Abildgaard, C.S. Knudsen, C.S. Hoejskov, E. Greibe, and T. Parkner, Reference intervals for plasma vitamin B12 and plasma/serum methylmalonic acid in Danish children, adults and elderly, *Clin Chim Acta* **525** (2022), 62-68.

[2] T. Azegami, K. Uchida, M. Tokumura, and M. Mori, Blood Pressure Tracking From Childhood to Adulthood, *Front Pediatr* **9** (2021), 785356.

[3] H.F. Chiu, K. Venkatakrishnan, O. Golovinskaia, and C.K. Wang, Impact of Micronutrients on Hypertension: Evidence from Clinical Trials with a Special Focus on Meta-Analysis, *Nutrients* **13** (2021).

[4] A.V. Chobanian, G.L. Bakris, H.R. Black, W.C. Cushman, L.A. Green, J.L. Izzo, Jr., D.W. Jones, B.J. Materson, S. Oparil, J.T. Wright, Jr., E.J. Roccella, L. National Heart, D.E. Blood Institute Joint National Committee on Prevention, P. Treatment of High Blood, and C. National High Blood Pressure Education Program Coordinating, The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure: the JNC 7 report, *JAMA* **289** (2003), 2560-2572.

[5] A.C. Flint, C. Conell, X. Ren, N.M. Banki, S.L. Chan, V.A. Rao, R.B. Melles, and D.L. Bhatt, Effect of Systolic and Diastolic Blood Pressure on Cardiovascular Outcomes, *N Engl J Med* **381** (2019), 243-251.

[6] J.T. Flynn, D.C. Kaelber, C.M. Baker-Smith, D. Blowey, A.E. Carroll, S.R. Daniels, S.D. de Ferranti, J.M. Dionne, B. Falkner, S.K. Flinn, S.S. Gidding, C. Goodwin, M.G. Leu, M.E. Powers, C. Rea, J. Samuels, M. Simasek, V.V. Thaker, E.M. Urbina, S. Subcommittee On, and C. Management Of High Blood Pressure In, Clinical Practice Guideline for Screening and Management of High Blood Pressure in Children and Adolescents, *Pediatrics* **140** (2017).

[7] L. Ge, B. Sadeghirad, G.D.C. Ball, B.R. da Costa, C.L. Hitchcock, A. Svendrovski, R. Kiflen, K. Quadri, H.Y. Kwon, M. Karamouzian, T. Adams-Webber, W. Ahmed, S. Damanhoury, D. Zeraatkar, A. Nikolakopoulou, R.T. Tsuyuki, J. Tian, K. Yang, G.H. Guyatt, and B.C. Johnston, Comparison of dietary macronutrient patterns of 14 popular named dietary programmes for weight and cardiovascular risk factor reduction in adults: systematic review and network meta-analysis of randomised trials, *BMJ* **369** (2020), m696.

[8] M. Huang, L. Long, L. Tan, A. Shen, M. Deng, Y. Peng, W. Yang, H. Li, Y. Wei, M. Li, F. Liao, C. Liu, A. Lu, H. Qu, C. Fu, and K. Chen, Isolated diastolic hypertension and risk of cardiovascular events: a systematic review and meta-analysis of cohort studies with 489,814 participants, *Front Cardiovasc Med* **8** (2021), 810105.

[9] S. Jang, S.T. Kim, Y.K. Kim, and Y.H. Song, Association of blood pressure and hypertension between parents and offspring: The Korea National Health and Nutrition Examination Survey, *Hypertens Res* **46** (2023), 368-376.

[10] A. Karmakar, A. Bhattacharyya, B. Biswas, A. Dasgupta, L. Bandyopadhyay, and B. Paul, Effect of educational intervention on risk factors of cardiovascular diseases among school teachers: a quasi-experimental study in a suburb of Kolkata, West Bengal, India, *BMC Public Health* **23** (2023), 2304.

[11] W. Kolanowski, K. Lugowska, and J. Trafialek, Increased Physical Activity at School Benefits Arterial Blood Pressure in Children-A Prospective Follow-Up Cohort Study, *Int J Environ Res Public Health* **19** (2022).

[12] C. Liu, Y. Liao, Z. Zhu, L. Yang, Q. Zhang, and L. Li, The association between serum copper concentrations and elevated blood pressure in US children and adolescents: National Health and Nutrition Examination Survey 2011-2016, *BMC Cardiovasc Disord* **21** (2021), 57.

[13] D.L. Mattson, J.H. Dasinger, and J.M. Abais-Battad, Dietary Protein, Chronic Salt-Sensitive Hypertension, and Kidney Damage, *Kidney360* **4** (2023), 1181-1187.

[14] J. Mitchell, Physical Inactivity in Childhood from Preschool to Adolescence, *ACSMs Health Fit J* **23** (2019), 21-25.

[15] H. Nakamura, H. Tsujiguchi, Y. Kambayashi, A. Hara, S. Miyagi, Y. Yamada, T.T.T. Nguyen, Y. Shimizu, D. Hori, and H. Nakamura, Relationship between saturated fatty acid intake and hypertension and oxidative stress, *Nutrition* **61** (2019), 8-15.

[16] J. Nikorowitsch, R. Bei der Kellen, A. Haack, C. Magnussen, J. Prochaska, P.S. Wild, M. Dorr, R. Twerenbold, R.B. Schnabel, P. Kirchhof, S. Blankenberg, M.R.P. Markus, and J.P. Wenzel, Correlation of systolic and diastolic blood pressure with echocardiographic phenotypes of cardiac structure and function from three German population-based studies, *Sci Rep* **13** (2023), 14525.

[17] S. Paruthi, L.J. Brooks, C. D'Ambrosio, W.A. Hall, S. Kotagal, R.M. Lloyd, B.A. Malow, K. Maski, C. Nichols, S.F. Quan, C.L. Rosen, M.M. Troester, and M.S. Wise, Recommended Amount of Sleep for Pediatric Populations: A Consensus Statement of the American Academy of Sleep Medicine, *J Clin Sleep Med* **12** (2016), 785-786.

[18] K.M. Redwine and B. Falkner, Progression of prehypertension to hypertension in adolescents, *Curr Hypertens Rep* **14** (2012), 619-625.

[19] L. Te Morenga and J.M. Montez, Health effects of saturated and trans-fatty acid intake in children and adolescents: Systematic review and meta-analysis, *PLoS One* **12** (2017), e0186672.

[20] B.H.R. Wolffenbuttel, M.R. Heiner-Fokkema, R. Green, and R.O.B. Gans, Relationship between serum B12 concentrations and mortality: experience in NHANES, *BMC Med* **18** (2020), 307.

[21] Y. Zhang, Z. Huang, M. Zhang, C. Li, Z. Zhao, X. Zhang, Y. Guan, N.G. Mudoti, and L. Wang, Sleep Status Among Children and Adolescents Aged 6-17 Years - China, 2016-2017, *China CDC Wkly* **5** (2023), 11-16.

dHealth 2024
D. Hayn et al. (Eds.)

doi:10.3233/SHTI240025

CareNet Tyrol - Information System Success Assessment for Case & Care Management Service

Sabrina B. NEURURER[a,b,1], Raphael KLEIN[c,d], Bernhard PFEIFER[a,b], Gabi SCHIESSLING[e], Eva SCHULC[f], Chiara MESSINA[f], Kristina REITER[c], Angelika RZEPKA[g] and Peter KASTNER[c]

[a] *Department of Clinical Epidemiology, Tyrolean Federal Institute for Integrated Care, Tirol Kliniken GmbH, Innsbruck Austria*
[b] *Division for Digital Health and Telemedicine, UMIT TIROL - Private University for Health Sciences and Health Technology, Hall i.T., Austria*
[c] *telbiomed Medizintechnik und IT Service GmbH, Graz, Austria*
[d] *Institute of eHealth, University of Applied Sciences FH JOANNEUM, Graz, Austria*
[e] *Tyrolean Federal Institute for Integrated Care, Tirol Kliniken GmbH, Innsbruck, Austria*
[f] *Division for Integrated Care, UMIT TIROL - Private University for Health Sciences and Health Technology, Hall i.T., Austria*
[g] *AIT Austrian Institute of Technology GmbH, Graz, Austria*

Abstract. Background: CareNet is the IT-based tool for Case and Care Management (CCM) in Tyrol, which facilitates standardised documentation of CCM activities. Objectives: Analysing the pilot usage of CareNet Tyrol. Methods: Evaluation of the success and user experience of CareNet, expert interviews and a questionnaire-based assessment. Results: Feedback from users in both phases indicated that the CareNet platform provides general benefits, but falls short of fully supporting the daily work of CCM experts and avoiding the need for parallel use of different documentation tools. Conclusion: This paper provides an insight into the ongoing transition to digital documentation for CCM at LIV Tyrol. While user feedback highlights areas for improvement, digital documentation is proved to be beneficial for the CCM team.

Keywords. case & care management, case management, ict in healthcare, patient care management, integrated care, ehealth, ict technologies

1. Introduction

Due to the aging population, healthcare is changing, resulting in increased cost [1]. The demand for services from various stakeholders, such as healthcare and social insurance providers, is continuously rising. The entitlement to assistance, which was initially defined purely in socio-legal terms, has now evolved into a working alliance that considers economic aspects and promotes cooperation among all parties involved [2].

[1] Corresponding Author: Sabrina B. Neururer, Department of Clinical Epidemiology, Tyrolean Federal Institute for Integrated Care, Tirol Kliniken GmbH, Innsbruck Austria, E-Mail: s.neururer@tirol-kliniken.at

The increasing complexity of caring for chronically ill people in different life situations and the complexity of care structures within the health care system make case management (CM) necessary [3]. The literature suggests that case and care management (CCM) is increasingly being used in various areas of the social and healthcare sector, including nursing, social work, and medicine [4, 5].

CCM combines a structured social and nursing care concept with individual coordination and organization of care and support services. The patient is supported throughout the entire course of the (chronic) condition wherever possible. The concept is described as cross-structural, networked care that is tailored to the needs of individual patients and optimizes the use of available services and resources [6]. The term 'care management' refers to the management tasks involved in providing comprehensive care, including assistance, nursing, inpatient treatment, and outpatient care [7]. Case management is a practical and widely applicable method for organising care for clients [4]. It involves different professions and actors in dealing with complex cases. The process supports independent client care and is tailored towards specific target groups. Further, it provides solution- and goal-orientation, and follows guidelines [6].

Supplementary, the ubiquitous use of information and communication technologies (ICT) is having an increasing impact on citizen's daily lives. The potential for developing and adapting ICT solutions in nearly all areas of society, including healthcare, is vast due to the wide range of possible uses of multimedia interfaces [8]. ICT acts as a bridge between healthcare and technology [9]. Further, cooperation and networking between stakeholders are necessary for planning public health strategies and implementing ICT tools to promote health and serve a better healthcare system [8].

This article analyses the pilot use of an ICT solution for CCM in Tyrol, Austria, and highlights the challenges and requirements for its implementation. The study also explores critical barriers and hurdles, as well as the advantages and disadvantages of using an ICT solution for CCM. Finally, the work addresses the acceptance of the implemented solution.

2. Methods

The initial stage of the projects involved examining the existing CCM procedures and processes at the Tyrolean Federal Institute for Integrated Care (LIV Tyrol) and conducting a comprehensive requirement analysis bases on an actual state analysis. The initial requirements for an integrated care network solution were identified through discussions on a regular basis with a group of experts from LIV Tyrol and the participants within the treatment network [10]. The results were utilised to establish and configure the CareNet Tyrol ICT solution, utilising core components of the KIT telehealth framework developed by the AIT Austrian Institute of Technology GmbH [11]. CareNet Tyrol comprises of the following features and functions (see to Figure 1): user administration, client view, and data export for evaluation.

Fundamental functions encompass (1) user management, (2) separate user groups for all districts in Tyrol, (3) a general overview, and (4) client and task lists.

The client view offers information on the client's administrative data, contact persons (such as family doctor, specialist, relatives, etc.), status of the care process (ongoing, evaluation, completed), notes with categories (such as contact information, goals, assessments, etc.), task creation and processing, task categories, appointment series and reminders, and document uploads in different formats with folder structure.

In March 2023, the 'CareNet Tyrol' service was launched on the IT network of Tirol Kliniken GmbH [12]. To assess the success of the information system in an applied environment, we established a two-step approach.

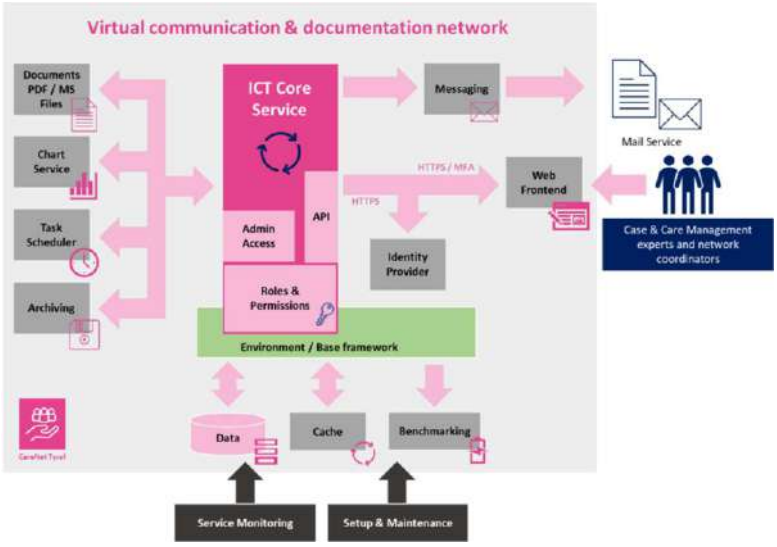


Figure 1. Visualization of the CareNet Tyrol system architecture showing the main components of the software platform that is based on the KIT telehealth framework [10].

Step 1 – Expert interviews after pilot application

After a three-month pilot application, expert interviews were conducted. The experts were selected based on their extensive experience and expertise in case and care management as well as their usage of CareNet during the pilot phase. To answer the questions, we conducted a comprehensive and systematic literature review and created a plan for guided interviews with experts. We evaluated the results using Mayring's qualitative content analysis [13]. For this study, we used transcript analysis as the summarizing method, which is the most appropriate for answering questions in qualitative content analysis. The material was reduced to the essentials, resulting in a summary of the relevant findings. We used both inductive and deductive methods, developing superordinate categories for the questions from the theory at the beginning. After completing the transcription, the transcripts were coded and the subcategories were categorized inductively. Finally, the theoretical and empirical results were summarised and the questions were answered [14].

Step 2 – Assessment questionnaire after six months of routine use

Following a six-month rollout phase, we assessed user experience using the Information System Success Model Survey, which is based on the Delone and McLean Information System Success Model [15]. The survey comprises questions in six dimensions: information quality (5 items), system quality (7 items), service quality (6 items), user satisfaction (5 items), net benefits (7 items), and intention to use (4 items). The questionnaire included six open-ended questions regarding benefits and potential areas for improvement. It was derived from previous research but has not undergone formal validation [16].

3. Results

In March 2023, after the system launch, four CCM experts used CareNet Tyrol to document the first 199 CCM cases in four pilot districts. 116 of all cases were female. At the end of the pilot phase, 76% (152) of the cases were closed, indicating that all stages (see Table 1) and functions of the system were utilised. After Phase 2, the roll-out to all Tyrolean districts, a total of 14 CCM experts documented an additional 399 cases of CCM (see Table 1), of which 223 were female. All clients in care were provided with a written informed consent, which had to be signed, to allow digital documentation in CareNet.

Table 1. Basic data of CCM clients in the pilot phase from March-June and rollout phase from July to December 2023 including gender (male (m)/female (f)) and stages.

Stage	Pilot phase (1) CCM cases (m/f)	Roll out phase (2) CCM cases (m/f)
Overall	199 (83/116)	399 (176/223)
Active	34 (13/21)	180 (70/110)
Pending	8 (4/4)	43 (20/23)
Finished	152 (63/89)	162 (77/85)
Paused	5 (3/2)	12 (8/4)
Canceled	0	2 (1/1)

After the pilot phase, and a few adjustments to the system configuration, based on feedback from phase one team, the system was rolled out to all nine Tyrolean districts. Fourteen CCM experts received individual and group training via video sessions. Table 2 displays the distribution of documented cases across the Tyrolean districts.

Table 2. Data from districts in Tyrol including gender (male/female), between July and December 2023.

Tyrolian districts	CCM cases (m/f)	Tyrolian districts	CCM cases (m/f)
Imst	52 (23/29)	Kufstein	14 (6/8)
Innsbruck	36 (14/22)	Landeck	38 (18/20)
Innsbruck Land-Ost	39 (16/23)	Lienz	76 (31/45)
Innsbruck Land-West	26 (11/15)	Reutte	59 (27/32)
Kitzbühel	25 (15/10)	Schwaz	34 (15/19)

Results of the expert interviews

The expert interviews were conducted with four early-stage users at the end of the pilot phase in June 2023. The interview questions were prepared to focus on the main goals of the CareNet System, which aims to provide a general tool for documentation and workflow support for CCM work. The experts' responses were summarised based on whether one aspect was evaluated as barrier or as advantage (refer to Table 3).

Table 3. Summary of results of expert interview after pilot phase

Feedback topics	Barriers	Advantage
Use of standardized assessment		x
Use of a flexible system architecture	x	
Consideration of the GDPR aspects		x
Change of documentation and workflow	x	x
Standardization of data collection and evaluation		x
Helps to improve teamwork		x
Helps to improve quality of care for clients		x

Overall, the responses indicated that the users perceived numerous benefits in using CareNet for CCM work. The main obstacles and challenges have been addressed in the areas of system architecture flexibility and changes to documentation and workflow. It is important to note that while a flexible system architecture may not directly benefit users, it can contribute to sustainable use and adaptation for various applications within the organization. It is important to note that the introduction of a new tool often faces resistance, particularly from users who already have established processes and procedures.

Results of the Information System Success Model Survey

In the survey of the six dimensions of the Information System Success Model (range 0%-100%), the mean results varied from 58.3% (intention to use) to 95.0% (service quality) (Figure 2), which indicated needs for improvement e.g. regarding intention to use, information quality, where the users see advantages in daily use. All users reported a high level of experience and confidence in using computers and smartphones in their daily work.

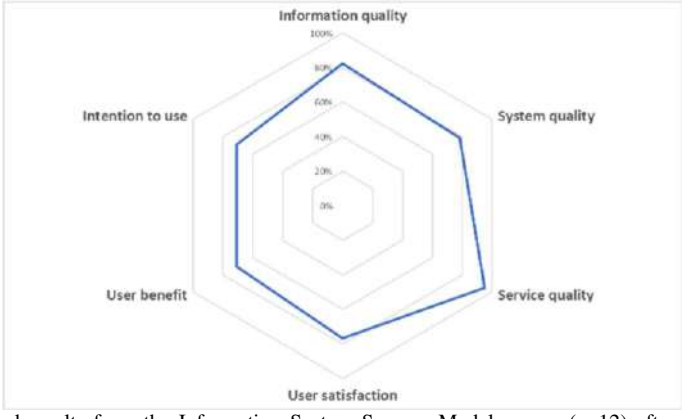


Figure 2. Aggregated results from the Information System Success Model survey (n=12) after 6 months (minimum=0%, maximum=100%). Information quality 76,9%; system quality 80,6%; service quality 95,0%; user satisfaction 76,8%; net benefit 70,7%; and intention to use 58,3%.

In terms of user benefit and intended use, the experts rated the CCM significantly lower than the other four topics. This was also reflected in the answers of the interview, which suggested (a) integrating additional functions to avoid using multiple documentation tools simultaneously, (b) providing an interface to Microsoft Outlook for joint calendar management and scheduling, and (c) enabling comprehensive data export for reporting. Regarding system quality, the majority rated the consideration of data protection measures and the use of two-factor authentication positively. However, it was noted that automatic logout, which requires re-login after a work interruption (e.g. due to a phone call), can be time-consuming.

4. Discussion

Today, leveraging digital applications is crucial for streamlining CCM workflows, offering numerous advantages and simplifications. Standardized data collection and

structure enable essential analyses and evaluations, aiding demand-oriented planning for future years in the field. Implementing ICT standards is essential for uniform data collection and evaluation, easing data exchange through stored information on a platform. This not only saves time and resources but also enhances the quality of care by providing necessary information in a structured and standardized manner. Ensuring secure storage of sensitive personal data is paramount, adhering to GDPR regulations.

Feedback from users highlighted that while the CareNet platform provided general benefits, it fell short in fully supporting daily CCM tasks, particularly in avoiding the parallel use of different documentation tools. To address this, it was recommended to implement an overview of all customer appointments post-evaluation and include functions for quick documentation of anonymous informal cases. Reducing redundant documentation and tool usage is crucial.

This work addresses a critical topic in the health and social care sector, shedding light on the significant shift to digital documentation in Tyrol's CCM. Despite areas for improvement highlighted by user feedback, digital documentation is proving advantageous for organization, execution, and documentation in CCM.

Acknowledgement

We gratefully acknowledge the CareManagers of CMT for their valuable contributions to this research.

References

[1] Wynia K. et al. Experiences of case managers in providing person-centered and integrated care based on the Chronic Care Model: A qualitative study on embrace, *PLoS ONE* **13**(11) (2018),
[2] Brinkmann V. *Case Management Organisationsentwicklung und Change Management in Gesundheits- und Sozialunternehmen*, 2nd ed. Kiel: Gabler, (2010).
[3] von Reibnitz C. *Case Management: praktisch und effizient*, **2**. Springer Berlin Heidelberg, (2009).
[4] Kollak I. and Schmidt S. *Fallübungen Care und Case Management*. Springer Berlin Heidelberg, (2015)
[5] Gessl AS, et al. Demand for community-based Case Management in Austria - a qualitative analysis. BMC Nurs. 2022 Jan 4;21(1):5.
[6] Kollak I. and Schmidt S. Instrumente des Care und Case Management Prozesses, Berlin, (2019).
[7] Thielscher C. *Medizinökonomie*, **1**. Gabler Verlag, (2012).
[8] Haluza D. and Jungwirth D. ICT and the future of health care: Aspects of health promotion, *International Journal of Medical Informatics* **84**(1) (2015), 48–57
[9] Annual IEEE Computer Conference et al., *10th IEEE International Conference on e-Health Networking, Applications and Services, 2008 HealthCom 2008 ; Singapore, 7-9 July 2008*. (2008).
[10] LIV Landesinstitut für Integrierte Versorgung Tirol - Care Management. https://www.caremanagement-tirol.at/ (accessed 2024-01-24).
[11] AIT Austrian Institute of Technology, KIT KEEP IN TOUCH. https://kit.ait.ac.at (accessed 2024-01-09).
[12] Tirol Kliniken GmbH, https://www.tirol-kliniken.at (accessed 2024-01-24).
[13] Mayring P. and Fenzl T. *Qualitative Inhaltsanalyse*. Springer, (2019)
[14] Klein R. Analyse des Piloteinsatzes einer IKT-Lösung für Case und Care Management in Tirol, Bachelor Thesis, FH Joanneum Graz, Graz, (2023)
[15] Delone W.H. and McLean E.R. The DeLone and McLean Model of Information Systems Success: A Ten-Year Update, Journal of Management Information Systems, 19:4, 9-30.
[16] Ammenwerth E, et al. HerzMobil, an Integrated and Collaborative Telemonitoring-Based Disease Management Program for Patients With Heart Failure: A Feasibility Study Paving the Way to Routine Care. JMIR Cardio. 2018 Apr 30;2(1):e11.

dHealth 2024
D. Hayn et al. (Eds.)

doi:10.3233/SHTI240026

HerzMobil Tirol PreOp – A Multidisciplinary Telemonitoring Project for Heart Failure Patients Prior to Cardiac Surgery

Karin HINTERBUCHNER[a,1], Elisabeth KLEINHEINZ[b], Bettina FETZ[b], Leo PÖLZL[a], Johannes HOLFELD[a], Bernhard PFEIFER[c,d], Sabrina NEURURER[c,d], Christian PUELACHER[e], Can GOLLMANN-TEPEKOEYLUE[a], Susanne KRESTAN[b], Andreas HUBER[b] and Gerhard PÖLZL[e]

[a] *Department of Cardiac Surgery, Medical University of Innsbruck, Innsbruck, Austria*
[b] *Tyrolean Federal Institute for Integrated Care, Tirol Kliniken GmbH, Innsbruck, Austria*
[c] *Department of Clinical Epidemiology, Tirol Kliniken GmbH, Innsbruck, Austria*
[d] *Division for Digital Health and Telemedicine, UMIT Tirol, Hall i.T., Austria*
[e] *Department of Internal Medicine III, Cardiology and Angiology, Medical University of Innsbruck, Innsbruck, Austria*

Abstract. Background: Patients with heart failure are at risk of perioperative complications with elective cardiac surgery. Objectives: Conception of a multidisciplinary telemedicine-assisted optimisation project for high-risk patients prior to elective cardiac surgery. Methods: Multidisciplinary concept design. Results: A pilot-project for 30 patients was developed. Conclusion: Design of the first preoperative telemonitoring-assisted optimisation project for high-risk patients undergoing cardiac surgery.

Keywords. Telemedicine, Preoperative Care, Heart Failure, Health Education

1. Introduction

Patients with heart failure undergoing cardiac surgery are at risk of perioperative complications and high mortality [1]. It is assumed that patients with a higher NT-proBNP value have a higher risk of intraoperative and postoperative complications and a higher mortality rate.

The aim of this paper is to present an approach of optimizing high-risk patients terms prior to their elective cardiac (bypass and/or valve) surgery based on the "HerzMobil Tirol" disease management program.

[1] Corresponding Author: Elisabeth Kleinheinz, Tyrolean Federal Institute for Integrated Care, Innsbruck, Austria, E-Mail: elisabeth.kleinheinz@tirol-kliniken.at

2. Methods

In order to adapt "HerzMobil Tirol" to supporting high-risk patients prior to their elective surgery, a multidisciplinary team developed a detailed concept for pre-operative optimisation. Additionally to the treatment process, the telemonitoring platform HerzMobil (telbiomed, HerzMobil CareManager) is used, which facilitates the monitoring of various relevant parameters, such as blood pressure, body weight, medication intake, well-being, etc.

3. Results

The project aims to optimize high-risk patients by optimising drug therapy and thus reducing NT-proBNP, improving quality of life and physical performance as well as self-management while decreasing depression and anxiety. In a first project phase, 30 patients will be enrolled in the pre-operative project three months prior to their surgery. Patients undergo disease-specific, device-specific, and preoperative training, alongside with comprehensive clinical assessments by specialist nurses.

This will be followed by a 3-month telemonitoring period. This phase is comparable to the classic HerzMobil Tirol approach, except for hospital physicians who will support the telemonitoring process from a clinical point of view. The telemonitoring platform captures vital data including blood pressure, heart rate, weight, medication adherence, and patient well-being. Additionally, information regarding medical parameters (history, laboratory parameters, vital signs, etc.) and several questionnaires (e.g. hospital anxiety and depression scale (HADS-D), EQ5D-5L, EHFScB-9) as well as the patients' satisfaction with the program are assessed at the beginning and the end of the program. A training therapist, who will help to increase the level of activity, accompanies the patients during these three months. Patient progress will be assessed through various tests including the 6-minute walk test and the sit-to-stand test, conducted at multiple time points throughout the project.

In addition to the parameters mentioned above, the length of stay in the intensive care unit, the post-operative need for ECMO and/or renal replacement therapy and the mortality rate will be assessed.

4. Discussion

To our knowledge, this concept is the first preoperative telemonitoring project for high-risk patients undergoing cardiac surgery. It is expected that patients will benefit from this project, as optimized/stable patients have a lower risk of severe complications during and after surgery. Preliminary data will be available within this year.

References

[1] Pichette M., Liszkowski M., Ducharme A., Preoperative Optimization of the Heart Failure Patient Undergoing Cardiac Surgery, *Canadian Journal of Cardiology* **33**(1) (2016), 72-79.

dHealth 2024
D. Hayn et al. (Eds.)

doi:10.3233/SHTI240027

Comparison of the Performance of the Fast Healthcare Interoperability Resources (FHIR) Subscription Channels

Nikola KIRILOV[a,1]

[a] *Institute of Medical Informatics, Heidelberg University Hospital, Germany*

Abstract. Background: The Fast Health Interoperability Resources (FHIR) standard was proposed and released to solve the interoperability problems of the electronic health records. The FHIR Subscription resources are used to establish real-time event notifications from the FHIR server to another system. There are several communication channels such as rest-hook and websocket. The objective of our work is to compare the performance of the FHIR subscription using the rest-hook and websocket channels. Methods: HAPI FHIR server, python websocket clients and HTTP endpoints were used to measure the processor and memory usage of the two subscription channels. Tests were performed with 5, 10, 15, 20, 30, 40, 50, 60, 70 and 80 clients. The performance was logged using windows performance monitor. Results: The rest-hook subscription showed near six-fold increase in resource utilization when increasing the clients from 5 to 80. On the contrary, the websocket subscription channel did not reach a two-fold increase. Conclusion: The type of the subscription channel should be carefully selected and load distribution should be considered when the number of clients grows.

Keywords. fhir, subscription, channel, websocket, rest-hook, performance

1. Introduction

The Fast Health Interoperability Resources (FHIR) standard was proposed and released to solve the interoperability problems of the electronic health records. It has been increasingly adopted and has been gaining popularity since its introduction. It is implemented as a Hypertext Transfer Protocol (HTTP)-based representational state transfer (REST) application programming interfaces (APIs) to perform operations on the resources. The resources are definitions of the common concepts in healthcare and include Patient, Observation, Encounter etc. [1]

Health Level 7 (HL7) application programming interface (HAPI) releases libraries and an open-source server under the Apache 2.0 license which can be used free of charge [2]. The FHIR Subscription resources are used to establish real-time event notifications from the FHIR server to another system. There are several communication channels such as rest-hook and websocket. Rest-hook channel sends notifications using Hypertext Transfer Protocol (HTTP) POST request and websocket channel sends messages over a websocket connection [3].

[1] Corresponding Author: Nikola Kirilov, Institute of Medical Informatics, Heidelberg University Hospital, Germany, E-Mail: Nikola.Kirilov@med.uni-heidelberg.de

Windows Performance Monitor is a powerful tool for examining system performance data in real time. Important performance parameters are the processor time and private bytes of a process. Processor time is the percentage of elapsed time that the processor spends to execute a non-idle thread. The Processor time counter can be considered a benchmark in understanding the utilization of the Central Processing Unit (CPU usage). The Private Bytes counter indicates the total amount of memory allocated by a process, excluding memory shared with other processes [4-6].

The objective of this work is to compare the performance of the FHIR subscription using the rest-hook and websocket channels.

2. Methods

HAPI FHIR server version 6.3.13, which was deployed on a Tomcat web server version 9 connected to a PostgreSQL 15 database, was used. Python script using the python requests library to send test data resources to the FHIR server was implemented. The data consisted of 1000 Observation resources and for each test the same resources were generated and used.

The Python websocket library was used to implement the websocket client and the Flask framework to create an HTTP endpoint for the rest-hook subscription. The clients were run on a separate machine, which was connected to the server using cable connection in a local area network. A test Patient and a Subscription resource with a websocket channel were created to notify connected clients when a new Observation resource is available. The websocket client establishes a connection with the server and sends a bind message with the id of the Subscription resource. After that it receives notifications when new resources are available. To warm up the server we sent 2000 Observation resources with 1 client connected. Hereafter 10 tests were carried out, where first 5 and then subsequently 10, 15, 20, 30, 40, 50, 60, 70 and 80 clients were connected and subscribed to the resource. Each measurement included the processor time (usage) in % and private bytes for the tomcat server process using the Windows' Perfomance Monitor. The data was written to a CSV file.

The same process was repeated for the rest-hook channel by cleaning the database and creating the test Patient and Subscriptions. The warm up procedure by generating 2000 Observation resources with 1 subscribed HTTP endpoint was again performed, followed by the tests. First 5 HTTP endpoints were subscribed, then 10, 15, 20, 30, 40, 50, 60, 70 and 80 respectively. The same performance parameters were measured and written to a CSV file. To obtain a full picture of the performance difference for each test case we calculated the full duration in seconds in which the CPU was under load by the tomcat server and the mean usage in %. Both subscription channels were compared multiplying the CPU usage by the duration.

To analyze the correlation between the number of subscribed clients and system resource utilization linear regression, Pearson correlation coefficient and the coefficient of determination were used.

3. Results

For the rest-hook channel a gradual increase in the mean CPU usage was identified, starting from 3.89 % for 5 clients and reaching 11.31 % for 40 clients. The duration

where the CPU was under load increased from 34 s to 45s. In the range 50-80 clients the mean usage did not change much maintaining values between 11.15% and 13.15%. Nevertheless, the duration rose from 45.00 s to 74.00 s (Table 1).

Table 1. Mean CPU usage (%) and duration (s) using rest-hook channel.

Subscribed clients	Mean usage	Usage duration
5	3.89 %	34.00 s
10	6.15 %	37.00 s
15	5.75 %	36.00 s
20	6.82 %	41.00 s
30	8.84 %	45.00 s
40	11.31 %	43.00 s
50	13.15 %	45.00 s
60	11.54 %	48.00 s
70	12.27 %	58.00 s
80	11.15 %	74.00 s

The websocket channel recorded mean CPU usage ranging from 1.03 % to 1.67 %, gradually increasing with the growth in the number of subscribed clients. The same tendency was observed for the usage duration, which increased from 35.00 s to 44.00 s. (Table 2)

Table 2. Mean CPU usage (%) and duration (s) using websocket channel.

Subscribed clients	Mean usage	Usage duration
5	1.03 %	35.00 s
10	1.17 %	35.00 s
15	1.22 %	34.00 s
20	1.22 %	35.00 s
30	1.48 %	36.00 s
40	1.53 %	43.00 s
50	1.47 %	44.00 s
60	1.51 %	44.00 s
70	1.67 %	37.00 s
80	1.91 %	42.00 s

The rest-hook usage × duration increased from 132.26 to 825.43, where websocket increased only from 41.5 to 70.17. A linear regression was used to test if the number of subscribed clients could predict the CPU usage × duration for both rest-hook and websocket. Subscribed clients were found to significantly predicted the usage × duration ($\beta_1 = 8.69$, $p < 0.00$ for rest-hook and $\beta = 0.43$, $p = 0.00$ for websockets). The correlation coefficient R^2 was 0.974 for rest-hook and 0.824 websocket (Table 3 and Figure 1).

Table 3. CPU usage (%) × duration (s). Rest-hook channel vs websocket channel. Correlation between subscribed clients and usage × duration values.

Subscribed clients	Rest-hook % × s	Websocket % × s
5	132.26	41.56
10	227.63	36.09
15	207.04	42.25
20	279.52	49.95

30	397.62	42.86
40	486.33	59.18
50	591.71	64.21
60	553.99	69.32
70	711.40	60.26
80	825.43	70.17
Correlation	$R^2 = 0.974$ ($p<0.00$)	$R^2 = 0.824$ ($p<0.00$)

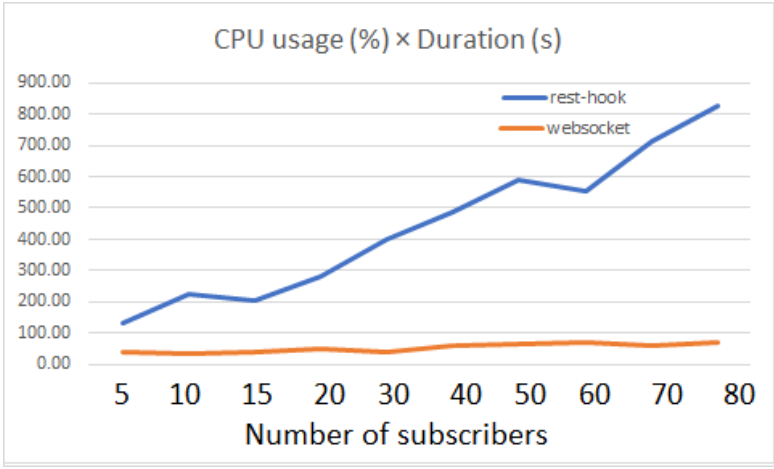


Figure 1. Line chart showing CPU usage (%) × duration (s). Rest-hook channel vs websocket channel.

The private bytes value resembling the allocated memory by the tomcat server did not change during the tests and stayed around 1.5 GiB.

The presented results may undoubtedly slightly vary depending on the hardware equipment and devices. The subscription services provide real-time notifications and are relevant to clinical use, especially to send information to doctors, nurses, physician assistants etc. inside an organization, as well as to general practitioners, medical specialists and patients. When developing such services load distribution should be taken into consideration. In order to distribute the load of the subscription service in the hospital network a dedicated subscription server could be used. This server could use FHIR Subscription over the websocket channel to receive notifications. In that way the load on the FHIR server will be reduced to a minimum. The subscription server can forward the subscriptions using REST hooks to dedicated HTTP endpoints for each department, which on the other hand can spread the subscriptions to websocket clients (doctors, nurses, physician assistants). A dedicated http endpoint for external users (general practitioners, medical specialists and patients) could also be connected. A model example to distribute the load of the subscription service in the hospital environment is given on Figure 2.

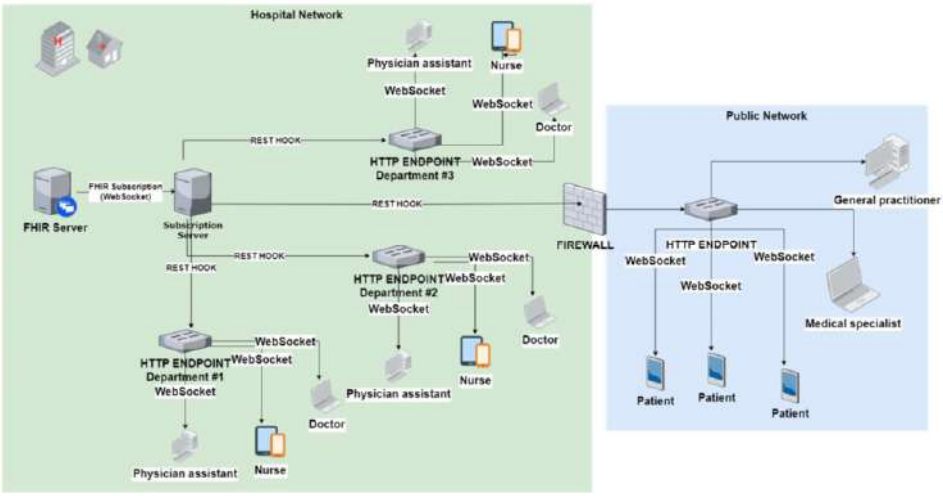


Figure 2. Model example to distribute the load of the subscription service in the hospital environment.

4. Discussion

The results showed that the rest-hook channel subscriptions consume more computing power. This is already apparent in the 5 clients test where rest-hooks use an average of 3.89 % CPU compared to the 1.03 % CPU of websockets for a comparable duration of 34 and 35 s. By increasing the number of clients for rest-hook subscriptions the usage jumps by almost 10 %, whereas websockets don't experience more than 1 % of increase in CPU usage. The same effect can by observed in the values of the CPU usage duration. Websockets put the CPU under load 10 s longer when the number of clients raises. Rest-hook, on the other hand, experiences a surge of 40 s.

The CPU usage × duration gives a full picture of the difference in the performance of the both subscription channels. The rest-hook values increase almost six-fold starting from 132.26 for 5 clients and going up to 825.43 for 80 clients. On the contrary, websocket subscription channel does not even show a two-fold increase with values of 41.56 and 70.17 for 5 and 80 clients respectively. Usage × duration and number of clients can fit a linear model for both web-socket and rest-hooks, but the coefficient of determination R^2 for rest-hook indicates a value of 0.974, which shows much stronger correlation with the number of subscribed clients compared to the websocket's 0.824.

The difference in the performance of the two subscription channels could be explained by the technology used to deliver the notification. Rest-hook channels use a HTTP request to send data, whereas websocket channels use websocket connections. To perform a HTTP request the server needs to create a request object, to send the request and wait for response. For Websocket, on the other hand, the server creates an object on client connection, which exists until the client disconnects. This object can be used to communicate. With the increase in the number of clients when using rest-hooks the number of requests that need to be generated raises and requires the creation of

many objects in the memory, which simultaneously are waiting for a response [7]. In the case where a websocket channel is used the server only needs to call the already existing object and to send a message [8]. This leads to big CPU usage difference between the two channel types. Despite of the better efficiency of the websocket channel, keeping a large number of open connections, as well as constantly sending keepalive messages to check whether the connection is still alive poses a considerable challenge for the server [9].

Difference in the memory usage could not be measured as the HAPI FHIR server as a Java Web Application uses a Java virtual machine (JVM) which allocates a fixed amount of memory, regardless of using it or not.

A limitation of our work is the lack of measurement of the processor usage during the websocket client connection.

5. Conclusion

FHIR subscription is an important feature, which offers real-time notifications over a subscription channel when new data is available. The type of the subscription channel should be selected carefully, due to the difference in the server resource usage, which changes with the increase of the number of connected clients. In such cases load distribution should be taken into consideration when designing subscription services inside an organization.

References

[1] Ayaz M, Pasha MF, Alzahrani MY, Budiarto R, Stiawan D. The Fast Health Interoperability Resources (FHIR) Standard: Systematic Literature Review of Implementations, Applications, Challenges and Opportunities. JMIR Med Inform. 2021 Jul 30;9(7):e21929. doi: 10.2196/21929. Erratum in: JMIR Med Inform. 2021 Aug 17;9(8):e32869. PMID: 34328424; PMCID: PMC8367140.

[2] Hussain MA, Langer SG, Kohli M. Learning HL7 FHIR Using the HAPI FHIR Server and Its Use in Medical Imaging with the SIIM Dataset. J Digit Imaging. 2018 Jun;31(3):334-340. doi: 10.1007/s10278-018-0090-y. PMID: 29725959; PMCID: PMC5959839.

[3] https://build.fhir.org/subscription.html (accessed on 31.01.2024)

[4] https://techcommunity.microsoft.com/t5/ask-the-performance-team/windows-performance-monitor-overview/ba-p/375481 (accessed on 31.01.2024)

[5] https://learn.microsoft.com/en-us/archive/technet-wiki/12984.understanding-processor-processor-time-and-process-processor-time (accessed on 31.01.2024)

[6] https://learn.microsoft.com/en-us/windows-hardware/drivers/debugger/using-performance-monitor-to-find-a-user-mode-memory-leak (accessed on 31.01.2024)

[7] S. A. Chowdhury, V. Sapra and A. Hindle, "Client-Side Energy Efficiency of HTTP/2 for Web and Mobile App Developers," 2016 IEEE 23rd International Conference on Software Analysis, Evolution, and Reengineering (SANER), Osaka, Japan, 2016, pp. 529-540, doi: 10.1109/SANER.2016.77.

[8] H. Nakajima, M. Isshiki and Y. Takefuji, "WebSocket proxy system for mobile devices," 2013 IEEE 2nd Global Conference on Consumer Electronics (GCCE), Tokyo, Japan, 2013, pp. 315-317, doi: 10.1109/GCCE.2013.6664841.

[9] https://dyte.io/blog/scaling-websockets-to-millions/ (accessed on 31.01.2024)

dHealth 2024
D. Hayn et al. (Eds.)

doi:10.3233/SHTI240028

Automated Clinical Trial Cohort Definition and Evaluation with CQL and CDS-Hooks

Anja SCHWAB[a,1], Anna M. LIN[a] and Stephan WINKLER[a]

[a] *Department of Medical and Bioinformatics, University of Applied Sciences Upper Austria, Hagenberg, Austria*

Abstract. Background: Patient recruitment for clinical trials faces major challenges with current methods being costly and often requiring time-consuming acquisition of medical histories and manual matching of potential subjects. Objectives: Designing and implementing an Electronic Health Record (EHR) and domain-independent automation architecture using Clinical Decision Support (CDS) standards that allows researchers to effortlessly enter standardized trial criteria to retrieve eligibility statistics and integration into a clinician workflow to automatically trigger evaluation without added clinician workload. Methods: Cohort criteria are translated into the Clinical Quality Language (CQL) and integrated into Measures and CDS-Hooks for patient- and population-level evaluation. Results: Successful application of simplified real-world trial criteria to Fast Healthcare Interoperability Resources (FHIR®) test data shows the feasibility of obtaining individual patient eligibility and trial details as well as population eligibility statistics and a list of qualifying patients. Conclusion: Employing CDS standards for automating cohort definition and evaluation shows promise in streamlining patient selection, aligning with increasing legislative demands for standardized healthcare data.

Keywords. Clinical Trial, Clinical Quality Language, Health Level 7, Clinical Quality Measures, Clinical Decision Support Systems

1. Introduction

1.1. Problem statement

Recruitment of patients for early-stage clinical trials is a complex and challenging process. Traditional methods often include manual evaluation of patients' eligibility by clinicians. The idea of automation is not new, but many solutions are specific to individual clinical domains and highly dependent on specific Electronic Health Record (EHR) systems. [1–4]

[1] Corresponding Author: Anja Schwab, University of Applied Sciences Upper Austria, Hagenberg, Austria, anja.schwab939@gmail.com

1.2. Background

Recent studies have highlighted the growing adoption of Clinical Decision Support (CDS) standards in tool development, underlining their potential to transform clinical trial recruitment processes [5]. Related Fast Healthcare Interoperability Resources (FHIR®)-based publications include estimations on the expected number of eligible patients [6], an automated recruitment system for a cardiology department [7], a preliminary trial eligibility assessment [8], and a recruitment registry [9]. The proposed solutions are either domain-specific or focus on a single aspect of the recruitment process.

1.3. Objective

The objective of this work is to design and implement an (a) EHR-independent, non-domain specific architecture that (b) enables researchers to enter standardized study criteria without technical authoring knowledge, (c) generates overall eligibility statistics and (d) obtains a list of qualifying patients to potentially contact them. For clinicians who participate in the enrolment process (e) all studies that a specific patient qualifies for should automatically be evaluated without any required knowledge of existing trials, and (f) provide automated evaluation and resulting display of relevant information integrated directly into their workflow avoiding any additional administrative effort.

2. Methods

The choice of standards is essential for ensuring interoperability while building EHR-independent, non-domain-specific systems. As a quickly evolving, EHR-independent standard for healthcare data exchange, the FHIR® standard [10] was chosen as the center of this architecture to ensure seamless, standardized data exchange for accurate clinical decision-making to attain objective (a).

2.1. CQL for Logic Definitions of Inclusion and Exclusion Criteria

The Clinical Quality Language (CQL) [11] allows researchers to define highly complex and specific inclusion and exclusion criteria that integrate directly with FHIR®-based medical records. In contrast to the limitations of assessing eligibility by regular FHIR® search [8], the CQL enables flexible and extensive authoring capabilities with the option of reusability of the same logic definitions for patient- and population-level evaluations. Narrative inclusion and exclusion criteria of the study cohorts can be authored into arbitrarily complex CQL expressions that are evaluated to assess an overall eligibility status of either *true* – a patient qualifies or *false* – a patient is not eligible for the trial. Existing CQL authoring tools allow for the generation of such CQL expressions without explicit knowledge of the underlying data retrieval syntax, e.g. via User Interface

(UI) prompts [12] to achieve the objective (b). FHIR® R4 is being used because the most current FHIR® release R5 is not yet integrated into a CQL engine.

2.2. CDS-Hooks for Patient-Level Triggers in the Clinician's Workflow

CDS rules can be defined by CQL expressions that are automatically applied to FHIR® data. CDS-Hooks [13] are a CDS standard that can trigger logic within an EHR system upon certain actions such as opening a patient chart and therefore, enabling a direct integration of clinical knowledge within a real-time healthcare system [11].

With a patient-view CDS-Hook implementation, the evaluation is triggered upon opening a patient file – allowing for automated checks that run in the background of the clinician's workflow. A dialog window could inform the clinician about available trials in real-time for one specific patient without additional interaction, and without the requirement that clinicians have explicit knowledge about existing trials accomplishing objectives (e) and (f) - ultimately increasing awareness of existing trials. CDS-Hooks act as the bridge between the abstract decision logic and its practical application within real-time clinical environments.

2.3. FHIR® Measures for Automated Population-Level Evaluations

The defined CQL expressions can be applied to an entire population by using a FHIR® Measure, invoked by the FHIR® Operation *$evaluate-measure*[2]. The resulting FHIR® MeasureReport returns a total number of eligible patients as well as a list of qualifying research subjects [14] targeting the objectives defined in (c) and (d).

2.4. Application of the Standards to the Clinical Trial Use Case

The UI is realized through a JavaFX prototype, which serves to demonstrate potential application within a research environment. Clinician workflows are simulated using manually created *hookInstances* that mimic automatic triggers in a real-world clinical context. The core component, the "CQL Factory," uses Java to generate necessary FHIR® resources and CQL through a template-oriented approach. This process is enhanced by additional functionalities for converting user-input criteria into executable CQL expressions. The architecture facilitates adaption to any language or tool that can interact with a FHIR® server via RESTful API. Figure 1 provides a detailed visualization of the architecture.

[2] http://hl7.org/fhir/R4/measure-operation-evaluate-measure.html

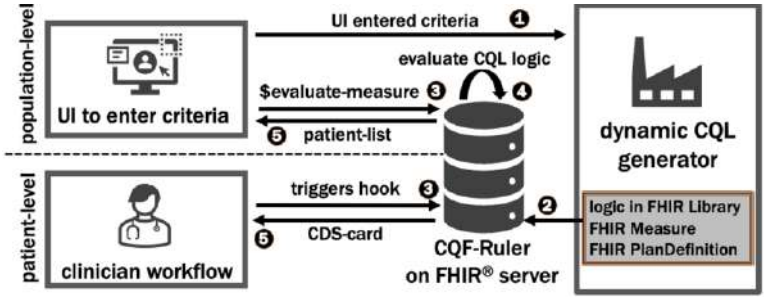


Figure 1. Architecture: New trials can be added by the researcher via the UI (1). The tool automatically translates the criteria into the corresponding CQL expressions targeting objective (b). The FHIR® Measure resource for population-level evaluation and the PlanDefinition[3] for the CDS-Hook are generated and include references to the CQL logic library. The resources are posted to a FHIR® server, along with any dependency libraries (2). The open-source project CQF-Ruler[4] generates the CDS-Hook from the corresponding FHIR® PlanDefinition with an embedded action trigger (3) to evaluate eligibility based on CQL expressions (4). A CDS-card returns any relevant information that should be displayed to the clinician (5). Within the clinical trial UI, the *$evaluate-measure* operation can be invoked (3). The CQF-Ruler again evaluates the CQL expressions (4) and generates statistics as well as optionally a list of all eligible Patient ids (5) depending on the report type.

The architecture uses key FHIR® resources for different aspects: Patient and related resources for CQL evaluations applicable to all use cases, the Library resource for CQL execution across all scenarios, PlanDefinition for CDS-Hooks in patient-level, and Measure for the *$evaluate-measure* operation in population-level evaluations.

3. Results

The proposed standards were applied to a real-world clinical trial scenario demonstrating the architecture's technical feasibility in patient eligibility evaluation and its potential for integration into clinical workflows by achieving all objectives defined in Section 1.

3.1. Application to a Real-World Clinical Trial[5]

The criteria of trial NCT04753502 as illustrated in Table 1 exhibit an ideal level of detail to demonstrate the framework's technical ability to accurately process and interpret the specific requirements of clinical studies. They are sufficiently detailed to test the system's capabilities but not so intricate as to overshadow the framework's functionality with the complexities of CQL expression generation. These five types of criteria have been selected: Age, Gender, Pregnancy Status, Tobacco Use, Condition inclusion and exclusion codes. The cohort criteria were simplified as the focus is on the application of CQL and not on authoring the logic itself. Therefore, the logic definition of the location of care, for example, is

[3] http://hl7.org/fhir/R4/cdshooksserviceplandefinition.html
[4] https://github.com/cqframework/cqf-ruler
[5] Relevant FHIR® Resources, CQL expressions as well as a postman collection with a complete walkthrough are publicly available on GitHub: https://github.com/1anja1/ClinicalTrialCQL

disregarded. The same workflow can be applied to trial criteria in any medical domain and any EHR that is FHIR®-based consequently achieving objectives (a) and (b). The Agency for Healthcare Research and Quality created a CDS-authoring tool that shows that the generation of more complex definitions is possible [12]. Table 2 gives an overview of three example FHIR® test patients.

Table 1. Simplified criteria of trial NCT04753502 from https://clinicaltrials.gov

Min Age	Max Age	Gender	Pregnancy Status	Tobacco Use	Inclusion Criteria	Exclusion Criteria
18	-	female	true	Not relevant	Appendicitis	-

Table 2. Overview of example test patients

Patient ID	Age	Gender	Pregnancy	Tobacco Use	Conditions
Patient1	5	male	false	false	Asthma (J45.909), Acute appendicitis (K35)
Patient2	25	female	true	false	Acute appendicitis (K35)
Patient3	44	female	true	true	Malignant neoplasm of breast (C50)

{"cards": [{ "summary": "Clinical Trial Qualification",
"detail": "The Patient qualifies for trial: Laparoscopic Treatment for Appendicitis During Pregnancy",
"source": {"label": "Clinical Trial Details",
"uri": "https://clinicaltrials.gov/study/NCT04753502"}, ...} (1)

GET server-base/Measure/ID/$evaluate-measure?reportType=subject-list (2)

{"resourceType": "MeasureReport",
"contained":[{…"entry": [{"item":{"reference":
"Patient2"}}]}], …
"group": [{"population": [{ … "count": 1, … }] … }]} (3)

Listing 1 shows the card that gets returned when the hook is invoked on the Patient resource with the id *Patient2* – who is eligible for the trial. In a clinician workflow, the CDS-Hook would get triggered upon opening a patient chart and the card would be displayed as a dialog window. This action gets simulated by sending a *hookInstance* of a *patient-view* hook to the FHIR® server which also serves as a CDS-Hooks endpoint consequently achieving the defined objectives (e) and (f). The HTTP request to the FHIR® server for population-level evaluation is shown in Listing 2. The reportType *subject-list* supports returning

a list of Patient ids. If a reportType is not specified, a summary report results in counting an anonymous total of eligible subjects. Listing 3 demonstrates the successful implementation of objectives (c) and (d) by supplying the number of qualifying patients ("count") as well as a list of their ids upon calling the *$evaluate-measure* operation on all three test patients shown in Table 2. Using this, researchers can gain insights into eligibility statistics and, if the research platform's consent guidelines permit, they may also proactively reach out to eligible participants.

4. Discussion

It is important to acknowledge that criteria for real-world clinical trials can be exceedingly complex. CQL possesses the capability to handle the level of complexity required for comprehensive clinical knowledge. Existing CQL authoring tools [12] already showed the creation of CQL definitions based on user input. The novelty of the approach presented in this paper, however, extends beyond merely generating CQL. It is primarily aimed to evaluate the feasibility of utilizing CDS standards for automating patient selection in clinical trials. It encapsulates the entire workflow from the translation of criteria to its application in clinical workflows. This comprehensive process, involving a FHIR® Measure for population assessment and a CDS-service for integration into a clinician workflow underlines the power and versatility of CDS standards. This work lays the foundation for a CDS-based architecture for clinical trial patient selection with the potential to expand this framework in the future to accommodate the intricate complexities and specific cohort requirements prevalent in various medical domains.

4.1. Limitations & Outlook

As demonstrated in Chapter 3, it is technically feasible to create CDS logic with CQL which can be applied both at the patient and population level. However, in the absence of a generic mechanism for parameterizing measure logic within the FHIR® R4 *$evaluate-measure* operation, the utilization of diverse criteria remains unstructured. This limitation necessitates direct modifications to the CQL expressions. As a workaround, a CQL template script gets directly modified by replacing the specified default values of the parameters with the values defined by the researchers for the cohort.

The ability to pass the FHIR® Parameters Resource to *$evaluate-measure* in a future FHIR® version is anticipated to streamline this process significantly. It will enable the input of criteria into the CQL logic via the Parameters Resource, thereby possibly eliminating the need for a new library for each clinical trial. It is important to note that this impacts only the architecture implementers, not the researchers entering trial criteria.

There is potential for combining the CDS-Hook with a *Substitutable Medical Applications, Reusable Technologies (SMART)* application. This integration could facilitate seamless utilization within EHRs and health portals, enhancing the utility and applicability of the presented approach. Future research should explore the integration of the framework into a real-world FHIR®-based hospital information system, providing valuable insights into its applicability and effectiveness in a clinical setting.

5. Acknowledgements

Special thanks to Bryn Rhodes and his clinical reasoning team for their exceptional mentorship and expertise in knowledge authoring and their feedback, which have been crucial to the development of this paper.

References

[1] B. Kankova and G. Duftschmid, Reusing Data of an EU-Wide EHR System for Clinical Trials: A Capabilities Analysis., *dHealth* (2020), 17–22
[2] M. Dugas et al., Routine data from hospital information systems can support patient recruitment for clinical studies, *Clinical Trials* **7**(2) (2010), 183–189
[3] G. Melzer et al., Leveraging real-world data for the selection of relevant eligibility criteria for the implementation of electronic recruitment support in clinical trials, *Applied Clinical Informatics* **12**(01) (2021), 17–26
[4] G. De Moor et al., Using electronic health records for clinical research: the case of the EHR4CR project, *Journal of biomedical informatics* **53** (2015), 162–173
[5] P. Taber et al., New standards for clinical decision support: a survey of the state of implementation, *Yearbook of medical informatics* **30**(01) (2021), 159–171
[6] A. Banach et al., APERITIF–automatic patient recruiting for clinical trials based on HL7 FHIR, in *Public Health and Informatics*, IOS Press, (2021), 58–62.
[7] C. Scherer et al., Implementation of a clinical trial recruitment support system based on fast healthcare interoperability resources (FHIR) in a cardiology department, *European Heart Journal-Digital Health* **3**(4) (2022), ztac076–2795
[8] C. Gulden et al., Investigating the Capabilities of FHIR Search for Clinical Trial Phenotyping, *Studies in health technology and informatics* **253** (2018), 3–7
[9] A. C. Cheng et al., ResearchMatch on FHIR: Development and evaluation of a recruitment registry and electronic health record system interface for volunteer profile completion, *Journal of Clinical and Translational Science* **7**(1) (2023), e222
[10] Health Level Seven International, FHIR v4.0.1, (2019),. https://hl7.org/fhir/R4
[11] Health Level Seven International, Clinical Quality Language (CQL) v1.5.2, (2023),. https://cql.hl7.org/
[12] Agency for Healthcare Research and Quality, CDS Connect Authoring,. https://cds.ahrq.gov/cdsconnect/authoring (accessed January 20, 2024)
[13] HL7 & Boston Children's Hospital, CDS Hooks 2.0 Release, (2022),. https://cds-hooks.hl7.org/2.0/ (accessed January 20, 2024)
[14] A. M. Lin et al., From Authoring to Evaluating an Electronic Health Quality Measure–Applying Logic to FHIR® with CQL for Calculating Immunization Coverage, in *dHealth 2023*, IOS Press, (2023), 12–17.

dHealth 2024
D. Hayn et al. (Eds.)

doi:10.3233/SHTI240029

Machine Learning-Based Prediction of Malnutrition in Surgical In-Patients: A Validation Pilot Study

Diether KRAMER[a,b,1], Stefanie JAUK[a,b], Sai VEERANKI[a,b], Michael SCHREMPF[b,c], Julia TRAUB[a], Eva KUGEL[a], Anna PRISCHING[a], Sandra DOMNANICH[a], Maria LEOPOLD[a], Peter KRISPER[a,c] and Gerald SENDLHOFER[a,c]

[a] *Steiermärkische Krankenanstaltengesellschaft m.b.H. (KAGes), Graz, Austria*
[b] *PH Predicting Health GmbH, Graz, Austria*
[c] *Medical University of Graz, Graz, Austria*

Abstract. Background: Malnutrition in hospitalised patients can lead to serious complications, worse patient outcomes and longer hospital stays. State-of-the-art screening methods rely on scores, which need additional manual assessments causing higher workload. Objectives: The aim of this prospective study was to validate a machine learning (ML)-based approach for an automated prediction of malnutrition in hospitalised patients. Methods: For 159 surgical in-patients, an assessment of malnutrition by dieticians was compared to the ML-based prediction conducted in the evening of admission. Results: The model achieved an accuracy of 83.0% and an AUROC of 0.833 in the prospective validation cohort. Conclusion: The results of this pilot study indicate that an automated malnutrition screening could replace manual screening tools in hospitals.

Keywords. Machine learning, clinical prediction models, clinical decision support, malnutrition.

1. Introduction

Malnutrition is present in 10-65 % of hospitalised patients [1]. It often causes complex pathophysiological changes that significantly increase the risk of surgical interventions, severe drug therapies or certain diseases and can significantly worsen the patient's quality of life. Studies have shown that malnourished patients have longer hospital stays, more complications, higher treatment costs and higher mortality rates [2].

Screening of malnourished patients is crucial in order to perform effective nutritional management in hospitals, but most screening tools require additional assessments or documentation. Due to the busy hospital routine, these screenings are often incomplete or missing. Presenting an alternative to manual screening, recently published machine learning (ML) models have achieved AUROCs of 0.835 [3] and 0.869 [4] when predicting malnutrition on test data of hospitalised patients.

The aim of this pilot study was to validate a ML-based malnutrition prediction on a convenience sample of surgical in-patients. An automated malnutrition screening should be enabled, leading to less workload for clinicians and thereby support to recognize patients at risk and complete malnutrition screening for all patients.

[1] Corresponding Author: Diether Kramer, KAGes, Graz, Austria, Email: diether.kramer@kages.at

2. Methods

A random forest model was trained on EHR data of 232,953 admissions from internal medicine, surgery, urology, dermatology and otorhinolaryngology, including 27,296 patients classified as malnourished. Patient data was extracted from the EHR system of KAGes from 2012 until 2021 and contained demographic, laboratory and nursing data, diagnoses, procedures and medication. The model achieved an AUROC of 0.887 (95 %-CI: 0.8838-0.8903) on an independent test data set of over 77,650 admissions.

From September 2023 until January 2024, dieticians assessed the malnutrition status of patients admitted to the division of plastic, aesthetic and reconstructive surgery of the University Hospital Graz. Depending on staff resources, as many patients as possible were evaluated according to the GLIM criteria [5] in a blinded setting. Experts' ratings were then compared to ML predictions, which were performed automatically and in real time in the evening of admission using the model with 338 features.

3. Results

The malnutrition status was assessed for a convenience sample of 159 patients. Clinical experts rated 27 patients (17.0 %) as malnourished (see Table 1). The ML model achieved a sensitivity of 70.4 %, a specificity of 85.6 %, an accuracy of 83.0 % and an AUROC of 0.833 (95 %-CI: 0.7421-0.9229) in the prospective validation data set.

Table 1. Comparison of expert malnutrition rating with risk prediction of the machine learning model.

	Model prediction		
Expert Rating	**Low risk**	**High risk**	**Total**
No malnutrition	113 (85.6%)	19 (14.4%)	132 (100.0%)
Malnutrition	8 (29.6%)	19 (70.4%)	27 (100.0%)
Total	121 (76.1%)	38 (23.9%)	159 (100.0%)

4. Discussion

This validation pilot study indicates that the developed ML model is able to detect malnutrition within the first hours of admission with an excellent accuracy. The prediction was conducted in a prospective setting, fully automatically and based only on existing EHR data. This algorithm could be used either as a pre-screening for the existing tool or replace it entirely. Anyhow, in order to achieve more reliable results, a higher number of patients needs to be included in the ongoing research.

References

[1] A.R. Cass, and K.E. Charlton, Prevalence of hospital-acquired malnutrition and modifiable determinants of nutritional deterioration during inpatient admissions: A systematic review of the evidence, *J. Hum. Nutr. Diet.* **35** (2022) 1043–1058. doi:10.1111/jhn.13009.

[2] Z. Sun, X.-J. Kong, X. Jing, R.-J. Deng, et al., Nutritional Risk Screening 2002 as a Predictor of Postoperative Outcomes in Patients Undergoing Abdominal Surgery: A Systematic Review and Meta-Analysis of Prospective Cohort Studies, *PLOS ONE*. **10** (2015) e0132857.

[3] P. Timsina, H.N. Joshi, F.-Y. Cheng, I. Kersch, et al., MUST-Plus: A Machine Learning Classifier That Improves Malnutrition Screening in Acute Care Facilities, *J. Am. Coll. Nutr.* **40** (2021) 3–12.

[4] B.T. Jin, M.H. Choi, M.F. Moyer, and D.A. Kim, Predicting malnutrition from longitudinal patient trajectories with deep learning, *PLOS ONE*. **17** (2022) e0271487. doi:10.1371/journal.pone.0271487.

[5] T. Cederholm, G.L. Jensen, M.I.T.D. Correia, M.C. Gonzalez, et al., GLIM criteria for the diagnosis of malnutrition–A consensus report from the global clinical nutrition community, *Clin.Nutr.* **38**(2019) 1–9.

dHealth 2024
D. Hayn et al. (Eds.)

doi:10.3233/SHTI240030

Feasibility of an Integrated Telemonitoring System for Home-Recorded EEG with Automated AI-Based EEG Analysis

Clemens LANG[a,b,1], Silvia WINKLER[c], Johannes KOREN[a,b], Gerhard GRITSCH[d], David POLLREISZ[e], Stefan KÖPL[e], Tilmann KLUGE[d] and Christoph BAUMGARTNER[a,b,f]

[a] *Department of Neurology, Clinic Hietzing, Vienna, Austria*
[b] *Karl Landsteiner Institute for Clinical Epilepsy Research & Cognitive Neurology, Vienna, Austria*
[c] *Sigma Software Solutions OG, Vienna, Austria*
[d] *AIT Austrian Institute of Technology, Center for Health & Bioresources, Vienna, Austria*
[e] *Brainhero GmbH, Vienna, Austria*
[f] *Medical Faculty, Sigmund Freud University, Vienna, Austria*

Abstract. Background: Self-recorded EEG by patients at home might present a viable alternative to inpatient epilepsy evaluations. **Objectives and Methods:** We developed a novel telemonitoring system comprising seamlessly integrated hard- and software with automated AI-based EEG analysis. **Results:** The first complete study participation results demonstrate feasibility and clinical utility. **Conclusion:** Our telemonitoring solution potentially improves treatment of patients with epilepsy and moreover might help to better distribute resources in the healthcare system.

Keywords. DICOM EEG, telemonitoring, automated EEG analysis, epilepsy

1. Introduction

Epilepsy is a common neurological disorder affecting 50 million people worldwide. To reduce waiting times for inpatient epilepsy evaluations and offer evaluation for patients who can't be admitted as inpatients, home-recorded electroencephalography (EEG) by patients has been proposed as an alternative. Feasibility of such self-recorded EEG using different methodological approaches and hardware, e.g. EEG systems, has been demonstrated by other authors. Yet, those systems had shortcomings in terms of usability and practical aspects, such as wired EEG systems and no online (live) transmissions to treating clinicians, no integrated system, thus limiting possible applications [1,2].
Therefore, to further investigate home-recorded EEG, we aimed to develop an integrated EEG home-recording solution with automated transmission to treating clinicians as well as AI based EEG analysis in the FFG funded project DigitalEPI (project number FO999887651). We are now reporting the initial results of the feasibility study of our project. However, due to the very limited space on the two pages of this manuscript, we could not cover all aspects of our study and system used, such as user and device management, recording times and data volume, AI pipelines and DICOM SR generation.

[1] Corresponding Author: Clemens Lang, Department of Neurology, Clinic Hietzing, Vienna, Austria, E-Mail: c.lang@gesundheitsverbund.at

2. Methods

A novel integrated EEG home-recording solution was developed by Brainhero, AIT and Sigma in close collaboration with clinician scientists from the Karl Landsteiner Institute for Clinical Epileptology and Cognitive Neurology (KLI). The EEG recording system was based on an existing medical device (MDR class IIa) by Brainhero with a semi-dry EEG electrode cap, an EEG amplifier connected via Bluetooth to a tablet with a LTE modem, and a simple and user-friendly recording app for guidance through the recording process. Upon completion of the recording, the EEG data was transmitted encrypted over the internet to the Clinic Hietzing. Subsequently, the EEG files underwent automated conversion into DICOM EEG format based on a conversion routine by Sigma [3]. The DICOM EEG was then processed by the encevis AI-based automated spike detector [4], whose results were stored as a DICOM structured report (DICOM waveform annotation SR, currently in process of standardization). Both, the recorded EEG and spike detections were then made available in the encevis EEG viewer for review by clinicians. After an initial on-site instruction, patients were instructed to record EEG at home twice daily.

3. Results and Discussion

Presently, the DigitalEPI feasibility study has started at the KLI and we report the first patient's successful study participation. In total, recruitment of 18 patients is planned.
To investigate recording quality and validate against the gold standard of EEG recordings by hospital staff, the first 20 home EEG recordings were visually examined for artifacts and signal quality by trained neurophysiologists. The initial results are promising in comparison with conventional EEG recordings and prove to be more than sufficient for clinical EEG reporting. Patient feedback is overwhelmingly positive in terms user-friendliness of the system. Nevertheless, patient recruitment is still ongoing and more data are of course needed for final assessment of the newly developed system.
In conclusion, we demonstrate initial feasibility results of our novel integrated telemedical system allowing laypersons to self-record EEG at home, automatically recording, transmitting, converting recordings to standardized DICOM, as well as analyzing using AI algorithms. Therefore, we believe that our telemonitoring system stands out from other previously described solutions [1,2]. Thus, it might improve treatment options for patients with epilepsy and save costs in the healthcare system.

References

[1] Baum U, Baum AK, Deike R, Feistner H, Markgraf B, Hinrichs H, et al. *Feasibility assessment of patient-controlled EEG home-monitoring: More results from the HOME(ONE) study*. Clin Neurophysiol 2022;140:12-20.

[2] Biondi A, Simblett SK, Viana PF, Laiou P, Fiori AMG, Nurse E, et al. *Feasibility and acceptability of an ultra-long-term at-home EEG monitoring system (EEG@HOME) for people with epilepsy.* Epilepsy Behav 2024;151:109609.

[3] Lang C, Winkler S, Koren J, Huber M, Kluge T and Baumgartner C. *DICOM® integrated EEG data: A first clinical implementation of the new DICOM standard for neurophysiology data.* Clin Neurophysiol. 2023;155():107-112

[4] Fürbass F, Kural MA, Gritsch G, Hartmann M, Kluge T, Beniczky S. *An artificial intelligence-based EEG algorithm for detection of epileptiform EEG discharges: Validation against the diagnostic gold standard.* Clin Neurophysiol. 2020;131(6):1174-1179.

dHealth 2024
D. Hayn et al. (Eds.)

doi:10.3233/SHTI240031

Telemedicine for Ketogenic Dietary Treatment in Refractory Epilepsy and Inherited Metabolic Disease: State of Play and Future Perspectives

Alexander HÖLLER[a,b,1], Stefan WELTE[b], Anna Katharina SCHÖNLAUB[c], Charlotte UHLISCH[d], Sabine SCHOLL-BÜRGI[d], Anastasia MALE-DRESSLER[e], Bernhard PFEIFER[f,g] and Günter SCHREIER[b]

[a] *Division for Nutrition and Dietetics, University Hospital Innsbruck, Austria*
[b] *Digital Health Information Systems, Center for Health & Bioresources, AIT Austrian Institute of Technology, Graz, Austria*
[c] *Department of Pediatrics, Division of Pediatric Neurology, University Hospital Innsbruck, Austria.*
[d] *Department of Pediatrics I, Inherited Metabolic Disorders, University Hospital Innsbruck*
[e] *Department of Pediatrics and Adolescent Medicine, Medical University of Vienna*
[f] *Division for Digital Medicine and Telehealth, UMIT TIROL - Private University for Health Sciences and Health Technology, Hall in Tirol, Austria*
[g] *Tyrolean Federal Institute for Integrated Care, Tirol Kliniken GmbH, Innsbruck, Austria*

Abstract. Ketogenic dietary therapies (KDT) are diets that induce a metabolic condition comparable to fasting. All types of KDT comprise a reduction in carbohydrates whilst dietary fat is increased up to 90% of daily energy expenditure. The amount of protein is normal or slightly increased. KDT are effective, well studied and established as non-pharmacological treatments for pediatric patients with refractory epilepsy and specific inherited metabolic diseases such as Glucose Transporter Type 1 Deficiency Syndrome. Patients and caregivers have to contribute actively to their day-to-day care especially in terms of (self-) calculation and (self-) provision of dietary treatment as well as (self-) measurement of blood glucose and ketones for therapy monitoring. In addition, patients often have to deal with ever-changing drug treatment plans and need to document occurring seizures on a regular basis. With this review, we aim to identify existing tools and features of telemedicine used in the KDT context and further aim to derive implications for further research and development.

Keywords. ketogenic diet, telemedicine, refractory epilepsy, inherited metabolic disease

[1] Corresponding Author: Alexander Höller, Division for Nutrition and Dietetics, University Hospital Innsbruck, Austria, E-Mail: alexander.hoeller@tirol-kliniken.at

1. Introduction

Ketogenic dietary therapies (KDT) are diets that bring on a metabolic condition comparable to fasting and comprise a reduction in carbohydrates whilst dietary fat is increased up to 90% of daily energy expenditure. The amount of protein is normal or slightly increased. These dietary adaptions yield in a metabolic switch from the utilization of glucose as primary energy source for the central nervous system to liver ketogenesis and thus utilization of ketone bodies as primary energy substrate [1]. KDT are effective, well-studied and established non-pharmacological treatments for pediatric patients with refractory epilepsy and specific inherited metabolic diseases (IMD) such as Glucose Transporter Type 1 Deficiency Syndrome (GLUT1DS) [2,3]. Further, the evidence for KDT in adolescents and adults is growing in recent years [4]. Patients and caregivers have to contribute actively to their day-to-day care especially in terms of (self-) calculation and (self-) provision of dietary treatment as well as (self-) measurement of blood glucose and ketones for therapy monitoring. Patients also have to deal with ever-changing drug treatment plans and need to document occurring seizures. Thus, KDT lead to a significant amount of paper based documentation as well as frequent e-mail and telephone contact with health care professionals (HCP).
However, complexity of treatment, low rates of participation and poor retention rates seem to be barriers that have to be overcome in the future [5,6]. Especially during the COVID 19 pandemic, interest in telemedicine aspects in the monitoring and treatment of patients with refractory epilepsy increased strongly [7]. In addition, the International League Against Epilepsy (ILAE) calls for action to promote telemedicine for individuals with epilepsy [8]. Currently, there is a lack of telemedicine solutions for the purpose of therapy monitoring and therapy management of patients with refractory epilepsy or IMD undergoing KDT although that might reduce complexity and contribute to facilitate access, adherence and quality of KDT. Hence, we want to outline the current role of telemedicine in KDT management and present a conceptual framework with regard to the requirements from a clinical and digital health point of view for future research and development.

2. Methods

A systematic literature review in MEDLINE was conducted in December 2023 to identify existing telemedicine strategies for the treatment of patients with refractory epilepsy and IMD undergoing KDT. PICO criteria (Population, Intervention, Comparator, Outcome) were defined as follows: Population (refractory epilepsy, glucose transporter type 1 deficiency), Intervention (ketogenic diet); Comparator (none); Outcome (feasibility). Derived keywords and MeSH (Medical Subject Heading) terms were combined using the Boolean operators “OR” and “AND”. As we were primarily interested in telemedicine methods applied in clinical settings for treatment of refractory epilepsy or IMD with KDT and further wanted to increase the sensitivity of our search, we focused on the criteria Population and Intervention. Considering the predefined inclusion and exclusion criteria, a title- and abstract screening was performed, followed

by an assessment of the remaining articles, focusing on specific telemedicine methods and addressed requirements. Articles were included for full text assessment if original data or a specific concept were reported.

3. Results

Our search in the MEDLINE database identified 11 articles eligible for title and abstract screening. As two articles did not provide original research data or specific concepts, further assessment was performed for 9 studies that were published between 2012 and 2023 and originated from Argentina (n=2) [9,10], USA (n=2) [11,7], Italy (=3) [12,13,14], Brazil [15], and the United Kingdom [16].

Armeno et al. published an observational study with the aim to explore feasibility, effectiveness and safety of online guided KDT initiation and follow up versus standard outpatient KDT. They included children (n=37) between 0 and 18 years of age with drug resistant epilepsy who started KDT between January 2020 and April 2021. 18 patients received online treatment and 10 patients received standard outpatient treatment. The use of Zoom™, WhatsApp™, e-mail and telephone was considered as telemedicine approach in the intervention group. Telemedicine was considered feasible and there was no difference between both groups regarding safety and effectiveness [9].

Semprino et al. conducted a questionnaire-based survey to assess parent satisfaction with telemedicine guided KDT management. The survey did address parents of children between 8 months and 18 years with refractory epilepsy (n=54), treated at a secondary-care center in a region with limited financial resources. The use of messaging and video call via WhatsApp™ was considered as telemedicine approach. Parent's satisfaction was evaluated using a 13-item questionnaire. High parent satisfaction (96.3%) and recommendation-rate (72.2%) as well as easy access and helpfulness of professional coordinated social-network support groups (90%) were reported [10].

Kossoff et al. reported data on their experience with KDT initiation and monitoring during the coronavirus disease 2019 pandemic. Experience from the pediatric and adult ketogenic diet centers at Johns Hopkins University was described on the basis of four respectively 38 patients receiving KDT due to refractory epilepsy or GLUT1DS. Strategies comprising telemedicine aspects included the use of e-mail, telephone, video via Polycom™, Zoom™, Doximity™ (live or prerecorded), SharePoint (a web based Microsoft™ platform for data exchange) and an electronic medical record secure messaging system. Those strategies were applied for initiation, follow-up, patient education, regular visitation and second opinion visits [7].

Cervenka et al. performed a prospective, open-label proof-of-principle study over 3 months investigating the feasibility, safety, and effectiveness of e-mail administered KDT over the course of three months in adults who had refractory epilepsy (n=25). Patients received a manual of instructions on how to self-administer a 20g carbohydrate per day MAD. Unrestricted access via e-mail to the treating neurologist was provided and patients were evaluated at baseline and 3 months. Seizure-frequency, urine ketones and weekly weights were recorded on a provided paper-based calendar and transmitted via e-mail or fax on a monthly basis. Potential adverse effects were also reported in that way. After 1 month, 9 (41%) showed a >50% seizure reduction including one individual (5%) with >90% seizure reduction. After 3 months, 6 (27%) showed a >50% seizure reduction including 3 (14%) with >90% seizure reduction. Median number of e-mails

sent were 6 (IQR 1-19). After one month 95% of patients remained on KDT, after three months 64% remained on the diet [11].

Costa et al. conducted a questionnaire-based survey, to assess the effects of a ketogenic diet management app and web materials in addition to paper-based materials (n=22) compared to paper-based materials only (n=18), designed for caregivers and pediatric patients in the context of KDT. Main features of the "KetApp" were to record and monitor dietary parameters, generate recipes and menus, record blood glucose and ketosis values, record the anthropometric data and monitor them over time, and record epileptic seizure data. Perception and satisfaction were assessed using the Information Satisfaction Questionnaire as well as additional multiple-choice questions with a five-point Likert scale. Overall findings on satisfaction, attitude towards treatment and awareness indicate that web-based materials were beneficial for caregivers of children with refractory epilepsy receiving KDT [12].

Zini et al. published their concept regarding a prototype mHealth application, named "Ketty" for education and monitoring purposes in the treatment of patients with KDT. The main features developed to be supported by "Ketty" were therapy assignment (nutritionist), coaching (dietitian), self-management (patient), and assessment (nutritionist, dietitian). Key roles were patient/caregiver, doctor and dietitian. Patients were not only able to use the application as source of information, but also to access food databases, to generate and calculate recipes, to document their diet history, and to track ketones, glycemia and epilepsy crises. HCP were able to remotely prescribe the dietary regimen, coach the patients, browse data and analyze trends [13].

Ferraris et al 2020. retrospectively evaluated e-mails regarding frequency and content sent to and received from their patients or caregivers (n=34) under KDT during the first year of therapy. Affected patients were 2-17 years old and either diagnosed with drug-resistant epilepsy (n=14) or GLUT1DS (n=20). Median mails sent were 36 (IQR 23.0-64.0) per family. GLUT1DS patients sent more e-mails than patients with refractory epilepsy (median 39.0 (IQR 25.5-56.5) vs. median 26.0 (IQR 19.0-65.0)). Higher e-mail exchange occurred in the group that showed increase in its linear growth (median 83.5; IQR 48.0-102.0) [14].

Lima et al. shared their perspective on the challenges of KDT in the public healthcare system in Brazil and provided information on challenges, perspectives, and their experience of telemedicine assisted KDT follow-up in seven patients during the COVID 19 pandemic. Applied telemedicine strategies comprised the use of telephone, e-mail, messaging applications and voice calls. They addressed hazards to treatment such as long travel distances for clinical visits, financial problems as well as anxiety and stress caused by the pandemic and discussed the use of telemedicine strategies to reduce costs and increase feasibility for affected individuals [15].

Bara et al. performed an online survey on patient and caregiver views from five centers in the United Kingdom about video consultations in the context of KDT. Patients and caregivers (n=40) affected by refractory epilepsy or metabolic conditions, where KDT was indicated, were included. The survey comprised 16 items with open and closed questions, including Likert scales. Video consultations were common (57,5%) and seen as at least partly preferential to on-site consultations (45%), whereas 22.5% of participants reported that they would not like to have video consultations. Benefits were reported in terms of saving travel time, reducing stress by not needing parking slots nearby, and not having to take time off work. In addition, the positive effect on the environment was stressed by 30% of participants. Not being able to get blood tests (55%), weight or height check as well as the less personal contact during video

consultation were frequently seen as disadvantages (42.5% each) [16]. Table 1 gives an overview on the tools and features of telemedicine strategies applied in the context of KDT for refractory epilepsy and IMD.

Table 1. Results from the systematic literature review in Medline – Tools and features

Study	Tools	Features
Armeno et al. 2022	telephone, e-mail, Zoom™, WhatsApp™	messaging, instant messaging, voice calls, video calls
Semprino et al. 2020	WhatsApp™	instant messaging, voice calls, video calls
Kossoff et al. 2020	telephone, e-mail, video conference, SharePoint (Microsoft™), electronic medical record secure messaging system	messaging, instant messaging, voice calls, video calls, video conference, store, organize, share, and access information
Costa et al. 2021	video, website, app (KetApp)	share, and access information, therapy monitoring, therapy management
Lima et al. 2020	telephone, e-mail, messaging applications and voice calls	messaging, instant messaging, voice calls, video calls
Zini et al. 2018	app (Ketty)	share, and access information, therapy monitoring, therapy management
Bara et al. 2023	video consultation	video calls
Ferraris et al. 2020	e-mail	messaging
Cervenka et al. 2012	e-mail	messaging, share, and access information

4. Discussion

Telemedicine for KDT is feasible and might be necessary during a pandemic. Further, it certainly bears potential benefits for patients, caregivers and the society with regard to stress reduction, cost savings, increased efficiency and the reduction of carbon emissions. Two studies included only adult patients [11,15] whereas six studies included caregiver-child dyads [9,710,12,14,16]. In one study, the concept of a specific app is described [13]. Published strategies primarily target the use of communication tools as add-ons or alternatives to on-site clinical visits [9,10,14,15,16]. Most commonly, telephone, e-mail and different types of messaging tools were applied. Four centers used several channels in parallel to communicate and provide information. Five centers relied on only one channel. Further, published efforts also aim at facilitating therapy management via online provision of information and providing easy-access to professional expertise via e-mail management [7,11]. Preliminary experiences from two centers exist regarding the use [12] and development [13] of specific applications trying to enable comprehensive therapy monitoring and management features such as seizure documentation, blood glucose and ketone monitoring, dietary prescription, calculation and tracking. Table 2 maps general, disease and KDT specific requirements for future digital health applications in therapy monitoring and management of patients treated with KDT based on our preliminary review and the authors extensive clinical experience. Additional benefits might be gained within the setup of clinical registries and the support of clinical studies. Telemedicine in the KDT context might not be suitable for all patients and reveals challenges such as technical difficulties (e.g. poor internet connection, inability to connect to virtual platform), necessity of training for patients, caregivers and HCP,

remuneration, data safety, cyber and legal security. At the same time opportunities like reduced travel time, easier planning of follow-up visits, multidisciplinary meetings and educational sessions as well as easier access to KDT will result from digital health innovation within this field. [9,10]

5. Conclusion

Feasibility of telemedicine in KDT was shown and further research might firstly need to assess existing requirements from patients, caregivers, and HCP dealing with KDT. In combination with a comprehensive workup of all existing literature and developments in the medical device sector, this might lead to the creation of potential multi-channel tools to facilitate KDT implementation, increased adherence and increased efficacy. From a clinical point of view it appears of utter importance, that patients planned for KDT receive upfront diagnostic workup that requires on-site visitation to a certain extent. A single messaging tool might not match the quality of personal contact between HCP and patients, respectively caregivers, but advanced telemedicine strategies will certainly have an important role in the future of KDT. However, the impact of telemedicine on feasibility, adherence and effectiveness has to be studied in future clinical trials in comparison to standard treatment.

Table 2. Requirements for telemedicine applications in the context of Ketogenic Dietary Treatment

Directly dietary therapy associated issues and requirements	
Issue	**Requirements**
Blood glucose monitoring	Documentation, visualisation, sensor integration
Blood ketone monitoring	
Breath ketone monitoring	
Dietary prescription	Wizard for calculation of energy and nutrient requirements Interface to database for foods for special medical purposes
Dietary tracking	Food protocol
Diet facilitation	Interface to food databases Wizard for recipe calculation
Growth and weight monitoring	Documentation, visualisation
Directly disease associated issues and requirements	
Seizure frequency	Documentation, visualisation
Medication prescription and monitoring	Documentation, visualisation Interface to medication database
Drug level monitoring	Documentation, visualisation
Quality of life	Integration of standard questionnaires
Cognition	Integration of standard questionnaires
General issues and requirements	
Patient registration	Basic data registration Registration of contact person
Patient management	Appointment- and task management
Communication	Unidirectional communication (HCP→Patient/Caregiver) (e.g. push messages, reminder) Bidirectional communication tools (HCP↔Patient/Caregiver) (e.g. messaging, video call) Interprofessional communication (HCP ↔ HCP)
Reporting	Generation of ad hoc and standard reports
Data safety and protection	Data storage at non-profit health care provider

References

[1] S. Scholl-Bürgi et al., Ketogenic diets in patients with inherited metabolic disorders. J Inherit Metab Dis. 38(4) (2015), 765-73.
[2] E.H. Kossoff et al, Optimal clinical management of children receiving dietary therapies for epilepsy: Updated recommendations of the International Ketogenic Diet Study Group. Epilepsia Open. 21;3(2) (2018), 175-192.
[3] J. Klepper et al., Glut1 Deficiency Syndrome (Glut1DS): State of the art in 2020 and recommendations of the international Glut1DS study group. Epilepsia Open 13;5(3) (2020), 354-365.
[4] M.C. Cervenka et al., International Recommendations for the Management of Adults Treated With Ketogenic Diet Therapies. Neurol Clin Pract 11(5) (2021), 385-397.
[5] F. Ye F et al, Efficacy of and patient compliance with a ketogenic diet in adults with intractable epilepsy: a meta-analysis. J Clin Neurol 11 (2015), 26–31.
[6] P. Klein et al, Dietary treatment in adults with refractory epilepsy: a review. Neurology 83 (2014), 1978–1985.
[7] E.H. Kossoff et al., Ketogenic diet therapy provision in the COVID-19 pandemic: Dual-center experience and recommendations. Epilepsy Behav 111 (2020), 107181.
[8] P. Samia et al., Telemedicine for Individuals with epilepsy: Recommendations from International League Against Epilepsy Telemedicine Task Force. Seizure 10 (106) (2023) 85-91.
[9] M. Armeno et al., Telemedicine- versus outpatient-based initiation and management of ketogenic diet therapy in children with drug-resistant epilepsy during the COVID-19 pandemic. Seizure 98 (2022) 37-43.
[10] M. Semprino., Telemedicine, drug-resistant epilepsy, and ketogenic dietary therapies: A patient survey of a pediatric remote-care program during the COVID-19 pandemic. Epilepsy Behav (2020) 112:107493.
[11] M.C. Cervenka et al., E-mail management of the modified Atkins Diet for adults with epilepsy is feasible and effective. Epilepsia 53(4) (2012) 728-32.
[12] A.M. Costa et al., Evaluation of E-Health Applications for Paediatric Patients with Refractory Epilepsy and Maintained on Ketogenic Diet. Nutrients 9;13(4) (2021), 1240.
[13] E.M. Zini et al., An mHealth Application for Educating and Monitoring Patients Treated with a Ketogenic Diet Regimen. Stud Health Technol Inform (247) (2018), 481-485.
[14] C. Ferraris et al., Use of Remote Monitoring by E-mail for Long-Term Management of the Classic Ketogenic Diet. Nutrients 19;12(6) (2020), 1833.
[15] M.C. Lima et al., Challenges in telemedicine for adult patients with drug-resistant epilepsy undergoing ketogenic diet treatment during the COVID-19 pandemic in the public healthcare system in Brazil. Epilepsy Behav 113 (2020) 107529.
[16] V.B. Bara et al., Patient and carer perspectives on the use of video consultations in the management of the ketogenic diet for epilepsy. Epilepsy Behav (145) (2023), 109280.

dHealth 2024
D. Hayn et al. (Eds.)

doi:10.3233/SHTI240032

Infection Control Through Clinical Pipelines Built with Arden Syntax MLM Building Blocks

Leonhard HAUPTFELD[a], Andrea RAPPELSBERGER[b] and Klaus-Peter ADLASSNIG[b,a,1]

[a] *Medexter Healthcare GmbH, Borschkegasse 7/5, 1090 Vienna, Austria*
[b] *Medical University of Vienna, Center for Medical Data Science, Institute of Artificial Intelligence, Spitalgasse 23, 1090 Vienna, Austria*

Abstract. Healthcare-associated infections (HAIs) may have grave consequences for patients. In the case of sepsis, the 30-day mortality rate is about 25%. HAIs cost EU member states an estimated 7 billion Euros annually. Clinical decision support tools may be useful for infection monitoring, early warning, and alerts. MONI, a tool for monitoring nosocomial infections, is used at University Hospital Vienna, but needs to be clinically and technically revised and updated. A new, completely configurable pipeline-based system for defining and processing HAI definitions was developed and validated. A network of data access points, clinical rules, and explanatory output is arranged as an inference network, a clinical pipeline as it is called, and processed in a stepwise manner. Arden-Syntax-based medical logic modules were used to implement the respective rules. The system was validated by creating a pipeline for the ECDC PN5 pneumonia rule. It was tested on a set of patient data from intensive care medicine. The results were compared with previously obtained MONI output as a suitable reference, yielding a sensitivity of 93.8% and a specificity of 99.8%. Clinical pipelines show promise as an open and configurable approach to graphically-based, human-readable, machine-executable HAI definitions.

Keywords. clinical decision support, intensive care, infection control, healthcare-associated infections, early detection of disease, fuzzy logic

1. Introduction

Healthcare-associated infections (HAIs) are those acquired by patients during their treatment at healthcare institutions, primarily hospitals. They may be caused indirectly through germs on a surface or in the air, or directly, as a consequence of a medical procedure such as the insertion of a catheter. The most common types of HAIs in the EU are pneumonia (PN), urinary tract infection (UTI), surgical site infection (SSI), bloodstream infection (BSI), and gastrointestinal infection.

The consequences for patients can be grave. BSI may lead to septic shock and sepsis, the 30-day fatality rate of which is around 25% in Western Europe [1]. HAIs cause an

[1] Corresponding Author: Klaus-Peter Adlassnig, Medical University of Vienna, Center for Medical Data Science, Institute of Artificial Intelligence, 1090 Vienna, Austria, E-Mail: klaus-peter.adlassnig@meduniwien.ac.at

estimated loss of two million disability-adjusted life years in the EU/EEA, a rate of 501 per 100,000 inhabitants [2]. An average of 5.5% of patients acquired an infection during their hospital stay in the EU from 2016–2017 [3].

Clinical decision support (CDS) software can be employed to interpret the vast quantity of patient data points generated by a patient data management system (PDMS), and determine the current HAI risk for a particular patient. This provides early warning and alerting for suspected HAI cases. Despite proven efficacy in detecting HAI [4], key challenges such as interconnectivity with other medical systems [5] hinder widespread adoption of the systems. An infection monitoring tool named MONI was installed at University Hospital Vienna, has been in routine use, and is still active for research purposes. It is based on rules that define clinical HAI conditions using medical logic modules written in Arden Syntax, an HL7 International standard for medical knowledge representation and processing. It rates patient vitals from the PDMS for HAI risk. A special feature of MONI is that it uses fuzzy logic. In other words, the results were never dichotomous but a risk percentage, which is also known as the applicability of a detection rule. MONI achieved high performance when it came to detecting cases, demonstrating a sensitivity of 90.3% and a specificity of 100%. The results were achieved well over six times faster than by conventional means [6]. MONI has been relegated to very limited use due to loss of interconnectivity with its present data sources. What we now need is a new approach that considers both MONI's successes as well as new paradigms and technologies in the CDS and IT spaces.

2. Methods

2.1. Requirements

The purpose was to create a rule processing engine that could assess the risk of HAI and explain the assessment after taking over patient data and special medical rule definition files. The rule definitions needed to be as easy as possible to create and modify, while also retaining the maximum degree of versatility. Every rule had to define a risk rating mechanism for one HAI condition and an explanation as to how the process works.

The rule definitions are implemented in the form of logic networks known as clinical pipelines. Every pipeline is a network of components with connected inputs and outputs. Components are assigned types that determine their function. A selection of the important types is listed below.

- **Trigger**: is the starting point for a pipeline, takes in external data, and makes it available to the network at its outputs.
- **MLM**: executes an MLM with input data and expresses the result data as output.
- **Logic**: takes fuzzy applicabilities at its inputs, applies configured logical operations to them, and expresses a single result as output.
- **Decision**: is the endpoint of a pipeline and takes a final applicability assessment at its input.

The resulting explanation is a tree structure with the applicability value from the decision at its root. Any component output can be inserted into the tree structure as desired. The rule processing engine, referred to as the runtime, was implemented in Java 21 using the

Spring Boot framework for robustness. A visual pipeline editor was created using JavaScript with the Next.JS framework.

2.2. Testing

To validate the feasibility of implementing an HAI rule as a clinical pipeline and determine its effectiveness in rating risk, a test rule needed to be implemented and run on patient data. The test rule needed to be relevant to state-of-the-art clinical operations, well defined enough to be implementable, a good fit for the test data, and comparable to a similar MONI rule for validation. The PN5 rule by the European Centre for Disease Prevention and Control (ECDC) for clinically defined pneumonia was used for validation. It originates in current ECDC guidelines, requires no microbiology data, and is implemented in MONI. To better fit the test data and improve comparability, a slightly altered version of PN5 from MONI was implemented. The altered version splits the rule into two types – for ventilated and non-ventilated patients (Figure 1).

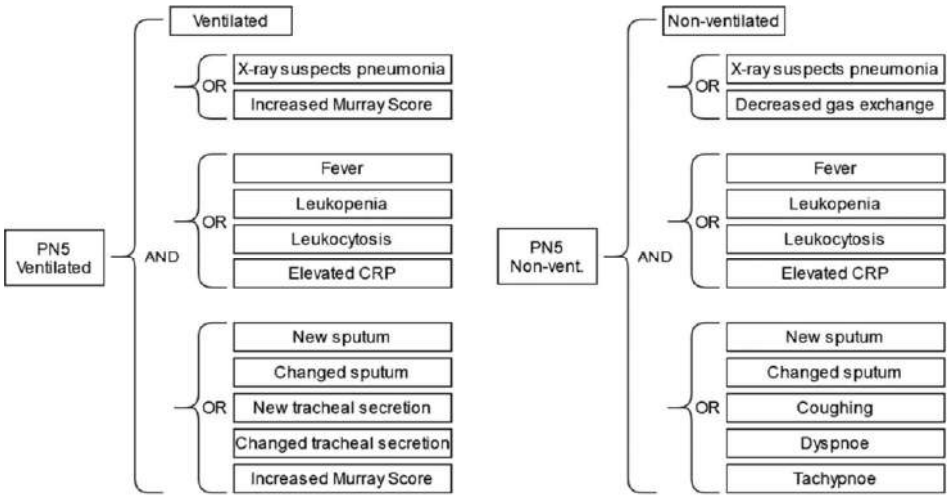


Figure 1. The PN5 pneumonia detection rules

2.3. Validation

Clinical validation was performed on the most recent backup of MONI's data and knowledge. First, patients for whom MONI generated a degree of compatibility with PN5 above 0% during their stay were selected. Then, the clinical pipeline PN5 rule was run over all days of the patients' stay. The applicabilities of *PN5 Ventilated* and *PN5 Non-ventilated* were then compared with the applicabilities of the equivalent MONI rules *HPN_KliBeatmet* and *HPN_KliNBeatmet*. For the comparison, MONI's applicabilities were regarded as the reference values.

As the implementations of the rules and the underlying software were not identical, the results were not expected to be an exact match. For clinical use, it is only relevant whether a rule has detected an HAI risk or not. Therefore, for a given day, as long as both the clinical pipeline and MONI results had a degree of compatibility greater than or equal to 20% (true positive (TP)) the results were considered a match. If both results were less than 20%, we recorded this case of being true negative (TN). All days of a

patient's stay were evaluated and a confusion matrix of TP, TN, false positives (FP) and false negatives (FN) was created from the results. Sensitivity and specificity were also calculated. This was done once for each rule pairing, yielding two sets of results.

3. Results

3.1. Testing

The PN5 pipeline was constructed by first placing its trigger component and working down towards the decision component. After triggering, the pipeline first retrieves the data it needs – temperature and leukocyte count, for instance. This data is then rated for its applicability in regard of related medical conditions. Leukocyte count can be rated for leukopenia and leukocytosis. These conditions are then linked by logic components based on how they relate to the PN5 condition. At the end of this decision network, the final decision regarding PN5 applicability is reached. This results in three distinct phases for the PN5 pipeline: data ingress, ratings, and logic. Visually, the pipeline follows a teardrop shape, where the retrieved data expands into many ratings and is finally combined back into a single decision component by the logic (Figure 2). An explanation based on all the rating values is also generated (Figure 3).

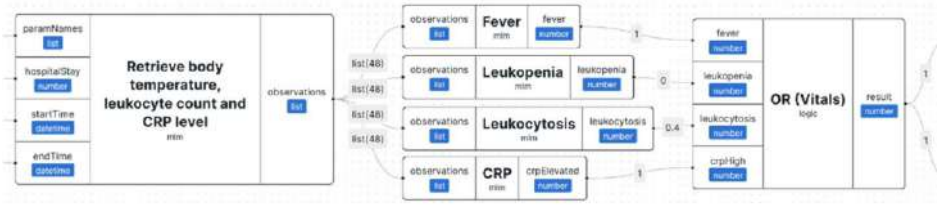


Figure 2. Excerpt from the PN5 pipeline definition. This diagram is a typical example of flow within such a pipeline. Vital parameters are retrieved on the very left, passed to rating MLMs in the middle, which then pass their ratings on to logic, in this case an OR operation. Fuzzy logic can also be seen in this example, with definite fever, an elevated CRP level (100% applicability), and a white blood cell count approaching leukocytosis (40%).

Figure 3. Excerpt from the PN5 pipeline explanation tree. It shows how the explanation is generated by building a tree from intermediate ratings of medical conditions. This explanation was generated from the pipeline shown in Figure 2.

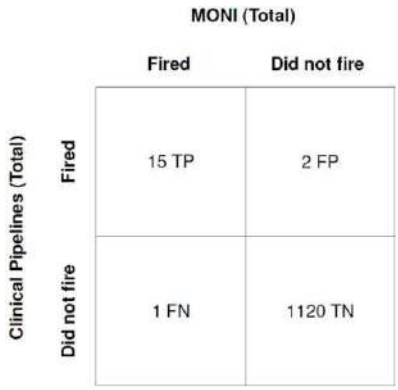


Figure 4. Validation results of the clinical pipelines PN5 rule

3.2. Validation

Twenty-two patients were identified for whom a PN5 applicability greater than 0% was calculated by MONI at some point during their stay. One of these patients was used while developing the PN5 rule to identify any errors in implementation and run testing evaluations against MONI's results for optimization. The patient used for optimizing the rule implementation was then removed from the validation set. The rule was run for each patient and every day of their stay. The 21 patients used for validation yielded a total of 603 days of patient stay for the evaluation.

The *PN5 Ventilated* rule generated 15 true positives, 2 false positives, 1 false negative, and 551 true negatives. This results in a sensitivity of 93.8% and a specificity of 99.6%.

The *PN5 Non-Ventilated* rule generated no true positives, false positives, or false negatives. It generated 569 true negatives, meaning it was never detected by MONI but also not by the clinical pipeline. The reason for this was the lack of any O^2 dispersion and saturation data. The specificity of this rule is 100%; the sensitivity cannot be calculated without true positives and false negatives.

The clinical pipeline rules generated a total of 15 true positives, 2 false positives, 1 false negative, and 1120 true negatives (Figure 4). A total sensitivity of 93.8% and a total specificity of 99.8% were achieved.

4. Discussion

4.1. Viability of Pipelines

Clinical pipelines with the implemented PN5 pipeline showed excellent detection performance when validated against MONI. This strongly supports the argument that this rule processing engine can be used as the basis for infection monitoring software. The rule illustrated in Figure 1 was implementable as a pipeline without much difficulty, and adjusting it after discovering inaccuracies in validation proved fairly easy as well. This shows that pipelines are not only accurate in detecting potential infection cases, but also easy to create and adjust.

4.2. Limitations and Possibilities

Clinical pipelines could not fully match MONI's detection performance. Furthermore, MONI itself is 90.3% sensitive [6]. Therefore, the detection rules need further refinement. The system is also currently untested with microbiological data and will require additional component types to process the complex taxonomy of bacteria and perform additional functions. Knowledge-based systems like this are also under threat by machine-learning-based detection. However, machine learning implementations usually suffer from low specificity [7] and a missing explanation system [8]. One possibility is to combine the pipeline system with machine learning models by creating corresponding component types.

4.3. Road to Adoption

Further clinical validation must be performed on the clinical pipeline system. More rules, such as the Robert-Koch-Institute's KISS rules [9] or the ECDC HAI-Net rules [10] need to be implemented and tested on a broader and more diverse set of patient data like the MIMIC-IV dataset maintained by MIT [11].

In summary, clinical pipelines are an easily configurable system for applying medical knowledge to patient data. Validations have shown its viability for developing a rule to detect pneumonia, a type of healthcare-associated infection, and its excellent results when applying that rule to patient data. Overall, pipelines could potentially improve the definition of clinical rules in software.

References

[1] M. Goto and M.N. Al-Hasan, Overall burden of bloodstream infection and nosocomial bloodstream infection in North America and Europe, *Clinical Microbiology and Infection* **19**(6) (2013), 501–509.

[2] A. Cassini, D. Plachouras, T. Eckmanns et al., Burden of Six Healthcare-Associated Infections on European Population Health: Estimating Incidence-Based Disability-Adjusted Life Years through a Population Prevalence-Based Modelling Study, *PLoS Medicine* **13**(10) (2016), e1002150.

[3] *Health at a Glance: Europe 2018: State of Health in the EU Cycle*. OECD Publishing, Paris / European Union, Brussels, 2018. Available at: https://doi.org/10.1787/health_glance_eur-2018-en.

[4] J.S. de Bruin, K.-P. Adlassnig, A. Blacky, H. Mandl, K. Fehre, and W. Koller, Effectiveness of an automated surveillance system for intensive care unit-acquired infections, *Journal of the American Medical Informatics Association* **20**(2) (2013), 369–372.

[5] R.T. Sutton, D. Pincock, D.C. Baumgart et al., An overview of clinical decision support systems: benefits, risks, and strategies for success, *NPJ Digital Medicine* **3** (2020), 17.

[6] A. Blacky, H. Mandl, K.-P. Adlassnig, and W. Koller, Fully Automated Surveillance of Healthcare-Associated Infections with MONI-ICU, *Applied Clinical Informatics* **2**(3) (2011), 365–372.

[7] *Definitionen nosokomialer Infektionen für die Surveillance im Krankenhaus-Infektions-Surveillance-System (KISS-Definitionen)*. Nationales Referenzzentrum für Surveillance von nosokomialen Infektionen, Robert Koch-Institut, Berlin, 2017. Available at: https://www.rki.de/DE/Content/Infekt/Krankenhaushygiene/Nosokomiale_Infektionen/Downloads/RKI_Definitionen_nosokomialer_Infektionen_E-Book.pdf.

[8] *Surveillance of healthcare-associated infections and prevention indicators in European intensive care units*. European Centre for Disease Prevention and Control, Stockholm, 2017. Available at: https://www.ecdc.europa.eu/sites/default/files/documents/HAI-Net-ICU-protocol-v2.2_0.pdf.

[9] A. Vellido, V. Ribas, C. Morales et al., Machine learning in critical care: state-of-the-art and a sepsis case study, *BioMedical Engineering OnLine* **17**(Suppl 1) (2018), 135.

[10] M. Moor, B. Rieck, M. Horn, C.R. Jutzeler, and K. Borgwardt, Early Prediction of Sepsis in the ICU Using Machine Learning: A Systematic Review, *Frontiers in Medicine* **8** (2021), 607952.

[11] A. Johnson, L. Bulgarelli, T. Pollard, S. Horng, L.A. Celi, and R. Mark, MIMIC-IV (version 2.2), *PhysioNet*. Available at: https://doi.org/10.13026/6mm1-ek67.

dHealth 2024
D. Hayn et al. (Eds.)

doi:10.3233/SHTI240033

Understanding IT in Healthcare: Relevance and Training Needs of IT in Private Medical Practice

Marco SCHWEITZER[a], Werner O. HACKL[a], Sabrina B. NEURURER[a,b] and Bernhard PFEIFER[a]
[a] *Division for Digital Health and Telemedicine, UMIT TIROL - Private University for Health Sciences and Health Technology, Hall in Tirol, Austria*
[b] *Department of Clinical Epidemiology, Tirol Kliniken GmbH, Innsbruck, Austria*

Abstract. Background: The integration of Information Technology (IT) into private medical practice is crucial in modern healthcare. Physicians managing office-related IT without proper knowledge risk operational inefficiencies and security. Objectives: This study determines the relevance of specific IT topics in medical practice and identifies the training needs of physicians for enhancing IT competencies in healthcare. Methods: In March 2023 a cross-sectional online survey was conducted with physicians comprising nine IT-related topics in Tyrol, Austria. Results: The survey results highlighted a strong perceived relevance and high demand for IT education among physicians working in their medical practice, especially in areas of core medical IT and security. The majority of responses indicated high relevance (76.7%) and high demand (69.7%) for IT topics in medical practice. Conclusion: The findings underscore a significant need for targeted IT training and support in medical practices, particularly in areas related to the medical practice and security. Addressing these needs could lead to improved healthcare delivery and better management of technological resources in the healthcare sector.

Keywords. Information Technology, Digital Health, Education

1. Introduction

The rapid digitalization of the healthcare sector has led to far-reaching changes in recent decades. Whether it is the electronic patient record or innovative telemedical procedures: The integration of information technology into the medical field promises to optimize patient care and make clinical processes more efficient [1,2]. Numerous studies emphasize the possibility of using these advances, among others, to improve the quality of care and extend access to healthcare services even in remote regions [3]. However, despite these recognizable benefits, a key challenge remains in the adequate knowledge and digital skills of medical staff.

Implementing new technologies in healthcare can be challenging [4], particularly for physicians in medical practice who often lack Information Technology (IT) knowledge and proper support. This can lead to errors, inefficient use of technology, and user frustration. Further, physicians working in private practice are fully responsible for managing their office's IT systems, making them directly accountable for any security issues. This necessity for IT proficiency is vital not only for operational efficiency but

also for patient care, telemedicine services, practice management, medical research, and ensuring cybersecurity. Consequently, there is a pressing need for effective IT training and support in the healthcare sector, especially for those in medical practice.

Building on these challenges, our study aims to precisely define and understand the scope of these issues. Specifically, the objective is to determine the relevance of certain IT topics in medical practice and identify the corresponding training needs for private medical practitioners. This focus will help in developing targeted strategies to enhance IT competency in healthcare settings, ensuring that physicians are well-equipped to manage their technological requirements efficiently and securely.

2. Methods

This study employed a cross-sectional survey approach to assess the significance of IT and the associated training requirements for physicians in medical practice. The online questionnaire assessed nine healthcare-related IT topics:

1. IT Infrastructure for Core Medical Activities: Systems for medical treatment and documentation.
2. Non-medical Specific IT Infrastructure: General practice organization systems including office tools and network management.
3. Digital Processes and Procedures: Automation of medical practice processes and efficiency tools.
4. Healthcare Interoperability (ELGA): Health information exchange and connection with other healthcare providers.
5. Data Protection: Compliance with GDPR and reliable handling of patient data.
6. IT Security: Measures against cyber threats and maintaining technical security.
7. Telemedicine: Digital physician-patient interaction and remote healthcare services.
8. Future Trends: Emerging technologies like artificial intelligence (AI), cloud services, and health apps.
9. Sustainability in IT: Energy-efficient and resource-conserving IT practices.

Participants rated the relevance of each area for their work and the demand of further training on a 5-point-likert scale of 1 (very relevant/high demand) to 5 (not relevant/no need). The questions concerning the training for each topic were only available when the answer of the related relevance for this topic was rated with 1-3. In addition, the survey collected general office information, such as the type of practice (affiliations to one or multiple statutory health insurances), the specialty (general practitioner or specialist or both), the number of years in the practice and the size of the team. The survey also explored preferences for online training features, including free training, continuing education points, and customizable options.

Initially, the survey was tested with a small group of medical professionals (n=20) during a professional meeting. Since this preliminary trial did not result in any further needed modifications, and all questions were deemed understandable and relevant, the results of this initial survey were incorporated into the final analysis. This integration ensured that the survey was refined and validated by a representative sample before wider distribution.

The survey was implemented as an online survey using LimeSurvey [5] and distributed to all registered physicians (general physicians and specialists) in Tyrol, Austria (approximately 1,400) via email through the Tyrolean Medical Association. The

survey remained open for three weeks in March 2023. The raw data was exported from LimeSurvey in CSV format and further processed in Jupyter Notebook [6], facilitating a comprehensive analysis that encompassed data cleaning, transformation, and visualization to derive essential qualitative and quantitative insights..

3. Results

Out of 235 initiated surveys, 199 were completed and submitted. To ensure a robust analysis, only these fully completed surveys were considered for further analysis, filtering out any incomplete responses.

The survey results indicated a significant high perceived relevance of IT across all examined domains in the medical practice setting with the highest scores in topics for general medical IT comprising the infrastructure for core medical activities and security. Notably, most of the responses (76.7%) rated the relevance for all topics within a range of 1-2 (high relevance) as depicted in Figure 1 and Figure 2, demonstrating a strong acknowledgment of IT's importance in medical practice.

Training needs mirrored this trend, with 69.7% of responses falling within the 1-2 range (high demand), underscoring a substantial interest in further IT education among practitioners. However, a divergence was observed in the relative significance attributed to certain topics. Specifically, themes for Telemedicine, Future Trends and Sustainability in IT were consistently deemed less relevant compared to other areas. This was reflected both in their lower relevance scores and in the diminished priority assigned to them in the context of further training. Although the topic of Future Trends shows lower relevance, the corresponding need for training is considerably higher, suggesting a disparity between current perceptions of importance and the actual demand for skill development in this area.

Of particular note were the areas of core medical IT and security, which emerged as the most relevant topics. IT Security, in detail, not only garnered high relevance ratings but also displayed a pronounced demand for training. This finding suggests a heightened awareness and concern among physicians regarding the security aspects of IT in their practices. The substantial interest in training for IT security likely reflects the escalating challenges and risks posed by cyber threats in the healthcare sector even for medical practices.

An interesting pattern emerged across different experience groups of physicians, where experience is measured along the years of working as a physician. The most significant disparities related to experience were in data protection relevance (less than 1 year of experience: 1.2; over 20 years of experience: 2.0) and in IT training needs in the area of processes (less than 1 year of experience: 1.1; over 20 years of experience: 2.0), though these differences were not substantial. Similarly, the comparison between general practitioners and specialists showed a consistent pattern with little variation between the groups. The study participants reported their associations with Austria's statutory health insurers operating in Tyrol, revealing a balanced representation among the three national insurers - BVAEB (62.8%), ÖGK (61.8%), SVS (63.8%) [7] - and the Tyrol-specific public health insurance, KUF (70.4%).

The preferences for training options were quite clear: Nearly 99% (n=198) of participants expressed their desire for free training courses and the same amount preferred online training courses. Additionally, 61% (n=121) indicated that earning continuing education points as proposed by the National Medical Association for such

training would be important to them. 15.1% of respondents (n=30) identified other significant criteria as free text for training, with a notably high frequency of references to "practice relevance".

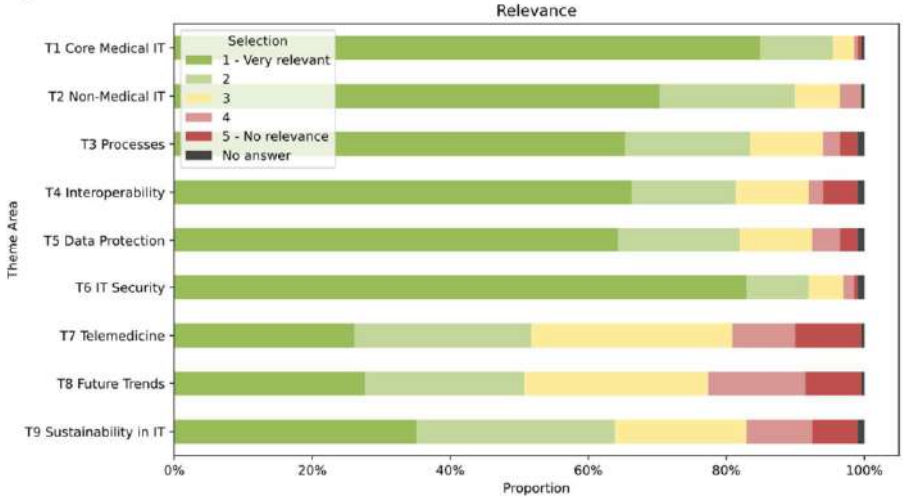


Figure 1. Stacked bar chart illustrating the perceived relevance of nine healthcare IT topics in a physician's practice. Responses were scored on a scale from 1 (indicating high relevance) to 5 (indicating no relevance), with an option for no response.

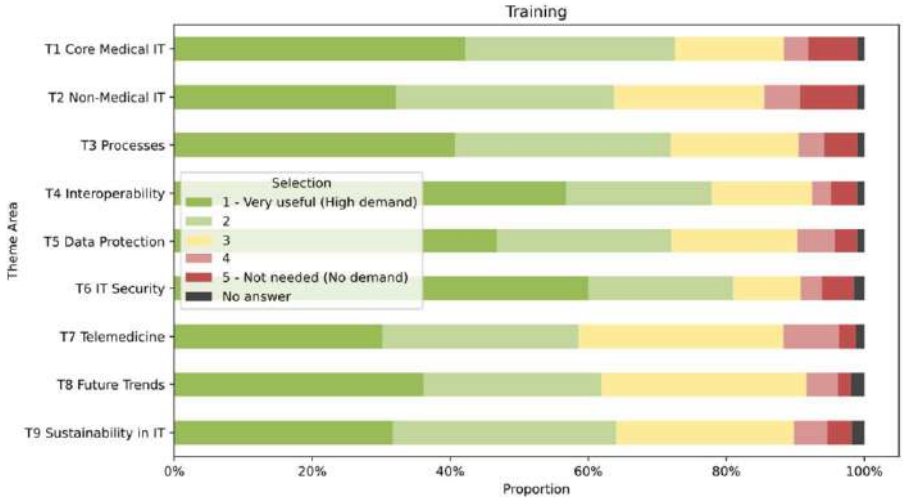


Figure 2. Stacked bar diagram depicting the training needs of physicians across various healthcare IT topics. Each topic's demand is rated on a scale from 1 (high demand for training) to 5 (no training need), visually represented through stacked bars.

4. Discussion

The analysis of our online survey yielded insightful findings regarding the IT competence and training needs of medical practices.

4.1. Discussion of Methods

The use of an online survey realized through LimeSurvey, provided a cost-effective and quick way to reach a large number of physicians. This type of data collection has proven popular in recent literature as a means of capturing opinions and perspectives from practitioners [8]. Despite being disseminated through the medical association, a typically credible and influential entity, our study experienced a lower response rate than expected. The reduced rate of participation in our study, despite the involvement of a respected professional body, could introduce potential biases. This outcome suggests that factors other than the source's familiarity might influence the willingness of respondents to participate in online surveys.

Analyzing the data with Jupyter Notebook allowed for a reproducible data analysis. Jupyter notebooks are well known in the research community for their ability to integrate both code and annotative text [6]. This not only provides a transparent representation of the data analysis but also a way for other researchers to reproduce or even extend the analysis. The topics chosen for investigation were meticulously selected to encompass a broad spectrum of information technology aspects relevant to medical practice, guided by established benchmarks for initiating a new practice [9] and the overarching field of medical informatics. Nevertheless, the number or specificity of topics might be a limitation. Future studies could integrate additional or different topic areas based on the evolving needs and trends in medical informatics [10,11].

In conclusion, the method of soliciting feedback directly from physicians regarding their preferences for training offerings has provided valuable insight also into how the training should best be offered and organized. The vast preference for free online training suggests that accessibility and costs are important factors that need to be considered.

4.2. Discussion of Results

It is notable that the vast majority of respondents consider topics such as the general relevance of medical IT and security to be particularly important. This is consistent with findings from previous studies, which also highlighted the increasing importance of digitalization in the healthcare industry [12,13].

The fact that the topic for future trends appeared less relevant but showed a high need for educational training is interesting and could indicate a discrepancy between the current level of knowledge and the desire to be prepared for future developments. This finding is also reflected in other studies, which emphasized that medical staff recognize the importance of future technology trends, but often do not feel adequately prepared for them also in how to integrate them in medical practice [4].

The fact that the relevance and need for training decreased with years of service is also a revealing finding. It could indicate that experienced doctors either feel more confident in their current knowledge or are less open to technological innovations. This is consistent with studies suggesting that acceptance and adaptation to new technologies can often vary with age and experience [14]. It is also worth noting that there were hardly any differences in the ratings between general practitioners and specialists.

Finally, the high preference for free online training and desire for continuing education credits is not surprising, given current trends in medical education and the limited time for physicians. Previous studies have already pointed to the growing popularity of online courses [11].

Overall, these findings underscore the need to continuously assess and adapt medical IT training needs to ensure that medical staff are best prepared to meet the challenges of digital transformation even in medical practices. Future studies should investigate appropriate learning strategies, including e-learning, to effectively meet the diverse IT training needs of physicians in the digital health landscape.

Acknowledgments

This work was supported by the Tyrolean Medical Association (Ärztekammer für Tirol).

References

[1] D. Setyadi, and M. Nadjib, The Effect of Electronic Medical Records on Service Quality and Patient Satisfaction: A Literature Review, *Journal Research of Social Science, Economics, and Management*. **2** (2023) 2780–2791. doi:10.59141/jrssem.v2i12.500.

[2] OECD, and E. Union, Adoption of Electronic Medical Records, OECD, Paris, 2022. doi:10.1787/76c7b6a2-en.

[3] A. Haleem, M. Javaid, R.P. Singh, and R. Suman, Telemedicine for healthcare: Capabilities, features, barriers, and applications, *Sens Int*. **2** (2021) 100117. doi:10.1016/j.sintl.2021.100117.

[4] J. Konttila, H. Siira, H. Kyngäs, M. Lahtinen, S. Elo, M. Kääriäinen, P. Kaakinen, A. Oikarinen, M. Yamakawa, S. Fukui, M. Utsumi, Y. Higami, A. Higuchi, and K. Mikkonen, Healthcare professionals' competence in digitalisation: A systematic review, *J Clin Nurs*. **28** (2019) 745–761. doi:10.1111/jocn.14710.

[5] LimeSurvey GmbH, LimeSurvey — Free Online Survey Tool, (n.d.). https://www.limesurvey.org/de (accessed January 25, 2024).

[6] T. Kluyver, B. Ragan-Kelley, F. Pérez, B. Granger, M. Bussonnier, J. Frederic, K. Kelley, J. Hamrick, J. Grout, S. Corlay, P. Ivanov, D. Avila, S. Abdalla, C. Willing, and Jupyter Development Team, Jupyter Notebooks—a publishing format for reproducible computational workflows, 2016. doi:10.3233/978-1-61499-649-1-87.

[7] Federal Ministry of Social Affairs, Health, Care and Consumer Protection, Structural Reform of the Social Insurance System, (n.d.). https://www.sozialministerium.at/en/Topics/Social-Affairs/Social-Insurance/Structural-Reform-of-the-Social-Insurance-System.html (accessed January 29, 2024).

[8] C. Audibert, D. Glass, and T.P. Johnson, Method and transparency of online physician surveys: an overview, *Survey Methods: Insights from the Field (SMIF)*. (2020). doi:10.13094/SMIF-2020-00001.

[9] Handbuch zur Gründung einer PVE, (2022). https://primaerversorgung.gv.at/sites/default/files/2022-09/20220912_Gr%C3%BCndungshandbuch_Version%20f%C3%BCr%20PPV_0.pdf (accessed January 19, 2024).

[10] S. Tomar, M. Gupta, M. Rani, and H.S. Shyam, Healthcare Digitalisation: Understanding Emerging Technological Trends, in: 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS), 2023: pp. 2459–2463. doi:10.1109/ICACCS57279.2023.10113106.

[11] I.A. Scott, T. Shaw, C. Slade, T.T. Wan, C. Coorey, S.L.J. Johnson, and C.M. Sullivan, Digital health competencies for the next generation of physicians, *Intern Med J*. **53** (2023) 1042–1049. doi:10.1111/imj.16122.

[12] J.P. DeMello, and S.P. Deshpande, Factors Impacting Use of Information Technology by Physicians in Private Practice, *IJHISI*. **7** (2012) 17–28. doi:10.4018/jhisi.2012040102.

[13] F. Vogt, F. Seidl, G. Santarpino, M. van Griensven, M. Emmert, G. Edenharter, and D. Pförringer, Healthcare IT Utilization and Penetration among Physicians: Novel IT Solutions in Healthcare - Use and Acceptance in Hospitals, *Eur Surg Res*. **59** (2018) 100–113. doi:10.1159/000490241.

[14] T. Nimjee, E. Miller, and S. Solomon, Exploring Generational Differences in Physicians' Perspectives on the Proliferation of Technology within the Medical Field: A Narrative Study, *Healthc Q*. **23** (2020) 53–59. doi:10.12927/hcq.2020.26172.

dHealth 2024
D. Hayn et al. (Eds.)

doi:10.3233/SHTI240034

Screening Automation in Systematic Reviews: Analysis of Tools and Their Machine Learning Capabilities

Elias SANDNER[a,b,1], Christian GÜTL[b] and Igor JAKOVLJEVIC[a,b] and Andreas WAGNER[a]

[a] *IT-Department, CERN, Geneva, Switzerland*
[b] *CoDiS-Lab ISDS, Graz University of Technology, Graz, Austria*

Abstract. Systematic reviews provide robust evidence but require significant human labor, a challenge that can be mitigated with digital tools. This paper focuses on machine learning (ML) support for the title and abstract screening phase, the most time-intensive aspect of the systematic review process. The existing literature was systematically reviewed and five promising tools were analyzed, focusing on their ability to reduce human workload and their application of ML. This paper details the current state of automation capabilities and highlights significant research findings that point towards further improvements in the field. Directions for future research in this evolving field are outlined, with an emphasis on the need for a cautious application of existing systems.

Keywords. Systematic Review Automation, Title and Abstract Screening, Research Tool, Machine Learning, Text Classification

1. Introduction

A systematic review (SR) aims to evaluate and interpret all existing studies related to a particular field of interest. Annual Review[2] releases SRs in more than 30 disciplines, including computer science and medicine. In medical contexts, SRs address research questions about the frequency of disease occurrence, their expected progression, the dangers involved in diagnosing them, and the strategies for their management, to name a few aspects [1]. SRs have their level of evidence confined to that of the studies they encompass. Nonetheless, by aggregating more data than individual studies, SRs enhance the precision of the overall findings. Consequently, they offer the most reliable evidence to address research questions [2]. However, this evidence comes at the cost of an enormous workload, as it typically takes several months to complete SRs [3,4]. Within the SR process, the screening phase is described as the most difficult and time-consuming aspect of the process and is the most urgent task that requires a reliable support system [5]. Therefore, this survey paper aims to answer the following research question: **How advanced**

[1] Corresponding Author: Elias Sandner, IT-Department, CERN, Geneva, Switzerland and CoDiS-Lab ISDS, Graz University of Technology, Graz, Austria, E-mail: elias.sandner@cern.ch
[2] https://www.annualreviews.org/

is the current level of automation in the title and abstract screening phase, which methods show promise for further advancement, and what key research findings should inform future improvements in this area?

Section 2 emphasizes the need for enhanced automation through an overview of the current SR process. Section 3 describes how machine learning is integrated into existing tools, and Section 4 describes recent research findings. The paper concludes with Section 5 which suggests promising paths for future work.

2. Background and Related Work

Conducting an SR involves multiple stages and support tools to streamline several steps are already in use. For example, in [6] it is demonstrated how such systems helped to conduct an SR within two weeks and the main tools used are listed in Table 1. This paper focuses on the screening phase within the SR process. Subsequently, this task is outlined in detail, followed by an in-depth analysis of the required time effort.

To determine studies that comply with the specified inclusion and exclusion criteria, retrieved studies undergo a thorough screening process. Initially, the title and abstract of each paper are assessed to eliminate the bulk of the nonrelevant papers and then a detailed examination of the full text is executed for comprehensive evaluation. To reduce bias and errors, it is recommended that the screening is conducted independently by a minimum of two reviewers. Discrepancies should be addressed through collaborative discourse or by incorporating the judgment of an additional expert. This approach ensures the comprehensive identification of relevant studies, although it is associated with a high workload. In [6] RobotSearch was utilized during the title and abstract screening phase (TiAb screening). It is capable to filter out documents that are definitely not randomized controlled trials (RCTs), the specific study type to which this SR was limited. However, it is important to consider that the exclusive focus on RCTs notably simplified the screening process. This is a significant point, as such a limitation might not apply to other SRs, where a broader range of study types could lead to a more complex and time-consuming screening effort. Additionally, the SRA Helper was employed in both the TiAb screening and full text screening phases. It offers a user-friendly interface that allows documents to be included or excluded using hotkeys.

SRs yield significant evidence, although the completion of all steps in the outlined process typically spans several months. To highlight the effort required to perform SRs, Table 2 sums up the results of three time analyzes. Given the substantial time needed to conduct SRs, a survival analysis of 100 SRs indicates that 7% of these reviews already showed signs of quantitative or qualitative obsolescence at the time of publication [8]. [5] highlighted that the screening phase is the most time-intensive aspect of the entire

Table 1. Overview of support tools employed in major tasks of the SR process as reported by Clark et al. (2020)[6].

Task	Support system
Project proposal design	template
Systematic search	SRA word frequency analyzer, SRA polgot search translator
Eligibility Screening	RobotSearch, SRA Helper
Data extraction	digital spreadsheet

Table 2. Time demands of SRs based on three analyses. Data sources and metrics are derived from the studies conducted by Beller et al. (2013)[9], Demetres et al. (2023)[3], and Borah et al. (2017)[4]; All values represented in Days.

	Beller et al. (2013)	**Demetres et al. (2023)**	**Borah et al. (2017)**
Data origin	Medline	Weill Cornell Medicine	PROSPERO
Data quantity	300	101	195
Time of SR conduction	2009-2011	2011-2021	before July 2014
Time measured from/to	last search / publication	Requesting librarian support / submission	Registration / publication
Min	0	42	42
Max	1314	930	1302
Median	153	N/A	461
Average	N/A	295	473

process. Additionally, this phase was recognized as the most challenging and the one that most urgently necessitates a dependable support system. As indicated in [4] most of the literature is removed during the TiAb screening with a median reduction in citations of 95% at this stage and 3.7% during full text screening. Furthermore, [3] outlined that abandoned SRs, most likely occur in the TiAb screening phase.

Data filtration carried out by highly compensated experts, a process uncommon in most fields, is currently a consistent component of the SR process. Therefore, tools designed to increase expert efficiency are commonly used to reduce human workload.

3. Current Level of Screening Automation

Despite the current inability to fully automate the screening process, numerous software tools significantly aid human experts. [10] analyzed 16 tools based on 21 features. The five tools that rank highest (Table 3) offer a stable and supported release, comprehensive documentation, active customer support and features for multiple users, importing and allocating references, and the inclusion/exclusion of references with labeled reasons for exclusion, and resolution of discrepancies. While the top four tools support the distinction between TiAb and full text screening, this is the only mandatory feature not fulfilled by Rayyan[3]. However, users can simply export relevant citations from the TiAb screening and import them into a new review. Therefore, we do not interpret this as a major drawback. Furthermore, the study (published in 2019) indicates that Covidence[4] does not offer any ML automation feature, but the current version does. As a result, 5 out of the 16 analyzed tools now provide some form of ML support, as summarized in Table 3. The typical workflow for using one of these tools during the TiAb screening phase is illustrated in Figure 1 and is subsequently described.

Search results, retrieved from various databases, are imported into the selected review management tool, typically in RIS, RevMan, or PubMed format. Initially, integrated solutions for deduplication are utilized. Subsequently, the user manually screens citations using the interface's hotkeys for including and excluding reasons. Once a number of decisions are made, the ML system uses labeled documents to estimate inclusion probabilities of unscreened citations and reorders the citation queue accordingly, prioritizing those with the highest likelihood of inclusion. With more labeled data, the rank-

[3]https://www.rayyan.ai/
[4]https://www.covidence.org/

Table 3. Feature analysis and AI support of selected review management tools. (*2 according to [10] but machine learning features are available in the current version.)

	Feature Analysis according to [10]			Machine learning support			
Tool	Mandatory features (out of 9)	Desirable features (out of 9)	Optional features (out of 3)	Deduplication support system	Relevance ranking	RCT classifier	Custom classifier
DistillerSR	9	8	3	yes	yes	No	Yes
EPPI-Reviewer	9	7	3	yes	yes	Yes	Yes
SWIFT Active Screener	9	7	2	yes	yes	No	No
Covidence	9	5	3*	yes	yes	Yes	No
Rayyan	8	6	2	yes	yes	No	Yes

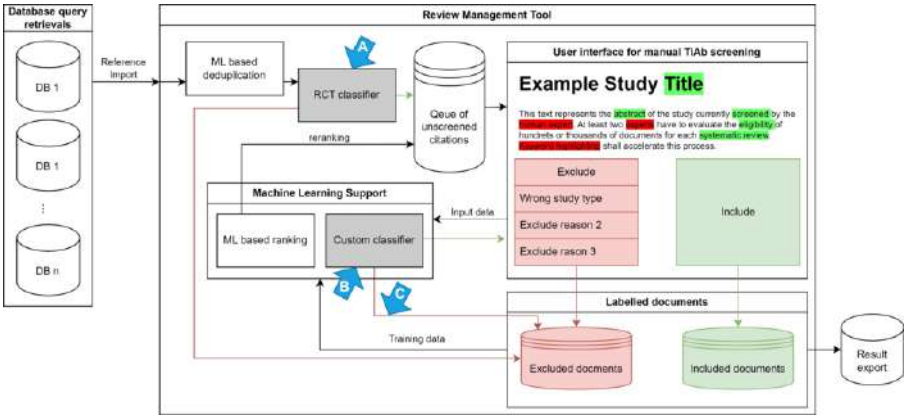


Figure 1. Title and abstract screening with support tool; A: only applicable to Covidence and EPPI-Reviewer[11,7]; B: only applicable to DistillerSR, EPPI-Reviewer and Rayyan; C: Only applicable to DistillerSR and EPPI-Reviewer as Rayyan provides suggestions without making autonomous decisions. [13,15,14]

ing's accuracy improves. For example, EPPI-Reviewer[5] updates its ranking every 25 citations. Swift-Active-Screener[6] also estimates the number of relevant citations left in the unscreened document list.

However, Swift Active Screener also stands out as the only one that does not use ML classification. As mentioned in Section 2, some SRs are confined to RCTs. The RCT classifier component (Figure 1 A) autonomously excludes citations with different study designs, considerably lessening the need for manual screening. The Cochrane RCT classifier[12] achieves a recall rate of 0.99. Consequently, Covidence has integrated this original classifier. EPPI-Reviewer also provides a solution with a recall rate of 0.99, based on data manually labeled by the Cochrane crowd, as explained in [7].

DistillerSR[7], Eppi-Reviewer and Rayyan furthermore provide classifiers that can be trained based on the initial manual decisions as illustrated with the custom classifier component (Figure 1 B). DistillerSR offers integrated classification software based on a statistical approach. Based on the application in three SRs, it is claimed to reduce human workload by 57.4%. The false negative rate is 1. 17%, and the recall is not disclosed.

[5]https://eppi.ioe.ac.uk/
[6]https://www.sciome.com/swift-activescreener/
[7]https://www.distillersr.com/

Furthermore, only a minimum of 10% of the citations were manually reviewed, leaving the number of relevant papers missing unclear. [13] Rayyan employs a random forest ensemble model, evaluated on 15 pre-labeled SRs. The best performance showed a 0.986 recall and reduced workload by 46.9%, but when applied to a different review, the recall dropped to 0.75 and workload reduction to 3%. In both instances, 50% of pre-labeled data was used for training and the rest for testing. Once enough training data is gathered, Rayyan activates its prediction model, offering suggestions for undecided studies. However, to maintain review quality, the final decision always rests with a human expert. [14] EPPI-Reviewer offers the capability to manually create bespoke classifiers using previously screened citations. This tool utilizes the scikit-learn Python library[8] and is adept at binary classification of new records. Additionally, it provides statistical insights before applying these created classifiers to new sets of citations.

4. Current Approaches Towards Further Automation

Building on the analysis of existing tools, this section highlights further relevant research findings. Custom classifiers, integrated in existing tools are especially suitable for living SRs as they require labeled data. For their effective use in living SRs, it is essential to maintain consistent scope, unchanged field terminology, and an original SR large enough to supply adequate training data. [15] explored this use case employing EPPI-Reviewer's classifier function, specifically using the stochastic gradient descent (SGD) classifier[9] with logistic regression in both instances. They first evaluated ML classifiers' performance in recall and screening reduction. Classifiers assigned relevance scores from 0 to 99 for each citation, using a threshold of 10 to filter out low-relevance citations. Recalls ranged from 92% to 100%, and screening reduction, measured by papers left for manual review, ranged from 40% to 74%. The study demonstrated improved classifier performance when supplemented with specific exclusion and inclusion criteria, rather than solely binary labels. They also highlighted that text preparation might impact classification performance more significantly than the choice of algorithm. Based on those findings, they continued to apply a custom classifier for the update search. EPPI-Reviewer's classification tool provides insights into the relevance score distribution of citation records, aiding in estimating the classifier's reliability for a particular citation set. This information assists in determining the classifier's integration into the screening workflow. Citations with a relevance score greater than 20 were manually screened. Citations scoring between 13 and 20 were batch screened in sets of 500, and if two consecutive batches lacked relevance, all remaining citations were deemed irrelevant. Citations ranking below 13 were automatically discarded. Implementing these rules resulted in a 61% decrease in screening efforts. 98% of the relevant references were identified in the top 21% of the citations, with relevance scores ranging from 20 to 99. Notably, a highly relevant study with a lower score of 14 was also included. This approach is estimated to save around 25 hours of screening time, considering an average 7-second review time for less relevant records.

While [15] applied a statistic approach, [16] applied a large language model to build a custom classifier for update search. Three models based on Bidirectional Encoder Rep-

[8] https://scikit-learn.org/

[9] https://scikit-learn.org/stable/modules/sgd.html

resentations from Transformers (BERT)[17] were tested. Model one utilizes text from Wikipedia and books for both vocabulary development and pre-training. The second model builds its vocabulary using text from Wikipedia and books and pre-trains with abstracts from selected articles. Meanwhile, the third model constructs its vocabulary from abstracts of articles acquired and uses abstracts from included articles for its pre-training phase. Each model underwent fine-tuning with the titles of included articles. The third model excelled in nearly all performance metrics, notably achieving an AUC above 90, surpassing other models which did not exceed 67. This implies that enriching language representation with domain-specific data boosts performance. To address performance distortion from imbalanced class composition, a common problem in applying ML classification for screening, dummy data were generated by altering keywords in excluded citations. This adjustment led to an increase of recall from 0.55 to 0.91.

Developing classifiers focues on standard eligibility criteria provides a promising alternative, applicable across various SRs, unlike those designed for specific SRs. The importance of study design in these criteria is notable, and [18] developed a classifier to categorize COVID-19 literature into one of 22 study designs. Five classifiers, based on both general and domain-specific corpora, were trained using manually annotated data records. The five classifiers were subsequently combined into an ensemble model. In the evaluated ensemble model, a voting strategy is employed, while another possibility is aggregating class probabilities. Study designs were classified into classes and subclasses. The ensemble model outperformed all standalone models, registering an AUC-ROC of 94.33 at the class level and 94.77 for specific study designs, compared to 91.77 and 92.07, respectively, by the best standalone model.

5. Conclusion and Discussion

Systematic review is the research methodology that provides the most evidence. The associated workload justifies the demand for efficient automation tools. This paper offers a detailed investigation of current tools and recent research aimed at automating the title and abstract screening process, particularly emphasizing the role of machine learning. Custom classifiers, informed by initial human decisions, are integrated into the available tools, but their accuracy for specific SRs often remains uncertain. Consequently, there is a need to integrate and report predefined rules transparently when using these systems. Furthermore, focusing on text preparation rather than selecting specific algorithms could lead to further improvements. Considering specific inclusion and exclusion reasoning instead of binary labels to train custom classifiers not only increases transparency but also performance. Although large language models significantly influenced other fields, this was not observed. However, for further research in this direction, the relevance of domain-specific vocabulary must be considered. Additional research on classification based on specific eligibility criteria should be preferred over focusing on specific reviews, as this enables the collection of more training data and expands its applicability to a broader range of reviews. This approach is already effective for SRs focused on RCTs, and efforts to classify other study types are in progress. It is proposed to identify additional common eligibility criteria and use ML classifiers to address them. In conclusion, the research underscores the need to advance automation in SRs, highlighting the potential and limitations of current solutions. It emphasizes the importance of continuous innovation and the cautious application of existing systems.

Acknowledgement

The World Health Organisation, in particular Alice Simniceanu and Luca Fontana, are acknowledged for providing invaluable advice and guidance, dedicating time to discuss existing solutions and current challenges in systematic review automation. The study was done as part of the joint CERN and WHO "CAiMIRA" project, that is funding the PhD project, in the context of which this paper was written.

References

[1] P. Glasziou, L. Irwig, C. Bain, and G. Colditz, Systematic Reviews in Health Care: A Practical Guide. Cambridge University Press, 2001.

[2] K. Coleman, S. Norris, A. Weston, K. Grimmer-Somers, S. Hillier, T. Merlin, P. Middleton, R. Tooher, and J. Salisbury, "NHMRC Additional Levels of Evidence and Grades for Recommendations for Developers of Guidelines," National Health and Medical Research Council (NHMRC), Canberra, 2009.

[3] M. R. Demetres, D. N. Wright, A. Hickner, C. Jedlicka, and D. Delgado, "A decade of systematic reviews: an assessment of Weill Cornell Medicine's systematic review service," Journal of the Medical Library Association: JMLA, vol. 111, no. 3, p. 728, 2023, Medical Library Association.

[4] R. Borah, A. W. Brown, P. L. Capers, and K. A. Kaiser, "Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry," BMJ Open, vol. 7, no. 2, p. e012545, 2017, British Medical Journal Publishing Group.

[5] J. C. Carver, E. Hassler, E. Hernandes, and N. A. Kraft, "Identifying barriers to the systematic literature review process," in 2013 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement, 2013, pp. 203-212.

[6] J. Clark, P. Glasziou, C. Del Mar, A. Bannach-Brown, P. Stehlik, and A. M. Scott, "A full systematic review was completed in 2 weeks using automation tools: a case study," Journal of Clinical Epidemiology, vol. 121, pp. 81-90, 2020.

[7] I. J. Marshall, A. Noel-Storr, J. Kuiper, J. Thomas, and B. C. Wallace, "Machine learning for identifying randomized controlled trials: An evaluation and practitioner's guide," Research Synthesis Methods, vol. 9, no. 4, pp. 602-614, 2018.

[8] K. G. Shojania, M. Sampson, M. T. Ansari, J. Ji, S. Doucette, and D. Moher, "How quickly do systematic reviews go out of date? A survival analysis," Annals of Internal Medicine, vol. 147, no. 4, pp. 224-233, 2007, American College of Physicians.

[9] E. M. Beller, J. K.-H. Chen, U. L.-H. Wang, and P. P. Glasziou, "Are systematic reviews up-to-date at the time of publication?," Systematic Reviews, vol. 2, no. 1, pp. 1-6, 2013.

[10] S. Van der Mierden, K. Tsaioun, A. Bleich, C. H. C. Leenaars, et al., "Software tools for literature screening in systematic reviews in biomedical research," Altex, vol. 36, no. 3, pp. 508–517, 2019. [Online]. Available: Springer International Publishing AG.

[11] Covidence, "Automation using the Cochrane RCT classifier," Covidence Knowledge Base, 2024. [Online]. Available: https://support.covidence.org/help/automatically-tag-studies-not-reporting-on-rcts. [Accessed: Mar. 8, 2024].

[12] J. Thomas, S. McDonald, A. Noel-Storr, I. Shemilt, J. Elliott, C. Mavergames, and I. J. Marshall, "Machine learning reduced workload with minimal risk of missing studies: development and evaluation of a randomized controlled trial classifier for Cochrane Reviews," Journal of Clinical Epidemiology, vol. 133, pp. 140-151, 2021, Elsevier.

[13] S. K. Venkata, S. Velicheti, V. Jamdade, S. Ranganathan, M. Achra, K. K. Banerjee, C. Dutta Gupta, M. Happich, and A. Barrett, "MSR99 Application of Artificial Intelligence in Literature Reviews," in Value in Health, vol. 26, no. 12, pp. S412, 2023, Elsevier. [Poster presentation].

[14] M. Khabsa, A. Elmagarmid, I. Ilyas, H. Hammady, and M. Ouzzani, "Learning to identify relevant studies for systematic reviews using random forest and external information," Machine Learning, vol. 102, pp. 465-482, 2016, Springer.

[15] C. Stansfield, G. Stokes, and J. Thomas, "Applying machine classifiers to update searches: Analysis from two case studies," Research Synthesis Methods, vol. 13, no. 1, pp. 121–133, 2022.

[16] S. Aum and S. Choe, "srBERT: automatic article classification model for systematic review using BERT," Systematic Reviews, vol. 10, no. 1, pp. 1–8, 2021.

[17] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," arXiv preprint arXiv:1810.04805, 2018.

[18] J. Knafou, Q. Haas, N. Borissov, M. Counotte, N. Low, H. Imeri, A. M. Ipekci, D. Buitrago-Garcia, L. Heron, P. Amini, et al., "Ensemble of deep learning language models to support the creation of living systematic reviews for the COVID-19 literature," Systematic Reviews, vol. 12, no. 1, p. 94, 2023.

dHealth 2024
D. Hayn et al. (Eds.)

doi:10.3233/SHTI240035

A Standardized Treatment Pathway for Telehealth-Based Care of Chronic Wounds in Austria

Hannah VINATZER[a,e], Angelika RZEPKA[a,1], Barbara BINDER[b], Viktoria GRUBER[b], Rainer HOFMANN-WELLENHOF[b], Patrick MAYR[a], Peter KASTNER[c], Alexander HÖLLER[a,d] and Günter SCHREIER[a,e]

[a] *AIT Austrian Institute of Technology, Graz, Austria*
[b] *Medical University of Graz Department of Dermatology and Venereology, Graz, Austria*
[c] *telbiomed Medizintechnik und IT Service GmbH, Graz, Austria*
[d] *University Hospital Innsbruck division for Nutrition and Dietetics, Innsbruck, Austria*
[e] *Graz University of Technology, Graz, Austria*

Abstract. Chronic wounds present a significant healthcare challenge in Austria as well as in other countries. The interdisciplinary approach to wound treatment involving various caregivers, doctors, and relatives, poses challenges in documentation and information exchange. To overcome these barriers and promote patient-centered care, a new telehealth-supported treatment pathway for chronic wounds has been developed. The primary focus was to regularly update the status of the chronic wound by responding to predefined questions and transmitted images of the chronic wound. This was achieved by an interdisciplinary team of experts in chronic wound care, providing a new perspective for digital implementation in the healthcare system.

Keywords. Telehealth, telemedicine, chronic wound, wound management, care guideline, patient-centered treatment

1. Introduction

Chronic wounds, as described in literature, are skin lesions that persist beyond a defined healing period, failing to restore the anatomical and functional integrity of the skin. The exact definition and, foremost, the specified period in which the wound should improve or ideally heal varies in the related literature. It is important to note that any wound has the potential to progress into a chronic wound, whereby a distinction can be made between several types of chronic wounds, based on the initial condition, for example, as a comorbidity of chronic diseases like diabetes mellitus. [1],[2]

Chronic wounds bring numerous challenges for patients, including severe pain, complications, and reduced quality of life, posing both physical and psychological burdens. [3],[4] However, the healthcare system faces a significant challenge in wound care, namely, the lack of a structured approach to wound management. Patients receive

[1] Angelika Rzepka, AIT Austrian Institute of Technology, Graz, Austria, Angelika.Rzepka@ait.ac.at

treatment from a variety of healthcare specialists, with treatment strategies constantly differing. The importance of uniform protocols in chronic wound care is emphasized in the literature, aiming to enhance wound healing. [5],[6] To enable a uniform approach in the treatment of chronic wounds, patient-centered treatment systems promise a positive outcome. Continuous communication between patients and wound care experts is crucial for implementing successful wound care, ensuring optimal treatment adjustments. [7],[8] The patient-centered approach could be combined with a telehealth service, to significantly enhance communication between patients and experts, leading to increased patient satisfaction and a corresponding improvement in quality of life. [9]

The treatment of chronic wounds also poses challenges for the healthcare system in Austria. As the "Wund?Gesund!" initiative shows from the data collected in 2015, around 70,000 people in Austria suffer from chronic wounds annually. However, only 15% of patients are treated according to current medical standards and around half of those affected are not sufficiently aware of the status of their wound during treatment. [10] These numbers underscore the urgency to reconsider routine wound care in Austria. Out of this need the "*Telewundmanagement*" project has been developed. The aim of the project was to establish a new tele-dermatology treatment pathway for routine use in the Austrian healthcare system regarding chronic wounds. The treatment of chronic wounds through a telehealth setting thus represents a new approach in wound care in Austria.

2. Methods

To develop a standardized care guideline for telehealth-based treatment of chronic wounds the following approach was made: First, a treatment pathway concept was developed in collaboration with four experts from the Department of Dermatology and Venerology, Medical University of Graz, Austria. This involved defining the use cases and inclusion criteria and key processes, that could be considered for the telehealth treatment of chronic wounds. After conducting a comprehensive analysis of those existing key processes, the concept of the treatment pathway had been defined. Based on the treatment pathway, a more detailed elaboration of the status update of the chronic wound was developed. Once this had been defined, an interdisciplinary workshop was held with several stakeholders from around Austria who work in the field of wound care management. We put emphasis on the diversity of the stakeholders in terms of professions and gender. This included two nurses with expertise in wound management, a doctor from the department of dermatology and venereology, a doctor from the department of wound care, a general physician, a surgeon and two experts in telehealth applications. The aim of this interdisciplinary meeting was to evaluate the concept of the treatment pathway and to implement new proposals or point out missing approaches, as well as to define core data fields for the documentation of chronic wounds.

3. Results

3.1. Definition of the use cases

The *"Telewundmanagement"* project addressed various use cases related to telehealth for chronic wounds using the Telewundmanagement (TWM) technical implementation of the treatment pathway:

- Healthcare Requirement: The general practitioner or home nurse seeks consultation for the therapy of a wound from a specialist working in the inpatient setting.
    - ICT Support Feature: Through the TWM application data can be exchanged between them and a summarizing report of the affected wound can be generated.
- The patient is discharged after an inpatient stay or outpatient care with a need for ongoing wound treatment.
    - Documentation and communication for post-treatment care are provided through the TWM System.
- A general practitioner would like to review the status of the wound treatment.
    - Relevant data can be accessed through the TWM System.
- A patient is admitted as an inpatient or comes to a care facility.
    - The current chronic wound data is accessible through the TWM System.

3.2. Telehealth treatment pathway for chronic wound patients

Enrolment Phase: Patients are included in the treatment pathway at the outpatient ward of the Department of Dermatology and Venerology, Medical University of Graz, Austria. The following pre-assignment processes, which are not part of the treatment pathway, are eligible: 1.) patient's own initiative, 2.) recommendation by caregiving relatives or home nursing care, 3.) referral from the general practitioner or specialist or 4.) referral from other departments of the hospital or other inpatient facilities.

- **Enrolment Process:** After being informed about the treatment pathway and obtaining consent, the patient can be enrolled in the treatment pathway for up to 6-month and, within the framework of the telehealth service, receive telemedical care in their home environment.
- **Registration Process:** A member of the outpatient expert team, such as a nurse, doctor, or coordinating personnel at the hospital, registers the patient in the TWM system. They also guide the patient or their designated trusted person through the treatment pathway, system operation, as well as their roles and responsibilities.

Monitoring Phase: The monitoring of the chronic wound is conducted via a status update including a photo and can be done by the patients themselves, caregiving relatives, home nursing care, general practitioners, or medical specialists.

- **Status Updates:** Over the 6 months of the treatment pathway, the patient, or the designated trusted person, telemedically transmits the status updates of the chronic wound. This status updates include a photograph of the wound and answering questions regarding the treatment and the status of the chronic wound. The expert is notified upon the status update and can respond to it, using predefined questions. After review by the doctor, the patient receives feedback, and, if needed, adjustments to the therapy are agreed upon.
- **Virtual Control:** Regular virtual controls and, if necessary, additional in-person contacts, are done. As part of this process, a comprehensive status report detailing the preceding wound care interventions can be generated and retrieved.
- **Follow-Up Examinations:** After 3 months, a doctor from the outpatient expert team coordinates a follow-up examination, followed by a quality control assessment of the care performance and a concluding examination after 6 months.

- **Virtual Checks:** Between these examinations, virtual checks are conducted every 2 to 4 weeks by the network doctors, a member of the outpatient expert team, and therapy adjustments are made, if necessary.

Completion Phase: After 6 months, a final discussion at the coordinating office, i.e., the Department of Dermatology and Venereology, Medical University of Graz, Austria, determines whether extending telemedical care by 3 months is needed.

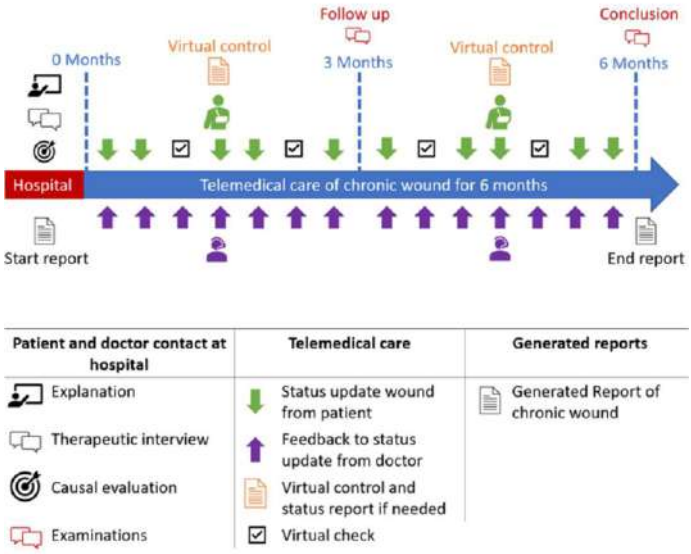


Figure 1. Graphic representation of the treatment pathway through the TWM system.

3.3. Status update of the chronic wound

The patient provides a status update of the chronic wound in regular intervals. A wound photo is taken alongside a specialized ruler to interpret coloration and size, which is shown in Figure 2.

Figure 2. Illustration of the wound ruler used. The image has been scaled so that the marked distances of 1 cm no longer show the actual width of 1 cm.

Besides the photo (taken with a standard smartphone or tablet), the status update of the chronic wound includes questions about the wound, well-being, and wound care. These questions are listed in Table 1. The question "Any other questions?" can be used to pose individual questions to the experts. All questions were mapped to the LOINC wound assessment panel[1].

[1] https://loinc.org/39135-9, last visited 2024-01-31.

Table 1. List of questions and possible answers that are posed to the patient as part of the status update. The questions and answers were translated from German into English.

Questions which can be answered with Yes or No:	
Has the previous treatment shown progress in wound healing?	
Was the wound treated as prescribed?	
Are there any limitations in daily activities due to the wound?	
Has emotional well-being worsened due to the wound?	
Is there any wound fluid leakage?	
Does the wound have a noticeable odor?	
Questions with multiple response options:	**Response options**
Is the wound painful?	Very severe pain, Severe pain, Moderate pain, Mild pain, No pain
Is changing the dressing painful?	Very severe pain, Severe pain, Moderate pain, Mild pain, No pain
How would you describe the surrounding area of the wound?	Blistered, Keratotic, Dried, Swelled, Reddened, Itchy, Flushed complexion, Hardened, Macerated
What is the condition of the wound?	Open wound, Wound healing, Covered wound, Yellow covered wound, Wound is flushed/healing, Wound is dried black
What color is the wound?	Red, Pink, Yellow, Black, Gray
How can the wound edges be described?	Poorly defined, Well defined, Rolled wound edges
Describe the color of the wound secretions.	Green, Yellow, No color, Brown, Red
How would you describe the wound fluid?	Bloody, Thin, Purulent, Viscous, No wound fluid
Open response questions:	**Limitations**
How wide is the wound?	Must be an integer
How long is the wound?	Must be an integer
Who conducted the status update?	
What is the current wound therapy?	
Any open questions?	

Table 2 lists the questions and answer options for the status update of the chronic wound that are answered by the doctor and communicated to the patient.

Table 2. List of questions which the doctor answers and is thus available as feedback for the patient. The questions and answers were translated from German into English.

Question which can be answered with Yes or No:
Shall the current wound care be continued?
Open response questions:
Recommended therapy (medication, dressing material, wound cleansing, compression, pressure relief)
Answer to open question

4. Discussion

Figure 1 illustrates the telehealth treatment pathway for the integrated care of chronic wounds, as designed for Austria's routine care infrastructure. This interdisciplinary guideline supports the discussed use cases, serving as a basis for further telehealth implementation. We acknowledge that initiating patient enrolment at the outpatient ward may not be optimal, as chronic wound patients ideally receive care in primary care settings. Several developments in Austria's healthcare system led to the fact that primary care, in general, does not feel responsible for chronic wounds at present. For the future we plan to provide for patient enrolment in primary care in the treatment pathway as well as to include outpatient specialists (e.g., dermatologists) as network doctors.

Generally, implementing this treatment pathway promises the following benefits for care and the patient. 1.) The standardized wound status update shown in Table 1 provides data on the wound in a standardized format. Questions and answer options were created to ensure clarity, considering non-medical users such as patients and caregivers. They are based on international documentation standards, thus providing the foundation for compiling a standardized document for Austria's electronic health record (ELGA). 2.) By taking a photo of the chronic wound with the colour ruler as shown in Figure 2, it is also possible for the expert to visually assess the chronic wound. Also, a series of photos can visualize the treatment success for both the experts and the patients. 3.) Another advantage is that the patient can provide feedback in the sense of Patient Reported Outcomes, see Table 2, and concerns can be clarified promptly by the expert. This is especially important since specific wound dressings cannot always be provided as prescribed in the home-care setting. 4.) The continuous status updates of the wound enable prompt adjustment and intervention if the wound care does not achieve the desired goals. 5.) Additionally, telemedical care might also reduce costs as well as CO_2 emissions for the patients, as the number of visits to the outpatient clinic can be minimized. [11]

The next step will be to implement the treatment pathway with telehealth interfaces for relevant stakeholder in a pilot region in Austria. A clinical trial with chronic wound patients will determine if the positive effects from test scenario will extend to real-world settings. Further development may involve integrating additional supportive treatments, such as standardized screening and dietary interventions for malnutrition, which are essential for both prevention and wound healing. [12]

Acknowledgment: The project was kindly funded by the Styrian Health Fund.

References

[1] Kyaw, B., Jaerbrink, K., Martinengo, L., Car, J., Harding, K., & Schmidtchen, A. (2017). Need for improved definition of "chronic wounds" in clinical studies. Acta dermato-venereologica, 98(1)
[2] Werdin, F., Tenenhaus, M., & Rennekampff, H. O. (2008). Chronic wound care. The Lancet, 372(9653)
[3] Eriksson, E., Liu, P. Y., Schultz, G. S., Martin -Green, M. M., Tanaka, R., Weir, D., Gurtner, G. C. (2022). Chronic wounds: Treatment consensus. Wound repair and regeneration, 30(2)
[4] Yan, R., Yu, F., Strandlund, K., Han, J., Lei, N., & Song, Y. (2021). Analyzing factors affecting quality of life in patients hospitalized with chronic wound. Wound Repair and Regeneration, 29(1)
[5] Morteza Mahmoudi & Lisa J Gould (2020) Opportunities and Challenges of the Management of Chronic Wounds: A Multidisciplinary Viewpoint, Chronic Wound Care Management and Research, 7
[6] Springer Vienna (2023). Keine Struktur für die Versorgung chronischer Wunden. ProCare 28
[7] Gethin, G., Probst, S., Stryja, J., Christiansen, N., & Price, P. (2020). Evidence for person-centred care in chronic wound care: a systematic review and recommendations for practice. Journal of wound care, 29
[8] Squitieri, L., Tsangaris, E., Klassen, A. F., van Haren, E. L., Poulsen, L., Longmire, & Pusic, A. L. (2020). Patient‐reported experience measures are essential to improving quality of care for chronic wounds: An international qualitative study. International wound journal, 17(4)
[9] Shelton, A., & Reimer, N. (2018). Telehealth Wound Applications: Barriers, Solutions, and Future Use by Nurse Practitioners. On - Line Journal of Nursing Informatics: OJNI; Chicago, 22(2).
[10] Herberger K, Rustenbach SJ, Haartje O, Blome C, Franzke N, Schäfer I, Radtke M, Augustin M (2011). Quality of life and satisfaction of patients with leg ulcers--results of a community-based study. Vasa. r;40(2).
[11] Dreher J, Meryk A, Riedl D, Kropshofer G, Hetzer B, Neururer S, Pfeifer B, Schneeberger-Carta V, Holzner B, Crazzolara R.(2023) Can virtual care reduce treatment burden in pediatric leukemia? A retrospective cost analysis. Pediatr Blood Cancer. 19
[12] Herberger K, Müller K,Protz K, Zyriax B-C, Augustin M, Hagenström K. (2020) Nutritional status and quality of nutrition in chronic wound patients.Int Wound;17

dHealth 2024
D. Hayn et al. (Eds.)

doi:10.3233/SHTI240036

Will 'Computable' Clinical Guidelines Be Compatible with Personalised Care?

Jack DOWIE[a,b,1], Mette Kjer KALTOFT[b] and Vije Kumar RAJPUT[a,c]

[a] *London School of Hygiene and Tropical Medicine, London, UK*
[b] *University of Southern Denmark, Odense, Denmark*
[c] *Stonydelph Health Centre, Tamworth, UK*

Abstract. Introduction. The potential benefits from digitalisation processes will only be fully realised if the conceptual challenges they uncover are accepted and addressed, alongside the technical ones such as interoperability. Will 'computable' clinical guidelines be compatible with personalised care if the definition of the relevant disease embeds preferences that pre-empt those of the individual patient? Method. As a case study we investigated the definition of diabetes in glycaemic management guidelines. Result. The dominant component of its definition – HbA1c ≥6.5% – embeds the consensus preference judgement of a 2009 International Expert Committee. Discussion. This preference-sensitive threshold for the diagnosis of diabetes has subsequently been endorsed in many guidelines relating to glycaemic management, though there are signs of awareness and concern with its implications. Conclusion. Those seeking to digitalise guidelines by making them 'computable' need to acknowledge and address their inbuilt preference-sensitivity - if they wish to further care that respects patient's preferences.

Keywords. clinical guideline, computable, diabetes, patient's preferences

1. Introduction

The possibility of Clinical Decision Support (CDS) being delivered in *computable* and, ultimately, computer-*executable* clinical guidelines, has led to some of the major ongoing projects in the digitalisation of healthcare. The Mobilizing Computable Biomedical Knowledge (MCBK) and OpenClinical projects are particularly noteworthy [1-4]. MCBK's stated mission is to disseminate biomedical knowledge in formats that can be shared and integrated into health information systems and applications. Making biomedical knowledge 'easily findable, universally accessible, highly interoperable and readily reusable' is seen as the way to further the joint goals of enhancing the care of the patient and the learning capacity of the healthcare system.

The multiple technical and organisational challenges in guideline digitalisation are well-appreciated, including by the UK chapter of the MCBK movement [5-7]. Most of the challenges are being tackled by the communities and stakeholders involved. The pervasive issues surrounding interoperability implementations and standards are the focus of HL7 FHIR (Health Level Seven - Fast Healthcare Interoperability Resources). Their clinical guideline implementation guidance [8] draws on the Multilayer

[1] Corresponding Author: Jack Dowie. E-Mail: jack.dowie@lshtm.ac.uk

Knowledge Representation Framework (MKRF) [9] with its four levels of guideline formatting, moving upwards from the currently dominant narrative level, initially to the semi-structured, then to the fully structured/computable, and finally to the fully computer-executable version.

While the fundamental aim of all these efforts is to improve the care of patients, the focus of most of this work is actually on providing support for the clinician. That CDS has been effectively *Clinician* Decision Support [10] led Sittig and colleagues to offer a 'lifecycle framework' to help guide digital initiatives towards patient-centered clinical decision support (PC CDS) [11]. The fundamental aim of a PC CDS would be to "ensure the right information is delivered to the right person, in the right format, via the right channel, at the right point in the workflow" - the so-called CDS Five-Rights. In keeping with this aim, the Sittig paper differs from all those previously cited in that it actually contains the word 'preferences' (whereas the others do not). Among its several appearances we find "In some circumstances, the patient's preferences for specific outcomes may be different from those of their clinicians. For example, a patient might prioritize the ability to drive over pain management for some musculoskeletal disorders, or even quality over quantity of life."

However, what does not appear to be appreciated, even by Sittig and colleagues, is that the 'knowledge' represented (at any level of structuring and computability) may be incompatible with the aim of a PC CDS. The patient's right to have *their* preferences make their decision on the basis of the information/knowledge delivered in a PC CDS is jeopardised if that information is not 'right'. And it will not be 'right' for that patient if it contains previously embedded preferences.

Pre-emptive preferences which make the knowledge within the guideline sensitive to preferences other than those of the patient have two possible origins. The simple and obvious level is in the compilation of the guideline, where movement from evidence and information (the 'is') to the recommendation (the 'ought') necessarily involves the introduction of preference (value) judgements. So long as the embedded preferences, and their basis, are transparent, the guideline remains useful in person-centred care, if used conditionally. If the embedded preferences come concealed (in the form of covert 'oughtism') this is ethically suspect – and legally questionable under a 'reasonable patient' standard for informed consent. The second possible source of pre-emptive preferences is the one investigated here: the preferences that may be embedded in the definition of the relevant disease (disorder, condition, syndrome).

2. Method

We took type 2 diabetes as our case study and traced the source and nature of the widely-accepted diagnostic criterion for diabetes and uncovered the reasoning behind it. We paid particular attention to whether the reasoning focused on the group level consequences of alternative possible cutoffs on a continuum, since the selection of one of these on this basis would potentially jeopardise person-centred care [12]. (To ensure the reasoning is reported accurately we quote extensively.)

3. Result

The currently dominant definition of diabetes can be sourced to an International Expert Committee (IEC), convened jointly by the American Diabetes Association, European Association for the Study of Diabetes, and International Diabetes Federation. Its 2009 report is clear and simple: "Diabetes should be diagnosed when A1C [i.e. HbA1c] is ≥6.5%" [13]. While there were other supporting considerations, the key reason provided for adopting this defining threshold was based on an analysis that included " ~28,000 subjects from nine countries and showed that the glycemic level at which the prevalence of "any" retinopathy begins to rise above background levels (any retinopathy includes minor changes that can be due to other conditions, such as hypertension), and for the more diabetes-specific "moderate" retinopathy, was 6.5% when the data were examined in 0.5% increments. Among the ~20,000 subjects who had A1C values <6.5%, "moderate" retinopathy was virtually nonexistent... the substantial increase in the prevalence of moderate retinopathy at A1C levels ≥6.5% supports a threshold level of glycemia that results in retinopathy most characteristic of diabetes... Any suggestion that the relationship between chronic glycaemic levels and the long-term complications of diabetes may be better expressed as a continuum, rather than as a strictly dichotomous relationship, is belied by the retinopathy findings presented herein."

So, the IEC placed the diagnostic threshold at the point where the prevalence of moderate or severe nonproliferative diabetic retinopathy (NPDR) was 2.6%, rejecting alternative thresholds, such as that associated with a prevalence of 0.7% (at 6.0% A1c) and of 4.3% (at 7.0% A1c) (data from [14]). That there were major possible health consequences from 'elevated' A1C, other than those of NDPR (kidney failure, peripheral neuropathy, heart disease) was emphasised, but the Committee seems happy to have used that one microvascular complication as a proxy for these.

The cutoff selection was not uncontroversial. "There is likely to be some initial debate concerning the cut point—A1C of 6.5%—chosen to define diabetes. This is, of course, a problem whenever one coerces a diagnosis, which by definition must be dichotomous, from a continuous variable... Concern will be compounded by the fact that the upper limit of normal for A1C is 6.0%, leaving something of a gray zone between this value and the 6.5% cut point for diabetes... The lack of an A1C value for a formal definition of "pre-diabetes" is likely to raise further and related concerns... The International Expert Committee is indeed careful to point out that the threshold does not identify an A1C level below which risk is nil but, instead, one below which risk is lower: an inflection point in a continuous positive relationship rather than a true step function" [15].

Surveying both the 2009 report and subsequent commentaries we conclude that reasoning has largely focused on the shape and position of functions for the consequences arising from alternative cutoffs, such as 2.6% NPDR at 6.5% A1c versus 4.3% at 7.0% A1c. Notably, the chosen disease-defining cutoff is presented as grounded in 'objective' medical/clinical data. The IEC does say that its decision "balanced the stigma and costs of mistakenly identifying individuals as diabetic against the minimal clinical consequences of delaying the diagnosis in someone with an A1C level <6.5%." But there is no other acknowledgement that the assignment of a cutoff installs, in the disease definition, a human preference for the consequences (harms and benefits) below the threshold as compared with those above it – in other words, it installs preferences in

relation to the possible false positive versus false negative errors that the IEC accept exist. This means the disease definition is preference-sensitive, specifically sensitive to the particular preferences installed in the Committee's cut-off selection. Despite a widespread contrary assumption, this conclusion holds up even if the function is of a strict 'hockey stick' shape, since patient's preferences may lead to a cut-off with medical consequences, even if ones without any are on offer.

4. Discussion

The UK chapter of MCBK has formed, in collaboration with the National Institute for Health and Social Care Excellence (NICE), a NICE Computable Implementation Guidance (NCIG) group with the ultimate objective of achieving HL7 FHIR standard-compliant digital guidelines. Its first major effort related to NICE Guideline (NG28) *Type 2 diabetes in adults: management* [16]**.** Of particular relevance here, this recommends that practitioners "Adopt an individualised approach to diabetes care that is tailored to the needs and circumstances of adults with type 2 diabetes, taking into account their personal preferences..." Quite quickly, NCIG found the diabetes management guideline contained significant challenges to its project in the form of "unstructured knowledge unlikely to be coded and some subjective judgement, for example, NG28 1.6.5 says 'Discuss and agree an individual HbA1c target with adults with type 2 diabetes. Encourage them to reach their target and maintain it unless any resulting adverse effects'" [17].

The NICE representative was acutely aware of the challenges of digitalisation [18]. "NICE is primarily still at level 1 of the [MKRF] knowledge hierarchy, producing much of its content as narrative text that, in computing terms, is unstructured... NICE has identified that adding structure and standard clinical codes to its guidelines, even to a semi-structured level, has significant methodological implications and an impact on the steps required to develop guidance... NICE understands that structured data and structured knowledge are crucial to enable the concepts of a continually learning healthcare system... Technically there are challenges of agreeing which existing formalisms, coding and information standards for representing clinical knowledge could be used to share knowledge effectively between systems, and where there are gaps, filling these by extending these standards or, if necessary, working to develop entirely new standards."

However, nowhere in the MCBK discussion or their later publication is the preference-sensitivity of the disease definition mentioned, or its major implications for digital structuring and computability explored. The definition of diabetes will, through the diagnostic threshold/s used in defining patients with it, embed the preferences of others over the probabilistic consequences of intervention and non-intervention. The preferences of a patient over those consequences are pre-empted in whole or substantial part by those of a group of medical experts. The guideline recommendation that the patient's preferences be 'taken into account' in the clinical consultation is therefore devoid of operational meaning.

Paradoxically, NICE can be interpreted as recognising the ontological problem, albeit in oblique fashion. In NG28 at 1.6.5 we find "Discuss and agree an individual HbA1c target with adults with type 2 diabetes. Encourage them to reach their target and maintain it, unless any resulting adverse effects (including hypoglycaemia), or their efforts to achieve their target impair their quality of life. Think about using the NICE

patient decision aid on weighing up HbA1c targets to support these discussions." The aid *Type 2 diabetes: agreeing my blood glucose (HbA1c) target* informs the patient that "For reducing the risk of long-term health problems, the evidence is unclear about how much extra benefit comes from aiming for a lower target HbA1c compared with aiming for a slightly more relaxed target. Discuss with your diabetes team how much benefit you might expect, thinking about your age, how long you have had diabetes and whether you already have some of the health problems that can come with it." There is no mention of what the 'slightly more relaxed target' might be, or of what the major long-term consequences of it would be. Instead, the remainder of the aid shifts attention solely to two possible downsides of 'aiming for a lower blood glucose target': having to take more medicines and being more likely to get side effects and being more likely to experience 'hypos'. The logic of NICE's acceptance of a 'relaxed target' (higher than 6.5%) for informed patients who decide to maintain glycemic control at (say) 7.2% HbA1c, is that the construction and diagnosis of a disease called diabetes is unnecessary.

5. Conclusion

The project to digitalise clinical guidelines for diabetes provides the opportunity to confront the currently undetected and undiagnosed challenge that follows from the preference-sensitivity of the disease definition. In making their decision on glycemic-related interventions, the preferences of the patient diagnosed with diabetes are currently pre-empted by those embedded in their diagnosis. Incidentally, the individualisation of care, as well as its personalisation, is jeopardised, because the knowledge about interventions being input into clinical decisions will often be compromised by the preference-sensitivity of the disease definition having affected the underlying research - such as trials being confined to persons with diabetes diagnosed at HbA1c $\geq$6.5%.

The informed consent process is jeopardised when the harms and benefits of interventions are not those associated with the observations for the individual patient, but ones mediated by their diagnosis. Specifically, we ask how can biomedical 'knowledge' in relation to a 'disease' support the personalised decision of a patient, if that 'knowledge' is sensitive to (more strongly, contaminated by) the preferences of the group of medical experts who created the disease, through the implicit installation of their consensual preferences in setting the threshold for its diagnosis? Patient's preferences should trump those of medical experts, whose expertise is confined to the medical consequences of options and does not extend to preferences over those - or any other - consequences.

Identifying a problem is not providing a solution, but it is a necessary condition for making progress towards one. What should digitalised Clinical Decision Support in the form of a Patient Decision Aid be doing? Pre-eminently it should be facilitating the elicitation of the patient's preferences in regard to the probabilistic consequences (benefits and harms) of available interventions at alternative cutoffs (e.g. 5.5, 6.0, 6.5, 7.0, 7.5), knowledge about which it is the aids function to provide, in conjunction with the clinician.

Most readers will have inferred that the argument has implications for all guidelines where the definition of the target condition involves a preference-based cut-off on a biophysical continuum (or instrument-based index). Osteoporosis and hypertension are just two of innumerable examples.

References

[1] Adler-Milstein J, Nong P, Friedman CP. Preparing healthcare delivery organizations for managing computable knowledge. *Learn Health Syst.* 2018 Oct 10;3(2):e10070. doi: 10.1002/lrh2.10070.
[2] Khan N, Rubin J, Williams M. Summary of fifth annual public MCBK meeting: Mobilizing computable biomedical knowledge (CBK) around the world. *Learn Health Syst*. 2023 Jan 12;7(1):e10357. doi: 10.1002/lrh2.10357.
[3] Fox J. Cognitive systems at the point of care: The CREDO program. *J Biomed Inform.* 2017 Apr;68:83-95. doi: 10.1016/j.jbi.2017.02.008. Epub 2017 Feb 15. PMID: 28232035.
[4] Fox J, Khan O, Curtis H, Wright A, Pal C, Cockburn N, Cooper J, Chandan JS, Nirantharakumar K. Rapid translation of clinical guidelines into executable knowledge: A case study of COVID-19 and online demonstration. Learn Health Syst. 2020 Jul 14;5(1):e10236. doi: 10.1002/lrh2.10236.
[5] Wyatt J, Scott P. Computable knowledge is the enemy of disease. *BMJ Health Care Inform.* 2020 Jul;27(2):e100200. doi: 10.1136/bmjhci-2020-100200.
[6] Walsh K, Wroe C. Mobilising computable biomedical knowledge: challenges for clinical decision support from a medical knowledge provider. *BMJ Health Care Inform*. 2020 Jul;27(2):e100121. doi: 10.1136/bmjhci-2019-100121.
[7] Wong D, Peek N. Does not compute: challenges and solutions in managing computable biomedical knowledge. *BMJ Health Care Inform.* 2020 Jul;27(2):e100123. doi: 10.1136/bmjhci-2019-100123.
[8] HL7 International / Clinical Decision Support. Clinical Practice Guidelines
[https://build.fhir.org/ig/HL7/cqf-recommendations/documentation-approach-06-01-levels-of-knowledge-representation.html]
[9] Boxwala AA, Rocha BH, Maviglia S, Kashyap V, Meltzer S, Kim J, Tsurikova R, Wright A, Paterno MD, Fairbanks A, Middleton B. A multi-layered framework for disseminating knowledge for computer-based decision support. *J Am Med Inform Assoc*. 2011 Dec;18 Suppl 1(Suppl 1):i132-9. doi: 10.1136/amiajnl-2011-000334.
[10] Rajput VK, Dowie J, Kaltoft MK. Are Clinical Decision Support Systems Compatible with Patient-Centred Care? *Stud Health Technol Inform*. 2020 270:532-536. doi: 10.3233/SHTI200217. PMID: 32570440.
[11] Sittig DF, Boxwala A, Wright A, Zott C, Desai P, Dhopeshwarkar R, Swiger J, Lomotan EA, Dobes A, Dullabh P. A lifecycle framework illustrates eight stages necessary for realizing the benefits of patient-centered clinical decision support. *J Am Med Inform Assoc*. 2023 Aug 18;30(9):1583-1589. doi: 10.1093/jamia/ocad122.
[12] Kaltoft MK, Nielsen JB, Dowie J. Risk Thresholds and Risk Classifications Pose Problems for Person-Centred Care. *Stud Health Technol Inform.* 2018 251:19-22. PMID: 29968591.
[13] International Expert Committee. International Expert Committee report on the role of the A1C assay in the diagnosis of diabetes. *Diabetes Care.* 2009 Jul;32(7):1327-34. doi: 10.2337/dc09-9033. Epub 2009 Jun 5.
[14] Colagiuri S, Lee CM, Wong TY, Balkau B, Shaw JE, Borch-Johnsen K; DETECT-2 Collaboration Writing Group. Glycemic thresholds for diabetes-specific retinopathy: implications for diagnostic criteria for diabetes. *Diabetes Care.* 2011 Jan;34(1):145-50. doi: 10.2337/dc10-1206.
[15] Fonseca V, Inzucchi SE, Ferrannini E. Redefining the diagnosis of diabetes using glycated hemoglobin. *Diabetes Care.* 2009 Jul;32(7):1344-5. doi: 10.2337/dc09-9034.
[16] NICE. Type 2 diabetes: agreeing my blood glucose (HbA1c) target Patient decision aid.
https://www.nice.org.uk/guidance/ng28/resources/patient-decision-aid-on-type-2-diabetes-agreeing-my-blood-glucose-hba1c-target-pdf-2187281198
[17] Scott P, Heigl M, McCay C, Shepperdson P, Lima-Walton E, Andrikopoulou E, Brunnhuber K, Cornelius G, Faulding S, McAlister B, Rowark S, South M, Thomas MR, Whatling J, Williams J, Wyatt JC, Greaves F. Modelling clinical narrative as computable knowledge: The NICE computable implementation guidance project. Learn Health Syst. 2023 Sep 28;7(4):e10394. doi: 10.1002/lrh2.10394.
[18] Mitchell A. A NICE perspective on computable biomedical knowledge. *BMJ Health Care Inform.* 2020 Jul;27(2):e100126. doi: 10.1136/bmjhci-2019-100126

dHealth 2024
D. Hayn et al. (Eds.)

doi:10.3233/SHTI240037

Exploring Opportunities for Clinical Data Warehouse Enhancement Through Data Catalog Integration

Andreas WALKER[a,1], Werner O. HACKL[a] and Bernhard PFEIFER[a,b]

[a] *Division for Digital Health and Telemedicine, UMIT TIROL - Private University for Health Sciences and Health Technology, Hall in Tirol, Austria*
[b] *Department of Clinical Epidemiology, Tirol Kliniken GmbH, Innsbruck, Austria*

Abstract. Secondary use of clinical health data implies a prior integration of mostly heterogenous and multidimensional data sets. A clinical data warehouse addresses the technological and organizational framework conditions required for this, by making any data available for analysis. However, users of a data warehouse often do not have a comprehensive overview of all available data and only know about their own data in their own systems - a situation which is also referred to as 'data siloed state'. This problem can be addressed and ultimately solved by implementation of a data catalog. Its core function is a search engine, which allows for searching the metadata collected from different data sources and thereby accessing all data there is. With this in mind, we conducted an explorative online market survey followed by vendor comparison as a pre-requisite for system selection of a data catalog. Assessment of vendor performance was based on seven predetermined and weighted selection criteria. Although three vendors achieved the highest score, results were lying closely together. Detailed investigations and test installations are needed for further narrowing down the selection process.

Keywords. Metadata management, Data Catalog, Clinical Data Warehouse, Secondary use, Health data

1. Introduction

The increasing digitization of health information goes hand in hand with great potential for secondary use. However, this indirect use of electronic health data, also known as reuse, implies a prior integration of usually very heterogeneous and multidimensional data sets [1, 2]. A clinical data warehouse (CDW) addresses the technological and organizational framework conditions required for this, by making any data available for analysis [3].

However, data warehouse users often do not have a comprehensive overview of all the data contained in the CDW. Furthermore, they do not know the organizational contexts in which this data was created. Hindered by the so called 'data siloed state', practitioners often have limited access to comprehensive knowledge of clinical health

[1] Corresponding Author: Andreas Walker, UMIT TIROL, Hall in Tirol, Austria, E-Mail: Andreas.Walker@umit-tirol.at

data beyond their domain or specific documentation systems. Overcoming this challenge is crucial for unlocking the full potential of a CDW. To achieve this, there is a need for detailed metadata to dismantle data silos and enable efficient secondary use of health data.

Effectively managing metadata is essential for transforming existing information into a valuable resource for data analytics. Achieving systematic metadata management is made possible through the implementation of a data catalog. Such a catalog serves as a centralized database, containing metadata from various data sources within the application system landscape. At its core, a data catalog functions as a search engine, enabling users to explore metadata collected from diverse data sources and gain access to a comprehensive dataset. In the realm of data analysis, organizations utilize data catalogs for data-driven innovations [4].

This transformative technology is not limited to corporate advantages; when integrated with a CDW, data analysts and medical IT specialists are empowered to work beyond their familiar datasets, broadening their access to valuable information. The global view of the CDW enables them to use the most suitable data for their analyses and thus efficient secondary use. In this context, the selection and implementation of a data catalog are underway for the CDW at an Austrian university hospital.

Our goal is to enhance the secondary utilization of clinical health data and elevate user interaction with CDW. This paper outlines the planned methodology as a pre-requisite for system selection, providing insights into the approach adopted for this optimization process.

2. Methods

Before specifying the concrete requirements for such a data catalog system, we conducted an initial online market survey. This survey focused on seven predetermined requirements based on a literature analysis, aiming to narrow down the numerous potential vendors. Out of these, two were established as mandatory criteria, while three were designated as target criteria and two as optional criteria. Table 1 gives an overview of the seven prioritized requirements.

Table 1. Prioritized requirements for pre-selection of potential vendors. 1 = optional criterion, 2 = target criterion and 3 = mandatory criterion. The first five criteria were taken from Olesen-Bagneaux's book "The Enterprise Data Catalog" [4]

Requirement	Description	Weighting (Scale: 1 – 3)
Data lineage	Central component for the automated documentation of data movements. Data lineage shows how data moves within the IT landscape (from system to system) and, ideally, is transformed in the process (in other words: the organizational context in which this data was created and what it is used for)	3
On premises	Server-based application that is installed and used locally (as opposed to cloud computing or software as a service)	2
Data intelligence	Use of the data catalog for data governance and data analytics purpose	2
Data Governance	Supports governance end users in the management of confidential and sensitive data	3

Knowledge graph-powered	Improves data search by modeling the knowledge universe as an ontology	1
Open Source	Free software and source code open	1
Health Data	Vendors of data catalogs that specialize in or have experience processing healthcare data to meet the needs of a healthcare organization	2

For vendor comparison we examined whether vendors meet the criteria completely, partially, or not at all. To assess vendor performance, we calculated the total score of the criteria met. For this purpose, the number of fulfilled mandatory, target and optional criteria was multiplied by their respective weighting value for each vendor. If a criterion was partially met, only half of the weighting value was considered. The total score was then calculated by adding those results together.

3. Results

A total of six different vendors were compared as part of an online market survey. Table 2 shows vendors meeting the predetermined criteria completely (C), partially (P), or not at all (N). For assessment of vendor performance, their total score was calculated based on the fully or partially fulfilled criteria.

In the result, three vendors achieved the highest score with ten points, two vendors took the second place (nine points) and one vendor scored lowest with eight points.

Table 2. Vendor comparison. C = Complete, P = Partial and N = None. Total score = value of vendor performance.

Criteria	**Compen-dium** [5]	**Infor-matica** [6]	**Amund-sen** [7]	**Alation** [8]	**Manta** [9]	**data.world** [10]
Data lineage	C	C	C	C	C	C
On premises	C	C	C	N	C	P
Data intelligence	N	C	N	C	N	C
Data Governance	C	C	C	C	C	C
Knowledge graph-powered	N	N	N	N	N	C
Open Source	N	N	C	N	N	N
Health data	C	N	N	N	P	N
Total score	**10**	**10**	**9**	**8**	**9**	**10**

4. Discussion

Implementing a data catalog for the Clinical Data Warehouse (CDW) can hold significant potential for enhancing data science efficiency, particularly in the context of healthcare. The fundamental purpose of a data catalog is to facilitate data discovery, offering data analysts and IT specialists a comprehensive overview of available datasets.

Vendor comparison showed comparable results for all vendors. With the highest score of ten points and the lowest score of eight points, all vendors lie very closely together. Although every vendor fulfills the mandatory criteria “Data lineage” and “Data Governance”, they vary greatly among each other regarding the other requirements.

Presumably, every vendor specializes in one or more capabilities. For example, 'Compendium' has been developed especially for the healthcare sector, whereas 'Alation' is a key-player in the field of data intelligence. Thus, detailed investigations and, where appropriate, test installations are necessary to narrow the selection further down before making a final decision.

Given the multitude of commercially available data catalogs with varying capabilities, a crucial step in our approach was to conduct an explorative market survey. This survey not only served as a preliminary exploration but also laid the foundation for the subsequent system specification of the data catalog. The survey involved a meticulous comparison of vendors, considering their offerings against a predefined set of criteria.

Before initiating the market survey, it was imperative to define core capabilities for the data catalog. Recognizing that no single data catalog fulfills every criterion, we selected seven requirements based on a literature analysis outlined in Table 1. As the system selection process unfolds, these criteria will be further complemented by additional requirements to ensure a comprehensive evaluation.

It is essential to emphasize that the preliminary market survey and vendor pre-selection were instrumental aids for the subsequent system specification. Their primary purpose was to provide an initial market overview and facilitate the identification of potential candidates. However, a comprehensive and systematic market analysis is reserved for the post-conceptualization phase.

In the subsequent, more in-depth analysis, vendors that may have scored lower in the initial survey are subject to reconsideration. This approach allows for more nuanced evaluation, considering factors that might not have been immediately evident in the first survey. For instance, clarity on whether individual vendors exclusively operate on a cloud basis or also offer on-premises solutions might become apparent during a more detailed examination.

In summary, our decision to proceed with a data catalog for the CDW involves a methodical and phased approach, ensuring that system specification is informed by a thorough understanding of the market and tailored to the specific needs of our healthcare context. The iterative nature of our process acknowledges the dynamic nature of the market and the evolving requirements of our data management goals.

References

[1] S. M. Meystre et al., Clinical Data Reuse or Secondary Use: Current Status and Potential Future Progress I Introduction, (2017),

[2] C. Safran et al., JAMIA Perspectives on Informatics Toward a National Framework for the Secondary Use of Health Data: An American Medical Informatics Association White Paper, *Journal of the American Medical Informatics Association* **14**(1) (2007), 1–9

[3] E. Tute et al., Modeling of ETL-Processes and Processed Information in Clinical Data Warehousing., *books.google.com* **248** (2018), 204–211

[4] O. Olesen-Bagneux, *The Enterprise Data Catalog*. O'Reilly Media, (2023).

[5] The Data Catalog for Healthcare | Compendium. https://compendiumdatacatalog.com/ (accessed Mar. 12, 2024).

[6] Data Catalog – Tools und Lösungen | Informatica Deutschland. https://www.informatica.com/de/products/data-catalog.html (accessed Mar. 12, 2024).
[7] Amundsen, the leading open source data catalog. https://www.amundsen.io/ (accessed Mar. 12, 2024).
[8] Alation Data Catalog | Alation. https://www.alation.com/product/data-catalog/ (accessed Mar. 12, 2024).
[9] Data Lineage Done Right | Data Lineage Tool | Manta. https://manta.io/ (accessed Mar. 12, 2024).
[10] Data Catalog | data. world. https://data.world/product/data-catalog/ (accessed Mar. 12, 2024).

dHealth 2024
D. Hayn et al. (Eds.)

doi:10.3233/SHTI240038

Quality and Effectiveness of AI Tools for Students and Researchers for Scientific Literature Review and Analysis

Martin DANLER[a,1], Werner O. HACKL[a] , Sabrina B. NEURURER [a,b] and Bernhard PFEIFER[a,b]

[a] *Division for Digital Medicine and Telehealth, UMIT TIROL - Private University for Health Sciences and Health Technology, Hall in Tirol, Austria*
[b] *Department of Clinical Epidemiology, Tyrolean Federal Institute for Integrated Care, Tirol Kliniken GmbH, Innsbruck, Austria*

Abstract. This study scrutinizes free AI tools tailored for supporting literature review and analysis in academic research, emphasizing their response to direct inquiries. Through a targeted keyword search, we cataloged relevant AI tools and evaluated their output variation and source validity. Our results reveal a spectrum of response qualities, with some tools integrating non-academic sources and others depending on outdated information. Notably, most tools showed a lack of transparency in source selection. Our study highlights two key limitations: the exclusion of commercial AI tools and the focus solely on tools that accept direct research queries. This raises questions about the potential capabilities of paid tools and the efficacy of combining various AI tools for enhanced research outcomes. Future research should explore the integration of diverse AI tools, assess the impact of commercial tools, and investigate the algorithms behind response variability. This study contributes to a better understanding of AI's role in academic research, emphasizing the importance of careful selection and critical evaluation of these tools in academic endeavors.

Keywords. Artificial Intelligence, Literature Analysis, Academic Writing

1. Introduction

The integration of Artificial Intelligence (AI) tools has significantly changed how students and researchers approach scientific literature in the ever-evolving landscape of academic research. This paper embarks on an exploratory journey to evaluate the quality and effectiveness of AI tools specifically designed to aid in scientific literature review and analysis. With an increasing reliance on AI to streamline research processes, it is imperative to understand the capabilities, limitations, and suitability of these tools for academic purposes [1].

Our research entailed a comprehensive investigation into various AI tools, focusing on those tailored for scientific research. The primary objective was to categorize these tools based on their functionality and effectiveness in assisting with different research related tasks. A critical aspect of our study involved creating a detailed table that

[1] Corresponding Author: Martin DANLER, UMIT TIROL, Hall in Tirol, Austria, E-Mail: martin.danler@umit-tirol.at

provides an overview of these tools, offering a clear and concise resource for students and researchers alike [1].

Our analysis was limited to freely available AI tools that allow users to pose direct research questions, in recognition of the accessibility challenges faced by many in the academic community. This constraint not only underscores the commitment to inclusivity but also reflects a realistic approach to resource availability for a significant portion of the academic community [2].

The following sections of this paper will discuss the methodology used to select and evaluate these tools, followed by the results and a detailed analysis of our findings. The objective of this work is to provide the academic community with valuable insights into the current state of AI tools in scientific research. This will enable informed decisions about their application in literature review and analysis. Through this endeavor, we hope to contribute to the broader discourse on the integration of AI in academic research, highlighting both its potential and its limitations.

2. Methods

The methodological approach adopted in this study was designed to systematically identify and evaluate AI tools suitable for scientific research and academic writing. This section outlines the steps taken to achieve a comprehensive overview of the available tools, culminating in a tabular presentation that categorises these tools based on their specific functionalities and applicability [3].

2.1. Internet Search and Keyword Selection

The primary method of data collection involved conducting extensive searches on Google, the world's most widely used search engine. To ensure a focused and relevant search, two sets of keywords were employed: "AI Tools for Scientific Research" and "AI Tools for Academic Writing." These keywords were chosen to capture a broad spectrum of AI tools that are specifically geared towards facilitating various aspects of academic research and writing.

2.2. Search Execution and Data Collection

Multiple searches were conducted using the selected keywords. Each search result was scrutinized for relevance, and only those websites, articles, and resources that directly addressed AI tools for scientific research and academic writing were considered for further analysis. This process involved evaluating the content of each source for its depth, accuracy, and applicability to the academic research context.

2.3. Tool Selection Criteria

The AI tools identified through the search were subjected to a set of selection criteria to ensure their relevance and utility for this study. The primary criteria included:

- **Accessibility:** The tool must be freely accessible, without requiring financial investment, thus ensuring inclusivity for all researchers and students, regardless of their institutional or financial resources.
- **Functionality for Direct Research Queries**: The tool must allow users to pose direct research questions or queries, a feature critical for academic research and literature analysis.

2.4. Tabular Representation and Categorization

A systematic table (Table 1) has been compiled of the AI tools found during the research. This tabular format was designed to provide a clear and concise overview of each tool, including its name, pricing model, primary functionalities that are suitable for specific research-related tasks, whether references are issued and whether a research question can be asked. The categorization aimed to assist researchers and students in easily identifying the tools most relevant to their specific needs.

Table 1. AI tools found during research.

Name	Pricing	Category	RQ	Ref
scite.at	Monthly fee	Search, Summarize, Analyze	No	Yes
Assistant by scite	Free (Beta)	Search, Summarize, Analyze	Yes	Yes
Iris.ai	Monthly fee	Visualize, Summarize, Manage, Analyze	No	Yes
Research rabbit	Free	Visualize, Summarize, Manage	No	Yes
Scispace	Free	Search, Summarize	Yes	Yes
ChatGPT	Free and monthly fee	Write, Summarize, Analyze	Yes	No
Consensus	Free	Search, Summarize	Yes	Yes
Elicit	Free and monthly fee	Search, Summarize	Yes	Yes
ChatPDF	Free	Summarize	No	Yes
Google Bard	Free (Beta)	Write, Summarize	Yes	Yes
Jenni AI	Free and monthly fee	Write, Summarize	No	Yes
Semantic scholar	Free	Search, Summarize	No	Yes
OpenRead	Free and monthly fee	Search, Summarize	No	Yes
Trinka	Free and monthly fee	Write	No	No
Microsoft Copilot	Free	Write, Summarize	Yes	Yes
Scholar GPT	Free	Search, Summarize	Yes	Yes

2.5. Analysis and Evaluation

Following the tools selection, each tool was evaluated for its effectiveness and quality in aiding academic research and writing. To scientifically evaluate these tools, we posed the following research question multiple times to each tool to test for consistency in the responses:

RQ: "How do Medical Data Warehouses influence the efficiency and quality of healthcare delivery?"

To ensure a thorough analysis of the AI tool outputs, our approach has expanded to include not only evaluating the outputs but also conducting a detailed assessment of their sources and volume.

Table 2 presents a systematic list of all AI tools investigated, along with the word count of their outputs and the number of sources they reference. Additionally, it indicates whether the cited sources are scientific in nature.

This enables a more nuanced analysis, offering a dual perspective on the content volume and source credibility — both critical factors in the evaluation of AI-assisted

literature review and analysis. Through this refined lens, we strive to present a clear and objective portrayal of each AI tool's capacity to support the expansive terrain of academic research.

The following section will explore the implications of these findings, explaining how the quantity of content and the quality of sources intersect to influence the usefulness of AI tools in academic contexts. This discussion aims to clarify the potential applications and benefits these technologies offer to the scholarly community.

Table 2. Outcome of each AI tool quantified.

Name	Word Count	References	Cited Sources
Assistant by scite	358	9	Yes
Scispace	153	5	Yes
ChatGPT	349	0	No
Consensus	328	5	Yes
Elicit	98	4	Yes
Google Bard	469	3	No
Microsoft Copilot	126	3	No
Scholar GPT	339	6	Yes

3. Results

The investigation into the effectiveness and quality of various AI tools for scientific research and academic writing yielded significant insights, primarily reflected in the diversity of responses to research queries. In our exploration of AI tools for scientific research, we uncovered a tapestry of variability in the responses to identical research questions. The outputs differed notably in scope and depth, a testament to the array of methodologies and algorithms at play within each tool. These differences pave the way for a multitude of interpretations and information presentations.

In delving into the provenance of information that the AI tools utilized, our analysis paid special attention to their source selection. Remarkably, the investigation revealed a nearly complete lack of shared sources among the tools, with one notable exception. The tools Elicit and Consensus stood out, as they both referenced the same source within their outputs. This was an isolated occurrence amidst a landscape where each AI tool otherwise appeared to access its own distinct repository of data. This singular overlap between Elicit and Consensus points to possible similarities in their design or source retrieval algorithms. Despite this instance of convergence, the general trend showed a wide-ranging array of sources across the tools, indicating a rich diversity of data inputs that contribute to the unique outputs each AI tool provides. This diversity highlights the breadth of information available to these tools and suggests a capacity for offering varied perspectives and insights, enriching the overall landscape of AI-assisted research.

When delving into the types of sources used, it became apparent that some tools, like Google Bard and Microsoft Copilot, were not limited to scholarly materials but also pulled from general websites and other non-academic domains. The integration of such materials raises important questions about the academic soundness and dependability of the output from these AI instruments.

In evaluating the AI tools, we repeatedly submitted the same research question to each system to assess the consistency of their responses. It was found that although the substance of the responses exhibited minor variations with each iteration, the word count of the provided answers remained strikingly similar across multiple queries. This suggests that while the AI's rephrasing capabilities introduce some degree of variability,

the underlying information processed by the tool is drawn from a stable set of sources. Furthermore, the constancy of the cited sources for each tool indicates a fixed reference database from which the AI retrieves information. This pattern of consistent word count and source usage provides an intriguing insight into the operational consistency of these AI tools, despite the superficial variation in their outputs.

Another intriguing finding was the temporal limitation in the data sourcing, particularly with tools like ChatGPT, which relied on information only up to April 2023. This raises concerns about the ability of such tools to provide the most up-to-date research findings or to reflect the latest advancements in a field.

Despite these variations, most AI tools were found to primarily utilize scientific sources. However, there was an observable opacity in how they selected the scientific papers they did use. This opacity could potentially introduce biases and calls into question the selection criteria for the information processed by these tools.

Among the findings, it was noteworthy that several AI tools provided responses of remarkable quality to the posed research questions, demonstrating a sophisticated understanding and synthesis of the available scientific literature. These high-quality responses showcase the potential of AI to contribute valuable insights and facilitate the initial stages of research.

However, it is crucial to underscore that despite the high quality of some responses, relying solely on AI-generated outputs for crafting a scientific paper is not advisable. The process of producing scholarly work involves critical analysis, interpretation, and a deep understanding of the subject matter, which extends beyond the capabilities of current AI tools. While these tools can serve as an effective starting point or aid in the research process, they cannot substitute the intellectual rigor and analytical depth required for academic writing. The variability in the quality and relevance of AIgenerated content further emphasizes the need for thorough review and supplementation with human expertise to meet the scholarly standards of scientific research.[4] In summary, while AI tools offer promising avenues for supporting literature review and analysis, our results indicate significant variability in their outputs and use of sources.

This variability highlights the need for careful consideration and scrutiny when using these tools for academic research purposes.

4. Discussion

This research has provided valuable insights into the capabilities and limitations of freely available AI tools in the context of scientific research and academic writing. However, it is important to acknowledge certain limitations of our study, which in turn suggest avenues for future research.

Our study had a significant limitation as we deliberately excluded commercial AI tools. This was due to our focus on freely accessible options. It raises the question of whether paid tools may offer better performance or more advanced features. To gain a more comprehensive understanding of the AI landscape in academic research, future studies should include a diverse range of commercial tools.

Our methodology was restricted to tools that allow for the direct input of research questions. We intentionally did not include commercial AI tools, focusing solely on freely accessible options. Future research should explore combinations of AI tools, such as pairing writing-focused tools with those specialized in analysis and research, to

uncover synergies and enhanced capabilities that may not be apparent when tools are used in isolation

Variability in the outputs of AI tools was another observation from our testing, with differences evident even when responding to identical queries. Future studies should delve into the reasons behind these discrepancies and their implications for reliability and usability in academic contexts.

Further complexity was observed in the responses to the same question posed by different users; despite largely consistent content, there was a variation in textual presentation and structure. Further research is warranted to investigate the algorithms and processing mechanisms that contribute to these variations, providing insights into how AI tools personalize responses and the extent to which this impacts the objectivity and consistency of the provided information.

Among the analyzed outputs, we found no critical perspectives toward medical data warehouses or their application in healthcare provision. All examined texts emphasized the benefits, such as improved healthcare efficiency, support for decision-making through data analysis, and advancement of research into new treatments. These positive outcomes are mirrored by contributions to cost reduction and improved operational efficiency, personalized patient care, and support for public health decisions and predictive analytics. While the potential and realized benefits of medical data warehouses are well-documented, the absence of explicit discussion around challenges, such as data privacy, data integration complexity, data quality limitations, or the need for careful implementation, points to a gap in the literature. Such critical perspectives are crucial for a balanced understanding of technology application in the real world and would encourage dialogue on necessary improvements, safety measures, and ethical considerations regarding the use of data warehouses in healthcare provision.

The potential of open-access AI tools in academic research has been clearly demonstrated in this study. As AI continues to advance, deepening our understanding of its applications and implications in academic research becomes increasingly important. Future research pathways are plentiful and should aim to broaden our knowledge of AI tools, enhance their functionality in academic research, and address the ethical and practical nuances of their use in real-world scenarios.

References

[1] A. Gabriel, Artificial intelligence in scholarly communications: An elsevier case study, *Information Services and Use* **39**(4) (2020), 319–333

[2] R. Buchkremer et al., The Application of Artificial Intelligence Technologies as a Substitute for Reading and to Support and Enhance the Authoring of Scientific Review Articles, *IEEE Access*, **7**. Institute of Electrical and Electronics Engineers Inc., 65263–65276, (2019),.

[3] J. I. Martínez-López et al., Which are the tools available for scholars? A review of assisting software for authors during peer reviewing process, *Publications*, **7**(3). MDPI AG, (2019),.

[4] M. Hosseini et al., Using AI to write scholarly publications, *Accountability in Research* (2023), 1–9

dHealth 2024
D. Hayn et al. (Eds.)

doi:10.3233/SHTI240039

Unveiling the Surge: A Comprehensive Analysis of the Stimulating Impact of the Covid-19 Pandemic on Telemedicine and Its Academic Landscape

Werner O. HACKL[a,1], Sabrina B. NEURURER[a,b] and Bernhard PFEIFER[a,b]

[a] *Division for Digital Health and Telemedicine, UMIT TIROL - Private University for Health Sciences and Health Technology, Hall in Tirol, Austria*
[b] *Department of Clinical Epidemiology, Tirol Kliniken GmbH, Innsbruck, Austria*

Abstract. The Covid-19 pandemic spurred an unprecedented shift towards digitalization, prompting a surge in telehealth practices. This paper explores the impact of the pandemic on telemedicine through a comprehensive analysis of scientific publications. Utilizing a bibliometric approach, the study examines trends in telemedicine research before and after the onset of Covid-19. The systematic search in PubMed yielded 8,454 pre-Covid-19 publications (2016-2019) and 16,633 post-Covid-19 publications (2020-2023). A total of 21,989 distinct keywords were extracted. Co-occurrence maps reveal evolving thematic clusters, with "mhealth" and "ehealth" dominating pre-Covid-19, while "Covid-19" emerges as a top keyword post-pandemic. The Top-10 keywords shift post-Covid-19, reflecting dynamic research priorities. The bibliometric approach illuminates a heightened exploration of telehealth solutions post-pandemic, emphasizing the enduring impact of the crisis on academic discourse. Changes in key terms and shifts in key term ranking indicate dynamic research priorities and a broader consideration of multidimensional healthcare challenges. Acknowledging study limitations, the analysis offers a high-level perspective, focusing on authors' keywords. Despite challenges, the study provides a systematic overview, revealing the emergence of new telemedicine application domains and the need for further in-depth analyses. Future research directions may explore the ecological impact of telemedicine applications and other intriguing aspects, contributing to a comprehensive understanding of telemedicine's scholarly trajectory.

Keywords. Telemedicine, Impact analysis, Covid-19

1. Introduction

The Covid-19 pandemic, marked by unprecedented lockdowns and restrictions, catalyzed a profound shift towards digitalization across various facets of daily life [1]. Suddenly, practices such as home office, homeschooling, online meetings, and remote work became ubiquitous, spanning all age groups from kindergarten to professional life

[1] Corresponding Author: Werner Hackl, UMIT TIROL, Hall in Tirol, Austria, E-Mail: werner.hackl@umit-tirol.at

and retirement. Amidst this transformation, the realm of medical care faced the urgent need to establish means of connecting patients with health professionals without physical interaction.

Virtually overnight, the healthcare landscape adapted, ushering in a plethora of telehealth, including telemedicine, telemonitoring, and teleconsulting applications [2]. This rapid evolution necessitated virtual consultations, remote diagnostics, treatment planning, and prescription dispensing. Notably, these changes rippled through the research and publication landscape within the field of telemedicine [3].

This paper delves into this phenomenon, seeking to explore whether the Covid-19 pandemic indeed had a stimulating effect on telemedicine. Through a thorough analysis of scientific publications, we scrutinize the impact before and after the pandemic-related disruptions to public life, providing insights into the evolving academic discourse and research activities in the realm of telemedicine.

2. Methods

This study conducted a systematic search in PubMed for telemedicine publications, utilizing the MeSH Major Topic "Telemedicine" as the primary search term. We supplemented this with the search terms "Telemedicine," "Telehealth," and "Telemonitoring" within titles and abstracts of publications that had not been annotated with MeSH terms yet (search string: "Telemedicine"[MeSH Major Topic] OR (("publisher"[SB] OR "inprocess"[SB]) AND ("Telemedicine"[OT] OR "Telehealth"[OT] OR "Telemonitoring"[OT]))).

The search was limited to publications from the four years preceding the onset of the COVID-19 pandemic (search term restriction “AND 2016:2019[DP]”) compared to the subsequent four years (search term restriction “AND 2020:2023[DP]”).

To handle the export of datasets from PubMed exceeding 10,000 publication entries, we employed command line tools within EDirect [4] to extract and export the complete PubMed records of all identified publications.

Considering the wealth of available publications, we found it necessary to establish a constraint on the units of analysis. Consequently, it appeared judicious to restrict the selection of keywords by the authors for their respective publications. From the identified publications, we thus extracted the authors’ keywords (annotated as “OT -” within the PubMed export files) using regular expressions. We subsequently conducted a bibliometric network analysis [5] using the software tool VOSviewer [6] and applying a procedure as described in [7]. We then created co-occurrence networks for the keywords in both two result-sets of publications as well as for the entire set of all scientific papers found in the years 2016 to 2023.

The settings for the co-occurrence mapping were chosen to result in networks comprising the top 200 terms. Due to the varying number of publications in the subsets, different settings for the selection threshold (i.e. the minimum number of publications in which a keyword must appear) had to be chosen for each subset.

3. Results

The literature search yielded a set of 8,454 publications from the pre-Covid-19 period and 16,633 from the equivalent post-Covid-19 period (years 2020-2023).

A total of 21,989 distinct keywords could be extracted from these publications, with 10,007 different keywords originating from pre-COVID-19 era publications and 16,310 different keywords from post-COVID-19 publications. The selection thresholds had been set to 16 (pre-Covid-19 set), 37 (post-Covid-19 set), and 51 (complete set 2016-2023).

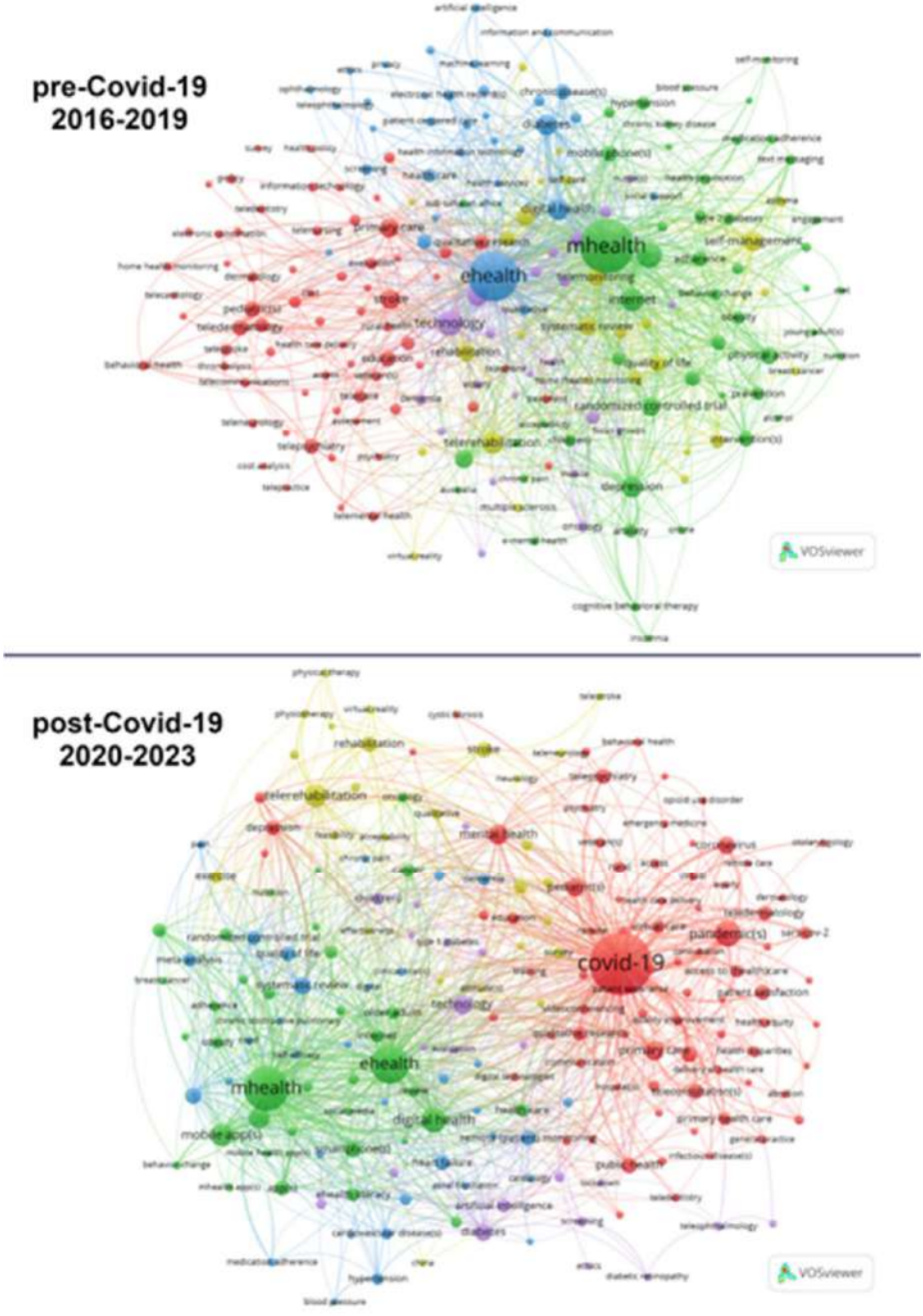


Figure 1. Clustered co-occurrence map of the Top-200 terms from the author's keywords of the 8,454 papers in the 2016-2019 query result set (top image) and of the 16,656 papers in the 2020-2023 set (bottom image).

We further removed the terms “telemedicine”, “telehealth” and “telemonitoring” before map creation because they explicitly appeared in the search string, naturally being the most common in all subsets, thereby occupying the most prominent positions in the resulting network maps and limiting the visibility of the subsequent terms.

Figures 1 and 2 depict clustered co-occurrence maps of the Top-200 authors’ keywords of the papers in the pre-Covid-19 (2016-2019, figure 1 - top image), post-Covid-19 (2020- 2023, figure 1 – bottom image) and combined (2016-2023, figure 2) result sets. Node sizes correspond to the frequency of the terms.

Edges indicate co-occurrence (figure 1 top: n= 4,060; figure 1 bottom: n=6,870; figure 2: n=7,967). Colors represent the different clusters (figure 1 top and bottom: n= 5; figure 2: n=3).

In the pre-Covid-19 publications, the terms “mhealth” (n=936) and “ehealth” (n=794) are the most prominent keywords. In the post-Covid-19 publications, these two keywords are also among the top keywords (“mhealth”: n=1,416; “ehealth”: n=1,053) but are clearly relegated to second and third place by the term “Covid-19” with a frequency of 2,799.

This order does not change when the two sets of resulting publications are combined. “Covid-19” remains in first place with 2,799 mentions, followed by “mhealth” with 2,294 and “ehealth” with 1,798. Except for the results of the clustering, where in the combined set only 3 clusters emerge from the 5 clusters in both separated sets, the co-occurrence map no longer changes substantially (cf. figure 2).

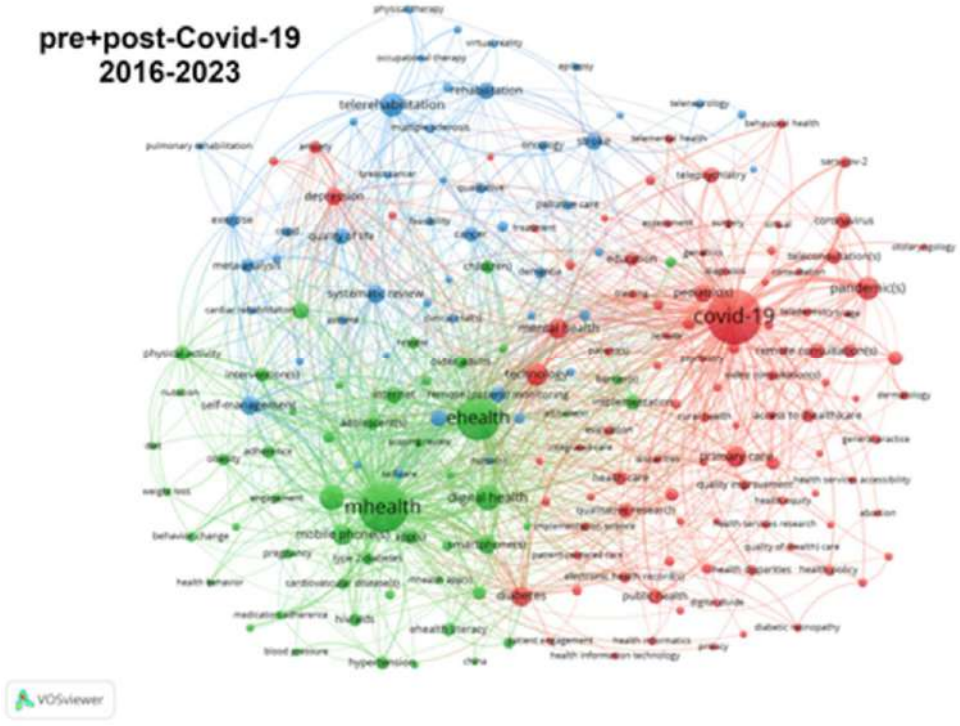


Figure 2. Clustered co-occurrence map of the Top-200 terms from the author’s keywords of the combined pre- and post-Covid-19 (2016-2023) query result set.

It is also noticeable that the following top terms also differ in the two result sets. For the pre-Covid-19 publications, these are "technology", "mobile app(s)", "internet", "telerehabilitation", "depression", "diabetes", "smartphone(s)", "digital health", and "self-management". In post-Covid-19 publications, "pandemic(s)", "digital health", "telerehabilitation", "mobile app(s)", "mobile phone(s)", "primary care" and "mental health" are among the most frequently chosen keywords.

Table 1 gives an overview of the Top-10 keyword-terms before and after the emergence of the Covid-19 pandemic and the associated restrictions.

Table 1. Overview on the Top-10 keyword-terms including absolute count and percentage in relation to all papers in the result set before, after and around the emergence of the Covid-19 pandemic and the associated restrictions.

Rank (Top-n)	Pre-Covid-19: Years 2016-2019 (count / percentage)	Post-Covid-19: Years 2020-2023 (count / percentage)	Pre+Post-Covid-19: Years 2016-2023 (count / percentage)
1	mhealth (936/11.1%)	covid-19 (2,799/16.8%)	covid-19 (2,799/12.7%)
2	ehealth (794/9.4%)	mhealth (1,416/8.5%)	mhealth (2,294/10.4%)
3	technology (213/2.5%)	ehealth (1,053/6.3%)	ehealth (1,798/8,2%)
4	mobile app(s) (212/2.5%)	pandemic(s) (503/3.0%)	digital health (603/2.7%)
5	internet (160/1.9%)	digital health (482/2.9%)	mobile app(s) (570/2.6%)
6	telerehabilitation (153/1.8%)	telerehabilitation (387/2.3%)	telerehabilitation (524/2.4%)
7	depression (144/1.7%)	mobile app(s) (364/2.2%)	pandemic(s) (503/2.3%)
8	diabetes (143/1.7%)	mobile phone(s) (318/1.9%)	technology (473/2.2%)
9	smartphone(s) (142/1.7%)	primary care (298/1.8%)	primary care (401/1.8%)
10	digital health (137/1.6%)	mental health (285/1.7%)	mobile phone(s) (396/1.8%)

4. Discussion

Our study employed a bibliometric approach to discern evolving telemedicine research trends before and after Covid-19. The surge in post-pandemic publications not only signifies a heightened exploration of telehealth solutions but also underscores the enduring impact of the crisis on the academic discourse within the field, as evidenced by the sustained prominence of "Covid-19" as a top keyword.

The shift in key terms, such as the rise of "mhealth," "ehealth," and "digital health," indicates dynamic research priorities, while the inclusion of terms like "pandemic(s)," "primary care," and "mental health" in post-Covid-19 publications reflects a broader consideration of multidimensional healthcare challenges.

In essence, this study not only quantifies the increase in telemedicine research but also illuminates nuanced shifts in thematic priorities, offering crucial insights into the lasting impact of the pandemic on healthcare practices and the ongoing scholarly exploration of telemedicine's pivotal role in modern healthcare delivery.

Acknowledging the major limitation of such a study, it is crucial to note that our analysis adopts a high-level perspective, focusing on authors' keywords. While a more granular examination of publication content as shown in [7] could provide deeper

insights, the sheer volume of papers makes this a challenging task. The presented approach, however, provides a thorough and systematically executed overview.

Delving into the papers, we observed the emergence respective the massive increase in new telemedicine application domains (e.g. teledermatology, telepsychiatry, telestroke, teledentistry, teleophthalmology, telenursing, telepharmacy, etc.), particularly in response to the challenges posed by the Covid-19 pandemic.

Furthermore, studies on the effects and benefits of these solutions have begun to surface. To gain more nuanced insights, further in-depth analyses are warranted, promising potentially intriguing results.

An additional compelling aspect worth investigating in more detail is the positive impact telemedicine applications can have on the ecological footprint [8,9]. Although some studies have touched upon this, systematic reviews and meta-analyses will provide evidence in this regard in the future.

Until then, many more fascinating aspects of telemedicine will undoubtedly come to light. Finally, as telemedicine continues to be integral to healthcare, our analysis significantly contributes to understanding its scholarly trajectory and may help shaping future research directions.

References

[1] Amankwah-Amoah J, Khan Z, Wood G, Knight G. COVID-19 and digitalization: The great acceleration. J Bus Res. 2021 Nov 1;136:602–11.

[2] Colburn DA. The Impact of Telehealth Expansion on Health Care Utilization, Access, and Outcomes During the Pandemic: A Systematic Review. Telemed J E Health. 2023;

[3] Hackl WO, Neururer SB, Pfeifer B. Telemedicine Research from Big Bang to 2022. Stud Health Technol Inform. 2023 May 2;301:220–4.

[4] Entrez Direct: E-utilities on the Unix Command Line - Entrez Programming Utilities Help - NCBI Bookshelf [Internet]. [cited 2024 Jan 31]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK179288/

[5] Waltman L, van Eck NJ, Noyons ECM. A unified approach to mapping and clustering of bibliometric networks. J Informetr. 2010;4(4):629–35.

[6] van Eck NJ, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics. 2010;84(2):523–38.

[7] Hackl WO, Neururer SB, Pfeifer B. Transforming Clinical Information Systems: Empowering Healthcare through Telemedicine, Data Science, and Artificial Intelligence Applications. Yearb Med Inform. 2023 Aug;32(1):127–37.

[8] Smith MW, O'Shea AMJ, Wray CM. Health Care and Climate Change-Telemedicine's Role in Environmental Stewardship. JAMA Netw Open. 2023 Jan 3;6(1):e2253794.

[9] Patel KB, Gonzalez BD, Turner K, Alishahi Tabriz A, Rollison DE, Robinson E, et al. Estimated Carbon Emissions Savings With Shifts From In-Person Visits to Telemedicine for Patients With Cancer. JAMA Netw Open. 2023 Jan 3;6(1):e2253788.

dHealth 2024
D. Hayn et al. (Eds.)

doi:10.3233/SHTI240040

Transforming Tele-Ophthalmology: Utilizing Cloud Computing for Remote Eye Care

Marco SCHWEITZER[a,b,1], Philipp OSTHEIMER[b], Arno LINS[c], Vito ROMANO[d], Bernhard STEGER[e], Daniel BAUMGARTEN[b] and Marco AUGUSTIN[c]

[a] *Division for Digital Health and Telemedicine, UMIT TIROL - Private University For Health Sciences and Health Technology, Hall in Tyrol, Austria*
[b] *Institute of Electrical and Biomedical Engineering, UMIT TIROL - Private University For Health Sciences and Health Technology, Hall in Tyrol, Austria*
[c] *Occyo GmbH, Innsbruck, Austria*
[d] *Eye Unit, ASST Spedali Civili di Brescia, Brescia, Italy*
[e] *Department of Ophthalmology, Medical University of Innsbruck, Innsbruck, Austria*

Abstract. Background: Tele-ophthalmology is gaining recognition for its role in improving eye care accessibility via cloud-based solutions. The Google Cloud Platform (GCP) Healthcare API enables secure and efficient management of medical image data such as high-resolution ophthalmic images. Objectives: This study investigates cloud-based solutions' effectiveness in tele-ophthalmology, with a focus on GCP's role in data management, annotation, and integration for a novel imaging device. Methods: Leveraging the Integrating the Healthcare Enterprise (IHE) Eye Care profile, the cloud platform was utilized as a PACS and integrated with the Open Health Imaging Foundation (OHIF) Viewer for image display and annotation capabilities for ophthalmic images. Results: The setup of a GCP DICOM storage and the OHIF Viewer facilitated remote image data analytics. Prolonged loading times and relatively large individual image file sizes indicated system challenges. Conclusion: Cloud platforms have the potential to ease distributed data analytics, as needed for efficient tele-ophthalmology scenarios in research and clinical practice, by providing scalable and secure image management solutions.

Keywords. Telemedicine, Ophthalmology, Cloud Computing.

1. Introduction

The advent of telemedicine has revolutionarily transformed healthcare delivery, making medical expertise accessible even in remotest areas. In ophthalmology, the impact is particularly pronounced, given the visual nature of the specialty and the technical advances in imaging technologies. Tele-ophthalmology has emerged as a critical tool for providing eye care services [1].

The demand for tele-ophthalmology services is driven by several factors: an aging global population with an increased prevalence of eye diseases, the shortage of ophthalmologists in rural and underserved areas, and the ongoing advancements in digital imaging and communication technologies [2,3]. Despite these advancements, current tele-ophthalmology practices continue to face regulatory and technological limitations. Practical issues such as reimbursement and acceptability remain significant

[1] Corresponding Author: Marco Schweitzer, UMIT TIROL - Private University For Health Sciences and Health Technology, Hall in Tyrol, Austria, E-Mail: marco.schweitzer@umit-tirol.at

barriers. On the technological front, limitations include a lack of operator independence, challenges in device interoperability, and the need for both efficient and accurate image annotation and analysis, which are critical for effective ophthalmic diagnosis [2,4].

To address these rising issues in ocular diagnostics, especially in anterior segment imaging, Occyo (Innsbruck, Austria) has developed an imaging technology designed to take standardized high-resolution photographs of the visible ocular surface in an operator-independent way, applicable to telemedical use [5,6]. Employing such devices in a location-independent manner and sharing them with remote ophthalmologic experts for detailed examination, it has the potential to shape the future of ophthalmologic diagnostics [4]. However, the challenge lies not only in image acquisition, but also in the efficient transmission, storage, and analysis of the high-resolution image data. Therefore, standards like DICOM (Digital Imaging and Communications in Medicine) [7] and best practices like IHE (Integrating the Healthcare Enterprise) [8] are used to facilitate the seamless exchange and management of medical images and related information to use case scenarios.

In recent years, cloud services have emerged as a pivotal resource in the healthcare sector, offering scalable, secure, and efficient solutions for managing and processing vast amounts of medical data [9]. These services are increasingly valued also in telemedicine for their ability to manage and process medical data efficiently, offering essential features like scalability, security, and efficiency. Despite their potential, the practical implementation of cloud services, particularly in the domain of teleophthalmology together with specific solutions like the Google Cloud Platform (GCP), remains underexplored [10]. This study aims to bridge this gap by analyzing how cloud-based solutions can significantly enhance tele-ophthalmology and provide a prototype on utilizing the GCP for telemedical applications.

2. Methods

The concept for the prototype is based on the principles of the Integrating the Healthcare Enterprise (IHE) Eye Care technical framework [11]. This approach involves leveraging the GCP as a Picture Archiving and Communication System (PACS) and utilizing the Open Health Imaging Foundation (OHIF) Viewer as the ophthalmic expert's primary tool for viewing and analyzing images. Occyo's anterior eye photography device is used as the imaging modality within this workflow. The core idea behind this methodology is to test and evaluate a system designed for storing all data efficiently and securely within the GCP and provide means for adding related information like annotations. Building upon this foundation, the proposed workflow is designed to support telemedical routine eye care needs, facilitating the capture of high-resolution eye images during patient visits. The intention is to enable remote ophthalmology experts to precisely review and analyze these images, thereby paving the way for the creation of structured reports with support of the system. This envisioned process aims to bridge the gap in eye care accessibility, offering a glimpse into the potential for remote diagnostics and treatment planning that combines expert analysis with the convenience and efficiency of tele-ophthalmology. An overview of the proposed prototype is depicted in Fig. 1 which is based on the previously published DICOM Simulation Platform [7].

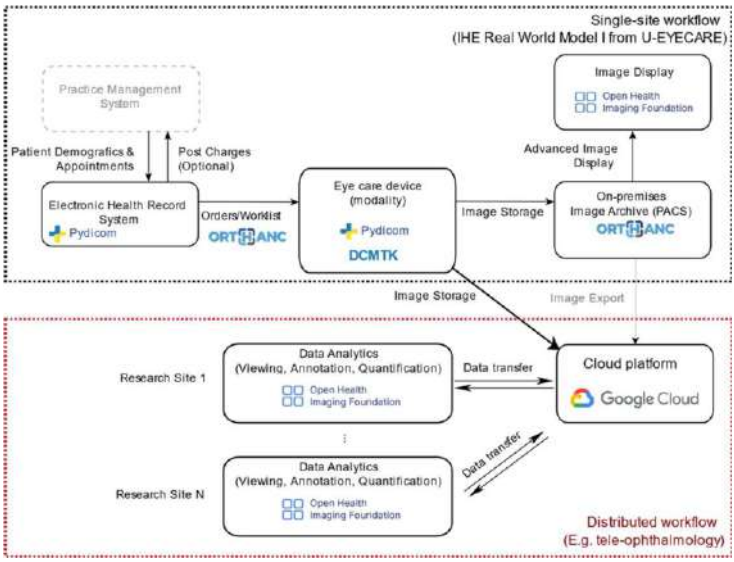


Figure 1. The workflow of the IHE Real World Model I from the U-EYECARE profile together with our DICOM simulation platform was expanded to enable the integration with a cloud platform, e.g., GCP.

The GCP offers a healthcare API for providing flexible, accessible and secure data stores such as DICOM or HL7 FHIR [12]. Further, the GCP is certified against several standards as outlined in Google Cloud's compliance offerings for the healthcare and life sciences industries within the European Union [13]. Beyond regulatory compliance the healthcare API offers data location control, and de-identification capabilities, ensuring utmost security and privacy adherence. We configured a DICOM store within GCP to accommodate the high-resolution ophthalmic images through DICOMweb. This web-based protocol facilitated seamless communication between our imaging modality and the GCP DICOM store. It allowed for the direct uploading of ophthalmic images from various locations, bypassing the limitations of traditional DICOM DIMSE networking. Strategies for optimal data transfer and memory optimization were analyzed. Further the imaging modality was integrated to utilize DICOMweb protocols, ensuring a streamlined process for transmitting images to the cloud. In parallel, we deployed an instance of the OHIF Viewer in the cloud [14], a well-established, open-source, web-based platform designed for DICOM. Known for its exceptional flexibility and seamless compatibility with DICOMweb, the viewer was configured to interface directly with the GCP DICOM store. This integration was pivotal in enabling relevant stakeholders to remotely access and analyze trial patient images.

Using this infrastructure, the OHIF Viewer provides the possibility to enhance its functionality. Individual so-called modes can be created in the OHIF Viewer, that offer a complete tailored functionality collection for the anterior segment photography use case. Each mode can load individual implemented extensions, e.g., an extension for annotating regions of interest (ROIs) within images to mark pathologies like lesions on the ocular surface and monitor them over time. Also, an extension within the OHIF Viewer is equipped with the capability to store annotations as DICOM structured reports which are exported to the GCP DICOM Store. These extensions will permit the input of general metadata, such as image quality assessments, and specialized scores like the

Efron grading for evaluating eye redness [15]. The integration ensures that all annotated information, along with crucial metadata, is cohesively maintained with the original imaging files, which is essential for maintaining the integrity and utility of the annotations and assessments in ongoing and future analyses.

This prototype is focused on the use case for data sharing for clinical research trials and involves using the platform to capture eye images and enable professionals to annotate and rate these images remotely. The ability to annotate and score images within the OHIF Viewer is anticipated to facilitate a more structured and comprehensive analysis, essential for research purposes.

3. Results

The prototype has successfully implemented a seamless integration of a GCP DICOM storage, which is specifically tailored to the management of high-resolution ophthalmic images. Problems arose due to the file size of these images, which initially exceeded acceptable waiting times when retrieving and displaying them in OHIF viewer with the default configuration. To tackle the challenge of large image file sizes and slow loading times in our tele-ophthalmology system, we applied JPEG2000 lossless compression, effectively reducing the average image size from 50 MB to ~30 MB per DICOM image. Additionally, we identified that the OHIF Viewer's simultaneous loading of images caused significant delays. By implementing a sequential loading strategy, we considerably improved the system's efficiency; where previously loading a single image from a set of five could take about 60 seconds with a 20 Mbit/s internet connection, the new approach allows the first image to be accessible in just 12 seconds. This enhancement, coupled with DICOMWeb's efficient and secure image transfer capabilities, has markedly optimized our tele-ophthalmology workflows. The successful integration of the OHIF viewer with our GCP DICOM storage has enabled a user-friendly interface for medical professionals to access, analyze, and annotate patient images remotely. Customizable through extensions, this setup enhances data sharing and analysis in research studies and lays the groundwork for future integration of machine learning and AI algorithms. These developments are crucial for advancing diagnostic capabilities and underscore our commitment to evolving a sophisticated, automated tele-ophthalmology solution.

4. Discussion

This study analyzes the potential of cloud-based solutions, particularly the GCP for teleophthalmology use cases. Hereby, we investigated the integration of the GCP into clinical workflows which potentially could be a step forward in increasing efficient and safe teleophthalmology practices. Hereby, modern cloud platforms are utilizing the DICOMWeb standard. In contrast, most medical imaging devices use the DICOM DIMSE protocol for data communication and transfer. Fulfilling this prerequisite is key for making an imaging device ready for web-based medical imaging and the interoperability with cloud platforms to be used in telemedicine use-cases.

In addition, DICOMweb protocols enhance tele-ophthalmology by streamlining image management and enabling remote analysis through compatible DICOM Viewers

such as the OHIF Viewer. The decision to use the OHIF Viewer as the DICOM viewer met all requirements, as it offers extensive flexibility for expansion and customization as well as easy connectivity with the GCP [16]. The development of a specific ophthalmic device-related extension in the OHIF Viewer, i.e. a mode, is especially promising, enabling customizable annotations and assessments, such as eye redness [15]. The ability for remote annotation and scoring of images within the OHIF Viewer and being synchronized with the GCP data storage backend may impact ophthalmic research by fostering collaborative multi-site studies.

Information communication and usability challenges associated with file size and loading time emerged unexpectedly throughout implementation. The functionality to generate and display DICOM thumbnails is typically supported by known PACS systems, but this feature is currently not provided by the GCP. While lossless compression represents a viable strategy to mitigate the issue of large file sizes, it inadvertently results in thumbnails being loaded at full-size quality, thus not effectively reducing loading times for initial image previews.

The importance of addressing general bottlenecks cannot be overstated; reliance on cloud services like GCP necessitates fast and reliable internet connectivity and introduces increased complexity in integration, alongside persistent concerns related to data security. Furthermore, the effective management and maintenance of cloud-based platform demands the expertise of cloud specialists to ensure the smooth and secure functioning of the services. For research and education this is highlighting the need for a strategic approach to resource allocation and skills development within the context of cloud computing.

Looking ahead, the planned integration of artificial intelligence (AI) based tools into this platform is an exciting development. Previous studies have shown the efficacy of AI in medical imaging, particularly in fields necessitating detailed image interpretation, like ophthalmology [17,18]. Integration with the GCP, as proposed in this work, offers the possibilities to further utilize Google's machine learning frameworks and build machine learning tools based on the data ingested using the same platform. As the OHIF viewer supports the incorporation of such tools, the platform's capabilities can be greatly enhanced with innovative functions. This integration ensures a seamless user experience, empowering medical professionals with advanced tools for analysis and diagnosis directly within their clinical and research workflows.

In conclusion, this research demonstrates that the GCP provides a robust foundation for integration within a tele-ophthalmology setting, taking into account certain challenges for an effective and efficient integration. The successful but somewhat problematic deployment indicates a promising direction for future developments in this field, potentially leading to more accessible, efficient, and comprehensive eye care worldwide. Continued development and research are essential to overcome current limitations and fully realize the potential of tele-ophthalmology. Our next steps will be to use this presented platform for the purpose of analyzing and annotating clinical trial data.

Acknowledgments
This work was supported by the federal state of Tyrol within the K-Regio program (project implEYE).

References

[1] S. Rathi, E. Tsui, N. Mehta, S. Zahid, and J.S. Schuman, The Current State of Teleophthalmology in the United States, *Ophthalmology*. **124** (2017) 1729–1734. doi:10.1016/j.ophtha.2017.05.026.

[2] M. Sharma, N. Jain, S. Ranganathan, N. Sharma, S.G. Honavar, N. Sharma, and M.S. Sachdev, Tele-ophthalmology: Need of the hour, *Indian J Ophthalmol*. **68** (2020) 1328–1338. doi:10.4103/ijo.IJO_1784_20.

[3] J. Dolar-Szczasny, A. Barańska, and R. Rejdak, Evaluating the Efficacy of Teleophthalmology in Delivering Ophthalmic Care to Underserved Populations: A Literature Review, *Journal of Clinical Medicine*. **12** (2023) 3161. doi:10.3390/jcm12093161.

[4] K.E. Goetz, A.A. Reed, M.F. Chiang, T. Keane, M. Tripathi, E. Ng, T. Nguyen, and M. Eydelman, Accelerating Care: A Roadmap to Interoperable Ophthalmic Imaging Standards in the United States, *Ophthalmology*. **131** (2024) 12–15. doi:10.1016/j.ophtha.2023.10.001.

[5] M. Augustin, P. Ostheimer, D. Baumgarten, U. Hausmann, V. Romano, and B. Steger, Standardized Imaging of the Ocular Surface Using a Novel External Eye Photography System, *Investigative Ophthalmology & Visual Science*. **64** (2023) 3405.

[6] M. Augustin, P. Ostheimer, U. Hausmann, D. Baumgarten, V. Romano, and B. Steger, An imaging system for standardized and enhanced photographs of the ocular surface, *Investigative Ophthalmology & Visual Science*. **64** (2023) PB0095.

[7] M. Schweitzer, K. Flórez, B. Steger, D. Baumgarten, V. Romano, and M. Augustin, Integrating a Novel Eye Imaging System into Clinical Practice: An Open-Source DICOM Simulation Platform, *Studies in Health Technology and Informatics*. **301** (2023) 198–203. doi:10.3233/SHTI230039.

[8] M. Schweitzer, B. Steger, A. Hoerbst, M. Augustin, B. Pfeifer, U. Hausmann, and D. Baumgarten, Data Exchange Standards in Teleophthalmology: Current and Future Developments, *Studies in Health Technology and Informatics*. **293** (2022) 270–277. doi:10.3233/SHTI220380.

[9] A. Tahir, F. Chen, H.U. Khan, Z. Ming, A. Ahmad, S. Nazir, and M. Shafiq, A Systematic Review on Cloud Storage Mechanisms Concerning e-Healthcare Systems, *Sensors (Basel)*. **20** (2020) 5392. doi:10.3390/s20185392.

[10] L. Faes, S.K. Wagner, D.J. Fu, X. Liu, E. Korot, J.R. Ledsam, T. Back, R. Chopra, N. Pontikos, C. Kern, G. Moraes, M.K. Schmid, D. Sim, K. Balaskas, L.M. Bachmann, A.K. Denniston, and P.A. Keane, Automated deep learning design for medical image classification by health-care professionals with no coding experience: a feasibility study, *The Lancet Digital Health*. **1** (2019) e232–e242. doi:10.1016/S2589-7500(19)30108-6.

[11] Integrating the Healthcare Enterprise, IHE EYECARE TF-1 Profiles Revision 4.0, (2016).

[12] DICOM | Cloud Healthcare API, *Google Cloud*. https://cloud.google.com/healthcare-api/docs/concepts/dicom (accessed January 30, 2024).

[13] Cloud Compliance - Regulations & Certifications, *Google Cloud*. https://cloud.google.com/security/compliance/offerings (accessed March 15, 2024).

[14] Open Health Imaging Foundation, https://ohif.org/ (accessed January 30, 2024).

[15] K. Sall, G.N. Foulks, A.D. Pucker, K.L. Ice, R.C. Zink, and G. Magrath, Validation of a Modified National Eye Institute Grading Scale for Corneal Fluorescein Staining, *Clin Ophthalmol*. **17** (2023) 757–767. doi:10.2147/OPTH.S398843.

[16] J.S. Wadali, S.P. Sood, R. Kaushish, S. Syed-Abdul, P.K. Khosla, and M. Bhatia, Evaluation of Free, Open-source, Web-based DICOM Viewers for the Indian National Telemedicine Service (eSanjeevani), *J Digit Imaging*. **33** (2020) 1499–1513. doi:10.1007/s10278-020-00368-4.

[17] O. Srivastava, M. Tennant, P. Grewal, U. Rubin, and M. Seamone, Artificial intelligence and machine learning in ophthalmology: A review, *Indian J Ophthalmol*. **71** (2023) 11–17. doi:10.4103/ijo.IJO_1569_22.

[18] D.S.J. Ting, V.H. Foo, L.W.Y. Yang, J.T. Sia, M. Ang, H. Lin, J. Chodosh, J.S. Mehta, and D.S.W. Ting, Artificial intelligence for anterior segment diseases: Emerging applications in ophthalmology, *Br J Ophthalmol*. **105** (2021) 158–168. doi:10.1136/bjophthalmol-2019-315651.

dHealth 2024
D. Hayn et al. (Eds.)

doi:10.3233/SHTI240041

Neural Network-Based Prediction of Perceived Sleep Quality Through Wearable Device Data

Martin BAUMGARTNER[a,b,1], Manuel GRÖSSL[c], Raphaela HAUMER[c], Katharina POIMER[c], Flora PRANTL[c], Katharina WEICK[c], Markus FALGENHAUER[a], Stefan BEYER[a], Andreas ZIEGL[d], Aaron LAUSCHENSKY[a], Fabian WIESMÜLLER[a,b,e], Karl KREINER[a], Dieter HAYN[a,e] and Günter SCHREIER[a,b]

[a] *AIT Austrian Institute of Technology, Graz & Vienna, Austria*
[b] *Institute of Neural Engineering, Graz University of Technology, Graz, Austria*
[c] *University of Applied Sciences Wiener Neustadt, Wiener Neustadt, Austria*
[d] *telbiomed Medizintechnik und IT Service GmbH, Graz, Austria*
[e] *Ludwig Boltzmann Institute for Digital Health and Prevention, Salzburg, Austria*

Abstract. Background: This study focuses on the development of a neural network model to predict perceived sleep quality using data from wearable devices. We collected various physiological metrics from 18 participants over four weeks, including heart rate, physical activity, and both device-measured and self-reported sleep quality. Objectives: The primary objective was to correlate wearable device data with subjective sleep quality perceptions. Methods: Our approach used data processing, feature engineering, and optimizing a Multi-Layer Perceptron classifier. Results: Despite comprehensive data analysis and model experimentation, the predictive accuracy for perceived sleep quality was moderate (59%), highlighting the complexities in accurately quantifying subjective sleep experiences through wearable data. Applying a tolerance of 1 grade (on a scale from 1-5), increased accuracy to 92%. Discussion: More in-depth analysis is required to fully comprehend how wearables and artificial intelligence might assist in understanding sleep behavior.

Keywords. sleep quality, wearable electronic devices, neural network, multi-layer perceptrons, electronic data processing

1. Introduction

Wearable devices represent a significant advancement in personal health monitoring, offering a convenient way to track various physiological parameters. These devices, which range from smartwatches to fitness trackers, are designed to measure and record health-related data such as heart rate, steps taken, calories burned, and even sleep patterns. In particular, the ability of these devices to assess sleep quality (SQ) is promising,

[1] Corresponding Author: Martin Baumgartner, AIT Austrian Institute of Technology, Vienna, Austria, E-Mail: martin.baumgartner@ait.ac.at

considering the vital importance of sleep in maintaining overall health and well-being [1] .

SQ, an essential determinant of physical and mental health, is influenced by various factors, including duration, continuity, and the composition of sleep stages [2]. Sleep is typically divided into three stages: light sleep, deep sleep, and rapid eye movement (REM) sleep. Each stage plays a unique role in the sleep cycle, contributing to the restorative processes of the body and mind [3]. The direct relationship between physiological measures and SQ is challenging to investigate since no definitive markers for objective SQ have been formulated. SQ is determined by a number of objective (e.g. duration) and subjective (e.g. restfulness) criteria [4].

Assessing SQ currently relies on polysomnography, a comprehensive sleep study conducted in specialized clinics with application of a multitude of measurement devices. This gold standard method includes monitoring brain waves via electroencephalogram, blood oxygen levels, heart rate, breathing, and eye and leg movements [2]. The application of this extensive use of equipment is often disturbing the SQ [5]. Further, sleeping in unfamiliar environments (i.e. sleep clinic) is associated with disturbed sleep especially in the first night (e.g. first-night effect [6]). However, wearable devices offer a more accessible and less intrusive way to track sleep patterns in comfortable environment, though they may not capture data with the same level of detail as polysomnography [7]. A study from 2019 demonstrated that – compared to polysomnography – wearable devices (Fitbit Alta HR and Philips Respironics Actiwatch 2) underestimated total sleep time by up to 47 minutes (Fitbit) and 38 minutes (AW2) and overestimated wake after sleep onset by up to 42 minutes. The sensitivity for detecting sleep was over 90% for both devices. Despite certain limitations in measurement accuracy, wearable devices have shown to be a viable alternative to polysomnography [8].

This leads to the research question, which vital parameters are ultimately needed to assess SQ. The purpose of this study was to explore the relationship between various factors affecting SQ, particularly focusing on the correlation between physiological measures captured by wearable devices and the subjective experience of SQ. By exploring these dynamics, this paper aspires to meaningfully contribute to the growing field of personal health monitoring, focused on its impact on sleep-related health results.

2. Methods

The approach encompassed data collection from a diverse group of participants, data processing procedures, and the early development of a neural network model to predict perceived SQ.

2.1. Subject recruitment & data collection

A total of 18 participants were recruited for the study, evenly distributed between the sexes (nine men and nine women) and ranging in age from 22 to 41 years. These participants consented to wear a variety of smartwatches, including models from Polar, FitBit, and Apple Watch, specifically for sleep tracking. The data collection phase spanned four weeks during the month of October 2023.

Participants were granted access to a dedicated online platform for data management. This platform enabled participants to record, visualize and review their sleep-related data and observe trends. Notably, in addition to manual data input at this platform, data from the smartwatches could be transmitted automatically through a dedicated mobile application that synchronized various accounts (e.g., FitBit, Polar). This flexibility allowed participants to record data with minimal effort but allowed for retrospective corrections.

Each participant recorded six data metrics daily: *Resting Heart Rate*, *Step Count*, *Physical Activity*, *Well-being*, *Device-measured SQ*, and *Perceived SQ*. Further granularity was provided for the *Physical Activity* metric by including details regarding the type of activity, a self-reported estimate of its intensity, and its duration in minutes. Well-being was evaluated through self-reported assessments, allowing participants to categorize their daily state of general well-being as "*good"*, "*average"* or "*poor". Device-measured SQ* was recorded as the devices' manufacturers provided it (typically a sleep score ranging between 0 and 100). Additionally, the *Perceived SQ* was collected using a grading system where 1 represents "*excellent sleep"* and 5 represents "*poor sleep"* according to subjective feeling of sleep restfulness.

2.2. Data processing

After the data collection phase, the acquired data was exported from the platform, imported into a Python programming environment, and processed based on data type. Preliminary statistical analyses were conducted to gain insights into the data and guided the subsequent steps of choice of model type and training strategy. The following steps were taken to develop a predictive model to estimate SQ:

Data preparation and exploration: We loaded the dataset from a CSV file and performed initial data cleaning by removing non-relevant or non-numeric columns. Descriptive statistics were computed to gain insights into the data's characteristics. Additionally, we visualized feature distributions and the distribution of the target variable (*Perceived SQ)*. The dataset occasionally included duplicated or multiple values (e.g. due to input or transmission errors). We applied different deduplication strategies according to data type. If multiple entries for heart rate or well-being data existed, the mean value was used. In case of multiple step counts, the sum was calculated if the counts were different (duplicated, identical counts of steps per day were considered erroneous and thus removed). If multiple physical activities were recorded, durations were summed up, the type was categorized as "*mixed*", and the highest reported intensity was chosen. If multiple recordings of SQ were present, the last one was selected (as it indicated a corrective input). If no value was available for a specific parameter, it was set to zero. All data was summarized on a daily level (referred to as "daily reports" from here on). These daily reports were included only if *Device-measured SQ* was non-zero and *Perceived SQ* fell within the range of 1 to 5.

Correlation analysis: A correlation matrix was computed to assess the relationships between different features in the dataset. This analysis helped identify feature dependencies and potential multicollinearity.

Data splitting: The data was split into training and testing sets, allocating 20% of the data for testing to ensure a robust evaluation of the model's performance. Standard scaling was applied to each feature to ensure consistent feature scaling.

Model experiments: To optimize our model, we conducted a series of experiments using a Multi-Layer Perceptron (MLP) classifier. We explored various aspects, including

different architectures, activation functions, learning rates, maximum iterations, and a range of alpha values (0.0001, 0.001, 0.01) to fine-tune regularization. Additionally, we experimented with feature transformation, data augmentation ("*digital twins*" with small random variation within 1 standard deviation of feature value), and ensemble methods to enhance model performance.

Feature engineering: In our feature engineering process, we introduced specific types of new features. We developed combined features like "*steps_to_hr_ratio*", where we merged steps and heart rate to understand the interplay between physical activity and cardiovascular response. Additionally, we employed aggregated features such as rolling averages over periods of 3 or 7 days to discern long-term trends from daily fluctuations. Furthermore, we explored feature interactions, exemplified by "*activity_product*", which multiplies activity duration and intensity to provide a holistic view of physical activity.

3. Results

3.1. Descriptive statistics and visualizations

Comprehensive descriptive statistics and visual analyses of data on measured and perceived SQ were conducted. Distribution of features and target variable (*Perceived SQ*) was depicted using histograms and graphical representations, followed by a detailed correlation matrix to visualize relationships among the different variables. As the original ascending scale of both *Perceived SQ* and *Well-being* naturally leads to negative correlation values with *Device-measured SQ*, they were inverted (e.g. 1 equals "*poor sleep*") to show positive correlation, which is more intuitive.

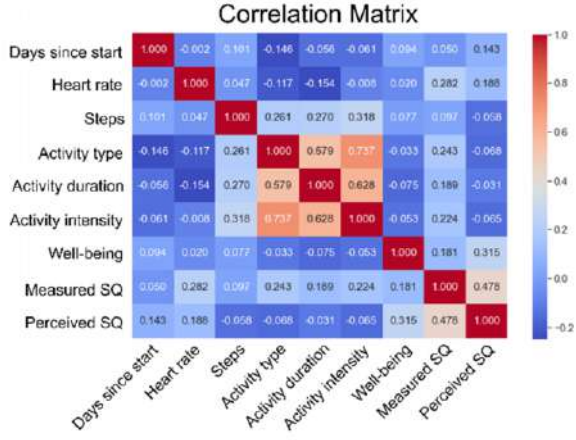


Figure 1. This matrix represents the Spearman correlations between nine variables. Perceived SQ and well-being values were inverted to display more intuitive positive correlation values.

Figure 1 shows the correlations between nine variables, of which the first six served as features for the model: days since start, heart rate, steps, activity type, activity duration, activity intensity, well-being, device-measured SQ and perceived SQ. In our case, the *Device-measured SQ* and *Perceived SQ* did not show strong correlations ($r > 0.7$) with any of the other values. *Device-measured* and *Perceived SQ* showed moderate correlation at 0.478. The *Device-measured SQ* exhibited a maximum positive correlation

with *Heart Rate* at 0.282. The *Perceived SQ* showed the highest positive correlation with *Well-being* at 0.315.

3.2. Modeling approach

Our modeling approaches encompassed a variety of experimental configurations and methods. We tested various architectures of MLPs, with an architecture of three layers (100, 50, and 25 neurons in respective layers) – achieving the highest accuracy (60%), while others like (150, 100, 50) and (200, 150, 100) reached 51% and 54%, respectively. Various activation functions (e.g., ReLU, tanh, logistic) were tested, with the logistic function, performing the best with an accuracy of 59%.

Alpha values and learning rates were tuned to control overfitting and enhance learning speed. Maximum iteration variations showcased sensitivity, where 500, 1000, and 1500 iterations achieved accuracies of 47%, 47%, and 44%, respectively. Feature transformation through PolynomialFeatures (degree 2) led to a reduced accuracy of 40%. Data augmentation with noise addition resulted in an accuracy of 46%. Ensemble methods, including VotingClassifier, achieved an accuracy of 46%. Hyperparameter tuning, model stacking, and ensemble approaches with RandomForest, GradientBoosting, SVM, and MLP were explored, with the VotingClassifier ensemble reaching the highest accuracy of 53%. Comprehensive feature engineering, encompassing additional, combined, and interaction features, along with RFECV for feature selection, contributed to a thorough evaluation. Visualization through bar charts provided a comprehensive overview of experiment and feature test accuracies, offering valuable insights into the model and feature dynamics. Furthermore, ensemble methods, including Stacking and Voting Classifiers, were employed to enhance prediction accuracy. An ensemble approach, combining MLP and RandomForest achieved an accuracy of 53%. For a more precise model configuration, hyperparameter tuning using scikit-learn modules GridSearchCV and RandomizedSearchCV was performed, integrating the best MLP model into ensemble methods.

As an additional experiment, we evaluated the best performing model with a tolerance of ±1 grade of perceived SQ. This increased accuracy significantly from 59% to 92%, which is visualized in Figure 2.

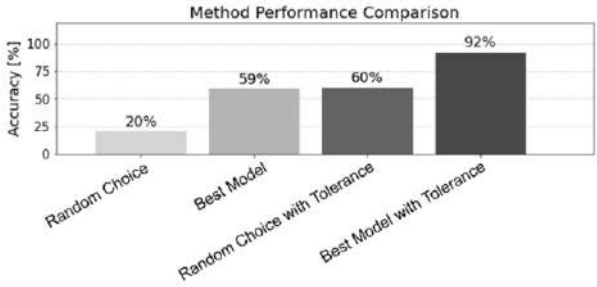


Figure 2. Comparison of different classification methods. Allowing a tolerance of ±1 grade increases accuracy significantly. Random choice is a simulated base line performance (1 out of 5 random selection).

4. Discussion

The purpose of this study was to find out 1) whether SQ is influenced by one or more specific vital parameters or activities and 2) whether the measured SQ corresponds to the perceived SQ. As the presented results clearly show, no strong correlation between measured data and SQ was identified. It is also open to question if more accurate recording devices would have resulted in a stronger correlation. Furthermore, for better comparability, limiting the subjects to a single device type would have stratified the measured SQ. Although the resulting sleep scores were given in already normalized 0-100% in all devices, the underlying algorithm for calculating this score likely differs between manufacturers and thus express different aspects of SQ. Evaluation of the engineered features did not significantly enhance prediction of SQ, they provided valuable insights into the complex nature of the data, contributing to our understanding and formulating future model refinements. These efforts aimed to capture additional information potentially relevant for predicting SQ. While accuracy scores between 40-60% percent were superficially underwhelming, allowing for a 1 grade tolerance in evaluation increased the accuracy dramatically to over 90% (see Figure 2), meaning that while the model did miss fine nuances of sleep, it also did not completely misjudge trends of perceived SQ.

Sleep is not only influenced by the factors already mentioned. Therefore, it is suggested to consider other parameters that influence SQ in future studies. For example, the timing of the measurement of SQ can have a significant impact on the results. Sleep patterns vary during different stages of sleep, making the assessment of SQ dependent on when it is measured. It therefore makes a difference whether one wakes up in a REM sleep phase or in a light sleep phase, for example. According to van Hasselt et al., waking up in the REM sleep phase might report to feel worse than subjects waking up in light sleep phases, although the *Device-measured SQ* is identical in both situations [9]. Furthermore, daylight hours are shorter during the winter which can affect the circadian rhythm and melatonin production. As a result, time of year during measurements has an impact on SQ [10]. This fact could be relevant for our analyses because the data was recorded in October and the nights in this month are longer than the days. Other parameters also vary over the course of the day. For example, during illnesses, perceived well-being can be lower at later hours of the day. The participants of this analysis recorded their well-being during earlier hours of the day and time of data collection was not controlled between subjects. This potentially led to suboptimal quality of physiological data like the resting heart rate that changes during the day. Illness can further affect the quality of sleep as pain, discomfort or medication can irritate sleep. For this reason, it is important to consider the health status of the participants when assessing SQ [11]. For future experiments, a more detailed self-assessed status of both physical as well as mental well-being could be helpful.

Sports watches are increasingly frequently used to measure SQ. However, their reliability and accuracy are debatable. Especially watches designed for wellness or sport applications are not certified as medical devices and thus their reliability is not guaranteed. Our findings of low correlation between measured and perceived SQ also support the assumption that their capabilities to accurately measure sleep are limited. Therefore, more research could be helpful to understand the capabilities of different technologies [12].

In conclusion, the analysis of SQ requires the consideration of various influencing factors over a longer period on more participants. In addition, the choice of equipment must carefully be considered to obtain accurate results.

References

[1] S. M. A. Iqbal et al., Advances in healthcare wearable devices, *NPJ Flexible Electronics* **5**(1) (2021), 9

[2] A. D. Krystal and J. D. Edinger, Measuring sleep quality, *Sleep Medicine* **9** (2008), S10–S17

[3] A. S. Tubbs et al., The basics of sleep physiology and behavior, in *Sleep and health*, Elsevier, (2019), 3–10.

[4] S. J. McCarter et al., Physiological markers of sleep quality: A scoping review, *Sleep Medicine Reviews* **64** (2022), 101657

[5] T. Blackwell et al., A novel approach using actigraphy to quantify the level of disruption of sleep by in-home polysomnography: the MrOS Sleep Study: Sleep disruption by polysomnography, *Sleep Medicine* **32** (2017), 97–104

[6] L. Ding et al., A meta-analysis of the first-night effect in healthy individuals for the full age spectrum, *Sleep Medicine* **89** (2022), 159–165

[7] A. J. Boe et al., Automating sleep stage classification using wireless, wearable sensors, *NPJ digital medicine* **2**(1) (2019), 131

[8] X. K. Lee et al., Validation of a consumer sleep wearable device with actigraphy and polysomnography in adolescents across sleep opportunity manipulations, *Journal of Clinical Sleep Medicine* **15**(9) (2019), 1337–1346

[9] L. W. A. Hermans et al., Representations of temporal sleep dynamics: Review and synthesis of the literature, *Sleep Medicine Reviews* **63** (2022), 101611

[10] S. J. van Hasselt et al., Seasonal variation in sleep time: jackdaws sleep when it is dark, but do they really need it?, *Journal of Comparative Physiology B* (2023), 1–11

[11] M. Kemple et al., Sleep quality in patients with chronic illness, *Journal of clinical nursing* **25**(21–22) (2016), 3363–3372

[12] E. Guillodo et al., Clinical applications of mobile health wearable–based sleep monitoring: systematic review, *JMIR mHealth and uHealth* **8**(4) (2020), e10733

dHealth 2024
D. Hayn et al. (Eds.)

doi:10.3233/SHTI240042

The Evolution of Telehealth in Heart Failure Management: The Role of Large Language Models and HerzMobil as a Potential Use Case

Hillary FARMER[a,b], Karl KREINER[b], Thomas SCHÜTZ[c], Gerhard PÖLZL[c], Christian PUELACHER[c] and Günter SCHREIER[a,b]

[a] *Institute of Neural Engineering, Graz University of Technology, Graz, Austria*
[b] *AIT Austrian Institute of Technology, Graz & Vienna, Austria*
[c] *Department of Internal Medicine III, Cardiology and Angiology, Medical University of Innsbruck, Innsbruck, Austria*

Abstract. The burgeoning domain of telehealth has witnessed substantial transformation through the advent of advanced technologies such as Large Language Models (LLMs). This study examines the integration of LLMs in heart failure management, with a focus on HerzMobil as a pioneering telehealth program. The technical underpinnings of LLMs, their current applications in the medical field, and their potential to enhance telehealth services, have been explored. The paper highlights the benefits of LLMs in patient interaction, clinical documentation, and decision-making processes. Through the HerzMobil case study, improvements in patient self-management and reductions in hospital readmission rates have been observed, showcasing the successful application of telehealth in chronic disease management. The paper also delves into the challenges and ethical considerations of LLM integration, such as data privacy, potential biases, and regulatory compliance, underscoring the need for a balanced approach that prioritizes patient safety and ethical standards.

Keywords. Telehealth, Natural Language Processing, Large Language Models

1. Introduction

The rapid evolution of telehealth, especially in the context of heart failure management, marks a significant stride in modern healthcare. The COVID-19 pandemic's push for remote care has notably advanced heart failure management, with LLMs enhancing patient education and understanding of their condition through continuous monitoring. [1] This paper focuses on the integration of advanced technologies like Large Language Models (LLMs) into heart failure telehealth, using HerzMobil [2] as a potential use case. It explores how tools like OpenAI's "Generative Pre-trained Transformer" (GPT) series and Alphabet's "Pathway Language Model 2" (PaLM2), known for their ability to emulate human-like text, can redefine patient-provider interactions specifically in heart failure care. This paper discusses how LLMs could transform heart failure telehealth, considering the promise of improved care against issues of privacy, bias, and compliance.

It examines LLMs' applications, their benefits for patient outcomes, and the unique challenges within heart failure management.[3]

2. Methods

This study was conducted through a comprehensive literature review, focusing on the recent advancements and applications of LLMs in telehealth, particularly for heart failure management. The analysis also included a detailed examination of the HerzMobil program as a case study to explore the practical implications of these technologies in a real-world setting. The report provides insights into the technical capabilities of LLMs, their practical use in the HerzMobil program, and the broader implications for the healthcare sector.

3. Results

3.1. LLM Technical Background & Current Applications

LLMs, such as GPT-3 and GPT-4, represent a significant leap in the field of artificial intelligence, particularly in natural language processing (NLP). At their core, LLMs are trained using a technique known as unsupervised learning, where they are fed vast amounts of text data. This training enables them to predict and generate text sequences, making them highly adept at understanding and generating human-like text. [4]

In the healthcare sector, LLM's are increasingly being leveraged for various medical applications. LLMs assist in synthesizing and analyzing vast amounts of medical literature, aiding researchers in keeping up with the latest studies and findings. Furthermore, LLMs offer considerable benefits to patients by dedicating time and patience to deliver adequate information, thereby empowering them to make more informed decisions regarding their healthcare. [5] "The growing mountain of required medical paperwork, formfilling, reporting, claims, orders, and so on creates so much friction, error and burnout. GPT-4 gives us hope that some of this can be reduced … [6]."

3.2. Telehealth & Virtual Care

Telehealth's role has significantly evolved, transitioning from a solution for remote areas to offering a broad spectrum of services, including virtual consultations and remote patient monitoring. This evolution accelerated with the COVID-19 pandemic, leading to widespread adoption in healthcare. The pandemic, with its lockdowns and distancing measures, along with relaxed regulations and technological progress, transformed telehealth into a key component of healthcare delivery. [1] Today, telehealth is valued for enhancing healthcare access, improving care quality, increasing patient satisfaction, and offering cost-effective solutions for chronic conditions and regular care needs. [5]

While telehealth has shown significant growth, it faces several challenges that must be overcome to maximize its effectiveness. The digital divide poses a significant barrier, as not all patients have equal access to the required technology and internet connectivity. This gap can lead to disparities in how patients engage with and adhere to their heart failure treatment plans via telehealth. Furthermore, ensuring a high standard of care

through accurate diagnosis and effective communication in a virtual setting is crucial [7]. Managing the large volume of data can benefit from AI support, with LLMs playing a distinct role in processing patient-provider interactions. For instance, LLMs can transcribe medical consultations via speech-to-text capabilities and then apply NLP to analyze the transcriptions, extracting valuable insights and aiding in the follow-up care process. Additionally, maintaining patient privacy and complying with stringent regulations such as the Health Insurance Portability and Accountability Act (HIPAA) [8] and the General Data Protection Regulation (GDPR) [9] introduces significant complexity.

3.3. Integration of LLMs in Heart Failure Telehealth

The use of LLMs in telehealth has the potential to greatly improve patient interaction. They enhance patient triage by efficiently analyzing symptom descriptions to prioritize care based on severity. They also elevate symptom checking by identifying possible conditions and asking follow-up questions for a thorough assessment. [10] Additionally, LLMs are useful in providing easily understandable health information, answering patient queries about diseases, treatments, and preventative measures, which is especially beneficial in telehealth where direct access to healthcare providers might be limited. [3]

To integrate LLMs into everyday clinical practice effectively, it is essential to focus on user-friendly interfaces that accommodate the diverse needs of patients, including those less tech-savvy or with varying levels of health literacy. Implementing voice recognition and natural language understanding capabilities can further enhance interaction, allowing patients to express their concerns in natural language and receive personalized, conversational responses.

LLMs streamline clinical documentation by automating the processing and summarization of medical records and clinical notes. This reduces the time healthcare professionals spend on documentation and minimizes human error. LLMs efficiently extract key information, such as prescription details and medical history, from these documents. When integrated with electronic health record systems, they provide healthcare providers quick access to essential patient information [11]. This facilitates more informed decision-making and enhances patient care. For the successful implementation of LLMs in clinical decision-making, it is crucial to ensure the models are trained on diverse and extensive datasets to accurately reflect the varied patient population they serve. This training helps in identifying patterns and insights from large datasets, contributing to the development of personalized treatment plans. Continuous monitoring and updating of these models are necessary to maintain their accuracy and relevance in the rapidly evolving medical field. This promising approach may enhance patient satisfaction and encourage adherence to treatments, although documented successes in this specific application are still emerging.

3.4. Potential Use Case: HerzMobil Program

The HerzMobil program is an innovative telehealth initiative designed for the management of heart failure patients. Its primary objective is to enhance patient care post-hospital discharge, reduce the frequency of hospital readmissions, and improve overall patient quality of life. The program employs a comprehensive approach, combining regular monitoring through mobile technology with guidance from healthcare professionals. Patients are provided with devices to monitor vital parameters like heart

rate and blood pressure, and the data is regularly reviewed by a team of healthcare providers, ensuring timely interventions and adjustments to treatment plans as necessary. [12] HerzMobil has shown notable successes, particularly in improving patient self-management and reducing hospital readmission as well as mortality rates [13]. The program's ability to provide real-time monitoring and feedback has been instrumental in these achievements. However, the scalability of the program and integration into broader healthcare systems presents logistical and financial challenges that need to be addressed for wider implementation. [2] [12]

Integrating Large Language Models into programs like HerzMobil could significantly enhance their effectiveness but require a nuanced understanding of the technical framework and potential challenges. LLMs could be used to analyze patient data more deeply, identifying subtle trends or warning signs that might not be immediately obvious, such as increased interaction with the system, questions asked by patients, or speech pattern analysis. [14] They could also assist in providing personalized health advice and answering patient queries in real-time, improving patient engagement and adherence to treatment plans. Moreover, LLMs could automate parts of the data analysis and reporting process, increasing the efficiency of healthcare providers and allowing them to focus more on direct patient care. The integration of LLMs into HerzMobil has the potential to make the program more responsive, personalized, and scalable, ultimately leading to better outcomes for heart failure patients by interacting in the patient's native language. Addressing the integration challenges involves examining the interoperability between LLMs and existing healthcare IT infrastructures, data privacy and security issues, and the adaptability of these models to the specific needs of HerzMobil's patient demographic. Solutions such as developing secure data pipelines, ensuring compliance with healthcare regulations, and customizing LLMs to address the linguistic and cultural diversity of patients will be critical. An in-depth analysis of how LLM integration could specifically benefit the HerzMobil program is essential. This could include improving real-time patient monitoring, enhancing personalized care through advanced data analysis, and automating routine administrative tasks to allow healthcare professionals to focus more on patient care.

3.5. Ethical Considerations and Challenges

In the integration of LLMs into telehealth, data privacy emerges as a paramount concern. The healthcare sector deals with sensitive personal information, and the incorporation of LLMs, which require extensive data for training and functioning, poses significant privacy and security risks. There is a critical need to ensure that patient data used in these models is anonymized and secure from breaches. [15] Furthermore, the transmission of data to locations with substantial LLM operations, particularly the United States, raises specific concerns under data protection regulations like GDPR, necessitating stringent measures to mitigate potential risks.

Another significant challenge is the potential for bias and the critical need for accuracy in medical recommendations, as well as the rapid evolution of models, which contrasts with the lengthy process of scientific evaluation. LLMs are trained on existing datasets, which may contain inherent biases, leading to skewed or unfair outcomes in patient care. For example, the model's effectiveness can be significantly limited by language barriers, as training data often lacks diversity in patient demographics and language. While the recommendations might be accurate in English, the amount of data

in other languages may be insufficient, potentially placing non-native speakers at a disadvantage due to a poorer quality of data for training purposes. [6]

The regulatory landscape for LLMs in healthcare presents its own set of challenges. The rapid pace of AI development often outstrips the speed at which regulations can be formulated and implemented. Current healthcare regulations may not adequately cover the nuances of AI and LLMs, leading to a grey area in terms of compliance and liability.[16]Collaboration between AI developers, healthcare professionals, legal experts, and regulatory bodies is essential to develop a framework that balances innovation with patient safety and ethical considerations. [15]

Since the common heart failure collective is over 60 years old and frailty of the patients increases with age, interaction with the LLMs by typing on displays or keyboards could be cumbersome. Input by voice control or dictation could overcome the limitation of typing, but presents challenges on its own with errors caused by stuttering or rephrasing of questions. These limitations may cause frustration on the patient's side and limit the usage of LLMs. Furthermore, reading of answers provided by LLMs on small screens such as mobile phones may be tedious, but bigger screens like tablet or desktop computers may be less available for socioeconomic reasons. [17]

4. Discussion

This paper has examined the evolving landscape of telehealth in heart failure management, with a specific focus on the integration of LLMs and the practical application within the HerzMobil program. The study reveals significant advancements and potential benefits brought by LLMs in enhancing patient care, particularly in terms of patient interaction, clinical documentation, and decision-making. The implementation of LLMs in telehealth has shown promise in improving patient-provider communication, making symptom assessment more accurate, and facilitating personalized healthcare. The HerzMobil program exemplifies the successful application of telehealth in managing heart failure, demonstrating improvements in patient self-management and reduced hospital readmission rates. However, the scalability of the program and integration into broader healthcare systems presents logistical and financial challenges that need to be addressed for wider implementation. [12] The potential integration of LLMs in programs like HerzMobil could further enhance these telehealth initiatives. LLMs' ability to process and analyze large volumes of data can lead to more personalized and efficient patient care. However, this integration is not without challenges. Data privacy concerns, potential biases in AI models, and regulatory hurdles present significant obstacles that need careful consideration.

In conclusion, the integration of LLMs in heart failure telehealth represents a significant advancement in healthcare technology. While challenges exist, the potential benefits in improving patient outcomes and healthcare efficiency are considerable. As we move forward, the initiation of the first experiments on a locally installed LLM in collaboration with clinical partners is a pivotal step. These initial experiments are primarily focused on ensuring the safety of patient inquiries. This marks the beginning of a new era in telehealth, where the emphasis is on safely integrating advanced technologies like LLMs into patient care.

References

[1] E. Monaghesh, A. Hajizadeh. The role of telehealth during COVID-19 outbreak: a systematic review based on current evidence. BMC Public Health 20. 2020.
[2] N. Bauer, S. Pötz. Steiermärkische Krankenanstaltengesellschaft m.b.H. https://www.kages.at/patienten-angehoerige/herzmobil-steiermark. last access: 26.01.2024.
[3] T. Greenhalgh, C. A'Court, S. Shaw. Understanding heart failure; explaining telehealth – a hermeneutic systematic review. BMC Cardiovascular Disorders 17. 2017.
[4] C. Peng, X. Yang, A. Chen. A study of generative large language model for medical research and healthcare, npj Digital Medicine 6. 2023.
[5] E. Topol. Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again. Basic Books, New York. 2019.
[6] P. Lee, C. Goldberg. The AI Revolution in Medicine: GPT-4 and Beyond. Pearson, 1st edition. 2023.
[7] J. Wang, S. Yu. Prompt Engineering for Healthcare: Methodologies and Applications. https://www.researchgate.net/publication/370417875_Prompt_Engineering_for_Healthcare_Methodologies_and_Applications. last access: 25.01.2024.
[8] T. Greenhalgh. Understanding heart failure; explaining telehealth – a hermeneutic systematic review. BMC Cardiovascular Disorders 17. 2017.
[9] U. Department of Health and Human Services. U.S. Department of Health and Human Services. https://www.hhs.gov/hipaa/index.html. last access: 22.01.2024.
[10] GDPR.EU. https://gdpr.eu/. last access: 23.01.2024.
[11] A. Bohr, K. Memarzadeh. The rise of artificial intelligence in healthcare applications, in: Artificial Intelligence in Healthcare. 2020.
[12] S. K. m. (KAGes). HerzMobil Steiermark Online Betreuung für Patientinnen/Patienten mit Herzinsuffizienz (Herzschwäche).
[13] G. Poelzl, T. Egelseer-Bruendl and B. Pfeifer. Feasibility and effectiveness of a multidimensional post-discharge disease management programme for heart failure patients in clinical practice: the HerzMobil Tirol programme. Clin Res Cardiol. 2022.
[14] Machteld J Boonstra, Davy Weissenbacher et al. Artificial intelligence: revolutionizing cardiology with large language models. *European Heart Journal*. 2024
[15] B. Mesko, E. Topol. The imperative for regulatory oversight of large language models (or generative AI) in healthcare, npj Digital Medicine volume 6. 2023.
[16] E. Sezgin. Redefining Virtual Assistants in Health Care: The Future with Large Language Models, Journal of Medical Internet Research 26. 2024.
[17] A. Mebazaa, B. Davison, O. Chioncel et al. Safety, tolerability and efficacy of up-titration of guideline-directed medical therapies for acute heart failure (STRONG-HF): a multinational, open-label, randomised, trial. Lancet. 2022.

dHealth 2024
D. Hayn et al. (Eds.)

doi:10.3233/SHTI240043

Agile eHealth Usability Evaluation: A Triangulative Approach to Promoting the Usability of eHealth Systems

Irina SINABELL[a,1] and Elske AMMENWERTH[a]

[a] *UMIT TIROL – Private University for Health Sciences and Health Technology, Hall in Tirol, Austria*

Abstract. Background: Usability evaluation is difficult to reconcile with agile software development for eHealth systems, because traditional usability evaluation is often complex and cumbersome to implement. However, obtaining prospective users' feedback during agile software development is crucial for improving the usability of eHealth systems, which is why there is an increasing need for agile eHealth usability evaluation. Objective: This study investigates whether agile usability evaluations are suitable to evaluate patient-centered eHealth systems being agile developed in health care and are applicable for prospective users, such as older persons suffering from age-related declines. Methods: A triangulation study was conducted combining iterative expert interviews with an exploratory case study. Results: The triangulation study revealed that the implementation of an agile eHealth usability evaluation with prospective users such as older persons proved to be possible. Conclusion: Established eHealth usability evaluation methods must be further evolved to address age-related impairments of older persons.

Keywords. telemedicine, user-centered design, medical informatics applications, consumer health informatics

1. Introduction

Information and communication technologies (ICT) are developing rapidly [1] as well as new health care innovations, particularly in the field of eHealth [2]. eHealth systems that promote a healthy lifestyle could improve the ability to recognize, monitor, and manage the own health condition [3]. ICT, such as eHealth systems, can thus contribute to improving the quality of life [4]. Besides the rapid development of eHealth systems, the population is ageing, leading to a digital divide between older and younger people in terms of the use of ICT [5]. eHealth systems can help older people to live independently [6], but rapid technological progress is leading to an ever-increasing variety of eHealth systems. Older people are willing to embrace new ICT [4] but sometimes have limited IT knowledge. Ensuring the quality of life of older people has become a priority of modern society [7], which demonstrates the need to improve the usability of eHealth systems.

[1] Corresponding Author: Irina Sinabell, UMIT TIROL – Private University for Health Sciences and Health Technology, Hall in Tirol, Austria, E-Mail: irina.sinabell@edu.umit-tirol.at

Expert- or user-based usability evaluation methods can be utilized for a usability evaluation of eHealth systems. During expert-based usability evaluation methods, the experts put themselves in the role of prospective users [8]. In user-based usability evaluation methods, prospective users are involved in the usability evaluation, which is usually carried out through usability tests in a usability laboratory [9]. eHealth systems are being developed using agile software development to incorporate user feedback obtained from usability evaluation early. However, traditional expert- or user-based evaluation methods are difficult to reconcile with agile software development [10], because they require a lot of time and resources [11].

Agile development requires the application of usability evaluation methods that are easy to implement. In order to keep up with the rapid development of eHealth systems and to obtain prospective users' feedback during agile software development, which is crucial for improving the usability of eHealth systems, there is an increasing need for agile eHealth usability evaluation.

2. Methods

In this study, a triangulation approach was chosen because several research methods were applied that are subject to qualitative research [12].

Figure 1 shows the chosen triangulative research design. Overall, the triangulation study comprised of step 1 to step 6 including two systematic literature reviews, phone-based as well as paper-based expert interviews, a formalization of the eHealth usability evaluation methods, and an explorative case study. In research phase 1, a systematic review was conducted complementary to the expert-based validation of eHealth usability evaluation methods. Research phase 2 encompassed the formalization of the eHealth usability evaluation methods. Research phase 3 comprised the context-based selection of appropriate eHealth usability evaluation methods and an explorative case study. In research phase 4 a second systematic literature review was carried out (in addition to the case study) to strengthen the results triangulatively.

Figure 1. Triangulative research design: research methods applied (presented in steps 1 to 6), in relation to the research phases (phases 1 to 4). The present study is based on three publications that deal with the research design of research phase 1 [13,14], research phase 2 [14], and research phases 3 as well as 4 [15].

2.1. Research phase 1

Research phase 1 comprised the systematically identification and expert-validation of rapidly deployable eHealth usability evaluation methods to support agile eHealth usability evaluations.

Step 1 - Systematic literature review I: This review considered 3,981 peer-reviewed and non-peer-reviewed articles [13]. These articles report on eHealth usability evaluation methods, the rapid deployment of eHealth usability evaluations, and the relationship with the applicability of eHealth usability evaluation methods. Following a systematic process, in the course of which the abstracts and then the full texts were analyzed, a list of 29 eHealth usability evaluation methods was extracted that were potentially useful to conduct an agile eHealth usability evaluation.

Step 2a - Expert-based validation: The second step involved the iterative expert-validation of the extracted list of 29 eHealth usability evaluation methods. These interviews aimed to prioritize the identified eHealth usability evaluation methods into recommended, potentially useful, or not recommended for use in agile eHealth usability evaluations [13]. These interviews included usability experts with at least 10 years of professional experience in field of user research. A total of 10 experts were interviewed. The transcribed text material originating from the expert-based validation was analyzed using inductive and deductive content analysis.

Step 2b - Developing a checklist: To strengthen the expert recommendations for conducting an agile eHealth usability evaluation, two additional expert interviews were performed. These additional paper-based expert interviews aimed at obtaining expert opinions on key aspects (e.g., number of test participants, documentation, hardware or software used, time required to conduct an agile eHealth usability evaluation) that may influence the deployment of an agile eHealth usability evaluation [14]. The paper-based interviews were analyzed via inductive content analysis.

2.2. Research phase 2

Research phase 2 comprised the development of a toolbox for agile eHealth usability evaluation.

Step 3 - Developing a toolbox: Based on step 1 and step 2, all identified eHealth usability evaluation methods were formalized by creating method cards [14]. The method cards include the name of the eHealth usability evaluation method, the number of the method card, a method description, any notes on related method cards, if applicable, and references. The described eHealth usability evaluation methods are complemented by information on strengths, weaknesses, similarities, as well as coherence to other eHealth usability evaluation methods presented in the toolbox.

2.3. Research phase 3

Research phase 3 involved investigating the applicability of eHealth usability evaluation methods to older persons and the feasibility of conducting an agile eHealth usability evaluation with older persons.

Step 4 - Context-based selection: Three appropriate eHealth usability evaluation methods were selected to implement the explorative case study: (1) co-discovery learning, (2) cooperative usability testing, and (3) remote user testing combined with think aloud. These eHealth usability evaluation methods were chosen from the previously developed toolbox for eHealth usability evaluations (step 3) [14].

Step 5 - Explorative case study: The aim of the explorative case study was to examine whether an agile eHealth usability evaluation can be conducted with older persons (n=7) who met the inclusion criteria. The inclusion criteria comprised criteria such as an age of at least 60 years, a basic computer literacy, and an e-mail account [15]. The older participants were selected regardless of their socioeconomic status. The case study was conducted using a web-based eHealth system that allows patients to retrieve diagnostics reports in a decentralized manner. Qualitative notes from the agile eHealth usability evaluation were systematically analyzed and the number of usability issues was systematically counted.

2.4. Research phase 4

Research phase 4 comprised the identification, dealing with, and how to overcome challenges that might be countered prior to, during, and after carrying out an eHealth usability evaluation with older persons.

Step 6 - Systematic literature review II: The review considered 300 scientific papers that report on the use of eHealth usability evaluation methods with older persons [15]. The titles and abstracts of the papers were systematically reviewed. Twenty papers were identified that met the inclusion criteria. Three additional papers were identified through hand search that reported especially on challenges of evaluating the usability of eHealth among older persons.

3. Results

3.1. Research phase 1

Step 1 - Systematic literature review on eHealth usability evaluation methods: Following the systematic literature review, a list of 29 eHealth usability evaluation methods consisting of eHealth usability evaluation methods was identified [13]. Most of the identified eHealth usability evaluation methods represent established eHealth usability evaluation methods that were used in the literature to accomplish rapidly deployable eHealth usability evaluations (e.g., expert review or feature inspection).

Step 2a - Expert-based iterative validation of eHealth usability evaluation methods to contrast expert knowledge with findings from literature to prioritize the identified eHealth usability evaluation methods (during step 1): Figure 2 shows that a total of 43 eHealth usability evaluation methods were prioritized into ten recommended eHealth usability evaluation methods, 22 potentially useful eHealth usability evaluation methods, and 11 not recommended eHealth usability evaluation methods. The potentially useful eHealth usability evaluation methods refer to eHealth

usability evaluation methods that were neither recommended nor not recommended by experts. The three most frequently recommended eHealth usability evaluation methods are remote user testing, expert review, and the rapid iterative testing and evaluation method.

Step 2b - Developing of a checklist for conducting an agile, easily applicable, and useful eHealth usability evaluation: Based on the experts' opinion from the supplementary paper-based expert interviews, a checklist for implementing an agile eHealth usability evaluation was formulated. The checklist comprises a total of five categories: data collection, location and number of test sessions, prospective user group, test participants, and test performance [14]. The checklist is intended for use by software developers or medical informaticians when conducting an agile eHealth usability evaluation.

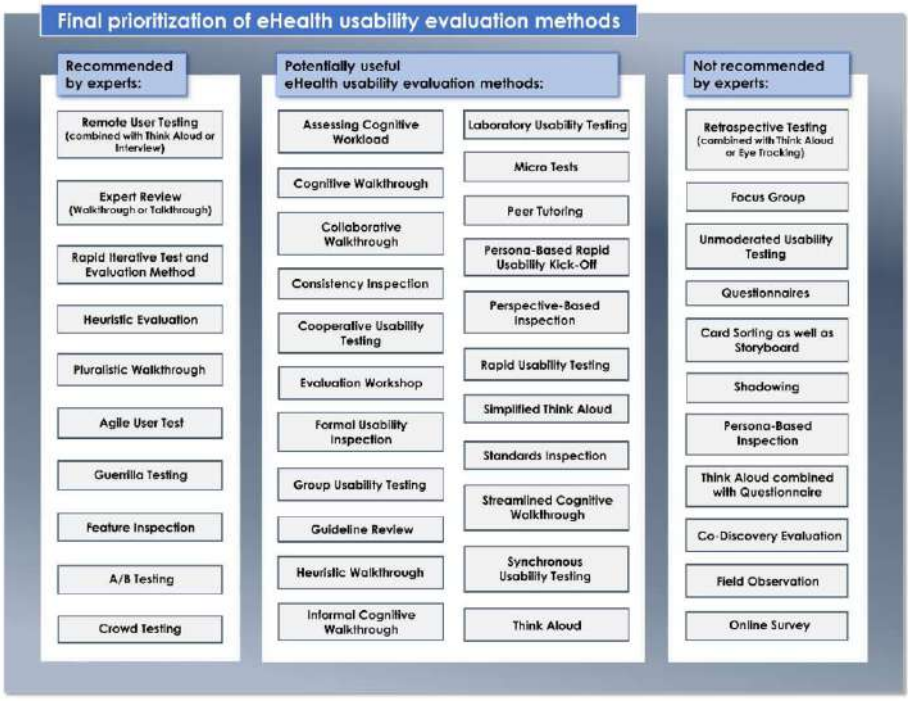


Figure 2. Final prioritization of eHealth usability evaluation methods [13]. Recommended eHealth usability evaluation methods and not recommended eHealth usability methods are ordered by the number of experts' choice. Potentially useful eHealth usability evaluation methods are arranged alphabetically.

3.2. Research phase 2

Step 3 - Based on the findings of step 1 and step 2 a toolbox consisting of rapidly deployable and useful eHealth usability evaluation methods was developed: A toolbox called ToUsE ("Toolbox for eHealth Usability Evaluations") was developed that comprises all 43 identified eHealth usability evaluation methods (step 1 and 2). ToUsE consists of eHealth usability evaluation methods that are useful for implementing agile eHealth usability evaluations. All eHealth usability evaluation

methods are listed on method cards with method descriptions [14]. The toolbox is intended for consideration by software developers, medical informaticians, usability experts, or medical professionals who aim to perform an eHealth usability evaluation.

3.3. Research phase 3

Step 4 - Context-based selection of appropriate eHealth usability evaluation methods to implement the case study: To accomplish the explorative case study three eHealth usability evaluation methods were chosen from ToUsE [15]. Co-discovery evaluation was selected to give older persons the opportunity to complete tasks together. Cooperative usability testing was chosen to jointly reflect with the older persons on their task processing by means of a video recording. Remote user testing combined with think aloud was used in order to accomplish the case study at older persons' homes and virtually involve the software developers of the eHealth system. Tasks were defined based on a fictitious scenario and explained to the older persons prior to the start of the case study.

Step 5 - Examination if the implementation of an agile eHealth usability evaluation is achievable with older persons: Remote user testing combined with think aloud could be successfully applied to evaluate the eHealth system with the older persons. The rationale is that the older persons felt comfortable in their familiar environment and were open-minded about using the new eHealth system. However, not all established eHealth usability evaluation methods are suitable for conducting an agile eHealth usability evaluation with older persons (e.g., cooperative usability testing and co-discovery evaluation) [15]. Cooperative usability testing failed because older persons preferred on completing the tasks alone. Co-discovery evaluation failed as well because the older participants did not agreed to a video recording due to privacy concerns.

3.4. Research phase 4

Step 6 - Systematic literature review on challenges of conducting agile eHealth usability evaluations with older persons and how these challenges can be overcome: From the case study and the systematic literature review, 24 recommendations were derived to address challenges prior to, during, and after the eHealth usability evaluation with older participants [15]. The recommendations were equally derived from the findings of the case study and the systematic review. Most of the recommendations addressed motivation, while the second largest number of recommendations was associated with cognition, and the third largest number of recommendations dealt with physical abilities.

4. Discussion

The results of the triangulation study revealed that the implementation of an agile eHealth usability evaluation in the specific context of the explorative case study and older persons as prospective users is possible. In total, 43 eHealth usability evaluation methods were identified as suitable for agile eHealth usability evaluations. Of these, ten were recommended by the experts based on their usefulness for agile eHealth

usability evaluations. The explorative case study revealed that remote user testing combined with think aloud could successfully be applied to evaluate the eHealth system with older persons. The recommendations for conducting an agile eHealth usability evaluation with older persons highlight the need to consider the challenges of this age group, as eHealth systems should be tested with prospective users who represent the intended users [15,16]. Although weak health conditions often prevent older persons from using eHealth systems [17], the exploratory case study showed that age-related health complaints are not a barrier to use. However, the established eHealth usability evaluation methods must be further evolved to address the specific needs of older persons that may be caused by age-related declines. This study contributed to the development of improved eHealth systems tailored to older users.

References

[1] T. Oikonomou, K. Votis, P. Korn, D. Tzovaras, and S. Likothanasis, *A Standalone Vision Impairments Simulator for Java Swing Applications*, in: Lecture Notes in Computer Science, vol 5889. Springer, Berlin, Heidelberg, 2009. pp. 387-398.

[2] M. Broekhuis, L. van Velsen, and H. Hermens, Assessing usability of eHealth technology: A comparison of usability benchmarking instruments, *International Journal of Medical Informatics* **128** (2019), 24-31.

[3] E. Spanakis, S. Santana, M. Tsiknakis, K. Marias, V. Sakkalis, A. Teixeira, J. H. Janssen, H. de Jong, and C. Tziraki, Technology-Based Innovations to Foster Personalized Healthy Lifestyles and Well-Being: A Targeted Review, *Journal of Medical Internet Research* **18**(6) (2016), e128.

[4] P. Jakkaew, and T. Hongthong, *Requirements Elicitation to Develop Mobile Application for Elderly*, in: Proceedings of the 11th Asia Pacific Conference, ICDAMT 2017, Chiang Mai, Thailand. IEEE, New York, 2017. pp. 464-467.

[5] B. J. Gilbert, E. Goodman, A. Chadda, A. Hatfield, D. Forman, and T. Panch, The Role of Mobile Health in Elderly Populations, *Current Geriatrics Reports* **4** (2015), 347-352.

[6] S. Sülz, H. J. van Elten, M. Askari, A. M. Weggelaar-Jansen, and R. Huijsman, eHealth Applications to Support Independent Living of Older Persons: Scoping Review of Costs and Benefits Identified in Economic Evaluations, *Journal of Medical Internet Research*, **23**(3) (2021)., e24363.

[7] G. Vitiello, and M. Sebillo, *The importance of empowerment goals in elderly-centered interaction design*, in: Proceedings of the International Conference on Advanced Visual Interfaces, AVI 2018, Castiglione della Pescaia, Grosseto, Italy. Association for Computer Machinery, 2018. pp. 1-5.

[8] J. Nielsen, *Usability inspection methods*, in: Proceedings of the Conference Companion on Human Factors in Computing Systems, Association for Computer Machinery, New York, 1994. pp. 413-414.

[9] E. Borycki, A. Kushniruk, C. Nohr, T. Takeda, S. Kuwata, C. Carvalho, M. Bainbridge, and J. Kannry, *Usability Methods for Ensuring Health Information Technology Safety: Evidence-Based Approaches. Contribution of the IMIA Working Group Health Informatics for Patient Safety*, in: Yearbook of Medical Informatics, **8**(1), Thieme, New York, 2013. pp. 20-7.

[10] T. Silva, M. Silveira, and F. Maurer, *Usability Evaluation Practices within Agile Development*, in: Proceedings of the 48th Hawaii International Conference on System Sciences. IEEE, New York, 2015. pp. 5133–5142.

[11] M. Pawson, and S. Greenberg, Extremely Rapid Usability Testing, *Journal of Usability Studies*, **4** (2009), pp. 124–135.

[12] M. Schreier, and Ö. Odag, Mixed Methods. In G. Mey, and K. Mruck (Eds.), *Handbuch Qualitative Forschung in der Psychologie*, VS Verlag für Sozialwissenschaften, Springer, Wiesbaden, 2010.

[13] I. Sinabell, and E. Ammenwerth, Agile, Easily Applicable, and Useful eHealth Usability Evaluations: Systematic Review and Expert-Validation, *Applied Clinical Informatics*, **13**(01) (2022), 67-79.

[14] I. Sinabell, *Agile eHealth Usability Evaluation*, Springer Vieweg, Wiesbaden, 2024.

[15] I. Sinabell, and E. Ammenwerth, Challenges and recommendations for eHealth usability evaluation with elderly users: systematic review and case study, *Universal Access in the Information Society*, (2022).

[16] G. A. Wildenbos, *Design speaks: Improving patient-centeredness for older people in a digitalizing healthcare context*, University of Amsterdam, Amsterdam, 2019.

[17] E. Airola, Learning and Use of eHealth Among Older Adults Living at Home in Rural and Nonrural Settings: Systematic Review, *Journal of Medical Internet Research*, **23**(12) (2021), e23804.

D. Hayn et al. (Eds.)

Author Index

Abanos, S. 28
Adlassnig, K.-P. 167
Aigner, C. 7
Akhlaghpour, S. 62
Alnajdawi, M. 74
Ammenwerth, E. 121, 234
Augustin, M. 215
Baranyi, R. 7, 87
Bauernfeind, S. 74
Baumgarten, D. 215
Baumgartner, C. 158
Baumgartner, M. v, 221
Berger, A. 101
Beyer, S. 107, 221
Binder, B. 186
Boonsorn, M. 129
Bott, O.J. 121
Brecher, A.-L. 55
Broeckx, L. 34
Buchgraber-Schnalzer, B. 9
Cattaneo, A. 68
Chinnakhot, P. 129
Currle, E. 15
Danler, M. 203
Denecke, K. 1, 22
Dietzel, N. 43
Domnanich, S. 156
Dowie, J. 192
Eberl, I. 41
Falgenhauer, M. 107, 221
Farmer, H. 228
Fatehi, F. 62
Fetz, B. 141
Garschall, M. 34
Glatz, D. 87
Gollmann-Tepekoeylue, C. 141
Graessel, E. 43
Grechenig, T. 7, 87
Gritsch, G. 158
Grössl, M. 221
Gruber, V. 186
Gütl, C. 179
Haas, P. 121
Hackl, W.O. 173, 198, 203, 209
Haider, M. 74
Hanke, S. 9
Harms, J. 87
Haug, S. 15
Haumer, R. 221
Hauptfeld, L. 167
Hayn, D. v, 221
Helsen, K. 34
Herrgen, J. 55
Herzmanek, D. 74
Hilberger, H. 9
Hinterbuchner, K. 141
Hoelbling, D. 7
Hoerner, W. 7
Hofmann-Wellenhof, R. 186
Holfeld, J. 141
Höller, A. 160, 186
Huber, A. 107, 141
Huber, S. 9
Jakovljevic, I. 179
Jauk, S. 156
Jonas, S. 113
Kaltoft, M.K. 192
Kappel, K. 87
Kareechum, W. 129
Kastner, P. 135, 186
Kirilov, N. 124, 143
Klein, R. 135
Kleinheinz, E. 141
Kluge, T. 158
Kolominsky-Rabas, P.L. 43
Köpl, S. 158
Koren, J. 158
Kramer, D. 156
Kramer, I. 93
Krastev, E. 28
Krasteva, R. 28
Krauss, O. 49, 74
Kreiner, K. 107, 221, 228
Krestan, S. 107, 141
Krisper, P. 156
Kropf, J. 34

Kugel, E. 156
Kuppelwieser, S. 107
Landi, D. 81
Lang, C. 158
Lauschensky, A. 107, 221
Leopold, M. 156
Lin, A.M. 149
Lins, A. 215
Lomthaisong, W. 129
Male-Dressler, A. 160
Markov, E. 28
Maul, L. 93
Mayr, P. 186
Messina, C. 135
Mirbeth, C. 41
Modre-Osprian, R. 107
Mueller, T. 55
Müller, J. 113
Namvar, M. 62
Neureiter, K. 34
Neururer, S. 107, 141
Neururer, S.B. 135, 173, 203, 209
Nghiep, L.K. 129
Noll, R. 101
Oberhoffer-Fritz, R. 113
Ohneberg, C. 41
Ostheimer, P. 215
Paula, H. 1
Pfeifer, B. v, 107, 135, 141, 160, 173, 198, 203, 209
Plaikner, S. 107
Poimer, K. 221
Pollreisz, D. 158
Pölzl, G. 141, 228
Pölzl, L. 141
Pool, J. 62
Prantl, F. 221
Prisching, A. 156
Prokosch, H.-U. 43
Puelacher, C. 141, 228
Rajput, V.K. 192
Rappelsberger, A. 167
Regazzoni, D. 68
Reichenpfader, D. 22
Reimer, L.M. 113
Reiter, K. 107, 135
Rettinger, L. 93
Rizzi, C. 68, 81
Roiss, M. 74
Romano, V. 215
Rösslhuemer, P. 22
Rzepka, A. 135, 186
Sandler, S. 74
Sandner, E. 179
Savoldelli, A. 81
Schaaf, J. 101
Schiessling, G. 135
Scholl-Bürgi, S. 160
Schönlaub, A.K. 160
Schreier, G. v, 107, 160, 186, 221, 228
Schrempf, M. 156
Schulc, E. 135
Schütz, T. 228
Schwab, A. 149
Schweitzer, M. 173, 215
Schweizer, F. 113
Seerainer, C. 9
Sendlhofer, G. 156
Sharifi, S. 62
Sinabell, I. 234
Sriphuwong, C. 129
Steger, B. 215
Storf, H. 101
Tanjga, N. 49
Tcharaktchiev, D. 28
Trajkovska, N. 74
Traub, J. 156
Tudpor, K. 129
Turnbull, N. 129
Uhlisch, C. 160
Untersteiner, H. 9
Veeranki, S. 156
Vinatzer, H. 186
Vitali, A. 68
Wagner, A. 179
Wagner, H. 107
Wagner, T.O.F. 101
Walker, A. 198
Weber, M.T. 101
Weick, K. 221
Weitlaner, T. 9
Welte, S. 160
Werner, F. 93
Widschwendter, M. 107
Wiesmüller, F. 221
Willinger, L. 113
Winkler, M. 74
Winkler, S. 149, 158

Zeiler, M. 43
Zeiringer, A. 87
Ziegl, A. 107, 221
Zivojinovic, M. 7

This page intentionally left blank