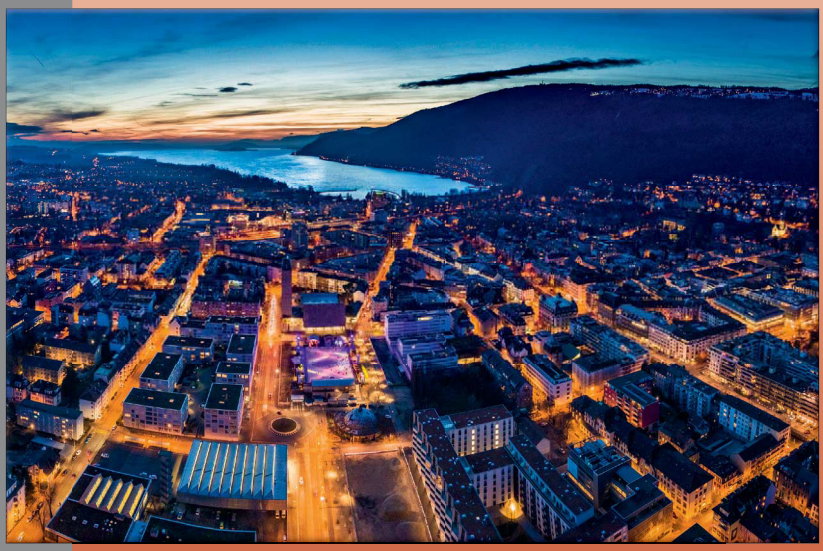
STUDIES IN HEALTH TECHNOLOGY AND INFORMATICS 292

Healthcare of the Future 2022

Digital Health – From Vision to Best Practice!



Editors: Thomas Bürkle
Kerstin Denecke
Jürgen Holm
Murat Sariyar
Michael Lehmann

IOS Press

There can be no doubt that digital technologies are set to become ever more intrinsic to many areas of healthcare in the future.

This book presents the proceedings of *Healthcare of the Future 2022*, held on 20 May 2022 in Biel/Bienne, Switzerland. This 2022 edition of the medical informatics conference has the subtitle and theme: *Digital Health – From Vision to Best Practice!* The conference explores recent advances in the deployment of digital technologies in areas such as eHealth, mHealth, personalized health and workflow-based health applications. The overarching aim of the conference is to bridge or eliminate current gaps in information with regard to outpatient care, inpatient care and the interfaces between them. The conference invited submissions for a main track and a young researchers track, and 19 papers are included here; 10 from the main track and 9 from young researchers. All papers have been peer reviewed by 2 reviewers. The papers are divided into 8 sections: advancing interoperability; semantic interoperability; medical informatics for medical research; evaluation of it influence; apps for patients and healthcare professionals parts 1 & 2; workflow based support in patient care; and research in medicine and medical informatics.

Presenting an overview of developments and research aimed at improving and accelerating healthcare processes, the book will be of interest to healthcare professionals from a wide range of disciplines.

ISBN 978-1-64368-280-8 (print)
ISBN 978-1-64368-281-5 (online)
ISSN 0926-9630 (print)
ISSN 1879-8365 (online)

HEALTHCARE OF THE FUTURE 2022

Studies in Health Technology and Informatics

International 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.

Studies in Health Technology and Informatics (HTI) series 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 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 292

Recently published in this series

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

Healthcare of the Future 2022

Digital Health – From Vision to Best Practice!

Edited by

Thomas Bürkle

Bern University of Applied Sciences, Biel, Switzerland

Kerstin Denecke

Bern University of Applied Sciences, Biel, Switzerland

Jürgen Holm

Bern University of Applied Sciences, Biel, Switzerland

Murat Sariyar

Bern University of Applied Sciences, Biel, Switzerland

and

Michael Lehmann

Bern University of Applied Sciences, Biel, Switzerland

IOS Press

Amsterdam • Berlin • Washington, DC

ISBN 978-1-64368-280-8 (print)
ISBN 978-1-64368-281-5 (online)
Library of Congress Control Number: 2022937264
doi: 10.3233/SHTI292

Publisher
IOS Press BV
Nieuwe Hemweg 6B
1013 BG Amsterdam
Netherlands
fax: +31 20 687 0019
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

Healthcare of the Future 2022
Digital health – From vision to best practice!

International Conference Biel/Bienne
20 May 2022

Thomas BÜRKLE[a,1], Kerstin DENECKE[a], Jürgen HOLM[a],
Murat SARIYAR[a] and Michael LEHMANN[a]
[a] *Bern University of Applied Sciences, Biel, Switzerland*

We are in the aftermath of a pandemic that triggered considerable advances in the digital tools supporting healthcare. Contact tracing apps supported the work of healthcare professionals while other apps collected symptoms to learn more about the disease and its development. Modern IT-based solutions scanning the web with bot technology for news about COVID cases, such as the COVID-19 Dashboard of the Johns Hopkins University,[2] initially delivered faster and better data than many national health authorities, who underestimated the efforts for collecting such data at the beginning of the pandemic.

We have also witnessed public authorities that were overwhelmed by the piles of infection reports sent on paper from GPs and hospitals using ordinary fax machines; authorities sometimes estimated the number of newly infected patients by weighing the incoming paper reports on a set of digital scales to deal with the data.

Each of us has attended many digital meetings using digital tools, and we were lucky that these tools were able to scale up to the sudden need and enable some continuity in communication when no one was allowed to travel or even leave the house. But we also experienced the fact that the tools can only partially replace personal contact and face-to-face discussions, which were not possible for considerable time spans. Now, in the transition back to normality, we are glad to invite you to the 2nd conference on the "Healthcare of the Future". Our goal is to enable networking and exchange among researchers, thus we had to delay the conference for a year due to the pandemic. We will be one of the first medical informatics conferences in 2022 to once again permit personal attendance.

In the first edition of "Healthcare of the Future" in 2019, we tried to forecast digitised healthcare in the year 2030. We foresaw a further increased life expectancy (now to be questioned) and an improved, self-determined life at home with the use of intelligent systems, wearable devices and telemedicine services. We defined digital workflows for our elderly Swiss lady, Elisabeth Brönnimann-Bertholet, who suffers from diabetes and hypertension. An integrated cross-institutional clinical pathway was

[1] Corresponding Author Thomas Bürkle, Bern University of Applied Sciences, Quellgasse 21, CH2501 Biel/Bienne, Switzerland; e-mail: thomas.buerkle@bfh.ch.

[2] https://gisanddata.maps.arcgis.com/apps/dashboards/bda7594740fd40299423467b48e9ecf6.

drawn up for her progredient hip arthrosis, which required surgery and a total hip endoprosthesis. We imagined a link between the patient at home, the GP, the specialist, the hospital and the rehabilitation centre by digital means.

In the pandemic however, we were forced to the realisation that the mix of infectious and healthy patients alone can be enough to put a modern hospital completely out of operation [1]. On the other hand, we also observed the use of modern robot technology for telepresence and physical distancing. It was even demonstrated that pre-diagnosis in public places using temperature checks and questionnaires could be performed by robots [2].

In 2019, we talked about personalised medicine. Now we have witnessed the development and worldwide availability of efficient vaccinations based on mRNA technology in record time, including the necessary clinical trials [3].

The first edition of the conference was held as part of the multi-stakeholder research project "Hospital of the future live" which had the goal of defining use cases for cross-sectoral treatment [4]. In its 2nd edition, we are just at the start of a novel competence centre «Hospital@Home» at the Institute for Medical Informatics of the BFH. This competence centre aims to examine if and what kinds of current inpatient treatment could be realised in the patient's home, and whether it is possible to integrate outpatient services to realise an improved care process. Thus, interoperability between the professional stakeholders in healthcare, and increasingly also with the patient, is an emerging topic.

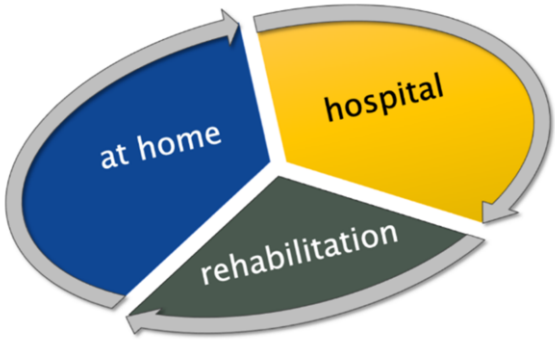


Figure 1. The patient journey – interoperability between caregivers and the patient at home

This is the setting for the 2022 edition of "Healthcare of the Future". The first conference had the goal of turning parts of the described visionary scenario into a tangible reality, with examples realised in our laboratory environment. In 2022, we will see which visions of the "Hospital of the future" have turned into reality and how digital interaction between nurses, caregivers, patients and healthcare institutions is to be realised.

The ultimate goal remains the same: improving and accelerating healthcare processes. Our three keynotes well reflect this by addressing key topics in medical

informatics: interoperability, data quality and artificial intelligence (AI) with its ethical implications.

- "Transforming the healthcare ecosystem towards better interoperability" by Silvia Thun.
- "Data quality – What do we want, what do we need, what can we pay for?" by Rainer Röhrig
- "Ethical implications of AI usage in healthcare" by Tanja Krones

The conference comprises two tracks: main track and young researchers' track. The four scientific sessions deal with the topics:

- Advancing Interoperability
- Semantic Interoperability
- Medical Informatics for Medical Research
- Evaluation of IT Influence

The young researchers' sessions will focus on apps to support patients and healthcare professionals, applications in research in medicine and medical informatics and methods and strategies for workflow-based support in patient care.

Biel/Bienne April 4th 2022

The Organising Committee

References

[1] M. Nacoti, A. Ciocca, A. Giupponi, P. Brambillascam F. Lussana, M. Pisano, G. Goisis, D. Bonacina, F. Fazzi, R. Naspro, L. Lonfghi, M. Cereda, C. Montaguti, At the Epicenter of the Covid-19 Pandemic and Humanitarian Crises in Italy: Changing Perspectives on Preparation and Mitigation. *NEJM catalyst innovations in Care delivery March (2020) 1–5. DOI: 10:1056*

[2] L. Aymerich-Franch, I. Ferrer (2020). The implementation of social robots during the COVID-19 pandemic. ArXiv preprint. ArXiv:2007.03941

[3] E. Checcucci, F. Piramide, A. Pecoraro, D. Amparore, R. Campi, C. Fiori, O. Elhage, P. Kotecha, A. Vyakarnam, S. Serni, P. Dasgupta, F. Porpiglia, The vaccine journey for COVID-19: a comprehensive systematic review of current clinical trials in humans. *Panminerva Med. 2022 Mar;64(1):72–79*

[4] T. Bürkle, K. Denecke, M. Lehmann, E. Zetz, J. Holm, Integrated Care Processes Designed for the Future Healthcare System. *Stud Health Technol Inform* **245** (2017), 20–24.

This page intentionally left blank

About the Conference

The Conference

Healthcare of the Future is a medical informatics conference which explores recent advances in the deployment of digital technologies in areas such as eHealth, mHealth, personalised health and workflow-based health applications. The overarching aim of the conference is to bridge or eliminate current gaps in information with regard to outpatient care, inpatient care and the interfaces between them.

The theme of the 2022 conference is “From vision to best practice!”.

Relevant topics include:

- communication and information exchange in care settings
- pandemics as a stress test for healthcare IT
- active living and assisted living (AAL), hospital@home
- case studies for the design of electronic health records
- workflow-based patient guidance in inpatient and outpatient settings
- digital medication processes in inpatient and outpatient settings
- disease management applications (e.g. diabetes)
- digital health interventions
- usability and human-centred design in healthcare.

Organising Committee / Chairs (alphabetic)

- Prof. Thomas Bürkle, Bern University of Applied Sciences, Biel/Bienne, Switzerland
- Prof. Kerstin Denecke, Bern University of Applied Sciences, Biel/Bienne, Switzerland
- Prof. Jürgen Holm, Bern University of Applied Sciences, Biel/Bienne, Switzerland
- Prof. Michael Lehmann, Bern University of Applied Sciences, Biel/Bienne, Switzerland
- Prof. Murat Sariyar, Bern University of Applied Sciences, Biel/Bienne, Switzerland

Peer Review Process

The Conference has a main track welcoming international scientific papers with a length of 6 pages. Students (Bachelor or Master) and those who have graduated within the last two years are invited to participate in the Young Researchers Track with a paper of 4 pages.

Each submission has been peer reviewed by 2 reviewers. Papers requiring major revisions underwent another peer review by a member of the Organizing Committee. We received 11 full papers and 10 young-researcher papers.

Peer Reviewers (alphabetic)

- Prof. Elske Ammenwerth, UMIT – University for Health Sciences, Medical Informatics and Technology, Hall / Austria
- Dr. Guillermo Lopez Campos, Queens University Belfast / UK
- Prof. Georg Duftschmid, Medical University Vienna / Austria
- Prof. Martin Dugas, University Heidelberg / Germany
- Prof. Thomas Ganslandt, University Erlangen-Nuremberg / Germany
- Prof. Ursula Hübner, University of Applied Sciences Osnabrück / Germany
- Prof. Markus Kreuzthaler, Medical University Graz / Austria
- Prof. Christian Lovis, University of Geneva / Switzerland
- Preetha Moorthy MS, Medical Faculty Mannheim of University Heidelberg / Germany
- Dr. Chris Paton, University Oxford / UK
- Carolyn Petersen MBI MS, Mayo Clinic, Rochester, MN / USA
- Prof. Hans-Ulrich Prokosch, University Erlangen-Nuremberg / Germany
- Prof. Octavio Rivera Romero, University of Sevilla / Spain
- Prof. Martin Sedlmayr, University Dresden / Germany
- Prof. Martin Staemmler, University of Applied Sciences Stralsund / Germany

Sponsors (as of April 7th, 2022)

Gold Sponsor

- HCI Solutions AG Bern / Switzerland

Silver Sponsors

- Spitalzentrum Biel-Bienne / Switzerland
- QUMEA AG Solothurn / Switzerland
- FMH (Foederatio Medicorum Helveticorum) Bern / Switzerland

Bronze Sponsors

- 3M Health Information Systems Rüschlikon / Switzerland
- Meierhofer Schweiz AG, Liebefeld / Switzerland

Contents

Advancing Interoperability

This page intentionally left blank

Healthcare of the Future 2022
T. Bürkle et al. (Eds.)

doi:10.3233/SHTI220310

SERO – A New Mobile App for Suicide Prevention

Lea MEIER[a,1], Caroline GURTNER[b], Stephan NUESSLI[a], Marko MILETIC[a], Thomas BÜRKLE[a] and Michael DURRER[c]
[a] *Bern University of Applied Sciences BFH, Biel, Switzerland*
[b] *Foundation Pro Mente Sana, Recovery and Social Politics, Zurich, Switzerland*
[c] *Lucerne Psychiatry, Lucerne, Switzerland*

Abstract. Mobile apps indicate a positive effect on suicidal ideation and potential impact on suicide attempts. As part of the SERO suicide prevention program, Lucerne Psychiatry in collaboration with partner organizations aims to reduce suicides and suicide attempts in its service area, and to improve the self-management of suicidal individuals with a mobile app. The concept for such an app was developed in a trialog with health professionals, persons at risk and their relatives and its functions were compared to six known essential app-based strategies for suicide prevention, such as the development of a safety plan, access to support networks and tracking of mood. We present the concept and architecture for the app and discuss potential added value, which may result from the intertwining of the strategies within the app, which will be available in its first version in late 2022.

Keywords. mental health, suicide prevention, self-management, self-monitoring, mobile app.

1. Introduction

Suicide causes more than 700'000 annual deaths worldwide and is the fourth leading cause of death among 15-29 year olds [1]. In Switzerland, two to three people a day end their lives by suicide [2] and due to Corona, emergency interventions at the University Hospital Zurich went up from 321 in the first 6 months of 2019 to 450 in the same time period in 2021 [3]. Management of patients at risk for suicide may involve combined modular approaches for prevention, treatment and follow-up, including hospitalization, pharmacotherapy and psychotherapy [4]. Algorithms exist listing recommendations for securing patient's safety, including limited access to lethal means, education and a safety plan [4].

Hundreds of mobile apps for depression and suicide prevention can be found [5,6] and some have been systematically evaluated [7,8]. Results suggest a positive effect on suicidal ideation [8] and potential effects on the prevention of suicide attempts, although not statistically significant [7]. Martinengo et al. [5] identified six essential app-based strategies for suicide prevention, namely 1) tracking of mood and suicidal thoughts, 2) development of a safety plan, 3) recommendation of activities to deter suicidal thoughts, 4) information and education, 5) access to support networks and 6) access to emergency

[1] Corresponding Author, Lea Meier, Bern University of Applied Sciences BFH, Institute for Medical Informatics I4MI, Quellgasse 21, 2501 Biel/Bienne, Switzerland; E-mail: lea.meier@bfh.ch.

counselling and found only 5 of 69 apps covering all six strategies. Most apps included only three strategies [5], with the most common being: Emergency contact information (65/69), direct access to a crisis helpline (46/69) and suicide-related education (35/69).

Lucerne Psychiatry with three psychiatric hospitals and numerous outpatient clinics, including two home treatment teams in Central Switzerland wishes to improve the self-management of suicidal individuals as the shortage of mental health professional is an immediate issue. The project "SERO: Suizidprävention Einheitlich Regional Organisiert[2]" (www.sero-suizidpraevention.ch) is a co-operative initiative with the goal to reduce suicides, suicide attempts and re-/hospitalizations due to suicide risk [9]. SERO is one of five suicide prevention projects funded by Health Promotion Switzerland's "Prevention in Healthcare" project grant [10]. Within this project, Lucerne Psychiatry collaborates with Health Promotion Switzerland and other partner organizations. As part of it, a mobile suicide prevention SERO-app is under development that strives to combine the above-mentioned six strategies within one single app and comprises the PRISM-S[3] instrument [11] for rapid self-assessment of suicidality.

2. Methodology

Literature research was conducted using the PubMed database. We were only interested in studies that focused on strategies implemented in apps for suicide prevention and its impact. Combination of keywords such as the following were used: (mobile app) AND ((suicide prevention) OR (suicide risk assessment)).

User requirements were collected in focus group discussions with health professionals, people with illness experience (suicidality) and relatives. A pilot study for the adaption of the PRISM-S assessment, which is usually performed with a mental health professional, was conducted in the inpatient setting. The aim was to develop a guidance for the patients in order to use the PRISM-S without being instructed by a mental health professional. To validate requirements and design, a formative usability test in a real-life setting with eleven psychiatric inpatients was performed within two focus groups and three individual sessions using a detailed clickable mock-up.

Implementation will be done cross-plattform and in parallel for iOS and Android devices with React Native (www.reactnative.dev). The app will be available to download for free in the respective stores in German, French and Italian language.

3. Results

3.1. The Whole May Be More than the Sum of Its Parts

When designing the SERO-app, we extensively analyzed the work of Martinengo et al. [5]. Our goal was to implement all six potential strategies for suicide prevention with analogue functions of the SERO-app. Fig. 1 illustrates the approach: Tracking of mood or suicidal thoughts is covered with the two functions "Suicide risk assessment" by means of the PRISM-S instrument [12] and a follow-up questionnaire to assess self-

[2] engl. "Suicide Prevention Unified Regionally Organized"

[3] "pictorial representation of illness and self management – suicidality"

reflection. Activities to deter suicidal thoughts are covered with the SERO-app functions to record own coping strategies and own distraction strategies.

The interesting point however is the desired interaction among those different strategies. Thus, the following added values are expected:

1. When tracking of mood or suicidal thoughts is completed, SERO suggests consulting the strategies recorded in the safety plan, e.g., coping or distraction strategies, or to consider talking to someone in the support network or an emergency number.
2. The tracking module and the safety plan development function are connected to the information screen function, which describes also intention and usage of both functions. Furthermore, the safety plan has outsourced its element "my personal helpers" to an own module to increase its accessibility.
3. Activities to deter suicidal thoughts are offered by suggesting emergency counselling and are proposed when creating the safety plan.

Figure 1. The six strategies for suicide prevention and the corresponding SERO-app functions. Numbered connections indicate interactions among strategies.

3.2. Conceptual Overview and Development Path

To enable the SERO-app to be distributed throughout Switzerland and to support later data analysis for research, secure data storage and access must be provided. Therefore, SERO-app uses MIDATA (www.midata.coop), a GDPR-compliant and citizen-owned health data platform, where each user has individual control over their personal data stored in their account [13]. Patient data is stored within FHIR-resources (www.hl7.org/fhir) in the secured MIDATA server environment under the personal account of the user. Data transmission between the SERO-app and MIDATA is encrypted, and further security aspects will be implemented, such as the app can only be used if the mobile phone is protected with a pin, fingerprint, or Face ID. App development is planned in two versions in order to rapidly provide a minimum viable product. SERO-app V1.0 will be released in 2022 and is limited to data and functionality for the suicidal person only. SERO-app V2.0 planned for 2023 will assist interaction of the suicidal person with his relatives and beloved ones and will permit data sharing for

parts of the safety plan and the suicide risk assessment. The MIDATA platform is able to support such interaction by means of an advanced consent management mechanism, where users can give third parties access to certain personal data. Fig. 2 shows the overall conceptual architecture for both versions.

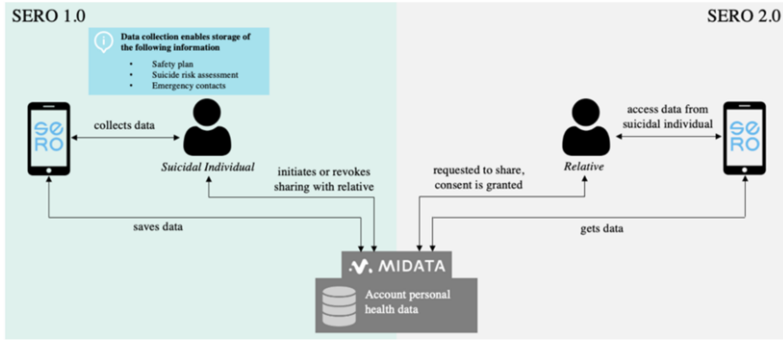


Figure 2. Conceptual architecture of SERO-app 1.0 and 2.0, and the MIDATA database and server.

3.3. Planned Architecture and features of SERO-App 1.0

SERO-app 1.0 will comprise four main modules which cover the app functions described in Fig. 1. The main modules "home", "safety plan", "suicide risk assessment" and "emergency contacts" comprise between one and six different app functions and support the suicidal individual in preventing, coping and reflecting on the crisis (see Fig. 3). In addition, SERO-app 1.0 will have two supporting modules which provide a) baseline information and education related to suicidality (information screen) and permit to b) personalize the app with the respective settings. The app will also have a function for onboarding and registration on the MIDATA server platform.

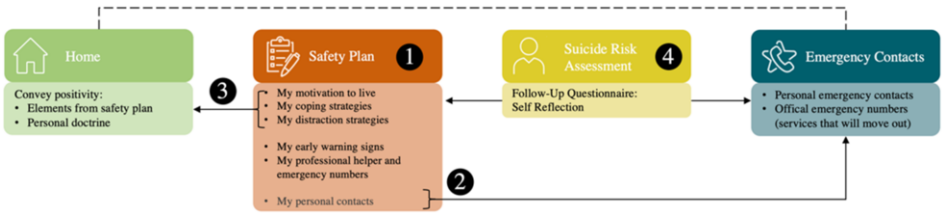


Figure 3. SERO-app modules, and their dependencies and interfaces.

The safety plan (Fig. 3, no. 1) is based on the plan developed within the SERO project on the basis of scientific findings, expert knowledge and the involvement of affected persons and their relatives [9]. The goal is to empower affected individuals and family members to use and reflect on their strengths and coping strategies. The plan is subdivided in the following sections: 1. My motivation to live, 2. My coping strategies, 3. My distraction strategies, 4. My early warning signs, 5. My personal contacts, and 6. My professional helper and emergency numbers. The plan will be implemented with a modular approach, meaning the sequence can be adapted individually to the needs of the users. While sections 1-4 and 6 are illustrated in the safety plan itself, a separate module has been created for section 5 (Fig. 3., no. 2). These emergency contacts are integrated

on the main screen; official emergency numbers can be accessed from any screen. Both can be directly called from within the app. While personal emergency contacts can be added individually by the user (incl. profile picture) and bring a personal touch to the app, official emergency numbers are predefined by the app. The latter are services that move out in emergencies around the clock, e.g., ambulance or police.

In addition to the emergency numbers, the home screen will always present a subsection of the safety plan (Fig. 3, no. 3) for the user to memorize his/her personal strategies. An example might be, "One of your distraction strategies is 'do relaxation exercises'. You should give this a try soon." Furthermore, push-notifications of subsections of the safety plan are being considered.

3.4. PRISM-S and Its Realization in SERO-App 1.0

Most current digital interventions [8] use the PHQ-9 ("Patient Health Questionnaire) [14] for risk documentation of suicidal ideation or self-harm, which is a self-administered diagnostic instrument for depression. However, in the SERO-app, we use the PRISM-S [12] for suicide risk assessment (Fig. 3, no. 4). It is a valid instrument for the visual clinical assessment of suicide risk [11] and can be used to assess how far a person is away from suicide (Fig. 4). The assessment is normally carried out together with a mental health professional on a physical board and was adapted for the app for self-application. Results of the conducted pilot study have shown that post-assessment self-reflection and access to emergency contacts are key aspects for successful realization. Therefore, the assessment is followed by a questionnaire and after completion, the consultation of the safety plan or to contact a relative is suggested. The usability test demonstrated that patients in treatment for acute suicidality feel supported by such a mobile app. Participants were able to use the mockup without experiencing panic attacks, but desired an emergency button to call relatives nearby.

The assessments performed are stored and the distance between the two circles is saved on MIDATA. Thus, the individual self-experience and progress can be discussed with a mental health professional during a therapy session.

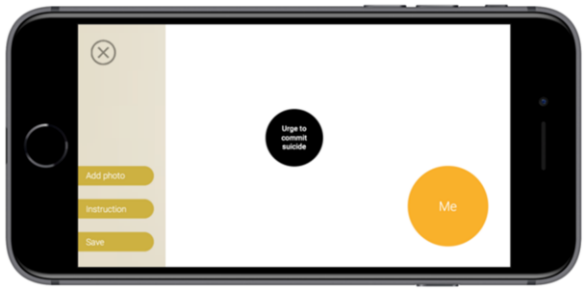


Figure 4. PRISM-S suicide risk assessment in the SERO-app. The user may move the black disc (Urge to commit suicide) towards or away from his "self" (Me). If the black disc overlaps the yellow one, this indicates the highest risk level.

4. Discussion/Conclusion

Our approach combining multiple suicide prevention strategies provides a simple and user-friendly mobile app and has been developed in close interaction with healthcare

professionals and persons in treatment for suicidality. Current studies indicate little knowledge about the effect of an app on suicidal behavior and thoughts [7,8], although an effect on suicidal ideation may be expected. Tracking and documentation of the person's status is recommended [4] and will be provided by means of the PRISM-S and questionnaire. Other recommended strategies include limiting access to lethal means, as well as appropriate pharmacological support [4].

Legal requirements have limited the functional scope of the SERO-app. For example, the PRISM-S does not provide feedback as this would fall within the scope of the Medical Device Regulation (REGULATION (EU) 2017/745) requiring financial resources not currently available. Nevertheless, we managed to avoid these limitations by adapting the methodology for self-assessment and adding elements for self-reflection.

We think that our combination of the six suicide prevention strategies from Martinengo et al. [5] and their interaction has the potential to support self-management of suicidality, but we will need to wait and see if SERO V1.0, planned for late 2022, will come up to our expectations and will be used on a continuous base, which is the utmost prerequisite for any positive effects.

References

[1] Suicide worldwide in 2019: global health estimates. World Health Organization; 2021.
[2] Schweizerisches Gesundheitsobservatorium. Suizid und Suizidhilfe [Internet]. 2021 [cited 2021 Mar 22]. Available from: https://www.obsan.admin.ch/de/indikatoren/suizid
[3] Wegen Corona leiden mehr Jugendliche an Suizidgedanken [Internet], watson.ch. 2022 [cited 2022 Jan 24]. Available from: https://www.watson.ch/!161738841
[4] The Assessment and Management of Risk for Suicide Working Group. VA/DoD clinical practice guideline for assessment and management of patients at risk for suicide. Washington: Department of Veterans Affairs, Department of Defense; 2013.
[5] Martinengo L, Van Galen L, Lum E, Kowalski M,Subramaniam M, Car J. Suicide prevention and depression apps' suicide risk assessment and management: a systematic assessment of adherence to clinical guidelines, BMC Med. 2019 Dec 19;17(1):231.
[6] Larsen ME, Nicholas J, Christensen H. A Systematic Assessment of Smartphone Tools for Suicide Prevention. PLoS ONE. 2016 Apr 13;11(4).
[7] Kennard BD, Goldstein T, Foxwell AA,McMakin DL, Wolfe K, Biernesser C, et al. As Safe as Possible (ASAP): A Brief App-Supported Inpatient Intervention to Prevent Postdischarge Suicidal Behavior in Hospitalized, Suicidal Adolescents. Am. J. Psychiatry. 2018 Sep 1;175(9):864–72.
[8] Witt K, Spittal MJ, Carter G, Pirkis J, Hetrick S, Currier D, et al. Effectiveness of online and mobile telephone applications ('apps') for the self-management of suicidal ideation and self-harm: a systematic review and meta-analysis. BMC Psychiatry. 2017 Aug 15;17(1):297.
[9] Durrer M, Fluder M, Fringeli J, Lerch M, Meinhof J, Maeck L, et al. Projekt SERO – Suizidprävention Einheitlich Regional Organisiert. Lead Opin Neurol Psychiatr. 2021 Jun;55–7.
[10] Gesundheitsförderung Schweiz. Projektförderung Prävention in der Gesundheitsversorgung [Internet]. 2021 [cited 2022 Jan 30]. Available from: https://gesundheitsfoerderung.ch/pgv/grundlagen-der-projektfoerderung/konzept.html
[11] Ring M, Harbauer G, Haas S, Schuetz C, Andreae A, Maercker A et al. Validierung des Suizidalitätseinschätzungsinstrumentes PRISM-S (Pictoral Representation of Illness Self Measure – Suicidality). Neuropsychiatrie. 2014 Dec 1;28(4):192–197.
[12] Harbauer G, Ring M, Schuetz C , Andreae A, Haas S. Suicidality assessment with PRISM-S - simple, fast, and visual: a brief nonverbal method to assess suicidality in adolescent and adult patients. Crisis. 2013;34(2):131–6.
[13] Hafen E, Personal Data Cooperatives – A New Data Governance Framework for Data Donations and Precision Health. In: J. Krutzinna, and L. Floridi, editors. The Ethics of Medical Data Donation. Cham: Springer International Publishing; 2019. p. 141–9. (Philosophical Studies Series).
[14] Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. J. Gen. Intern. Med. 2001 Sep;16(9):606–13.

Healthcare of the Future 2022
T. Bürkle et al. (Eds.)

doi:10.3233/SHTI220311

FHIR2BPMN: Delivering Actionable Knowledge by Transforming Between Clinical Pathways and Executable Models

Emmanuel HELM[a,1], Andreas POINTNER[b], Oliver KRAUSS[b], Andreas SCHULER[a], Barbara TRAXLER[a], Klaus ARTHOFER[c], Gerhard HALMERBAUER[c]
[a] *University of Applied Sciences Upper Austria, Department of Medical and Bioinformatics, 4232 Hagenberg, Austria*
[b] *Research Department Advanced Information Systems and Technology, University of Applied Sciences Upper Austria, 4232 Hagenberg, Austria*
[c] *University of Applied Sciences Upper Austria, Department of Process Management and Business Intelligence, 4232 Hagenberg, Austria*

Abstract. Healthcare processes have many particularities captured and described within standards for medical information exchange such as HL7 FHIR. BPMN is a widely used standard to create readily understandable processes models. We show an approach to integrate both these standards via an automated transformation mechanism. This will allow us to use the various tools available for BPMN to visualize and automate processes in the healthcare domain. In the future we plan to extend this approach to enable mining and analyzing executed processes.

Keywords. Clinical Guidelines, Model Transformation, HL7 FHIR

1. Introduction

Health Level 7 (HL7) Fast Healthcare Interoperability Resources (FHIR)[2] is a well-established standard in healthcare, which is used to support the exchange of data between information systems [1]. FHIR describes data formats as well as resources and provides interfaces for exchanging them.

One of these resources is the PlanDefinition, which allows the description of clinical artifacts. The FHIR R4 specification describes it as follows [2]:

> "This resource allows for the definition of various types of plans as a sharable, consumable, and executable artifact. The resource is general enough to support the description of a broad range of clinical artifacts such as clinical decision support rules, order sets and protocols." [2]

Business Process Model and Notation 2.0 (BPMN) on the other hand is a well-established standard by the Object Management Group (OMG). It is widely used in the

[1] Corresponding Author: Emmanuel Helm, University of Applied Sciences Upper Austria, Department of Medical and Bioinformatics, Softwarepark 11, 4232 Hagenberg, E-Mail: emmanuel.helm@fh-hagenberg.at

[2] HL7, FHIR and the FHIR logo are the registered trademarks of Health Level Seven International and their use does not constitute endorsement by HL7.

community and there is a lot of tooling available for modelling, simulating, and executing business processes. The use of BPMN as well as CMMN (Case Management Model and Notation) and DMN (Decision Model and Notation) for healthcare is described in the *Field Guide to Shareable Clinical Pathways* [4] and summed up in the term *BPM+*.

We provide an approach to make BPMN and FHIR interoperable. Based on this interchange between two well-established standards, this provides a basis to define, document, and mine processes in healthcare. We showcase this by manually modelling clinical guidelines in the FHIR PlanDefinition resources, which in turn are then automatically transformed using our presented approach. The results are then compared with a BPMN tool and validated by domain experts.

1.1. Problem Statement

Evidence based Clinical Pathways (CPs) are an important factor in modern healthcare. Typically presented in the form of narrative text, they aim to provide knowledge for healthcare professionals, guide the allocation of resources, reduce the risk of liability for negligence in the duty of care, and enable the assessment and assurance of quality in healthcare [3]. Note that the terms *Clinical Pathways*, *Clinical Guidelines*, or *Clinical Practice Guidelines* are often used interchangeably in literature.

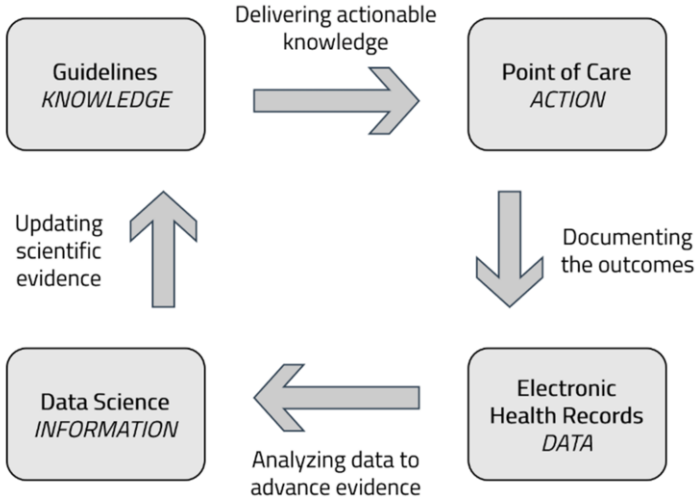


Figure 1. The data lifecycle with clinical guidelines as representation of knowledge (from [6]).

Due to their narrative nature, CPs are open to large variations of interpretation. To be viable as a basis for, e.g., computer-based workflow orchestration, decision support systems, or automated compliance checks, these CPs must be made explicit, including formal, repeatable semantics [4].

Figure 1 shows the basic idea how CPs/Guidelines as representations of scientific evidence can be used to deliver actionable knowledge at the point of care. Of course, there are attempts to tackle the challenge of computable CPs. Computerized Clinical Decision Support Systems (CDSS) are among the most successful solutions for this challenge [5]. Although decision support systems in medicine have been developed for the better part of three decades, they mostly “exist as cumbersome stand-alone systems, or exist in a system that cannot communicate effectively with other systems” [5].

Moreover, the knowledge embedded in these systems is often described using formal programming languages, not readily understandable by healthcare professionals and business analysts [4].

BPMN provides the means to model complex clinical processes, showing who does what, where and in what sequence [1]. It distinguishes between the flow of activities of single actors and the flow of messages and information between them. It enables visualization and has a formal notation where models can be executed by a workflow engine. However, it provides only a very generic notion of activities, not considering the specificities of healthcare. While BPMN is exchangeable via the XML format, the standard lacks mechanisms for lifecycle-management, i.e., storing, versioning or updating the model. It also allows for *dialects*, i.e., different ways of modelling the same process, thus making interoperability hard.

HL7 FHIR, on the other hand, comes with a rich information model for clinical data and mechanisms for versioning, provenance, access control, and messaging, among others. Certain resources are even designed to store CPs or elements of clinical workflows [6], i.e., the PlanDefinition resource. However, FHIR does not define how to visualize resources and while they can contain the information necessary for execution, resources do not define how to make them executable.

1.2. Proposed Solution

We aim to make computable CPs both, interoperable *and* readily understandable. We do not present a new standard to combine all features of BPMN and HL7 FHIR but develop an automated transformation approach between these formats. This can be seen as a basis for better computable CPs.

Since both models, the HL7 FHIR Plan Definition and the BPMN model, can be represented as graphs, we can define rules on how to map between the various elements in the corresponding models and use a graph transformation framework (GTF) [7] to implement and apply these rulesets.

We enable HL7 FHIR's capabilities to manage and transport clinical information in a structured, standardized way and we enable BPMN's potential to visualize models and to fuel them into workflow engines.

2. Method

Figure 2 depicts the top-level transformation approach. Our approach relies on the fact, that both the BPMN Metamodel and the FHIR Resource specification are mappable to a Graph-structure. This in turn allows us to transform the BPMN Model, which is defined by the BPMN Metamodel, and the FHIR workflow resource specification (PlanDefinition), which is defined by the FHIR Resource specification (StructureDefinition), into an interim graph structure, which conforms to the graph model.

This interim structure reduces the complexity of the overall transformation, as it neither relies on input nor output models. The interim graph structure is defined by the respective transformation implementation in the GTF. The transformation is then defined by the set of rules that allow to transform from and to the interim graph.

Finally, on the instance level, we see the BPMN Clinical Pathway, which conforms to the BPMN Model as well as the FHIR Clinical Pathway, which in turn conforms to

the FHIR Workflow resources. By applying the previously defined rule sets, a transformation between these instances is enabled.

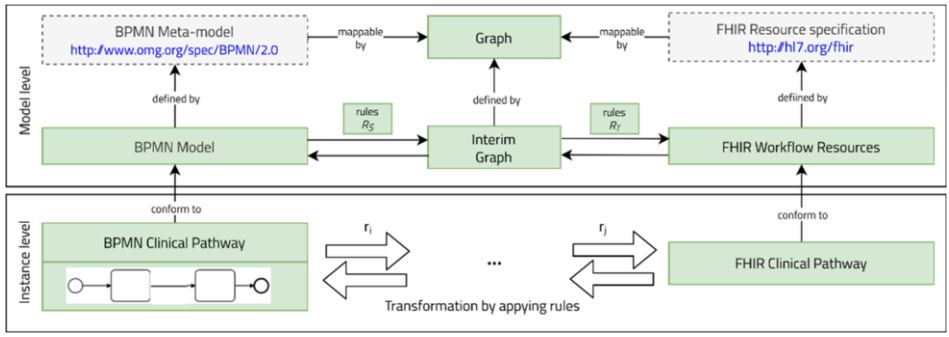


Figure 2. The basic transformation architecture based on an interim graph.

The main contribution of this work is the combination of HL7 FHIR with BPMN, enabling definition and visualization of processes, while also being able to consider the intricacies of healthcare systems. This is enabled by our transformation approach. It identifies patterns for the various concepts and defines transformation rules for each pattern. Both representations offer various concepts that can be used to define processes.

3. Results and Discussion

To be able to create transformation rules, it is necessary to identify patterns in both model domains that can be matched. We focused on six core concepts, namely (1) sequence flow, (2) exclusive split and (3) parallel split, (4) data flow, (5) triggers and (6) actors or participants. In this section, we will show the mappings for one of the six core concepts defined above, (1) the sequence flow.

All transformation patterns developed in this work can be found on the GitHub repository[3] also referenced in [8].

Table 1. Representation of sequence flow in FHIR PlanDefinition and BPMN [8].

HL7 FHIR PlanDefinition	**BPMN**

`<action>`
`  <id value='t1'/>`
`  <title value='Task 1'/>`
`  <relatedAction>`
`  <actionId value='t2'/>`
`  <relationship`
`  value='before-start'/>`
`  </relatedAction>`
`</action>`
 	
`<action>`
`  <id value='t2'/>`
`  <title value='Task 2'/>`
`</action>`
	`<userTask name="Task 1" id="t1">`
	`  <outgoing>sf_t1_t2</outgoing>`
	`</userTask>`
	
	`<userTask name="Task 2" id="t2">`
	`  <incoming>sf_t1_t2</incoming>`
	`</userTask>`
	
	`<sequenceFlow sourceRef="t1"`
	`targetRef="t2" id="sf_t1_t2"/>`

[3] https://fhooeaist.github.io/MSBPMN/transformation.html

In BPMN, the sequence flow is modeled by defining a *sequenceFlow* element that references the two *task* objects it connects. In the FHIR PlanDefinition we can achieve this by adding a nested *relatedAction* element into one of the two actions and by defining the ID of the other action as well as the relationship type (see Table 1).

The remaining concepts (2)-(6) require the interaction of the tasks with other elements, such as gateways, data elements, triggers, or actors. In BPMN this is modelled by defining these elements on the top-level structure and having linking elements between these elements and the tasks, building a flat hierarchy inside the model. For the FHIR PlanDefinition this is modelled differently, with elements as nested structures inside the tasks. Gateways, for example, are modelled by nesting tasks inside other tasks and specifying *groupingBehaviour* and *selectionBehaviour* for the parent element to define the relationships [2]. This leads to potentially deep, tree-like hierarchies. These structures need no separate linking elements, but due to their hierarchical structure they come with limitations, e.g., regarding loops and jumps.

We tested the approach on the basic patterns listed in Section 2 and on more complex CPs. The resulting diagrams including the initial FHIR PlanDefinition resources can be found in high resolution in the GitHub repository[4]. Figure 3 shows an automatically generated BPMN model based on the transformation result of the Austrian Federal Quality Guideline for Preoperative Diagnostics (BQLL PRÄOP) [9]. All subprocesses were expanded.

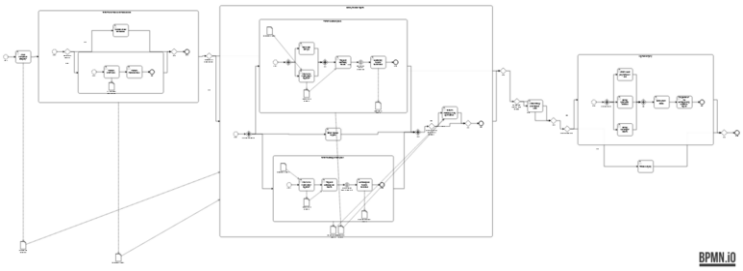


Figure 3. BPMN representation of the Federal Quality Guideline for Preoperative Diagnostics [9].

Preoperative diagnostics is used for the early detection of risks that may arise during and after surgery. It also serves to assess the basic fitness of patients for surgery. The guideline includes steps such as completing a structured questionnaire for medical history, performing cardiopulmonary testing, or determining various standardized scores in advance of deciding whether to perform a surgery [9].

To validate our transformation approach, we first created a FHIR PlanDefinition resource based on the narrative description in [9]. In the next step we automatically applied our transformation rules to generate a BPMN representation. The resulting XML file was then rendered (Figure 3) and compared to the narrative CP by domain experts.

The main finding was that the general flow (i.e., key activities and their sequence) was clearly identifiable and correctly represented in the model. However, some shortcomings of our approach were also identified: (A) a lack of formalization of decision tables for XOR gateways, (B) certain meaningful modeling elements that BPMN would provide, e.g., Swimlanes, were not used, and (C) loops, e.g., to repeat uncertain laboratory tests, are currently not possible due to the limitations of the hierarchical model described in Section 3.

[4] https://fhooeaist.github.io/MSBPMN/fhir2bpmn.html

To tackle (A) we plan to use DMN to formalize the decision points. However, we must determine whether the approach is compatible with the formal logic currently built into the HL7 FHIR resources described in [6]. (B) requires further research and coordination with the FHIR Workflow project as to how these elements can be modeled with the existing FHIR resources. For (C), we discuss the following two approaches: (C1) flatten the hierarchy in the PlanDefinition resource and use a referencing mechanism instead, or (C2) add the referencing mechanism despite the hierarchical base structure. We need to further investigate the consequences of these approaches and discuss them within the FHIR Workflow community.

4. Outlook

While the transformation still has its limitations, applying it to existing CPs, already represented as PlanDefinition resources, could be useful to enable the broader tooling of the BPMN domain. The understanding of the structure and semantics of the building blocks of healthcare workflows will also be helpful to provide inputs for the further development of the PlanDefinition resource, making it more suitable to represent CPs.

We will continue our work on the analysis of standards-based event logs by the means of process mining [10]. The transformation patterns and the interplay of FHIR workflow resources are also relevant for the analysis of FHIR audit record repositories.

Acknowledgements

This project is financed by research subsidies granted by the government of Upper Austria. We also thank the participants of the weekly HL7 International FHIR Workflow project calls for the fruitful discussions.

References

[1] Benson, T. and Grieve, G., Principles of Health Interoperability. Springer, 2021.
[2] HL7 International, FHIR Specification (v4.0.1: R4) Resource PlanDefinition - Content. 2019, https://www.hl7.org/fhir/plandefinition.html, last access: 05.01.2022
[3] Peleg, M.: Computer-interpretable clinical guidelines: a methodological review. Journal of biomedical informatics, 46(4), 744-763. 2013. https://doi.org/10.1016/j.jbi.2013.06.009
[4] OMG Healthcare Domain Taskforce, Field Guide to Shareable Clinical Pathways. 2020, https://www.trisotech.com/wp-content/uploads/OMG-Field-Guide-to-sharable-Clinical-Pathways.pdf, last access: 05.01.2022
[5] Sutton, R.T., Pincock, D., Baumgart, D.C. et al. An overview of clinical decision support systems: benefits, risks, and strategies for success. NPJ Digit. Med. 3(1). 2020.
[6] HL7 International, Clinical Decision Support Working Group, FHIR Clinical Guidelines. 2021, https://build.fhir.org/ig/HL7/cqf-recommendations/, last access: 05.01.2022
[7] Schuler, A., Pointner, A., Praschl, C.: GTF - Graph Transformation Framework, Fhooeaist/gtf: 1.0.0. 2020, https://doi.org/10.5281/zenodo.4314722
[8] Pointner, A., Helm, E., Krauss, O., Aldea, P.: MSBPMN: Model and Standards-Based Process-simulation in MediciNe, 2021, https://doi.org/10.5281/zenodo.4590013
[9] BMASGK: Bundesqualitätsleitlinie Präoperative Diagnostik (Version 2018). Bundesministerium für Arbeit, Soziales, Gesundheit und Konsumentenschutz, Wien, 2018
[10] Helm, E., Krauss, O., Lin, A., Pointner, A., Schuler, A., Küng, J., Process Mining on FHIR - An Open Standards-Based Process Analytics Approach for Healthcare, International Conference on Process Mining (ICPM) Workshops, 2020. https://doi.org/10.1007/978-3-030-72693-5_26

Healthcare of the Future 2022
T. Bürkle et al. (Eds.)

doi:10.3233/SHTI220312

Participatory Development of an Image-Based Communication Aid for Migrant Patients and Emergency Nurses

Kerstin DENECKE[a], Friederike J.S. THILO[b], Christopher LUEG[a], Loraine OLALIA[c], Anouk HALDEMANN[b] , Beatrice KAUFMANN[c]

[a] *Institute for Medical Informatics, Bern University of Applied Sciences, Bern, Switzerland*
[b] *Applied Research & Development in Nursing, Department of Health Professions, Bern University of Applied Sciences, Bern, Switzerland*
[c] *Institute of Design Research, Bern University of Arts, Bern, Switzerland*

Abstract. Language barriers hamper or delay delivery of urgent and emergency care to migrant children when they or their parents don't speak any of the languages commonly spoken in Switzerland. In such situations, nurses often fall back to use ad hoc communication aids, including translation apps and visual dictionaries, to collect information about a patient's medical history. In this paper, we report on the participatory design process for a novel image-based communication aid. It is specifically tailored to the needs of migrant patients and nurses within Swiss pediatric clinics. We collected requirements in surveys and in-depth interviews with pediatric nurses. A prototype app was developed and tested with users in a scenario-based usability test. The results clearly show that the images developed, especially for symptoms, accidents or nutrition and excretion, are well comprehensible for triage and anamnesis. In contrast, a temporal classification or chronological occurrence of health incidents is difficulty to express with images.

Keywords. Communication barriers, language, emergency care, migrant patients, pediatrics, communication aid, visual support

1. Introduction

Language barriers often affect the quality of treatment and care [1,2]. They increase the risk of adverse events and fatal outcomes in both hospital and primary care settings [3]. A Europe-wide survey of pediatric accident and emergency staff found that language barriers are the most common obstacle to emergency care for refugee children and young people [4]. Understandable communication in care is therefore essential for safe, high-quality and equitable healthcare for all patients. Existing solutions are limited in their effectiveness. Visual representations can be used to overcome language barriers [5,6].

In this paper, we describe the development of an image-based tool aiming at addressing communication barriers experienced in Swiss pediatric emergency care. Existing tools currently used in practice include translation apps, some of them visually enhanced, and visual dictionaries, both digital and traditional paper-based ones. Images used are often insufficiently expressive, the number of included images is very limited or aesthetically not very appealing so far. Following market analysis, we closely

examined four apps found to be relevant in terms of their suitability in Swiss pediatric emergencies. MediPicto [1] allows to make entries with visual support. Selection of language is restricted. medilang Pro[2] does not allow for patient questions. Available languages do not cover all languages necessary for usage in Switzerland. tipdoc emergency[3] contains the most relevant questions for collecting the medical history in pediatric emergencies. Questions can only be asked by nurses. Most of the existing apps are of limited utility. Pure translation apps use predefined phrases translated in a limited number of languages and fail when a language is unavailable. We will go beyond existing tools by supporting interaction by a non-verbal communication aid relying only upon images. We address the research question: What are the needs and requirements of a non-verbal, image-based digital communication for pediatric emergency anamnesis which will be perceived useful, feasible and acceptable for nurses, foreign-language patients and their relatives.

2. Methods

The communication aid was developed using a participatory design process. The approach adheres to the general principles of human centered design and interaction design. An interdisciplinary team of researchers from communication design, nursing research and medical informatics was formed and potential users (pediatric emergency nurses, migrant parents and children) were involved throughout the development. The development comprised several steps and was based on prior research by Kaufmann et al. [5,6]: 1) Collect requirements for the communication aid in discussion with nurses and by identifying limitations of existing approaches by literature search and market analysis. 2) Design images tailored to the most common communicative situations and validate understandability of images with migrant parents, children, and nurses. 3) Develop a prototype. 4) Conduct a usability test with nurses and migrant parents. Methodology and results of developing communication design images (step 2) will be described in another publication. In this work, we focus on the other three steps.

2.1. Requirements collection and prototype development

To gather requirements for our communication aid, we identified limitations and best practices in healthcare communication aids for the treatment of migrants, focusing on design issues. To analyze existing communication apps suitable for emergency settings, we searched for relevant apps in Google Search and in major app stores (Google Playstore, Apple App store). Apps were considered relevant when they aimed at supporting medical communication in emergencies and mentioned that they were aimed at supporting migrant patients. We also checked whether they would be suitable for use in pediatric emergency care in Switzerland. Assessment criteria included coverage of languages frequently occurring in Switzerland, focus on pediatrics or emergency, implementation of Swiss-specific administrative conditions like health insurance card, vaccination report or medication. Furthermore, we collected information needs of

[1] https://www.aphp.fr/medipicto
[2] https://medilang.com/index.html
[3] http://www.setzer-verlag.com/Neuigkeiten/Tip-doc-emergency-App

nurses in pediatric emergency care and their current communication strategies when language barriers occur.

We used a questionnaire to collect information from the heads of nursing of 9 pediatric emergency units across German speaking Switzerland and conducted interviews with 10 nurses working in emergency units for children. We collected information on challenging communication situations when dealing with language barriers, and existing strategies and communication aids they use to cope with these barriers. Finally, we asked for features they would like to see in a communication aid. To learn more about challenges of the current process in emergency communication, we used case examples. We also asked in detail which information is difficult to obtain in cases of language barriers and which is relevant for treatment. Based on the collected requirements, we designed a mock-up for our communication aid using Figma. In several iterations and in strong collaboration with nurses, the mock-up was improved until a version was implemented as functional prototype.

2.2. User testing with prototype

We tested the prototype with potential users in a scenario-based study realized as role-play and with cognitive walkthrough. Taking into account the perspectives of nursing, design and informatics, we aimed at answering the following questions: 1) How do nurses and parents generally cope with the application? 2) How do nurses use the application in a first triage? 3) What images do nurses and parents miss in initial triage? 4) What difficulties do nurses and parents encounter in using the app in initial triage? 5) Which factors promote the integration of the app in everyday care? Participants were recruited from former master students and from pediatric emergency department nurses at a university children's hospital (nurses) and from personal network of contacts (parents). A tablet was available for the user testing. All conversations were recorded for analysis purposes. Observations were collected in a handwritten observation log. From the recordings and observation logs, we draw conclusions on the interaction and use of the application. Interviews were conducted after completing a particular scenario and participants were asked for their experiences and suggestions.

The test setting was as follows: a pediatric nurse and a parent sat together across a corner at a table. The tablet was on the table in front of them. A researcher explained the roles and had both test subjects read the scenario. Then the nurse was asked to start the role play. The participating parents had a mother tongue other than German, but could understand German little to well. The caregiver should refrain from using language as much as possible and the testing parent should try not to react to language, but only to pictures and non-verbal communication. If the nurse was missing a picture, could not find something in the app or had other thoughts about the process that did not run as smoothly as she is used to in practice, she should comment aloud. One researcher each from nursing, design and medical informatics observed the nurse and a parent with the imaginary child in role play. As soon as the nurse came to a decision regarding further treatment, the role play was stopped. This was followed by a questioning of the nurse and parent about the role play by the nurse researcher and the design researcher. The four scenarios for emergency cases were:

- Scenario 1: 3-year-old girl, high temperature
- Scenario 2: 7-year-old boy, stomachache, and lack of appetite
- Scenario 3: 3-month-old girl, diarrhea, heart surgery, penicillin allergy

- Scenario 4: 6-year-old boy, fall from climbing frame, vomiting

An internal pre-test with two employees of the Bern University of Arts and of the Bern University of Applied Sciences that were not involved in the project was conducted on November 16, 2021. The aim of the pre-test was to test understandability of case scenarios and to test the entire test setting. As a result, an additional feature was implemented in the prototype which is the possibility to enter data such as the date of birth.

3. Results

3.1. Requirements for a digital communication aid

A future communication aid should be a simple tool that is quickly available, covers the most common situations and enables a dialogue with both the non-German-speaking children and their families. If utilizing translations, it should cover the most frequently spoken languages by patients in Swiss emergency units. The tool should base on pictures or pictograms and has to provide well and clearly designed illustrations. This means that they must be unambiguous, of high contrast, scalable, medically correct, and understandable by people with different cultural background. Content-wise the communication aid should cover the topics of the SAMPLE schema (Symptoms, Allergies, Medication, Past medical history, Last oral intake, Events prior to incident) and consider the most frequent children's diseases and accidents. The structure should be flexible and the content individualizable to address the patient's current condition. We recognize a certain contradiction between the need to have a very simple tool and one that covers as many languages and as much information as possible. From the results, we came up with three solutions for our communication tool: (1) A tool that offers a collection of images that can be shown to the patient like in a "Point it: Traveller's language kit" book, (2) A tool that comprises translated phrases along the SAMPLE interrogation schema accompanied by explaining images, (3) A chat tool allowing to combine emojis or pictograms to picture phrases.

3.2. Prototype

We decided for solution 1 for prototype development, because an image collection to show during the conversation is easy to use. Translations would not be necessary since it is only based on images. Our communication aid comprises the following features: Responsive design for tablets and mobile phones, search function to identify relevant images, automatic suggestion of related images and translation of phrases. A screenshot can be seen in Fig. 1. Images are grouped along 10 categories: administrative issues, allergies, general condition, nutrition and excretion, material, medication, symptoms, accidents, existing medical conditions, time and numbers. The current version can be accessed through the web: https://sprechendebilder.ch/. The final prototype is developed as web app using Ionic/Vue.js. This means the structure of the app is similar to a Vue.js-Project; rendering and building of the app is realized by the Ionic framework. Image information is stored as JSON. All images included in the prototype were designed by members of the project team (BK, LO). For each image, meta data has been created manually including names / describing terms, categories, keywords, related images.

Google Cloud Translations API was chosen since it provides translations for the 11 most frequently used main languages[4] in Switzerland (the often-mentioned Tigrinya is missing, however).

3.3. User testing results

Five user testings were conducted between November 19 and 22, 2021. Three tests were conducted directly in the children emergency department (but in a controlled setting) of two hospitals. Two testings were realized within the university. Among the parents, we had 3 mothers (Argentine, Italian, Philippine) and two fathers (Brasilian, Syrian). Three nurses were working in children emergency care, two were former master students, one of whom currently works on the pediatric emergency unit. The user tests lasted one hour in average. Between two and three case scenarios were considered per user testing.

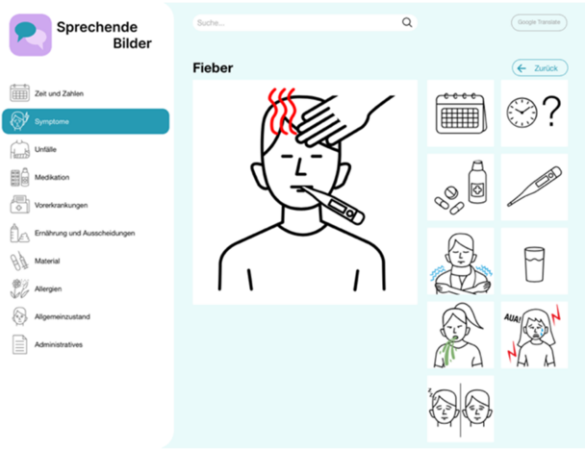


Figure 1. Tablet view of the prototype: On the left, a category can be selected (e.g. symptoms, medication, allergies). In the center, the image referring to fever has been selected and is shown enlarged. On the right, related images are shown. The search function is accessible through the data entry field on top.

During the role play, the images from the tool were more frequently used over time to communicate with the parent. The tablet was placed in the middle of the table in a way all persons could look at it. Looking at the pictures together allowed for an empathetic dialogue. Feedback on the images was entirely positive. They were found to be appealing, easy to understand and helpful, especially for the quick assessment of symptoms or the circumstances of an accident. The use of the translation functionality interrupted the interaction since only one person typed while the other person had to wait. Translations were complex or wrong; certain languages were unavailable; auto correct was enabled and corrected foreign language terms. The translation functionality seems to be not helpful. Nurses confirmed afterwards that they would use the tool in daily routine right from the beginning of the parent interrogation and not only when communication problems occur.

The testings showed that multiple images could be added, for example images on additional types of accidents or existing medical conditions. A category that is in the

[4] https://www.bfs.admin.ch/bfs/de/home/statistiken/bevoelkerung/sprachen-religionen/sprachen.html

current prototype entirely missing is the follow-up procedure, i.e. images that describe which examinations will be conducted.

We identified three groups of information that were difficult to express using the images. First, the participants had problems in showing time ranges using the available images. Second, it was difficult for nurses to ask for previous illnesses or chronic diseases, i.e. illnesses or accidents that took place in the past or symptoms that occurred in the past. Third, the app and its images do not allow to express quantities. Information on allergies was also difficulty to express using the images. The nurses suggested to include a protocol functionality to the app storing which images had been selected to be able to re-capture the communication afterwards. Another extension is to enable combining certain images with other images or with numbers (e.g. to express how often the child vomited). In this way more complex phrases could be expressed. To facilitate communication, each symptom could be combined with a yes/no selection option.

4. Conclusions

An image-based communication tool to support interaction between migrant parents and nurses in emergency cases is perceived useful. It allows nurses to quickly assess the situation and provides the basis for an appreciative conversation. Many aspects can be communicated using images, however, there are limitations in purely image-based communications. We learned that it is non-trivial to create understandable images related to time or images of difficult-to-present subjects, e.g. general condition. Our project clearly shows that the images developed, especially for symptoms, accidents or nutrition and excretion, seem to be well comprehensible for triage and anamnesis, but the temporal classification or chronological occurrence of health incidents was difficult. We are planning to extend the tool by images on the follow-up procedure. A next step will be a real-world user testing, i.e. the tool will be used in real emergency cases.

Acknowledgements: We acknowledge funding of this project from the foundation Bangerter Rhyner and Bern University of Applied Sciences. We thank all study participants for their contributions. Beyond, we would like to thank Lea Meier, Joshua Drewlow and François von Kaenel for implementing the prototype.

References

[1] T. Langer, S. Wirth. Overcoming language barriers with telephone interpreters: first experiences at a German children's hospital. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen*. 2014 Jan 1;108(5):278-82.

[2] F.N. Jaeger, L. Kiss, M. Hossain, C. Zimmerman. Migrant-friendly hospitals: a paediatric perspective-improving hospital care for migrant children. *BMC health services research.* 2013 Dec;13(1):1-4.

[3] L. Schrier et al. Medical care for migrant children in Europe: a practical recommendation for first and follow-up appointments. *Eur J Pediatr.* 2019 Sep;178(9):1449-1467.

[4] R.G. Nijman, J. Krone, S. Mintegi, et al. Emergency care provided to refugee children in Europe: RefuNET: a cross-sectional survey study. *Emergency Medicine Journal*. 2021 Jan 1;38(1):5-13.

[5] B. Kaufmann, T. Helfer, D. Pedemonte, M. Simon, S. Colvin. Communication challenges between nurses and migrant pediatric patients. *Journal of Research in Nursing.* 2020 May;25(3):256-74.

[6] B. Kaufmann, M.A. Simon, T. Helfer. "It somehow worked in the end": Managing demanding communication situations between nurses and migrant families in the pediatric hospital setting through the use of communication aids. In: *Proceedings of the 6th International Conference on Design4Health 2020*: 2, pp. 47-56.

Semantic Interoperability

This page intentionally left blank

Healthcare of the Future 2022
T. Bürkle et al. (Eds.)

doi:10.3233/SHTI220314

Evaluation of Domain-Specific Word Vectors for Biomedical Word Sense Disambiguation

Dennis TODDENROTH [1]
Chair of Medical Informatics, University Erlangen-Nuremberg, Germany

Abstract. Among medical applications of natural language processing (NLP), word sense disambiguation (WSD) estimates alternative meanings from text around homonyms. Recently developed NLP methods include word vectors that combine easy computability with nuanced semantic representations. Here we explore the utility of simple linear WSD classifiers based on aggregating word vectors from a modern biomedical NLP library in homonym contexts. We evaluated eight WSD tasks that consider literature abstracts as textual contexts. Discriminative performance was measured in held-out annotations as the median area under sensitivity-specificity curves (AUC) across tasks and 200 bootstrap repetitions. We find that classifiers trained on domain-specific vectors outperformed those from a general language model by 4.0 percentage points, and that a preprocessing step of filtering stopwords and punctuation marks enhanced discrimination by another 0.7 points. The best models achieved a median AUC of 0.992 (interquartile range 0.975 - 0.998). These improvements suggest that more advanced WSD methods might also benefit from leveraging domain-specific vectors derived from large biomedical corpora.

Keywords. Word sense disambiguation, word vectors, linear classifiers.

1. Introduction

The increasing electronic availability of various biomedical texts presents novel opportunities to reuse such information resources with methods from computational linguistics, or natural language processing (NLP). Medical or scientific NLP applications may for example verify the completeness of clinical documentation [1], or compute and evaluate concise summaries of clinical trial descriptions [2].

Relevant NLP subtasks include word sense disambiguation (WSD), which aims to estimate intended homonym meanings from surrounding text. Typically, human readers can intuitively infer from context whether a mentioned word such as *cortex* refers to a part of the brain or to a substructure of the adrenal glands. Due to the variety of enclosing syntax and semantics, however, implementing such a function in an NLP algorithm can be rather difficult.

Recently developed NLP methods include word vectors or embeddings that derive high-dimensional numeric representations from processing local co-occurrence patterns in large corpora. Since related words tend to be used in similar contexts, such word vectors reflect semantic similarity as proximity in vector space. Word vectors thus encode meaning in a form that combines computability with nuanced semantics, which

[1] Corresponding Author; E-mail: dennis.toddenroth@fau.de.

may exceed the capabilities of discrete vocabularies and explicitly defined NLP algorithms.

The effectiveness of word vectors depends on the size of the training corpora, as well as on their representativeness for subsequent textual inputs. The authors of *scispaCy* [3], a modern NLP library that includes word vectors trained in vast biomedical corpora, therefore propose that their package should be particularly suitable for processing domain-specific texts. Below we explore the utility of *scispaCy* word vectors for biomedical WSD by training and evaluating a set of simple linear classifiers.

2. Methods and Data

To explore the utility of domain-specific word vectors for WSD, we consider homonyms that are prevalent in Pubmed abstracts. Pubmed queries that leverage the controlled vocabulary used to index the underlying literature database (Medical Subject Headings, or MeSH) can produce highly selective results when search terms are represented in MeSH terms.

To disambiguate words with technical meanings that are increasingly used in publications from the field of digital medicine, we manually labeled 1,493 abstracts. These annotations resolve words such as *Matlab*, which is not part of MeSH, and which can either denote a programming language or a region in Bangladesh. Other labels differentiate whether *Java* refers to a programming language or to an island in Indonesia, whether *Python* denotes a serpent, and whether *R* is used in the sense of a statistical correlation. While alternative meanings may occur with different frequencies in the biomedical literature, annotations were assigned so that the frequencies of alternative meanings are rather similar (see legend in figure 1); this procedure aimed to optimize the efficiency of the analysis by maximizing the number of pairs of abstracts from which linguistic differences between alternative contexts can be derived. Another part of the analysis involved a subset of four medical homonyms that had been semi-automatically annotated in 923 abstracts from the MSH WSD data set in order to support the development and assessment of such algorithms [4].

Word vectors are originally associated with single tokens, but document-specific vectors can be computed by aggregating vectors weighted by the frequency of mentioned words. As a consequence of their high dimensionality, these aggregate vectors may encode the “average meaning” of a document or context remarkably well, even when the information contained in word sequences is entirely discarded. The left side of figure 1 shows the first two dimensions of a principle component analysis of abstract-specific vectors for the first one of the described data sets, computed via the R function *prcomp()*, colored according to the respective WSD labels.

Even if this image obscures substantial information from the other dimensions with lower variance, we can see some label disaggregation, which can in turn be used to estimate annotations. The arrows depicted in figure 1 represent pairwise differences between projected average locations or gravity centers, so that the scalar product of each (un-projected) directional vector and those of separate samples optimizes the association with the respective labels.

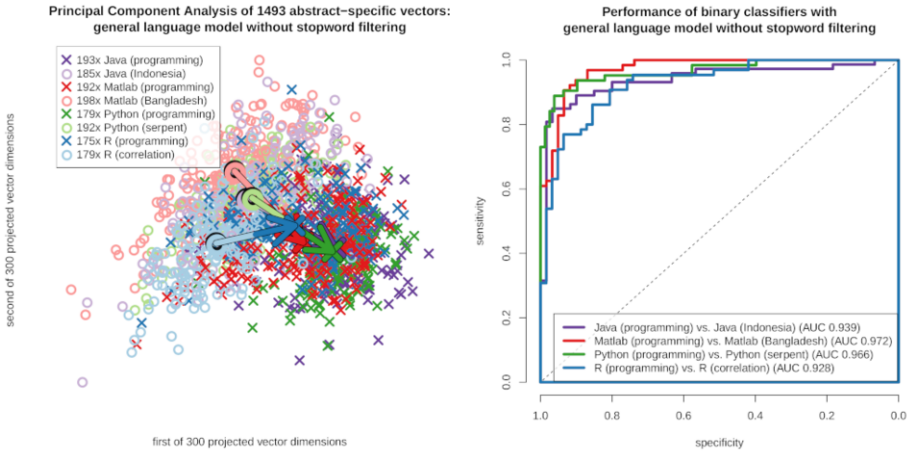


Figure 1. Principle components of word vectors for 1493 abstracts, colored according to manually defined annotations; arrows highlight pairwise differences between projected label-specific average locations (left). Receiver Operating Characteristic curves demonstrate how well a linear classifiers trained in a bootstrap subsample reproduces disjunctive held-out labels (right).

Natural language contains frequent words such as articles and conjunctions that are important for comprehending coherent texts, while their sheer presence of absence is relatively unspecific for alternative homonym contexts. Since the described aggregation already ignores word sequences, punctuation marks and so-called stopwords may be seen as noise in terms of the "average meaning" in vector space. Therefore, we evaluated the effect of removing stopwords on WSD performance.

From these data sets with 1,493 and 923 pairs of vectors and labels, partitions with disjunctive training and evaluation subsamples were repeatedly selected via bootstrapping. This procedure trained classifiers on rows that were randomly sampled with replacement, while disjunctive hold-out samples served to evaluate the respective classifier (200 repetitions). WSD performance was then measured as the area under Receiver Operating Characteristic curves that summarize the achievable constellations of sensitivity and specificity (ROC-AUC as in figure 1).

Word vectors were calculated under Python 3.8 with *spaCy* 3.0 and its general *en_core_web_md* language model (vectors with 300 dimensions) as well as *scispaCy*'s domain-specific *en_core_sci_lg* model in version 0.4 (200d). Classifier training and evaluation was implemented in R 3.6. Annotation data, script sources, and computed vectors are available at https://www.github.com/dtoddenroth/embeddingswsd/.

3. Results

For eight homonyms, four algorithmic configurations and 200 bootstrap repetitions, WSD effectiveness as measured by the median ROC-AUC in a total of 6,400 models was 0.976, with an overall range from 0.748 to 1.000 and an interquartile range from 0.945 to 0.991. Figure 2 summarizes the respective ROC-AUC distributions for WSD tasks and different algorithmic configurations. Note that the displayed vertical scale in figure 2 is restricted to ROC-AUC ranges between 0.8 and 1.0.

When aggregating model performance across tasks and bootstrap repetitions, we find that classifiers trained on domain-specific vectors outperformed those from a general language model by 4.0 percentage points (percentages of the unit square under a ROC curve), while a preprocessing step of filtering stopwords and punctuation marks enhanced discrimination by another 0.7 points. After pooling AUC values from WSD tasks and bootstrap repetitions in order to compare algorithmic configurations, the best constellation that combined stopword filtering with the specialized biomedical language model achieved a near-optimal median AUC of 0.992 (interquartile range 0.975 to 0.998).

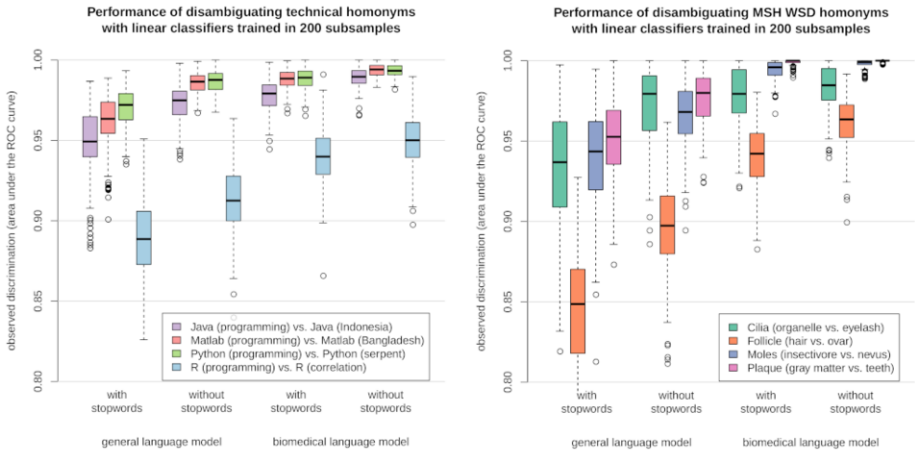


Figure 2. Tukey's boxplots visualize the variance of classifier performance (ROC-AUC) in 200 bootstrap partitions for disambiguating four technical homonyms (left) and four homonyms from the MSH WSD data set (right) for different vector sets and algorithmic configurations.

4. Discussion

Our findings indicate that linear classifiers based on domain-specific vectors outperformed those from the general language model. A preprocessing step of filtering stopwords and punctuation marks also improved model discrimination. The difficulty of the tested disambiguation tasks seemed to vary, and the most effective classifiers achieved near-optimal performance on the easier tasks.

The described method may seem naïve in the sense that it ignores all information contained in the particular word sequences around homonyms. The corresponding computational simplicity based on linear algebra, on the other hand, could facilitate implementing the approach for wider usage. When meaningful vector representations of documents can be stored in a relational database or in processing units that are optimized for parallel matrix multiplication, it seems feasible that vector-based document filtering and ordering can be realized with favorable flexibility and scalability.

Suppose that a researcher prepares a manuscript and wants to round out her set of cited references with additional relevant sources. While explicit search terms might of course be useful for finding candidate publications, homonyms among feasible criteria such as *follicle* might retrieve heterogeneous suggestions from dermatology as well as from gynecology, while her research may be concerned with only one of these topics. A

word vector derived from her manuscript or from the abstracts of cited references might then discriminate relevance more precisely than any constellation of search terms. Due to their nuanced high-dimensional representation, task-specific word vectors could even outperform explicit queries that do not include any homonyms. Since explicit MeSH labels are characteristically available for a subset of publications in the entire Pubmed literature database, future research might study the comparative or complementary value of vector-based predictions for WSD tasks, or could explore how well vector-based NLP models can impute MeSH labels for un-indexed articles.

Beyond the biomedical literature, querying textual patient records or clinical trial descriptions might also constitute promising use cases for classifiers based on word vectors. Medical informatics researchers have for some time developed and evaluated computerized trial recruitment systems that aim to support patient enrollment in ongoing medical studies, including with NLP methods [5] and machine learning models [6]. While we cannot expect word vectors to adequately capture all details such as intricate temporal criteria [7], part of the difficulty of matching patients to suitable trials may be attributable to ambiguous phrases in textual descriptions. If an eligibility criterion for example refers to a condition such as *plaque*, a word vector from its context could automatically suggest fitting neurological or dental patient records. Future studies might therefore explore whether the WSD effectiveness that we observed in scientific texts can be reproduced in individual patient-level clinical documents.

Previous research has deployed vector calculus for biomedical WSD [8], including in conjunction with recurrent convolutional neural networks based on self-trained embeddings [9]. The advanced performance that we observed when using simple linear classifiers based on domain-specific vectors suggests that future research might investigate whether such refined sequence-aware WSD models will also benefit from using vectors that had been pre-trained on large biomedical corpora.

References

[1] Schulz S, Seddig T, Hanser S, Zaiss A, Daumke P. Checking coding completeness by mining discharge summaries. Stud Health Technol Inform. 2011;169:594-8.

[2] Gulden C, Kirchner M, Schüttler C, Hinderer M, Kampf M, Prokosch HU, Toddenroth D. Extractive summarization of clinical trial descriptions. Int J Med Inform. 2019 Sep;129:114-121. doi: 10.1016/j.ijmedinf.2019.05.019.

[3] Neumann M, King D, Beltagy I, Ammar W. ScispaCy: Fast and Robust Models for Biomedical Natural Language Processing. Proceedings of the 18th BioNLP Workshop and Shared Task 2019. doi: 10.18653/v1/W19-5034.

[4] Jimeno-Yepes AJ, McInnes BT, Aronson AR. Exploiting MeSH indexing in MEDLINE to generate a data set for word sense disambiguation. BMC Bioinformatics. 2011 Jun 2;12:223. doi: 10.1186/1471-2105-12-223.

[5] Köpcke F, Prokosch HU. Employing computers for the recruitment into clinical trials: a comprehensive systematic review. J Med Internet Res. 2014 Jul 1;16(7):e161.

[6] Köpcke F, Lubgan D, Fietkau R, Scholler A, Nau C, Stürzl M, Croner R, Prokosch HU, Toddenroth D. Evaluating predictive modeling algorithms to assess patient eligibility for clinical trials from routine data. BMC Med Inform Decis Mak. 2013 Dec 9;13:134.

[7] Mate S, Bürkle T, Kapsner LA, Toddenroth D, Kampf MO, Sedlmayr M, Castellanos I, Prokosch HU, Kraus S. A method for the graphical modeling of relative temporal constraints. J Biomed Inform. 2019 Dec;100:103314.

[8] McInnes B. An Unsupervised Vector Approach to Biomedical Term Disambiguation: Integrating UMLS and Medline. Proceedings of the ACL-08: HLT Student Research Workshop, Columbus, Ohio. 2008.

[9] Festag S, Spreckelsen C. Word Sense Disambiguation of Medical Terms via Recurrent Convolutional Neural Networks. Stud Health Technol Inform. 2017;236:8-15.

Healthcare of the Future 2022
T. Bürkle et al. (Eds.)

doi:10.3233/SHTI220315

Classification of Patient Portals Described in Evaluation Studies Using the TOPCOP Taxonomy

Ines LEB[a,1], Selina MAGNIN[b], Martin BOEKER[c], Hans-Ulrich PROKOSCH[a], Elske AMMENWERTH[d] and Michael GLÖGGLER[d]

[a] *Friedrich-Alexander-University Erlangen-Nürnberg, Chair of Medical Informatics*
[b] *Medical Data Integration Center, Dept. for IT and Applied Medical Informatics, University Hospital Tübingen, Translational Bioinformatics, University of Tübingen*
[c] *Technical University of Munich, School of Medicine, Medical Center rechts der Isar, Institute for Artificial Intelligence and Informatics in Medicine – AIIM*
[d] *Institute of Medical Informatics, UMIT – Private University for Health Sciences, Medical Informatics and Technology, Hall in Tirol, Austria*

Abstract. Many patient portals have been introduced and evaluated in recent years. The results of evaluation studies are difficult to compare, however, as the evaluated patient portal is often not clearly or only incompletely described in the publication. This problem is common to evaluations in health informatics. We evaluated the completeness of descriptions of patient portals in 15 exemplary evaluation publications using the TOPCOP taxonomy. Our results show that core functionalities such as portal design, patient communication, educational features, or system notifications were quite clearly described in all 15 evaluation studies. Other descriptions, such as web accessibility or data management, were often not provided. We conclude that taxonomies such as TOPCOP should be used and even required for describing interventions in evaluation papers.

Keywords. Patient portal, taxonomy, evaluation, health informatics, information management

1. Introduction

A patient portal is a web-based application that allows patients to access their health-related data stored in the Electronic Health Record of a healthcare organization [1]. Due to their expected benefit for patient empowerment and quality of care, healthcare institutions have introduced patient portals in recent years, and researchers have evaluated their impact.

The results of evaluation studies are difficult to compare, however, as the evaluated patient portal is not clearly or only incompletely described in the publication. This problem is common to evaluations in health informatics, which has led to the development of recommendations for reporting evaluation studies [2]. Classifying patient portals using a taxonomy may help to reduce the problem [1].

[1] Ines Leb, Friedrich-Alexander-Universität Erlangen-Nürnberg, Wetterkreuz 15, 91058 Erlangen, E-Mail: ines.leb@fau.de

A taxonomy is a classification system to assign similar objects of a domain into groups based on distinct characteristics and offers a set of decision rules [3,4]. The reduction of complexity, identifying similarities and differences among objects, and the understanding of interrelationships are major advantages of taxonomies [5,6]. Furthermore, taxonomies help to systematize and enhance knowledge by observing and analyzing a domain, thus contributing to providing knowledge [7].

In the absence of a specific taxonomy for comparing evaluation studies of patient portals, we considered the TOPCOP taxonomy the best approach for our research. With 25 dimensions based on 65 characteristics, the TOPCOP taxonomy has a sufficient number of dimensions and characteristics to discriminate among patient portals and sufficiently explain the patient portals independent of clinical setting or country [4,8].

The aim of this study was to evaluate the completeness of the patient portal descriptions in evaluation studies by using the TOPCOP taxonomy.

2. Methods

For our study, we selected 15 exemplary patient portal evaluation studies. We took all ten studies from a most recent Cochrane Review on Randomized Controlled Trials on patient portals [1], as this review comprised the most recent complete lists of published RCTs on patient portals. We added five non-RCT studies to cover other study designs with assumed different publication quality.

To evaluate the studies, we used the TOPCOP taxonomy [8,9] (Figure 1). We read the full text of the 15 evaluation studies and searched for a description of the intervention. TOPCOP was used to classify this description as best as possible. Ambiguous decisions were discussed by all authors until consensus was reached.

3. Results

We classified all 15 studies according to the TOPCOP taxonomy, although many of these dimensions were not specified in the evaluation papers. Some studies mentioned a dimension but gave insufficient information to fully classify a paper according to the characteristic. In these cases, the classification was done by interpreting the available information. In cases where not enough information was available, we opted for the characteristic which indicates the non-existence of the respective feature. Hence, a feature defined as missing does not necessarily mean that the patient portal does not provide this feature, but only that the evaluation paper did not provide enough information.

We found ten described patient portals which were tethered to an EHR [10–19] and five that were integrated [20–24]. Eight patient portals were designed for secondary care [11–13,17,18,20–22], three for primary care [19,23,24], and one for tertiary care [16]. Another eight studies stated a disease-specific portal specialization [11,14,16–18,21,22,24], two universal [20,21], two extended [12,19], and another seven had a medical specialization [11,13,17,18,20,22,24]. One study described an insight to the activity monitoring [15].

ASPECTS	DIMENSIONS	CHARACTERISTICS			
Portal Design	D1: Portal Type	tethered		integrated	
	D2: Care Sector Target	primary care	secondary care	tertiary care	generic
	D3: Patient Target	outpatient		in & outpatient	
	D4: Portal Specialization	universal	extended	disease-specific	
	D5: Medical Specialty	generic		specialized	
	D6: Web Accessibility	not supported		supported	
	D7: App Expandability	not expandable		expandable	
	D8: Activity Monitoring	no insight		with insight	
Management	D9: Appointment Booking	no booking	request	schedule	hybrid
	D10: Prescription Renewal	no renewal		with renewal	
	D11: Portal Customizability	not customizable		customizable	
Communication	D12: E-Consult	no e-consult	asynchronous	synchronous	both
	D13: System Notifications	no notifications	notifications	reminder	alerts
Instruction	D14: Patient Education	no education	non-personalized	personalized	
	D15: Therapy Instructions	no instructions	non-protocol-based	protocol-based	
Self-Management	D16: Medication Summary	no summary		with summary	
	D17: Health Monitoring	no monitoring	self-reported	self-tracked	combined
	D18: Visit Preparation	no preparation		with preparation	
Self-Determination	D19: Declaration of Will	no registration		with registration	
	D20: Second Opinion	no inquiry		with inquiry	
	D21: Study Sign-Up	no sign-up		with sign-up	
Data Management	D22: Record Access	no control	shared control	full control	
	D23: Records Management	no management		with management	
	D24: Health Data Amend	review	correct	delete	
	D25: Health Data Upload	no upload		with upload	

Figure 1. TOPCOP taxonomy of patient portals [8].

For the portal design, the dimensions Web Accessibility and App Expandability were not specified by any study.

For the aspects of management, six studies reported a function for prescription renewal [12,16,18,21–23], three an appointment booking request [12,16,22], and three an appointment booking scheduling [15,21,23]. Only two described the ability to customize the patient portal [15,16].

One of the most reported specifications concerned the aspect of communication. An asynchronous e-Consult was provided by twelve patient portals [10,12,13,16–18,20–23], one synchronous [20], and one both [15]. In the dimension of system notifications, eight patient portals provided a reminder function [10,13,17,19,21–24], four alerts [11,16,18,20], and two notifications [15,18].

Another frequently specified aspect was instructions with seven patient portals which provided non-personalized education [13,15–17,20,21,24], five personalized education [10,11,18,19,22], seven had protocol-based therapy instructions [15,17–22], and three non-protocol-based [10,11,13].

Several studies reported self-management aspects as medication summary provided in twelve patient portals [10–12,14–18,20–23] and three visit preparation [15,22,23]. Health monitoring was specified as self-reported by five studies [10,11,21–23], combined by three studies [15,16,24], self-tracked by two [17,18], and one had no monitoring [19].

Only a few studies reported on self-determination aspects. Three provided a study sign-up [10,15,24] and two papers were assessed as lacking this function [13,21]. One

patient portal had a function for a declaration of will registration [23] and one an inquiry for a second opinion [15].

The last classified aspect was data management, which was not specified very often. Two studies stated shared control to record access [15,24], two no control, and one full control [16]. One patient portal enabled record management [15] and four described no available management [10,12,21,22]. A health data amend review was available in five patient portals [10,12,19,21,24] and one allowed the correction [22]. One portal described that no upload of health data is possible [22].

4. Discussion

Already in 2013, Goldzweig et al [25] concluded that, among other things, a better understanding of the implementation factors for patient portals is required. Kruse et al [26] reported in their systematic review of patient portals that often only specific features were analyzed instead of the full patient portal. Ammenwerth et al [27] stated in their review of patient portals that an aggregation of evidence is needed which could be achieved by a taxonomy.

We used the TOPCOP taxonomy to evaluate the completeness of descriptions of the patient portal in evaluation studies. We found that the description of core functionalities such as portal design, patient communication, educational features, or system notifications were quite clearly described in all 15 evaluation studies. Other descriptions, such as web accessibility or data management, were not provided in many studies.

We could not see any difference in completeness between RCTs and non-RCT studies. Both had gaps in describing their intervention.

The TOPCOP taxonomy was developed for health information managers to classify and compare patient portals to help them choose the most suitable solution for their needs [8,9]. We used the taxonomy for classifying patient portals used in evaluation studies. We can recommend this taxonomy for specification and comparison of evaluation studies. Since most taxonomies are built for a specific purpose, a different purpose may lead to a different taxonomy structure and other characteristics. This was the first research for evaluating the completeness of patient portal descriptions in 15 evaluation studies applying the TOPCOP taxonomy. Therefore, additional research may assess the suitability of TOPCOP for this purpose on a broader approach.

Another direction for subsequent research could be to assess whether further concepts should be added to the taxonomy to extend its usefulness, e.g. to classify evaluation studies. Van Mens et al [28] adapted the Clinical Adoption Framework (CAF) to expand the analysis of EHR adoption with the patients as an end-user. CAF classifies categories about people, organization, system quality, system use, and net benefits. Future work may compare if aspects described in CAF could be integrated into the TOPCOP taxonomy and so improve the taxonomy's usefulness.

Since the evaluation studies were often unclear in many dimensions, we interpreted the paper in a team of two authors, reaching consensus in all cases. We focused our study on 15 papers. A larger sample of evaluation studies could now be reviewed to confirm our results.

We conclude that the completeness of portal description was often incomplete. Health informatics should increase the use of taxonomies such as TOPCOP to better describe intervention in evaluation studies. This would better make it possible to compare and summarize the published evidence and as a basis for Evidence-Based

Health Informatics [29]. Further, all patient portals are in continuous development and change. Therefore an institution similar to HON (Health On the Net) which promotes deployment of useful and reliable internet-based health information, enabling its appropriate and efficient use [30] would also be useful in the context of patient portals to tackle the challenge of sustained and regular evaluation of such portals.

REFERENCES

[1] E. Ammenwerth, S. Neyer, A. Hörbst, G. Mueller, U. Siebert, and P. Schnell-Inderst, Adult patient access to electronic health records. Cochrane Database Syst Rev **2** (2021), CD012707.

[2] J. Talmon, E. Ammenwerth, J. Brender, N. de Keizer, P. Nykänen, and M. Rigby, STARE-HI--Statement on reporting of evaluation studies in Health Informatics. Int J Med Inform **78** (2009), 1–9.

[3] J.A. HAMPTON, Women, Fire, and Dangerous Things. Mind & Language **4** (1989), 130–137.

[4] R.C. Nickerson, U. Varshney, and J. Muntermann, A method for taxonomy development and its application in information systems. European Journal of Information Systems **22** (2013), 336–359.

[5] R.C. Nickerson, J. Muntermann, and U. and Varshney, Taxonomy Development in Information Systems: A Literature Survey and Problem Statement. In *AMICS 2010 Proceedings. 125.*, pp. 1–11.

[6] K. Bailey, *Typologies and Taxonomies,* SAGE Publications, Inc, 2455 Teller Road, Thousand Oaks California 91320 United States of America, 1994.

[7] J.R. Krobock, A taxonomy: hospital information systems evaluation methodologies. J Med Syst **8** (1984), 419–429.

[8] M. Glöggler and E. Ammenwerth, *Improvement and Evaluation of the TOPCOP Taxonomy of Patient Portals: Taxonomy-Evaluation-Delphi (TED) Approach,* 2021.

[9] M. Glöggler and E. Ammenwerth, Development and Validation of a Useful Taxonomy of Patient Portals Based on Characteristics of Patient Engagement. Methods Inf Med **60** (2021), e44-e55.

[10] K.P. McCarrier, J.D. Ralston, I.B. Hirsch, G. Lewis, D.P. Martin, F.J. Zimmerman, and H.I. Goldberg, Web-based collaborative care for type 1 diabetes: a pilot randomized trial. Diabetes Technol Ther **11** (2009), 211–217.

[11] S. Ahmed, P. Ernst, S.J. Bartlett, M.-F. Valois, T. Zaihra, G. Paré, R. Grad, O. Eilayyan, R. Perreault, and R. Tamblyn, The Effectiveness of Web-Based Asthma Self-Management System, My Asthma Portal (MAP): A Pilot Randomized Controlled Trial. J Med Internet Res **18** (2016), e313.

[12] L. Sulieman, B. Steitz, and S.T. Rosenbloom, Analysis of Employee Patient Portal Use and Electronic Health Record Access at an Academic Medical Center. Appl Clin Inform **11** (2020), 433–441.

[13] S.E. Ross, L.A. Moore, M.A. Earnest, L. Wittevrongel, and C.-T. Lin, Providing a web-based online medical record with electronic communication capabilities to patients with congestive heart failure: randomized trial. J Med Internet Res **6** (2004), e12.

[14] K. Kashiwagi and S. Tsukahara, Impact of patient access to Internet health records on glaucoma medication: randomized controlled trial. J Med Internet Res **16** (2014), e15.

[15] G. Di Tosto, A.S. McAlearney, N. Fareed, and T.R. Huerta, Metrics for Outpatient Portal Use Based on Log File Analysis: Algorithm Development. J Med Internet Res **22** (2020), e16849.

[16] P.J. Wagner, J. Dias, S. Howard, K.W. Kintziger, M.F. Hudson, Y.-H. Seol, and P. Sodomka, Personal health records and hypertension control: a randomized trial. J Am Med Inform Assoc **19** (2012), 626–634.

[17] J.D. Ralston, I.B. Hirsch, J. Hoath, M. Mullen, A. Cheadle, and H.I. Goldberg, Web-based collaborative care for type 2 diabetes: a pilot randomized trial. Diabetes Care **32** (2009), 234–239.

[18] P.C. Tang, J.M. Overhage, A.S. Chan, N.L. Brown, B. Aghighi, M.P. Entwistle, S.L. Hui, S.M. Hyde, L.H. Klieman, C.J. Mitchell, A.J. Perkins, L.S. Qureshi, T.A. Waltimyer, L.J. Winters, and C.Y. Young, Online disease management of diabetes: engaging and motivating patients online with enhanced resources-diabetes (EMPOWER-D), a randomized controlled trial. J Am Med Inform Assoc **20** (2013), 526–534.

[19] A.H. Krist, S.H. Woolf, S.F. Rothemich, R.E. Johnson, J.E. Peele, T.D. Cunningham, D.R. Longo, G.A. Bello, and G.R. Matzke, Interactive preventive health record to enhance delivery of recommended care: a randomized trial. Ann Fam Med **10** (2012), 312–319.

[20] K.O. Schnock, J.E. Snyder, T.E. Fuller, M. Duckworth, M. Grant, C. Yoon, S. Lipsitz, A.K. Dalal, D.W. Bates, and P.C. Dykes, Acute Care Patient Portal Intervention: Portal Use and Patient Activation. J Med Internet Res **21** (2019), e13336.

[21] S.S. Woods, C.W. Forsberg, E.C. Schwartz, K.M. Nazi, J.H. Hibbard, T.K. Houston, and M. Gerrity, The Association of Patient Factors, Digital Access, and Online Behavior on Sustained Patient Portal Use: A Prospective Cohort of Enrolled Users. J Med Internet Res **19** (2017), e345.

[22] R.W. Grant, J.S. Wald, J.L. Schnipper, T.K. Gandhi, E.G. Poon, E.J. Orav, D.H. Williams, L.A. Volk, and B. Middleton, Practice-linked online personal health records for type 2 diabetes mellitus: a randomized controlled trial. Arch Intern Med **168** (2008), 1776–1782.

[23] J.B. Jones, J.P. Weiner, N.R. Shah, and W.F. Stewart, The wired patient: patterns of electronic patient portal use among patients with cardiac disease or diabetes. J Med Internet Res **17** (2015), e42.

[24] A. Holbrook, L. Thabane, K. Keshavjee, L. Dolovich, B. Bernstein, D. Chan, S. Troyan, G. Foster, and H. Gerstein, Individualized electronic decision support and reminders to improve diabetes care in the community: COMPETE II randomized trial. CMAJ **181** (2009), 37–44.

[25] C.L. Goldzweig, G. Orshansky, N.M. Paige, A.A. Towfigh, D.A. Haggstrom, I. Miake-Lye, J.M. Beroes, and P.G. Shekelle, Electronic patient portals: evidence on health outcomes, satisfaction, efficiency, and attitudes: a systematic review. Ann Intern Med **159** (2013), 677–687.

[26] C.S. Kruse, K. Bolton, and G. Freriks, The effect of patient portals on quality outcomes and its implications to meaningful use: a systematic review. J Med Internet Res **17** (2015), e44.

[27] E. Ammenwerth, A. Hoerbst, S. Lannig, G. Mueller, U. Siebert, and P. Schnell-Inderst, Effects of Adult Patient Portals on Patient Empowerment and Health-Related Outcomes: A Systematic Review. Stud Health Technol Inform **264** (2019), 1106–1110.

[28] H.J.T. van Mens, R.D. Duijm, R. Nienhuis, N.F. de Keizer, and R. Cornet, Towards an Adoption Framework for Patient Access to Electronic Health Records: Systematic Literature Mapping Study. JMIR Med Inform **8** (2020), e15150.

[29] Ammenwerth, E., Rigby, M.., *Evidence-Based Health Informatics: Promoting Safety and Efficiency Through Scientific Methods and Ethical Policy,* Stud Health Tech Inform. Vol 222. 2016. IOS Press.

[30] Boyer C, Appel RD, Ball MJ, van Bemmel JH, Bergmans JP, Carpentier M, Hochstrasser D, Lindberg D, Miller R, Peterschmitt JC, Safran C, Thonnet M, Geissbühler A. *Health On the Net's 20 Years of Transparent and Reliable Health Information.* Stud Health Technol Inform. 2016;228:700-4. PMID: 27577475.

This page intentionally left blank

Medical Informatics for Medical Research

This page intentionally left blank

Healthcare of the Future 2022
T. Bürkle et al. (Eds.)

doi:10.3233/SHTI220317

Aligning Biobanks and Data Integration Centers Efficiently (ABIDE_MI)

Hans-Ulrich PROKOSCH[a,1], Ronny BABER[b,c], Paula BOLLMANN[b], Marie GEBHARDT[d], Julian GRUENDNER[a] and Michael HUMMEL[e,f]
[a]*Chair of Medical Informatics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany*
[b]*Leipzig Medical Biobank, Leipzig University, Leipzig, Germany*
[c]*Institute of Laboratory Medicine, Clinical Chemistry and Molecular Diagnostics, University Hospital Leipzig, Leipzig, Germany*
[d]*TMF - Technology, Methods, and Infrastructure for Networked Medical Research, Berlin, Germany*
[e]*German Biobank Node, Charité Universitätsmedizin Berlin, Berlin, Germany*
[f]*Central Biobank Charité (ZeBanC), Institute of Pathology, Charité Universitätsmedizin Berlin, Berlin, Germany*

Abstract. ABIDE_MI is a complementary funded 18 months project within the German Medical Informatics Initiative (MII), which aims to align IT infrastructures and regulatory/governance structures between biobanks/biobanking IT and the MII data integration centres (DIC) at German university hospitals. A major task in 2021 was the systematic collection of all documents describing rules, as well as proposal/contract templates for data and biosample use and access at each of the participating 24 university hospitals and their comparison with MII-wide consented data sharing principles, documents and governance structures. This comparison revealed large heterogeneity across the ABIDE_MI sites and further, redundant structures/regulations currently established at the German university hospitals. A second task was the design and stepwise development of an IT network infrastructure with central components (data and biosample query portal) and decentralized standardized FHIR servers to capture the standardized FHIR-based core data set modules (resources) defined within the MII working group "Interoperability". Subsequent steps in the project are the harmonization of the data and biosample sharing governance/regulation frameworks at each ABIDE_MI site, creating synergies for the research infrastructures at the German university hospitals and to link those resources to the German Portal for Medical Research Data and with the BBMRI-ERIC Directory and Negotiator tools.

Keywords. Biobank networks, real world data, medical informatics initiative, data and biosample sharing

1. Introduction

Real world data analysis in medicine today relies on the availability of clinical data as well as information about biospecimen collected during clinical care processes [1].

[1] Corresponding Author, Hans-Ulrich Prokosch, Chair of Medical Informatics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Wetterkreuz 15, 91058 Erlangen, Germany; E-mail: Hans-Ulrich.Prokosch@fau.de.

The harnessing and cross-consortial research use of such data is one of the major goals of the Federal Ministry of Education and Research (BMBF)-funded Medical Informatics Initiative (MII) and its four consortia [2]. Further, networking and collaboration of German biobanks to leverage high quality biobanking throughout Germany has been funded in Germany since 2011, starting with a program for the establishment of centralized biobanks (cBMB Program), the subsequent initiation of the German Biobank Node (GBN, 2013) and its German Biobank Alliance (GBA, 2017) [3, 4]. Both German initiatives (MII and GBN/GBA) aim at establishing federated data networks with distributed data repositories and central feasibility platforms. Unfortunately, until 2020, almost no major synergies between GBA and MII had been realized and partially parallel structures (on technical as well as organizational/regulation level) existed within the German university hospitals. Thus, it is the aim of the ABIDE_MI project (a complementary BMBF-funded project within the MII) to align such structures between biobanks and biobanking IT as well as the data integration centres (DIC), to create synergies for the research infrastructures at the German university hospitals and to link those resources to a central feasibility portal. The objective of this publication is to describe the results achieved after half of the project duration and to illustrate the plan for the second project phase.

2. Methods

ABIDE_MI started in May 2021 and comprises 24 German University Hospitals with 24 DIC and 25 biobanks (one of the University hospitals owns two biobanks at its separate locations in two cities). The project comprises central IT development tasks (e.g. developing a graphical user interface for cohort identification and biosample search (feasibility tool) linked via middleware components to distributed FHIR servers), deployment of such IT tools for all participating university hospitals, and decentralized work packages to be pursued within all university hospitals.

The latter are split into

- organizational tasks, like analyzing all governance and regulatory documents as well as committees associated with data and biosample sharing established at each hospital locally (typically based on previous biobank establishments as well as DIC establishments), and to compare them with the respective documents and regulations agreed upon in the MII working groups consent and data sharing.
- technical developments, such as developing the ETL processes to provide clinical data formatted according to the FHIR implementation guidelines of the basic modules of the MII core data set (person, encounter, diagnosis, procedure, laboratory, medication; see https://www.medizininformatik-initiative.de/en/medical-informatics-initiatives-core-data-set for details), the MII consent information module and biosample data, based on the MII biosample FHIR module.

Further, the IT framework should be designed in such a way, that the central feasibility tool should smoothly fit into the German Portal for Medical Research Data (Deutsches Forschungsdatenportal für Gesundheit = FDPG), currently being implemented by the MII coordination centre, and it should also provide an integration

pathway into the European biobank network BBMRI ERIC, with its federated search tools, the BBMRI Directory and the Sample Negotiator [5, 6].

For the feasibility tool development we built on previous project experiences and developments, e.g. the German Biobank Alliance, the German AKTIN emergency ward registry [7], the MIRACUM project [8] and the development of a national Covid-19 data exchange platform (CODEX) within the German network university medicine (NUM) [9]. The software development is pursued in an agile development process by a team of developers spread over six locations. The comprehensive software architecture relies on a set of micro services interacting with each other to translate the feasibility query user input in a predefined structured query syntax, transfer the queries securely to the local implementations of the network partners and execute the queries on a locally installed FHIR server.

The complete project is coordinated using the Atlassian® confluence collaboration platform and biweekly web conferences. In such regular web conferences biobanking representatives, DIC representatives of all partner sites and the coordination team met to discuss the status of project deliverables/milestones and particularly focus on the alignment of regulations/governance structures established on one side in the local biobank environments and on the other side within the data integration centre environments. Those meetings were especially helpful to present the arguments for and reasons behind all such regulations and structures and achieve a common understanding of the respective historical development within both communities. All preexisting documents from each of the partners (DICs and biobanks) describing e.g. consenting procedures, governance structures, application for the use and material transfer agreements for the exchange of biospecimen and associated clinical data were collected in the collaboration platform and then systematically compared to the central documents/regulations established within the MII, to identify gaps between the needs of the biobanking community and the regulations defined within the MII.

3. Results

3.1. Organizational Tasks

The major result of the first project phase is a technical report describing the current status of all regulations and governance structures established at the 24 participating university hospitals, and their comparison with their counterparts centrally agreed upon in the MII working groups. The documents identified to be important in this context are summarized in table 1.

The gap analysis depicts a heterogeneous situation, with some university hospitals having established biobanks (and the respective governance structures and use/access rules) many years ago and others, which had just recently established a centralized biobank at their university hospital, which enables to directly rely on the structures and regulations defined within the MII. One of the ABIDE_MI partners did neither have a central biobank nor a data integration centre (because up until now it was only a networking partner without DIC in a MII consortium) established until 2021 but uses the ABIDE_MI project to establish both such institutions at its location (this partner was not included in the following statistics).

With respect to the local adoption of the MII overall standardized use and access rules, 16 DIC reported, that they have already implemented this governance instrument,

6 were currently in the process of implementing it and one did not have any data/biosample use and access rules implemented yet. Compared to this, only 8 biobanks had already implemented use and access rules for biosample use in research projects, 3 were in the process of implementing local use and access rules. Only 3 biobanks had adapted the MII overall standardized use and access rules and 8 were in the process of implementing this standardized MII governance instrument. One biobank has just started to work on defining its use and access rules. 3 biobanks were not implementing any use and access rules yet (November 2021). In summary at 5 university hospitals DIC and biobank had already adapted MII overall standardized use and access rules. Use and access committees with already joined boards for biosamples and data existed at 8 university hospitals whereas in 15 university hospitals the data use and access committee and the biosample use and access committee were separate boards.

Table 1. Data/biosample use and access documents analyzed locally and compared with MII consented documents

local documents	MII consented documents
local "broad consent" based on the biobanking broad consent template (template from Arbeitskreis Medizinischer Ethikkommissionen e.V.)	MII "broad consent" template see https://www.medizininformatik-initiative.de/en/template-text-patient-consent-forms for details
local "broad consent" based on the MII broad consent template	
local biobank/DIC by-laws and/or statutes	MII data sharing process model
local biobank/DIC use and access rules	MII overall standardized use and access rules
local biobank/DIC templates for data/biosample use proposals	MII template for data use proposals
local biobank/DIC templates for data/biosample use contracts	contract template governing the use of data and biosamples in the MII

The local board of director approval of the MII broad consent template (which also includes a module for biosample use) was achieved at 20 university hospitals, nevertheless only at 7 of those hospitals the respective biobanks had started to use this broad consent template in their routine processes. 15 biobanks were still using a local biobanking consent form and in 1 biobank the consent process implementation was work in progress.

3.2. Technical Development

The IT development could benefit from earlier developments already pursued within the CODEX project and established a very similar network architecture. While in CODEX the focus was only on Covid-19 patients and thus the data items to be provided were reduced to the GECCO (German Corona Consent) dataset [10], the ABIDE_MI projects aims at providing access to clinical data and biosample information from all hospital patients with a much larger dataset, defined by the six basic modules of the MII core dataset, the consent module and the biosample core dataset module. Therefore, also the ontology used to build queries for ABIDE_MI was changed and created from the FHIR implementation guides of the above mentioned MII dataset modules. Further, the query

UI was enhanced with new features such as combination logic for linked data items and temporal restrictions.

The ABIDE_MI IT framework currently comprises

- a user interface (feasibility UI),
- a backend service which translates the user input into a standardized format (structured query) based on an ontology service [11] and
- an execution service, which can process the standardized format, convert it to queries for a FHIR server and execute the query (this execution service is distributed to all partners in the network).
- middleware components to provide a secure transportation of queries and query results between the central component and the decentral execution services.

The developed tools were built to support any FHIR server, which provides either a FHIR search interface or the ability to execute CQL queries, allowing the participating sites to choose which FHIR server to use.

4. Discussion and outlook

Health care integrated biobanks are usually closely cooperating with the institutional departments of pathology and clinical laboratory. Data governance of the biobanks is typically limited to the direct biosample information (managed e.g. in a biobank management system) and a small set of clinical data arising from the respective departments "own" departmental IT system and other data sources. Therefore, sample search tools, such as e.g., the GBA sample locator [12] are restricted to only a few clinical items. However, as illustrated e.g., by Castro et al. [13], Geiger et al. [14] and Lawrence et al. [15] precision medicine research in the future will require high quality assembled biosamples annotated with a comprehensive spectrum of clinical and molecular data. The integration of those data from historically separated data silos is one of the major goals of the MII. Thus, close coordination between biobanks and data integration centers is inevitable for future innovative medical research. In times of limited resources there is an utmost need to eliminate redundancies at the organizational and technical level.

This process was successfully initiated within ABIDE_MI with the described gap analysis concerning the organizational regulations and structures within the German university hospitals. It is the challenge now for the ABIDE_MI partners to synergistically align their data and biosample frameworks locally and to integrate them into the overall MII data/biosample sharing framework, especially the future German Portal for Medical Research Data. On the other hand, the review of the documents and intense discussions between the biobanking and MII/DIC communities have identified particular needs arising from the more complex process of sharing biosamples (as a limited resource) which are now brought into the MII working groups "consent" and "data sharing" to implement them in the MII framework in order to fully satisfy also the respective biobanking needs. As an example, an ABIDE_MI taskforce was initiated to develop a template for a material transfer agreement (MTA) to be added to the contract template governing the use of data and biosamples in the MII.

In February 2022 the first release deployed for the IT infrastructure was implemented at all ABIDE_MI sites and filled within a small projectathon with the respective MII core data set modules extracted and harmonized at each university hospital. At this point in time six DIC were capable to connect to the feasibility tool with their routine data FHIR server. In parallel we also pursued a usability evaluation of the new feasibility tool in order to receive user feedback and then further enhance the UI and the implemented search ontology during the second half of the project.

References

[1] Maissenhaelter BE, Woolmore AL, Schlag PM. Real-world evidence research based on big data: Motivation-challenges-success factors. Onkologe (Berl). 2018;24 (Suppl 2): 91-98.
[2] Semler SC, Wissing F, Heyder R. German Medical Informatics Initiative. Methods Inf Med. 2018;57(S 01):e50-e56. doi:10.3414/ME18-03-0003.
[3] Schüttler C, Prokosch HU, Hummel M, Lablans M, Kroll B, Engels C; German Biobank Alliance IT development team. The journey to establishing an IT-infrastructure within the German Biobank Alliance. PLoS One. 2021 Sep 22;16(9):e0257632. doi: 10.1371/journal.pone.0257632.
[4] Baber R, Hummel M, Jahns R, et al. Position Statement from the German Biobank Alliance on the Cooperation Between Academic Biobanks and Industry Partners. Biopreserv Biobank. 2019;17(4):372-374.
[5] Holub P, Swertz M, Reihs R, van Enckevort D, Müller H, Litton JE. BBMRI-ERIC Directory: 515 Biobanks with Over 60 Million Biological Samples. Biopreserv Biobank. 2016;14(6):559-562. doi:10.1089/bio.2016.0088.
[6] Reihs R, Proynova R, Maqsood S, et al. BBMRI-ERIC Negotiator: Implementing Efficient Access to Biobanks. Biopreserv Biobank. 2021;19(5):414-421. doi:10.1089/bio.2020.0144.
[7] Brammen D, Greiner F, Kulla M, Otto R, Schirrmeister W, Thun S, Drösler SE, Pollmanns J, Semler SC, Lefering R, Thiemann VS, Majeed RW, Heitmann KU, Röhrig R, Walcher F; [AKTIN - The German Emergency Department Data Registry - real-time data from emergency medicine : Implementation and first results from 15 emergency departments with focus on Federal Joint Committee's guidelines on acuity assessment]. Med Klin Intensivmed Notfmed. 2020 Dec 21:1–10. German. doi: 10.1007/s00063-020-00764-2. Epub ahead of print. PMID: 33346852; PMCID: PMC7750913.
[8] Prokosch HU, Acker T, Bernarding J, et al. MIRACUM: Medical Informatics in Research and Care in University Medicine. Methods Inf Med. 2018;57(S 01):e82-e91.
[9] CODEX | COVID-19 Data Exchange Platform, Netzwerk Universitätsmedizin. https://www.netzwerk-universitaetsmedizin.de/projekte/codex [last visited on January 3rd, 2022]
[10] Sass J, Bartschke A, Lehne M, Essenwanger A, Rinaldi E, Rudolph S, Heitmann KU, Vehreschild JJ, von Kalle C, Thun S. The German Corona Consensus Dataset (GECCO): a standardized dataset for COVID-19 research in university medicine and beyond. BMC Med Inform Decis Mak. 2020 Dec 21;20(1):341.
[11] Rosenau L, Kroll B, Kiel A, Majeed RW, Szimtenings L, Köhler T, Ingenerf J, Prokosch HU, Gruendner J. Generation of a Fast Healthcare Interoperability Resources (FHIR)-based Ontology for federated Feasibility Queries in the context of COVID-19. JMIR Medical Informatics accepted for publication.
[12] Schüttler C, Prokosch HU, Sedlmayr M, Sedlmayr B. Evaluation of Three Feasibility Tools for Identifying Patient Data and Biospecimen Availability: Comparative Usability Study. JMIR Med Inform. 2021 Jul 21;9(7):e25531. doi: 10.2196/25531. Erratum in: JMIR Med Inform. 2021 Oct 8;9(10):e33105.
[13] Castro VM, Gainer V, Wattanasin N, Benoit B, Cagan A, Ghosh B, Goryachev S, Metta R, Park H, Wang D, Mendis M, Rees M, Herrick C, Murphy SN. The Mass General Brigham Biobank Portal: an i2b2-based data repository linking disparate and high-dimensional patient data to support multimodal analytics. J Am Med Inform Assoc. 2021 Nov 28:ocab264. doi: 10.1093/jamia/ocab264. Epub ahead of print.
[14] Geiger, J, et al. 2018 Hospital-integrated Biobanking as a Service – The Interdisciplinary Bank of Biomaterials and Data Wuerzburg (ibdw). Open Journal of Bioresources, 5: 6.
[15] Lawrence E, Sims J, Gander A, Garibaldi JM, Fuller B, Davidson B, Quinlan PR. The Barriers and Motivators to Using Human Tissues for Research: The Views of UK-Based Biomedical Researchers. Biopreserv Biobank. 2020 Aug;18(4):266-273. doi: 10.1089/bio.2019.0138. Epub 2020 Jun 17. PMID: 32551838; PMCID: PMC7415886.

Healthcare of the Future 2022
T. Bürkle et al. (Eds.)

doi:10.3233/SHTI220318

Burnout and Depression Detection Using Affective Word List Ratings

Sophie HAUG[a] and Mascha KURPICZ-BRIKI[a,1]
[a] *Applied Machine Intelligence, Bern University of Applied Sciences, Biel, Switzerland*

Abstract. Burnout syndrome and depression are prevalent mental health problems in many societies today. Most existing methods used in clinical intervention and research are based on inventories. Natural Language Processing (NLP) enables new possibilities to automatically evaluate text in the context of clinical Psychology. In this paper, we show how affective word list ratings can be used to differentiate between texts indicating depression or burnout, and a control group. In particular, we show that depression and burnout show statistically significantly higher arousal than the control group.

Keywords. psychology, natural language processing, affective word lists, augmented intelligence

1. Introduction

Burnout syndrome and depression are prevalent mental health problems in many societies today. Whereas currently inventories including multiple-choice questions are used in the clinical intervention and research (e.g., [1] [2]) it has been shown that these have limitations and that novel methods using text data such as interview transcripts or free-text answers are promising. However, they generate a large overhead when evaluated manually [3]. Furthermore, it can be difficult to demarcate different mental health conditions, such as for example depression and burnout that might have overlapping symptoms [4]. Most of the related work in detecting indication for depression on text is done for the English language. In this paper, we provide insights into how affective word list ratings can be used to differentiate between texts indicating depression or burnout, and a control group, for the German language. In particular, we show that depression and burnout show statistically significantly higher arousal than the control group. The promising results of our work enable the development of new technologies to support clinical practitioners to automatically find indication for the presence of depression or burnout in a text written by a patient or an interview transcript. Previously, NLP methods have been applied to different clinical use cases, such as e.g., information extraction for cancer-related electronic health record (EHR) notes [15] or computational phenotyping [16]. Such new technologies can in the future provide new smart tools to support clinical practitioners in their daily work (so called *Augmented Intelligence*).

[1] Correspondence: Mascha Kurpicz-Briki, Applied Machine Intelligence, Bern University of Applied Sciences, Höheweg 80, CH-2502 Biel/Bienne, Switzerland; E-mail: mascha.kurpicz@bfh.ch.

2. Background: Affective Word Lists

Affective word ratings are normative ratings of words in a given language relating to the emotional content of a word. For example, say we have the category *valence*, which describes how strongly a word is associated with positive emotions. Then the word "happy" will have a very high score in the *valence* category and the word "sad" a very low one. Affective Word Lists (AWL) are dictionaries with word ratings for different emotional categories. They have been previously used both for studying emotional and physical response to language as well as for studying emotional meaning within language itself, e.g. [5]. Other works, while not having directly used affective word lists, have used features such as "percentage of positive words" [6] or "percentage of words associated with body image" [7] as specified features in their classifiers.

The first corpus of affective ratings for the English language was presented by Bradley and Lang in 1999 [13] and since then, various such corpora have been compiled in different languages. For our analysis we used two existing affective word lists for the German language: The first is the Affective Norms corpus, a dictionary of 350'000 lemmatised German words, that was introduced by Köper and Schulte im Walde [8] in 2016. For this list, all ratings were obtained from a supervised machine learning algorithm, which is why the list is substantially larger than other affective word lists which were labelled by hand.

The second corpus, the Berlin Affective Word List Reloaded (BAWL-R), introduced by Võ et al. [14], consists of 2900 German words and has been rated by 200 human annotators. For the final rating the average of the individual ratings was taken. Unlike the Affective Norms corpus, the BAWL-R has a part-of-speech column and thus allows, e.g., to only consider verbs or nouns in the analysis.

The affective categories used in both Affective Norms and BAWL-R are the following:

- *Imageability,* which describes the degree of visual imageability.
- *Valence*, which determines the positiveness or pleasantness that a word is associated with.
- *Arousal,* which relates to the intensity of the emotional activation level, ranging from excitement to relaxation. High arousal is "perceived as a sensation of being reactive to stimuli and mentally awake" [5].

Moreover, the Affective Norms corpus has an additional category, *Abstractness/ Concreteness*, which is somewhat related to *imageability* and describes the degree of sensual perceivability of a word (the higher the value the more concrete a word is). The work of Mäntylä et al. [5] uses a slightly different categorization, the *Valence-Arousal-Dominance (VAD)* model, which is frequently used in psychology for emotion categorization. The categories used are *valence*, *arousal* and *dominance*. *Dominance* describes the degree to which we have control on a stimulus, ranging from submission to feeling in control. The authors use this model to categorize emotions in Jira[2] issues and ultimately assess productivity and burnout in software engineering, with burnout being associated with low valence, low dominance and high arousal [5]. Unfortunately, there is no available German lexicon which uses the *dominance* category, which would have been especially enlightening, to compare the findings directly with [5].

[2] Jira is a software used for project management and collaboration, often used for IT projects.

3. Experimental Setup

Data set: An extended version of the data set from [9] consisting of texts of the following three classes was used: (i) Burnout (ii) Depression (online testimonials, transcripts of documentaries, online forums) and (iii) Control. Each category contains anonymized texts based on publicly available data originating from or having a strong relation to individuals suffering from burnout, depression or none of them (control group). We extended the data set, and in particular added the category (ii) for depression. For this work, the data set had to be lemmatized, so that when looking up words in the affective word lists, we would not miss a match due to a different case, affix or grammatical form.

Analysis of the Affective Word Lists (AWL): From analyzing the AWLs, the following could be inferred

- Both dictionaries use the infinitive form for verbs
- Both dictionaries use the singular form for nouns; the Affective Norms corpus stores nouns in capitalized form, whereas in BAWL-R all words are stored as both uppercase and lowercase but not capitalized (however this is superfluous because BAWL-R has a part-of-speech column which allows us to isolate nouns)
- In the Affective Norms corpus, participle forms (e.g., *laufend*) are kept as is and not reduced to their base verb form (unlike existing NLP libraries for German). BAWL-R, being significantly smaller, does not contain any participle forms.
- The Affective Norms corpus contains a small number of punctuation tokens; however, all punctuation was ignored in further analysis. BAWL-R does not contain any punctuation tokens.
- The Affective Norms corpus includes stop words, on the other hand, BAWL-R contains only nouns, verbs and adjectives.

In a first step, we extended the AWLs by an additional column *p_word* (pre-processed word) which stores the lowercase lemma of that word. Additionally, the Affective Norms corpus was extended by a column which stores the *part-of-speech*. This would allow a more refined analysis of e.g., only verbs used in the text. However, we found that computing a word's part of speech using SpaCy [3] out of context, i.e., processing the standalone word and not a sentence where that word occurs, only works for nouns and verbs, as SpaCy seems to consistently tag all adjectives as adverbs. Thus, using this method, it is only possible to tag nouns and verbs correctly.

Categorization: When using affective word lists for categorizing text one must decide on how to attribute a score to a piece of text based on the individual word scores from the affective word list. To do so, we used the approach presented by Mäntylä et al. [5] to define a VAD (Valence-Arousal-Dominance) score based on the so-called SentiStrength algorithm [10]. When looking up words in the AWL we always converted a word to its lowercase lemma and looked it up in the *p_word* column to ensure we do not miss a word due to cased spelling or grammatical inflection. For each category, *valence*, *imageability*, *arousal* and, in the case of the Affective Norms dictionary, *abstractness-concreteness*, we extended the data set by an additional column which stores the score for the respective category.

[3] SpaCy is a Python library for Natural Language Processing

Statistical Tests: Using the Shapiro Wilk Test we established non-Gaussianity for all affective word list categories. For verifying the statistical significance of the differences found in the scores across the categories Control, Burnout and Depression, we used the Kruskal-Wallis Test. For example, if the p-value is significantly smaller than $\alpha = 0.05$ for *valence*, we can assume the differences found in *valence* across the classes Burnout, Depression and Control are statistically significant and not just by chance.

4. Results

4.1. BAWL-R

The results on the smaller corpus BAWL-R are shown in Table 1 (for all categories the average score was used, scores range from -5 to 5).

Table 1. Mean and standard deviation (σ) for all affective word categories in the BAWL-R corpus (Valence, Arousal, Imageability) for the data set, with the classes Control Group, Burnout and Depression.

Class	Val. mean	Val. σ	Arou. mean	Arou. σ	Img. mean	Img. σ
Control	4.234	0.525	0.811	0.324	2.282	0.723
Burnout	4.247	0.610	0.936	0.389	2.573	0.825
Depress.	4.246	0.421	0.900	0.310	2.352	0.576

Table 2 presents the results for on the Affective Norms corpus (for all categories the average score was used, scores range from -5 to 5).

Table 2. Mean and standard deviation (σ) for all affective word categories in the Affective Norms corpus (Valence, Arousal, Imageability and Abstract-/Concreteness) for the data set, with the classes Control Group, Burnout and Depression.

Class	Val. mean	Val. σ	Arou. mean	Arou. σ	Img. mean	Img. σ	Abstr. mean	Abstr. σ
Contr.	3.303	0.682	2.971	0.519	3.370	0.689	3.144	0.624
Burn.	3.617	0.818	3.046	0.570	3.555	0.811	3.215	0.641
Depr.	3.518	0.653	2.928	0.520	3.331	0.570	3.046	0.504

For BAWL-R, we found that only the differences found for *arousal* and *imageability* are statistically significant, whereas the differences for *valence* yielded a high p-value of 0.856 in the Kruskal-Wallis test. Therefore, the differences in the *valence* category have to be considered coincidental. For the Affective Norms corpus, we found that all of the differences were statistically significant, i.e., each yielded p-value < 0.05 in the Kruskal-Wallis test.

Table 3. p-values for BAWL-R corpus and for the Affective Norms corpus using the Kruskal-Wallis Test. The p-value is <0.05 in all cases except BAWL-R *valence*, which means that the differences in this case have to be considered coincidental.

Corpus	Valence	Arousal	Imageability	Abstract-/Concreteness
Affective Norms	$4.29 \times 10{-}9$	$4.05\times 10{-}3$	$1.57 \times 10{-}4$	$2.53 \times 10{-}4$
BAWL-R	**0.856**	$3.62\times 10{-}5$	$2.13\times 10{-}6$	-

5. Discussion

BAWL-R: The results show that the Burnout texts scored highest in both the *imageability* and the *arousal* category. The latter is in accordance with [5], where Burnout is associated with high arousal. It has to be noted that the differences are on the small scale, being smaller than the average standard deviation for both cases. The higher value for *arousal* compared to depression would be in accordance with the notion of the "upward" cycle of burnout and the "downward" cycle of depression by [11]. There is little to be found in literature on the interpretation of the *imageability* score. A study by Raghunath et al. [12] has found an association between high imageability in certain word classes and an elevated anxiety level. However, the words whose usage was found to be correlated to anxiety were not just high in imageability but also low in valence, therefore we cannot transfer these findings directly to our scenario. Also, and maybe more importantly, the authors seem to have a different understanding of imageability than e.g., the creators of the Affective Norms corpus [8], for whom high imageability relates to "things we can actually see" [8], whereas Raghunath et. al. [12] define imageability as "the degree to which words evoke mental images" [12] - which is clearly not the same.

Affective Norms Corpus: We recall that this corpus differs from the BAWL-R corpus, in that it is much larger and contains most words used in daily conversation. Thus, in this case virtually every word in a sentence has a score, of which again we took the average score. Again, the Burnout class scored highest in all categories. Notably, both Burnout and Depression scored higher in the *valence* category than Control group, which is counterintuitive and contradicts the findings of [6], which found that essays written by depressed college students contained a lower ratio of positively valanced words than those written by students who do not suffer from depression or have recovered from it (note that [6] used a different dictionary for the English language and also a different score). The high valence score for Burnout also contradicts the findings of [5], who used the same score as we did. The higher *arousal* score of the Burnout class compared to Depression is again clearly visible and confirms the Burnout-Depression-dichotomy that is advocated by [11].

Limitations: One problem that stands out in the analysis of *valence* scores, is the AWL approach's inability to capture negation or negative modifiers. For example, the following high scoring text for the Valence category illustrates this well:

«Ungenügende Kontrolle und Wertschätzung führten bei mir zu einer hohen Selbstüberforderung.»

The word "Selbstüberforderung" (a state where one has overburdened oneself and is struggling to meet self-imposed or external demands) has the lowest valence score (1.435) and the word "Wertschätzung" (appreciation) the highest one (8.493). The negative modifier "ungenügend" (insufficient), which basically negates the presence of the word it modifies, is present twice. However, this is completely missed by this model ("ungenügend" is not even factored in because there is another word with a lower valence score) and this sentence's valence score is dominated by the outlier "Wertschätzung", which is in fact semantically negated. It stands to discussion why the works by [6] and [5] seem not be affected by this problem. We do have to take into account that these works analyzed texts in a different language, and usage of negation of a positive word vs. usage of a negative word could also depend on language and culture.

6. Conclusion

By analyzing the affective scores of texts from three different groups (Burnout, Depression and Control group), it was found that both Burnout and Depression show significantly higher *arousal* than the Control group. This gives first indication that this research direction is promising to enable new methods for clinical intervention in the future using text-based data. However, limitations have been identified that will need to be addressed in future work. Further research is required in the field of NLP for mental health in order to extend the technology and – in collaboration with clinical partners - to define the clinical requirements for tools based on such technologies.

The authors gratefully thank the Swiss National Science Foundation (SNSF), since the Bachelor thesis leading to this paper was conducted in the context of a SNSF Spark project.

References

[1] Maslach, C., Jackson, S. E., & Leiter, M. P. (1997). Maslach burnout inventory. Scarecrow Education.
[2] Demerouti, E., & Bakker, A. B. (2008). The Oldenburg Burnout Inventory: A good alternative to measure burnout and engagement. Handbook of Stress and Burnout in Health Care, 65, 78.
[3] Jaggi, F. (2019). Burnout praxisnah. Lehmanns Media.
[4] Schonfeld, I. S., & Bianchi, R. (2016). Burnout and depression: two entities or one?. Journal of Clinical Psychology, 72(1), 22-37.
[5] Mäntylä, M., Adams, B., Destefanis, G., Graziotin, D., & Ortu, M. (2016, May). Mining valence, arousal, and dominance: possibilities for detecting burnout and productivity?. In Proceedings of the 13th International Conference on Mining Software Repositories (pp. 247-258).
[6] Rude, S., Gortner, E. M., & Pennebaker, J. (2004). Language use of depressed and depression-vulnerable college students. Cognition & Emotion, 18(8), 1121-1133.
[7] Spinczyk, D., Bas, M., Dzieciątko, M., Maćkowski, M., Rojewska, K., & Maćkowska, S. (2020). Computer-aided therapeutic diagnosis for anorexia. Biomedical Engineering Online, 19(1), 1-23.
[8] Köper, M., & Im Walde, S. S. (2016, May). Automatically generated affective norms of abstractness, arousal, imageability and valence for 350 000 german lemmas. In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16) (pp. 2595-2598).
[9] Nath, S., and Kurpicz-Briki, M. (2021). BurnoutWords - detecting burnout for a clinical setting. In Proceedings of the 10th International Conference on Soft Computing, Artificial Intelligence and Applications (SCAI 2021), CS & IT Conference Proceedings, Zurich, Switzerland.
[10] Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., & Kappas, A. (2011). Sentiment strength detection in short informal text. Journal of the American Society for Information Science and Technology, 62(2), 419.
[11] Brühlmann, T. O. N. I. (2010, February). Burnout und Depression Überschneidung und Abgrenzung. In Swiss Medical Forum (Vol. 10, No. 08, pp. 148-151). EMH Media.
[12] Raghunath, B. L., Mulatti, C., Neoh, M. J. Y., Bornstein, M. H., & Esposito, G. (2021). The associations between imageability of positive and negative valence words and fear reactivity. *Psychiatry International*, *2*(1), 32-47.
[13] Bradley, M. M., & Lang, P. J. (1999). *Affective norms for English words (ANEW): Instruction manual and affective ratings* (Vol. 30, No. 1, pp. 25-36). Technical report C-1, center for research in psychophysiology, University of Florida.
[14] Võ, M. L., Conrad, M., Kuchinke, L., Urton, K., Hofmann, M. J., & Jacobs, A. M. (2009). The Berlin affective word list reloaded (BAWL-R). *Behavior Research Methods*, *41*(2), 534-538.
[15] Datta, S., Bernstam, E. V., & Roberts, K. (2019). A frame semantic overview of NLP-based information extraction for cancer-related EHR notes. *Journal of biomedical informatics*, *100*, 103301.
[16] Zeng, Z., Deng, Y., Li, X., Naumann, T., & Luo, Y. (2018). Natural language processing for EHR-based computational phenotyping. *IEEE/ACM transactions on computational biology and bioinformatics*, *16*(1), 139-153.

Evaluation of IT Influence

This page intentionally left blank

Healthcare of the Future 2022
T. Bürkle et al. (Eds.)

doi:10.3233/SHTI220320

Potential of Electronic Medical Record Data for National Quality Measurement

Niklaus S BERNET[a,1,2], Silvia THOMANN[a,1], Mascha KURPICZ-BRIKI[b], Leonie ROOS[a], Irma HJ EVERINK[c], Jos MGA SCHOLS[c], Sabine HAHN[a]
[a]*Applied Research in Nursing, Bern University of Applied Sciences Bern, Switzerland*
[b]*Applied Machine Intelligence, Bern University of Applied Sciences, Biel, Switzerland*
[c]*Maastricht University, Maastricht, the Netherlands*

Abstract. National quality measurements with risk-adjusted provider comparison in health care nowadays usually require administrative or clinically measured data. However, both data sources have their limitations. Due to the digitalisation of institutions and the resulting switch to electronic medical records, the question arises as to whether these data can be made usable for risk-adjusted quality comparisons from both a content and a technical point of view. We found that most of the relevant information can be exported with little effort from the electronic medical records. In using this data source an even more sophisticated operationalization of the data of interest is needed.

Keywords. Computerized Medical Records Systems, Risk Adjustment, Quality of Health Care, Data Collection, Hospitals

1. Introduction

Demographic changes are increasing the demand for hospital services (1), while concurrently growing cost pressures and a lack of qualified staff threaten patient safety (2). To monitor patient safety and provide a data basis for comparing hospital and quality improvement, national quality measurements are carried out annually in Switzerland. Up to now, the data sources for national quality measurements have been predominantly based on "primary clinical data" (survey or direct observation) or (secondary) administrative data. However, both approaches are associated with limitations (3). Primary clinical data collection is associated with a possible non-response bias and significant personnel burden. Administrative data often lack detailed clinical information (e.g. variables necessary for risk adjustment) because they are usually generated for payment purposes and not for quality measurements. A promising alternative or supplementary data source for national quality measurements is (electronic) medical record data ([E]MRD, comprising in this study medical, nursing and other clinical records). MRD have a high level of detail in terms of clinical information, such as health status information and results of assessments, and are increasingly available electronically as hospitals become more and more digitised (3). A major challenge in the use of these data

[1] Niklaus Bernet and Silvia Thomann contributed equally and are considered joint first authors

[2] Corresponding Author, Niklaus Bernet, Bern University of Applied Sciences, Applied Research & Development in Nursing, Murtenstrasse 10, 3008 Bern, Switzerland; E-mail: niklausstefan.bernet@bfh.ch

is the lack of standardisation and the different clinical information systems in which they are stored and, accordingly, the different interfaces. These are possible reasons why feasible and scientifically sound EMRD-based indicators for national quality measurements are sparse, as shown by the National Quality Forum's measurement portfolio, in which only 2 out of 76 endorsed inpatient safety measures are EMRD-based (4). The impetus to investigate the potential of EMRD for national quality measurement is additionally underlined by experience with the National Prevalence Measurement[3] (NPM) of Falls and Pressure Ulcers, which has been conducted annually in all Swiss acute care hospitals since 2011, except for 2020 and 2021 (due to COVID-19). In this national quality measurement, qualified nurses collect defined data from the medical records as well as directly at bedside on one day per year for all inpatients who have given their oral consent to participate. Although this approach is considered the gold standard at outcome level, the personnel costs are viewed critically, as are the rather low patient participation rate of 75% and the limited clinical relevance of the results due to the cross-sectional design. It is regarded even more critically, as hospitals state that all necessary information is also available in their EMRD. Therefore, we used the NPM as a reference to investigate whether it is feasible in terms of content and technology to use EMRD for national quality measurement purposes including risk-adjusted hospital provider comparisons.

2. Methods

Design: A feasibility study was conducted with a stakeholder-centric design using quantitative and qualitative methods. Stakeholder involvement in the architecture and design of software has shown to be important, as this allows the definition of a realistic and suitable potential architecture to use existing EMRD for quality measurement, and thus makes the implementation feasible (5).

Sample: The stakeholders were recruited by applying a gatekeeper procedure (6). Contacts of the contracting authority and the research group were used to select possible participants. Study information and an invitation were sent by e-mail. Initially, all stakeholders invited agreed to participate. During the project, however, one hospital withdrew due to limited staff resources, and the person representing the patients withdrew for unknown reasons. Thus, representatives from 3 hospitals (incl. nursing experts and managers, IT specialists) and 7 other stakeholders (from health insurers [1], the hospital association [1], national regulatory authorities [3] and regional regulatory authorities [2]) participated in the project.

Data collection and analysis: The data collection and analysis can be described in 3 phases. In the 1st phase, qualitative methods were used to explore experiences with and expectations of national quality measurement, including the possibilities for using different data sources. For this purpose, an online survey with 11 open questions defined by the project team was sent to the participating stakeholders as a preparatory task in February 2021 and the summarised results were jointly discussed and validated in a two-hour online workshop in March 2021. This enabled us to review pre-existing assumptions and to jointly define the procedure followed in this study. In the workshop, the extent to which the data used in the NPM for risk adjustment are available in their electronic medical records, and which data elements that have been missing so far might be additionally available was also discussed with the representatives of the hospitals.

[3] See www.anq.ch and www.lpz-um.eu for further information on the National Prevalence Measurement

Based on the findings, the project team developed a data structure in which each variable to be exported was defined (content and technical aspects such as the data format [comma-separated values, specifying the expected separators] or the handling of non-applicable variables [empty fields vs. "none/not applicable"]. In the 2nd phase, the hospitals exported the EMRD from their systems into a Microsoft Excel file according to the defined data structure. The data were then transmitted to the project group via a secure connection, merged into one data set and analysed according to the methodological approach of the NPM to enable a comparison of the different data sources. In addition to the descriptive analysis, a so-called risk adjustment model was created. Risk adjustment is of central importance in national quality measurements to enable a fair comparison of hospital performance (7), as the total variance of outcomes between hospitals can be explained by the following variance components (V) (8): V(outcome)= V(definitions/data quality)+V(case-mix)+V(clinical care quality)+V(chance). Risk adjustment thus means controlling for differences in patient-mix between hospitals so that the remaining variability in outcome between hospitals can be attributed with some degree of certainty to differences in the clinical quality of care provided. In the 3rd phase, the results were presented by the project team to the participants and discussed and validated with them in unstructured individual online interviews (via MS-Teams) of about 30 minutes in November/December 2021.

Ethics: The EMRD was completely anonymised, which means that data are not sensitive in the sense of the Swiss Data Protection Act. It was recommended that hospitals only export EMRD from patients who have given their consent to the use of their data as part of the NPM 2019, and to have the permissibility of a data export approved internally. Given that no health-related data were collected in the survey, workshops and interviews with participants formal ethical approval and authorization from the ethics committee was not required according to the Swiss Human Research Act. All participants received written information and gave their consent for participation.

3. Results

3.1. Findings on content-related aspects

In the 1st phase, all participants agreed that the digitization efforts of the hospitals should be rendered usable for national quality measurements to efficiently use already existing data for quality development purposes. In addition, they confirmed that risk-adjusted hospital comparison is and should remain a central element of national quality measurements. Concerning the current NPM, the following limiting factors were highlighted: prevalence instead of incidence measurement, the time-delayed publication of the results and the staff effort. It was also mentioned that more quality indicators will probably be measured in the future, which further underlines the need to use EMRD.

The variables used for risk-adjusted hospital comparison in the NPM (general and risk variables listed in Table 1, column 1) were confirmed as being the important ones. No additional data items were proposed. Most of the variables are available in EMRD, although some of them are not uniformly operationalized, especially clinical variables related to falls and pressure ulcer risk assessment (Table 1, column 1). No comparable variable could be identified for care dependency. In favour of obtaining data sets that are as comprehensive as possible, minor deviations in the operationalisation of individual variables were allowed in the data structure agreed upon for data export.

Table 1. Variables of the NPM, availability of the variables in EMRD, comparison of descriptive results and comparison of fall risk factors in risk adjustment models based on NPM and EMRD.

Variables		Descriptive results		Fall risk factors in risk adjustment models	
NPM	EMRD	NPM[a] (n=13240)	EMRD (n=1094)	NPM[b]	EMRD[c] OR (95% CI)
General variables					
Age (years)	☑	71.0 (m)	72.0 (m)	↑	1.96 (1.19–3.21)
Length of stay (days)	☑	4.0 (m)	5.0 (m)	↑	1.47 (1.20–1.81)
Care dependency (CDS sum score)	☒	70.0 (m)	NA[f]	↑	NA[f]
Sex (female)	☑	49.1%	48.6%	↓	0.59 (0.31–1.12)
Surgical procedure (yes)	☑	43.9%	NA[f]	↓	NA[f]
DG: Diseases of the circulatory system (yes)	☑[d]	57.5%	64.4%		
DG: Diseases of the musculoskeletal system (yes)	☑[d]	40.0%	37.2%		
DG: Endocrine, nutritional and metabolic diseases (yes)	☑[d]	36.3%	53.4%		
DG: Diseases of the genitourinary system (yes)	☑[d]	33.2%	37.7%		
DG: Diseases of the digestive system (yes)	☑[d]	27.9%	33.2%		
DG: Diseases of the respiratory system (yes)	☑[d]	26.4%	26.6%		
DG: Neoplasms (yes)	☑[d]	22.7%	32.7%	↑	
DG: Mental, behavioural and neurodev. disorders (yes)	☑[d]	20.5%	20.3%	↑	2.28 (1.23–4.25)
DG: Diseases of the blood and blood-forming organs (yes)	☑[d]	18.0%	22.0%		
DG: Diseases of the nervous system (yes)	☑[d]	14.7%	19.8%		
DG: Certain infectious and parasitic diseases (yes)	☑[d]	13.8%	18.6%		
DG: Diseases of the skin and subcutaneous tissue (yes)	☑[d]	8.3%	9.7%		
DG: Factors influencing health status and contact with health services (yes)	☑[d]	7.3%	29.4%		
DG: Injury, poisoning and certain other consequences of external causes (yes)	☑[d]	6.7%	29.4%		
DG: Diseases of the eye and adnexa (yes)	☑[d]	6.6%	5.1%		
DG: Symptoms, signs and abnormal clinical and laboratory findings (yes)	☑[d]	5.8%	32.9%		2.10 (1.10–4.00)
DG: Diseases of the ear and mastoid process (yes)	☑[d]	2.8%	4.0%		
DG: External causes of morbidity (yes)	☑[d]	2.2%	5.6%		
(DG: Codes for special purposes) (yes)	☑	NA[f]	15.6%	NA[f]	2.03 (1.04–3.97)
Fall specific risk variables					
Risk for fall (yes)	☑[e]	29.7%	33.4%	↑	4.22 (2.01–8.84)
Sedative/psychotropic medications intake (yes)	☑[d]	37.3%	72.6%	↑	2.43 (0.82–7.20)
Outcome Variable					
Fall in hospital (yes)	☑	3.7%	4.7%		
Pressure ulcer specific risk variable					
Risk for pressure ulcer (yes)	☑[e]	31.6%	21.4%		
Outcome Variables					
Nosocomial pressure ulcer (yes)	☑	3.9%	NA[f]		
Nosocomial pressure ulcer category 2 and higher (yes)	☑	1.7%	NA[f]		

Abbreviations: NPM = National Prevalence Measurement; EMRD = electronic medical record data; OR = odds ratio; 95% CI = 95% confidence interval; CDS = Care Dependency Scale; DG = diagnosis groups according to International Statistical Classification of Diseases and Related Health Problems 10th Revision; m = median; ☑ = variable available; ↑ = fall risk increasing factor; ↓ = fall risk decreasing factor; NA = not available.

[a]The descriptive results reported here are based on the NPM 2019[4].

[b]Only factors that were selected as significant variables ($p<.05$) in the risk adjustment model in at least two of three most recent NPMs[4] are reported to ensure comparison with the more reliable factors.

[c]All variables that were selected into the risk adjustment model are reported.

[d]Information in EMRD in more detail available (per patient lists of ICD-10 Diagnosis Codes and ATC-codes).

[e]Information available in EMRD but not uniformly operationalized.

[f]Variables are either not available in the NPM or EMRD or are available in the EMRD but could only be exported correctly by 2 of the 3 hospitals and could therefore neither be considered in the descriptive (data protection reasons) nor in the risk adjustment model comparison.

[4] See NPM reports from 2017, 2018 and 2019, available at www.anq.ch

In the 2nd phase, during the data export from the electronic medical records, some content-related questions emerged. While in the NPM, for example, the pressure ulcer risk is recorded on the day of the survey, the timeliness of the risk assessment in the EMRD can vary from hospital to hospital (assessment at admission, re-assessment intervals). Further ambiguities, which will have to be further specified in the future, arose in relation to the handling of paused medications or inactive diagnoses during data export.

The descriptive analyses based on the EMRD and the NPM 2019 show - with a few exceptions - comparable results (Table 1, column 2). Five ICD-10 diagnosis groups are found much more frequently in EMRD (plus 10.0% to 27.1%), an indication that secondary diagnoses are comprehensively mapped in EMRD. The biggest difference involved "sedative/psychotropic medications intake". Since each medication is recorded in the EMRD with the corresponding Anatomical Therapeutic Chemical (ATC) classification, a much more precise and automated assignment of each medication to the combined group of sedatives/psychotropics was possible. In the NPM, the allocation is done manually based on medication list review, so an underestimation is conceivable.

Comparison of the fall risk adjustment models (Table 1, column 3) also provides insight into the data quality of EMRD. Although two important risk variables could not be included in the EMRD-based risk adjustment model (care dependency and surgical procedure), the results are comparable. Six of the selected variables proved to be relevant in previous NPMs and point in the same direction (decreasing or increasing risk) regardless of the data source used. However, there are differences in the selected ICD-10 diagnosis groups. Only one diagnosis group, "mental and behavioural disorders", corresponds to the previous NPM findings. These differences may be related to the large differences in sample sizes. However, even in the NPM, varying diagnosis groups are selected every year. As one hospital was unable to provide EMRD on pressure ulcers, the model could only be tested for the indicator of falls.

3.2. Findings on technical feasibility

The hospitals had to invest between 0.5 and 1 day to generate the data export. Different methods, partially manually and partially automated, were used. The decisive factor was whether several data sources (e.g. system for medical records and/or billing data) had to be merged, or whether all information of interest was available in the same system. Despite the jointly defined data structure, certain variables could not be exported (in the desired format). In one institution, this was due to migration to a new clinical information system. In another, it was probably due to an incorrect data query, as the data were available in the system. In the current approach, the comma-separated values (CSV) files were transferred using a Microsoft Excel file. This was assessed in the individual interviews as an appropriate method. The extent to which building a direct interface would generate a return commensurate with the effort involved was doubted.

4. Discussion

With the involvement of relevant stakeholders including representatives of 3 hospitals, it was possible to explore for the first time how EMRD related to falls and pressure ulcers can be used for national quality measurements. The results are promising from a content and technical point of view. Most of the relevant variables can be exported with little effort from existing EMRD, and data analysis yields results similar to previous

NPMs. However, the results also reveal two relevant challenges. Firstly, it became apparent that certain variables are not available in the systems (degree of care dependency) and are operationalised differently (risk assessments). A data- and literature-based definition of a minimum dataset under these conditions (e.g. proxy variable for care dependency) thus appears to be crucial. Clear specifications as to which data must be available, how operationalised and how recorded in the systems are indispensable so that the EMRD can be used for national quality measurements including risk-adjusted provider comparisons. To promote consistent and sustainable system adaptation, broad stakeholder buy-in to draw up viable national recommendations and involvement of staff who use the systems in daily practice is essential. If professionals agree on what needs to be documented to reflect real life quality of care, this will promote acceptance of system adaptations as well as data quality and completeness. Secondly, a concept to ensure data quality and comparability (incl. e.g. time of recording) needs to be developed.

In general, however, the potential of EMRD for national quality measurements is considered high, as various limitations of the NPM could be eliminated: (i) staff effort is reduced, as most data already exist digitally and only need to be converted into an exportable format by means of a query, (ii) regular data extracting and/or continuous measurement will become possible, (iii) the better use of digital opportunities allows for automation, in which results can be made available in a timely manner, up to continuous monitoring, (iv) the non-response bias is reduced if anonymised data are allowed to be exported for quality measurements without patient consent. If an automated system with interfaces to the hospitals can be set up, this system would also be flexibly adaptable to other quality indicators, which will become quite important in view of the increasing awareness of patient safety and quality of care. In the future, a promising resource for automated national quality measurements based on hospitals' EMRD could be the new decentralised data infrastructure currently being developed and implemented as part of the Swiss Personalised Health Network initiative to make relevant health data interoperable and shareable for research by defining data formats, semantics and exchange mechanisms (9). In particular, as soon as a large number of hospitals participate in the network.

References

[1] Bundesamt für Gesundheit. Evaluation der KVG-Revision im Bereich der Spitalfinanzierung - Schlussbericht des BAG an den Bundesrat. Bern: Bundesamt für Gesundheit, 2019.
[2] Vincent C, Staines A. Enhancing the Quality and Safety of Swiss Healthcare. Bern: Federal Office of Public Health, 2019.
[3] Busse R, Klazinga N, Panteli D, Quentin W. Improving healthcare quality in Europe: Characteristics, effectiveness and implementation of different strategies. 2019.
[4] National Quality Forum. National Quality Forum: Find Measures [15.03.2022]. Available from: https://www.qualityforum.org/Qps/QpsTool.aspx.
[5] Saiedian H, Dale R. Requirements engineering: making the connection between the software developer and customer. Inf Softw Technol. 2000;42(6):419-28. doi: https://doi.org/10.1016/S0950-5849(99)00101-9.
[6] Andoh-Arthur J. Gatekeepers in Qualitative Research. In: Atkinson P, Delamont S, Cernat A, Sakshaug JW, Williams RA, editors. SAGE Research Methods Foundations. London: SAGE Publications Ltd.; 2019.
[7] Bernet NS, Everink IHJ, Schols JMGA, Halfens RJG, Richter D, Hahn S. Hospital performance comparison of inpatient fall rates; the impact of risk adjusting for patient-related factors: a multicentre cross-sectional survey. BMC Health Services Research. 2022;22(1):225. doi: https://doi.org/10.1186/s12913-022-07638-7.
[8] Lilford R, Mohammed MA, Spiegelhalter D, Thomson R. Use and misuse of process and outcome data in managing performance of acute medical care: avoiding institutional stigma. The Lancet. 2004;363(9415):1147-54. doi: https://doi.org/10.1016/S0140-6736(04)15901-1.
[9] Swiss Personalized Health Network. About SPHN [15.03.2022]. Available from: https://sphn.ch/organization/about-sphn/.

Healthcare of the Future 2022
T. Bürkle et al. (Eds.)

doi:10.3233/SHTI220321

Digi-Care: Exploring the Impacts of Digitization on Nursing Work in Switzerland

Thomas BÜRKLE[a,1], Christopher LUEG[a], Patrizia SALZMANN[b], Deli SALINI[b], François von KAENEL[a], Kezia LÖFFEL[b], Lea MEIER[a], Stephy-Mathew MOOZHIYIL[b], Sandro PERRINI[b], Ines TREDE[b], Andrea VOLPE[b] and Jürgen HOLM[a]
[a]*Bern University of Applied Sciences BFH, Institute for Medical Informatics I4MI*
[b]*Swiss Federal University for Vocational Education and Training SFUVET*

Abstract. In this paper we present first findings of the Digi-Care project, a multidisciplinary, multi-stakeholder research project investigating the impacts of digitization on nursing work practices and in particular the transmission of patient care information within and beyond nursing work practices. We completed the initial data collection of the funded 3-year research project and report on a plethora of significant and critical IT-related events. Some of them can be attributed to usability issues.

Keywords. Information Exchange in care settings, IT usability, evaluation, EPR systems

1. Introduction

Poor alignment between IT systems, workflows and actual work practices in clinical environments has been shown to have adverse effects upon patient care [1]. Poor software usability has been described as a contributing factor to such problems (e.g. [2]). Traditionally, usability is assessed using formative methods such as think aloud [3] and cognitive walkthroughs [4] in laboratory environments. Heuristic evaluation [6] and user-centered approaches including questionnaires such as SUS [7] have been successfully used for summative evaluation approaches.

These traditional approaches tend to underestimate the number of software issues, and it is difficult to assess the impact upon experienced versus inexperienced software users. Thus, Kushniruk et al. have advocated layered approaches of system testing with clinical simulations and naturalistic studies in real-world environments for level 3 evaluation to determine the organizational impact of IT systems [8,9]. Such studies in real-world environments do not only reveal pure usability problems, but a broad range of potential errors as described by [10]. In this paper, we report preliminary results of such a study performed in six Swiss hospitals with the intention to

[1] Corresponding author: Thomas Bürkle, Berne University of Applied Sciences, Quellgasse 21, 2501 Biel/Bienne, Switzerland; E-mail: thomas.buerkle@bfh.ch.

1. identify and characterize situations in which the transmission of patient care information (PCI) with digital devices is perceived as significant or even critical, and derive required nursing competencies to deal with such situations
2. use selected situations as bases for stimulating learning and development related to the sharing of clinical information

2. Methods

Digi-Care has a cooperative study design which comprises several project stages in a three-year time frame.

The first stage (ethnographical study and data collection) consists of semi-structured context interviews with nurses and IT responsible persons and an extensive workplace observation to collect data about nursing work practices and the transmission of PCI in six Swiss hospitals sized between 200 and 400 beds. We conducted observations in orthopedic units, gynecology, urology, interdisciplinary and also surgical departments. Four observers (KL, S-MM, SP, AV) observed four different nurses in each hospital. The nurses participated voluntarily, they hold a tertiary-level degree, they are responsible for direct patient care and work at least 80%.

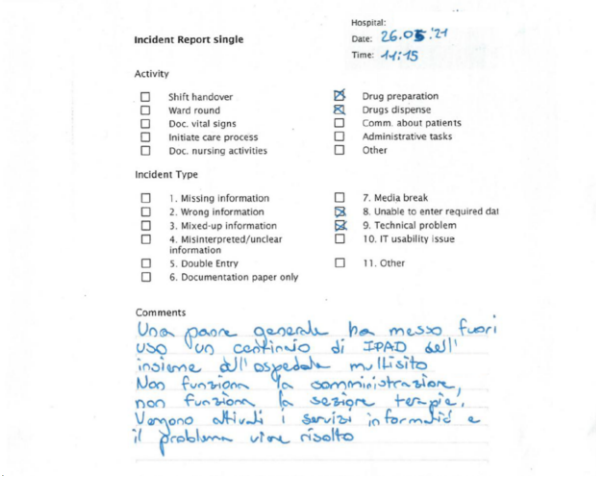
After a six-day familiarization in the respective hospital and ward, the observer followed a registered nurse for three consecutive work shifts and recorded crucial moments of the use of digital tools as well as non-digital tools and forms of transmission of PCI on video. A GoPro Hero 8 Black camera with extra power supply was carried in a chest harness to capture the entire situation. In addition, each observer carried a Nokia 8.1 smartphone featuring dual rear cameras for occasional close-up shots, e.g. data displayed on computer screen. To ensure high quality sound recording the observed nurses were equipped with RØDE Wireless GO microphone transmitters and the corresponding receiver was connected via USB-cable for sound input to the GoPro camera of observer. Subsequently, the video footage of these three shifts was cut into a one-hour video with Adobe Premiere Pro. This video was discussed with the observed nurse in a so-called self-confrontation interview, aimed to identify the PCI situations and the situated components of nurse's experience.

To identify IT-related incidents, we derived a paper-and-pencil template for the observer to identify exactly when and at what location the incident happened, and to characterize it in relation to the performed activity (e.g. shift, handover, drug preparation, etc.) and in relation to an incident type (e.g. missing information, double entry, media break, etc.). An exemplary incident report is given in Figure 1. Additional video footage was recorded for IT-related incidents. The incidents were coded (e.g. D3 for the third observed incident in the hospital D) and collected in an Excel spreadsheet with date, incident type, activity, short description, categorization and priority.

The second stage of the project will cover the selection and validation of interesting PCI situations and IT-related incidents in an iterative stepwise process. In the first step the researchers themselves will group and classify PCI situations and IT-related events. In the second step these findings will be discussed with participants in the observed hospitals. In the third step workshops with external experts will be held to further condense the data.

In a coming third stage both prototypical educational instruments for improved coping with selected situations and prototypical technical mockups/solutions to prevent selected IT-related events which concern bad IT usability will be developed. Further

project stages include the validation of these prototypes in workshops and the dissemination of the results.



Incident Report single

Hospital:
Date: 26.05.'21
Time: 11:15

Activity

- ☐ Shift handover
- ☐ Ward round
- ☐ Doc. vital signs
- ☐ Initiate care process
- ☐ Doc. nursing activities
- ☒ Drug preparation
- ☒ Drugs dispense
- ☐ Comm. about patients
- ☐ Administrative tasks
- ☐ Other

Incident Type

- ☐ 1. Missing information
- ☐ 2. Wrong information
- ☐ 3. Mixed-up information
- ☐ 4. Misinterpreted/unclear information
- ☐ 5. Double Entry
- ☐ 6. Documentation paper only
- ☐ 7. Media break
- ☒ 8. Unable to enter required dat
- ☒ 9. Technical problem
- ☐ 10. IT usability issue
- ☐ 11. Other

Comments

Una pausa generale ha messo fuori uso un continuo di IPAD dell' insieme dell' ospedale multisito
Non funziona la somministrazione, non funziona la sezione terapie,
Vengono attivati i servizi informatici e il problema viene risolto

Figure 1. Documentation template for critical IT incidents.

3. Preliminary Results – IT-related events in the first 5 hospitals

At the time of writing this paper, the 3-year project was still in its early stages, with 14 of 37 planned months completed. 320 hours of job shadowing time have been recorded until November 2021. 16 contextual interviews and 16 self-confrontation interviews have been completed.

IT-related events are assessed separately from the PCI situations which may or may not involve any digital tool, paper, phone or oral communication. To begin the second stage, the medical informatics researchers (TB, JH, FvK, CL and LM) started analyzing the first IT-related events. 144 such events have been recorded in five of the six hospitals, with data of the last hospital still under analysis. For future prototyping activities we will concentrate mainly on usability issues involved in the IT–related event.

Table 1 represents an example of such an event. It could have had considerable adverse effects for the patient. This event cannot be uniquely attributed to bad IT usability. Initially, the prescribing physician erroneously documented an incorrect amount of diluting liquid for the antibiotic. An inappropriate default list of available solutions for antibiotics could have contributed to this error, but other reasons such as work overload may be likely as well. If both 500mg and 1g dosage of Vancomycin are common in the hospital, an intelligent decision support rule would be required to assure in all cases, that the combination of antibiotic and dilution is always correct. Thus, the event was classified as technical and/or usability related, since additional decision support modules would be needed to catch such situations. In addition, this event was classified as highly critical. Typical for many such critical situations, the experienced nurse noticed that the administration of 1g Vancomycin with the prescribed dilution would be incorrect. She corrected the error herself and notified the physician so that no harm was done.

Table 1. Example of a recorded IT-related event.

Code	How often	Incident type	Activity	Text	Assessment
B3	Once	Mixed up information	Drugs dispense	Wrong prescription of Vancomycin: Vancomycin needs to be sufficiently diluted to be administered intravenously. The prescription was for Vancomycin 1g, but the indications for the dilution based on Vancomycin 500mg. The nurse was experienced enough to notice the wrong prescription and administered the IV antibiotic correctly and called the responsible person from the medical staff to adjust the prescription. The IT system does not have a built-in alert for this case of mixed-up or wrong prescription.	Usability / technical

Other recorded events included e.g. unexpected system downtime as well as duplicate documentation due to missing or insufficient IT interfaces between different department systems. Numerical results for the first five hospitals which have been analyzed are summarized in Table 2. The 144 events are distributed unevenly, resulting in between 1.25 to 3.92 events on average per shift.

Only 23 (16%) of the 144 observed IT-related events have been classified unanimously as pure usability problems. Many events had a mixed nature in terms of being classified as a) technical and usability, because they could be remedied e.g. by additional software modules, additional or better interfaces or they were considered b) mixed case vignette and usability, because education and training should or did help to overcome such situations.

Table 2. IT-related events in the first five hospitals.

	IT-events total	IT-events per shift	IT events per hour
Hospital A	24	2.00	0.25
Hospital B	36	3.00	0.38
Hospital C	15	1.25	0.16
Hospital D	22	1.83	0.23
Hospital E	47	3.92	0.49
Average	28.8	2.40	0.30

Some of the recorded IT-related events will need discussion in subsequent project stages not only with involved staff but also with external specialists, e.g. psychologists or specialists in learning, helping us to understand how and why human memorize relevant information. We noticed, for example, that in four out of five hospitals, nurses use a paper-based sheet (either printed from the EPR system or completely handwritten) to record the essential facts of the patients they care for in the respective shift. Part of this documentation was later copied back into the EPR. This process or habit leads to a higher degree of duplicate paper-based documentation and media breaks. On the other hand, handwritten notes may help the nurses to memorize important facts and synthesize the main information concerning the patients for their care activity [11].

4. Discussion

Our approach differs from that exemplified in [8] since we did not perform a full think-aloud study with screen recording of user actions. Instead, we performed an unobtrusive job shadowing study in real-world environments and asked our observers to take additional notes of observed IT-related events. Furthermore, we asked them to do a close-up recording of the computer screen with the smartphone camera if possible. Thus, in accordance with [12], we cannot be sure that we have captured all IT-related incidents in the observed time period and we are prone to some inter-observer bias.

On the other hand, we performed a multi-centric study, involving several different EPR systems in the observed hospitals. This broad approach of a field study should deliver interesting insights not only in terms of IT-related problems but also regarding how experienced staff members cope with such problems. We encountered, e.g. system breakdown times in one site and recorded how nurses dealt with this situation. As mentioned, other parts of the study design [13] will supply a systematic review of PCI situations and the way nurses deal with them using the methods of self-confrontation interviews and semiological analyses [14,15]. That part of stage 2 is still ongoing.

Initially, we expected significantly fewer IT-related events, e.g. one to three events in all three shifts of one ward, and were thus overwhelmed seeing four and more IT-related events in a single shift. This seems to be an indicator that digitization of documentation in inpatient care is not yet in a desired stage. As mentioned above, the observed events are not purely usability problems of a software product in use, but often a combination of missing interoperability, media breaks and duplicate documentation, which could be remedied by building better interfaces between different IT applications as part of the respective EPR system. Such effects were observed e.g. when patients went to the theatre for surgical interventions and documentation was therefore split among different IT-systems in anesthesia or the ICU.

In our observation, several potentially critical IT-related events have been prevented by alert staff noticing and correcting the error. This was the case, for example, when nurses noted, that physicians had entered a drug order in the EPR for the wrong time (before instead of after surgery). Here, different reactions could be observed. In some cases, nurses corrected the problem themselves and notified the physician, in other cases they were unsure and had difficulties to reach the ordering physician. Thus they undertook extra efforts to clarify the situation. This will be a topic for further examination and discussion in future stages of our study.

The level of digitization in the observed hospitals varies, but there is currently no clear indicator that a higher degree of digitization corresponds to less IT-related events. We plan to classify the observed level of digitization at the respective hospital to enable further investigation of this issue.

Summarizing, we observed many of the situations described in [10], such as fragmentation of data, problems with transfers, workarounds, or “misinterpretation of communication as information transfer”, often combined with loss of feedback. The latter two are the basis for our further work and should provide opportunities for discussion and development of better IT solutions. Usability is too important to be left to managers and software designers. For actual change we need to bring a sense of good usability to the people who benefit the most which are in this case the nursing staff [15].

5. Outlook

The essential steps of stage 1 comprising data collection and job shadowing in all hospitals were completed by the end of 2021. Stage 2 analysis of the IT-related events has started and analysis of the PCI-related situations is underway in parallel.

For further consideration of the IT-related incidents, we distributed the spreadsheet with all incidents among the medical informatics researchers (TB, JH, FvK, CL and LM) to prioritize those issues to be investigated further. In three workshops, the findings were discussed and a condensed list of 26 incidents for the first five hospitals emerged.

The next steps of stage 2 include the feedback process hospital staff and other professionals working in the field to create a shortlist of those PCI situations and IT-related events which are considered particularly relevant to most sites. For these IT events we will in stage 3 develop prototypical alternative IT designs and/or workflows to demonstrate how improved usability could help avoid the observed problem. These designs will be validated in a subsequent feedback loop in stage 4.

References

[1] Han YY, Carcillo JA, Venkataraman ST, Clark RS, Watson RS, Nguyen TC, Bayir H, Orr RA. Unexpected increased mortality after implementation of a commercially sold computerized physician order entry system. Pediatrics. 2005 Dec;116(6): 1506-12.

[2] Horsky J, Kuperman GJ, Patel VL. Comprehensive analysis of a medication dosing error related to CPOE. J Am Med Inform Assoc. Jul-Aug 2005;12(4): 377-82.

[3] Lewis C. Using the 'thinking-aloud' method in cognitive interface design, Research Report RC9265, IBM T.J. Watson Research Center, 1982.

[4] Rieman J, Franzke M, Redmiles D. Usability evaluation with the cognitive walkthrough. Proceedings of the Conference on Comp. Human Factors in Computing Systems (07-11), 1995, 387–388.

[5] Peute LWP, Jaspers MWM. The significance of a usability evaluation of an emerging laboratory order entry system. Int J Med Inform. Feb-Mar 2007;76(2-3): 157-68.

[6] Nielsen J. Heuristic evaluation. In: Nielsen J, Mack RL, editors. Usability inspection methods. Wiley, New York, 1994.

[7] Brooke J. SUS: A 'Quick and Dirty' usability scale. In: Usability Evaluation In Industry. Boca Raton, Florida, United States. CRC Press; 1996: 194.

[8] Kushniruk AW, Borycki EM, Kuwata S, Kannry J. Emerging approaches to usability evaluation of health information systems: towards in-situ analysis of complex healthcare systems and environments. Stud Health Technol Inform. 2011;169: 915-9.

[9] Kushniruk A, Senathirajah Y, Borycki E. Effective Usability Engineering in Healthcare: A Vision of Usable and Safer Healthcare IT. Stud Health Technol Inform. 2017;245: 1066-1069.

[10] Ash JS, Berg M, Coiera E. Some unintended consequences of information technology in health care: the nature of patient care information system-related errors. J Am Med Inform Assoc 2004;11(2): 104-112.

[11] Colligan L, Potts HW, Finn CT, Sinkin RA. Cognitive workload changes for nurses transitioning from a legacy system with paper documentation to a commercial electronic health record. Int J Med Inform. 2015 Jul;84(7): 469-76.

[12] Kanis H. Estimating the number of usability problems. Appl Ergon. 2011 Jan;42(2): 337-47.

[13] Salini D, Salzmann P, Löffel K, Moozhiyil S-M, Perrini S, Volpe A, Trede I, Bürkle T, von Kaenel F, Lueg C, Holm J. Observed situations involving transmission of patient care information: A basis for promoting learning and development among nursing staff. Accepted for 7th international Vet congress Zollikofen, 2022.

[14] Theureau J. Course-of-action analysis and course-of-action centred design. In Hollnagel E, editor. Handbook of cognitive task design (pp. 55-81). Lawrence Erlbaum, Mahwah, NJ 2003.

[15] Poizat G, Durand M, Theureau J. The challenges of activity analysis for training objectives. Le travail humain 2016;3(79), 233-58.

[16] Twidale MB, Nichols DM, Lueg CP. Everyone everywhere: A distributed and embedded paradigm for usability. J Assoc Inf Sci Technol 2021; 1–13.

Healthcare of the Future 2022
T. Bürkle et al. (Eds.)

doi:10.3233/SHTI220322

The Digital Transformation of Mental Health Care and Psychotherapy – A Market and Research Maturity Analysis

Alexander M. SCHMID[a,1] and Mascha KURPICZ-BRIKI[b]
[a] *Digital Technology Management, Bern University of Applied Sciences, Bern, CH*
[b] *Applied Machine Intelligence, Bern University of Applied Sciences, Biel, CH*

Abstract. Digital technology trends for mental health, instantiated with only emerging use cases or already established applications, offer significant potential to improve clinical therapy and care. In this paper, we identify five major trends, mHealth/eHealth, telehealth, artificial intelligence (AI), big data, and biosensors/wearables; describe seven specific technology use cases for mental health care and psychotherapy; and provide an overview of their maturity in practice.

Keywords. augmented intelligence, blended therapy, digital mental health, digital technology use cases, transformation, technology trends

1. Introduction

Digital transformation of mental health brings new opportunities and challenges to the sector. Emerging digital technology applications and use cases promise new treatments, diagnostic means, and management capabilities for mental disorders.

The evolution of digital technologies has the potential to reduce the prevalence of mental health disorders, for both individuals and populations as a whole [1], and to disrupt health care systems and their related policy-making efforts around the globe [2]. However, the speed of innovation adoption in the highly regulated sector of mental health care and psychotherapy is slow compared to the fast and fluctuant technological progress.

In this study, we investigate what technological trends and applications are emerging and assess their associated levels of market and research maturity. We aim to answer the following research questions: a) Which *major digital technology trends* impact mental health interventions?, b) Which *exemplary use cases* are available for the use of digital technologies in a clinical setting?, and c) What is the *maturity of evidence and adoption* for the efficacy of available digital technology products/services?

We have identified five major technology trends from technology studies and research in general healthcare with particular relevance for mental health [3–5]: mobile and electronic mental health (*mHealth/eHealth*), *telehealth*, *artificial intelligence (AI)*, *big data*, and *biosensors/wearables*. Table 1 characterizes those trends in detail and outlines their transformative potential for mental health care and psychotherapy. Trends in digital technology innovation are hardly fully discernible from one other, as they are

[1]Correspondence: Alexander M. Schmid, Bern University of Applied Sciences, Brückenstr. 73, CH-3005 Bern, Switzerland; E-mail: alexander.schmid@bfh.ch.

often intertwined with varying degrees of progress elements. Categorizations therefore can only be tentative.

Based on our investigations, we describe exemplary use cases for these technology trends and propose a map to illustrate their market and research maturity. We contribute an overview of use cases and assess their maturity, informing stakeholders – including mental health professionals, researchers, entrepreneurs, and patients – in technology-adoption decisions and policy-making.

2. Research Design

This report compiles the findings from a pragmatic, exploratory, qualitative multimethod inquiry. During the data collection and analysis process, we followed a semi-structured approach. We gathered intelligence from market reports; searched literature databases from fields such as healthcare, information systems research, medical informatics, and computer science; and conducted interviews. For the data analysis, we iteratively aggregated the artifacts from each of the three data collection streams and synthesized the key findings into our results, as documented herein.

We interviewed 21 stakeholders of mental health care transformation using a semi-structured protocol that covered five main themes: i) *technology trends*, ii) *use cases*, iii) *economic and regulatory developments* in the industry, iv) *state of research and practice* in digital mental health, and the v) *impacts of digitalization on therapeutic practices and institutions*. We documented the interviews with field notes.

In the market research stream, we took a technographic approach. We focused on specific segments of the global market, including diagnostics; behavioral mental health; and IT for therapy, care, and treatment. Based on case studies, use case descriptions, technology trend monitors, and the observation of mental health tech startup pitches, we synthesized vignettes of digital *technology in use* in the given context.

Furthermore, we conducted an exploratory literature research about two core topics: i) verification of digital technologies in therapy and self-help applications, and ii) AI technologies in diagnostics.

Table 1. Digital Technology Trends with Transformative Potential for Mental Health Care and Psychotherapy

Name: Description	Transformative Potential
AI: The use of algorithms for the intelligible solution of complex decision problems. Computers can efficiently and accurately analyze various data formats, including images, videos, audio, text, and statistics. These algorithmic analyses enable machine learning and the synthesis of models for correlations in the data and the prediction of conditions. Accordingly, the trend is accelerated through data accruing from *mHealth/eHealth*, *bio-sensors*, and *big data*.	Evidence-based decision-making in mental health care can be drastically strengthened through complementary or automated data analyses with computers. This trend transforms mental health because it enables new and more effective means of diagnosis (early detection and accuracy), prognosis (reference comparison), and treatment (tailoring), both at the individual and whole population levels [6,7]. The findings gained through AI research on mental health data and management can also inform better policy-making for health systems.
Big Data: The analysis of large data sets that are too complex to be handled manually or with traditional software. Sources encompass clinical, biological, administrative, or imaging data. It closely integrates with *AI*, as large data sets are a prerequisite for machine learning.	An important and growing source for mental health data is the electronic health record (EHR). The trend promises unprecedented potential for scientific exploration; descriptive observation; hypothesis generation; and predictions in clinical, research, and mental health management [8,9].

Biosensors/Wearables: Analytical devices that track physiological, behavioral, or biochemical body signals. Sensors can be integrated into wearables, that is, devices worn on the body (e.g., as extensions to phones or watches, or in clothes). The accumulated information fuels *big data* and *AI* and integrates with *mHealth*.	Sensors allow the identification of behaviors/conditions as consequences of underlying physiological alterations related to mental health disorders [10]. This trend transforms mental health by making continuous monitoring of patients and medical states possible. Data provide unprecedented insights for research and make therapeutic just-in-time/tailored interventions possible.
mHealth/eHealth: The delivery of mental health resources and services (information, treatment, management, etc.) by digital means via the internet using a variety of devices, including smartphones, tablets or personal computers, and remote monitoring devices. *mHealth* feeds into *AI* and improves accessibility to services, tailoring, adherence, and the flexibility of all parties [11].	Patients get broad access to resources and services, including information and therapy plans, fostering self-help. Devices and apps continuously generate data for research, diagnosis, and customized treatment, which enables monitoring and connections with peers, therapists, and care professionals. Offerings become more scalable than on-site, and in-person resources/ services, fostering patient autonomy with complements to traditional patient–therapist relationships.
Telehealth: The remote delivery of clinical and non-clinical mental health services, via audio, video, or text. It can be synchronous or asynchronous. Telehealth has been established for decades but has become considerably more versatile with the diffusion of the internet, including e-mail, smartphones, and video-chat systems. Telehealth closely blends with *mHealth/eHealth*.	Digital devices mediate the relationship between therapists and patients, allowing decentralized service provisioning. Through the developments in *mHealth/ eHealth*, *telehealth* often blends into web and mobile therapy applications and care management platforms. It decouples service provisioning from geographical restrictions, simplifies the match-making of patients with therapists and care workers, and offers important escalation pathways for *mHealth/eHealth* treatments.

3. Results

3.1. Digitally-Mediated Provisioning of Mental Health Care and Therapy

Use cases (UC) for *mHealth/eHealth* applications have gained considerable attention among the general public in recent years. Larger, mostly US-based companies such as BetterHelp, Talkspace, and Ginger offer full-service, online, and app-based mental health care services (**UC1**). Offerings encompass self-help programs and monitoring functionalities as well as *telehealth* functionalities, including peer support, guided therapy through professional coaches, or direct interaction with licensed therapists. In Europe, smaller companies, such as Minddistrict, GAIA, Selfapy, and the like, offer similar services, but these are mostly focused only on specific diagnoses (**UC2**). Such offerings are provided through smartphone applications via app stores or other online platforms.

Many *mHealth/eHealth* and *telehealth* applications are backed by academic evidence indicating their efficacy for the treatment of various disorders [e.g., 12], making them candidates for inclusion in clinical guidelines [13]. Accordingly, and although the process is relatively slow, regulators in Western countries are certifying a growing number of such applications as medical devices, rendering them eligible to be prescribed and expenses for their use to be covered by health insurers [14].[2]

Beyond treatment and therapy, *mHealth/eHealth* and *telehealth* applications are also popular for patient management, including health data management, the coordination of treatment pathways in clinics, or recovery journeys in general (**UC3**). Applications can

[2] Germany, for example, lists *mHealth/eHealth* and *telehealth* applications that are treated as certified medical devices (Digitale Gesundheitsanwendungen DiGA) in a web catalogue: https://diga.bfarm.de/de/verzeichnis?category=%5B%2277%22%5D.

be used for the management of both inpatients and outpatients [e.g., 15]. While such use cases do not require certification because they do not include therapeutic elements, they are an important element of integrated mental health care and therapy. Such applications can be used to store and manage EHR, therapy plans, and so on. Those functionalities allow for seamless integration with UC1 or UC2 applications and feed into the trend of *big data*.

3.2. Therapy Support and Self-Help Using Chatbots

Chatbots are computer programs that allow patients to interact with applications. They can address different purposes, such as therapy, training, and screening, and can be differentiated by their rule-based or machine learning approaches. A recent survey showed that most chatbots are rule-based (**UC4**); only 4 out of 41 solutions used technologies from the field of *big data* and *AI* such as machine learning (**UC5**) [16], though they demonstrated promising results regarding the processing of human language.

In chatbot use cases based on machine learning, *AI* is used for different purposes: virtual counseling for brief alcohol interventions with a smart virtual agent [17], the Wysa chatbot supports patients' mood management [18], the Tess chatbot engages in brief conversations for mental health support [19], and SimCoach provides an empathic "virtual human" (i.e., computer-generated characters) that provides information about post-traumatic stress disorder (PTSD) and depression to veterans [20].

Rule-based chatbots (UC5) can be seen as an extension of UC2. However, transition and overlap between UC5 and UC4 are continuous (e.g., such as with Woebot).

3.3. Diagnostics with Natural Language Processing

In the field of natural language processing, different works have investigated how social media posts or online support forums (OSF) can be used to detect mental health disorders or syndromes (**UC6**). A recent survey identified 29 studies investigating depression detection in OSFs [21]. A survey from 2017 reviewed mental illness detection methods on social media [22] and showed that mentally ill users were distinguishable from the control group and identifiable based on screening surveys, their public sharing of their diagnosis, or by their membership in an online forum.

In recent years, this subject has been further explored (e.g., to connect social media data to offline depression patients [23]). While these technologies could be valuable for clinical practitioners, scant research exists in the field of connecting these innovative technologies to clinical data and/or clinical trials. One first attempt to use interview data from burnout patients to train a machine learning model to detect burnout delivered promising results [24].

3.4. Condition Monitoring and Diagnosis Support through Wearable Sensors

The analysis of data generated by *biosensor/wearable* devices provides wide-ranging insights into mental health indicators such as emotions, mood states, stress, activity levels, and the like. In addition to wearable sensors, the use of smartphones has spread rapidly in the last decade, making their sensors—e.g., accelerometers, gyroscopes, thermometers, microphones, cameras, GPS, etc.—useable for tracking not only physiological and behavioral but also environmental parameters [10]. Through these capabilities, such use cases may integrate with UC2.

Coupled with the relevant analytical capabilities, such as those provided by *AI* and *big data* applications, use cases of *biosensors* and *wearables* serve to detect and monitor mental disorders and provide complementary insights into self-report instruments (**UC7**). For example, the unobtrusive measurement of activities of daily living and social rhythm, and even voice signals detected through sensors, can inform health care providers about the state of patients' chronic diseases, including depressive symptoms or bipolar disorders [e.g., 25–27]. Further studies suggest accurate monitoring opportunities for stress and anxiety, schizophrenia, and PTSD [10].

4. Discussion and Conclusion

Along with the emergence of digital technology use cases and evolving evidence for the effectiveness of their application in mental health care and therapy, bold promises are made while reservations linger that such applications might replace therapists or care workers. However, as appealing automation might appear from a naïve economic and techno-optimistic stance, the findings of our maturity analysis suggest that this scenario is unlikely to unfold.

According to the primary and secondary evidence we studied, a more realistic scenario for maximal diagnosis and treatment effectiveness and efficiency is the blending of human capabilities with digital technology application offerings, so-called *augmented intelligence*. Our findings also suggest that the use cases for these blended therapies and care applications vary significantly in their research and market maturity. While use cases of *mHealth/eHealth* and *telehealth* exhibit mature evidence for their effectiveness and certified market offerings, use cases of *AI*, *big data*, and similar applications are only now emerging technology trends, at most in a stage with promising experimental evidence (see Figure 1).

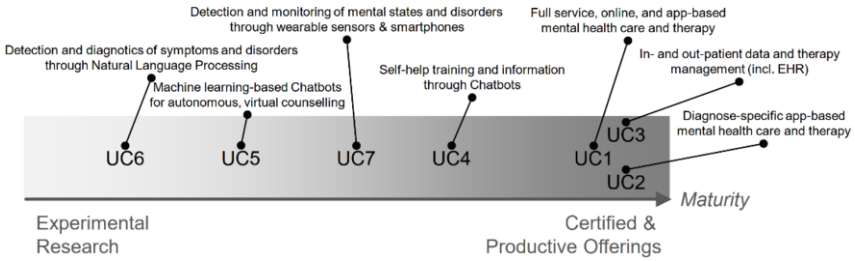


Figure 1. Maturity levels of the described digital technology use cases (UC1–7).

Our research provides an overview of the state of innovation for the application of digital technologies in mental health. Decision- and policy-makers should carefully observe the emerging evidence and offerings to continuously harvest the potential of these digital technology use cases and make them available for public use.

Acknowledgments

We are thankful to our interviewees. Furthermore, we acknowledge the valuable contribution of Dr. Markus Wyss to the data collection and selected parts of the analysis.

References

[1] Taylor CB, et al. Using Digital Technology to Reduce the Prevalence of Mental Health Disorders in Populations: Time for a New Approach. J Med Internet Res. 2020;22(7):e17493.
[2] Ben-Zeev D. The Digital Mental Health Genie Is Out of the Bottle. PS. 2020;71(12):1212-1213.
[3] Tailor K, The Patient Revolution: How Big Data and Analytics Are Transforming the Health Care Experience. Hoboken, NJ: John Wiley & Sons; 2016.
[4] Kim YK, editor. Frontiers in Psychiatry: Advances in Experimental Medicine and Biology. Singapore: Springer; 2019.
[5] Copeland B, Raynor M, Shah S. Top 10 health care innovations: Achieving more for less. deloitte.com/content/dam/Deloitte/us/Documents/life-sciences-health-care/us-lshc-top-ten-health-care-innovations.pdf [cited 2017 January 31].
[6] Fakhoury M. Artificial Intelligence in Psychiatry. In: Kim YK, editor. Frontiers in Psychiatry: Advances in Experimental Medicine and Biology. Singapore: Springer; 2019. p. 119-125.
[7] Ebert DD, Harrer M, Apolinário-Hagen J, Baumeister H. Digital Interventions for Mental Disorders: Key Features, Efficacy, and Potential for Artificial Intelligence Applications. In: Kim YK, editor. Frontiers in Psychiatry: Advances in Experimental Medicine and Biology. Singapore: Springer; 2019. p. 583-627.
[8] Monteith S, Glenn T, Geddes J, Bauer M. Big data are coming to psychiatry: a general introduction. Int J Bipolar Disord. 2015;3(1):21.
[9] Ressler KJ, Williams LM. Big data in psychiatry: multiomics, neuroimaging, computational modeling, and digital phenotyping. Neuropsychopharmacology. 2021;46(1):1-2.
[10] Sheikh M, Qassem M, Kyriacou PA. Wearable, Environmental, and Smartphone-Based Passive Sensing for Mental Health Monitoring. Front Digit Health. 2021;3:662811.
[11] Hollis C, et al. Technological innovations in mental healthcare: harnessing the digital revolution. The British Journal of Psychiatry. 2015;206(4):263-265.
[12] Berger T, et al. Evaluating an e-mental health program ("deprexis") as adjunctive treatment tool in psychotherapy for depression: Results of a pragmatic randomized controlled trial. J Affect Disord. 2018;227:455-462.
[13] Aitken M, Nass D. Digital Health Trends 2021: Innovation, Evidence, Regulation, and Adoption. iqvia.com/-/media/iqvia/pdfs/institute-reports/digital-health-trends-2021/iqvia-institute-digital-health-trends-2021.pdf [cited 2021 August 17].
[14] Weir K. The age of digital interventions. Monitor on Psychology. 2021 Oct 1;52:22.
[15] Rodriguez-Villa E, et al. The digital clinic: Implementing technology and augmenting care for mental health. Gen Hosp Psychiatry. 2020;66:59-66.
[16] Abd-Alrazaq AA, et al. An overview of the features of chatbots in mental health: A scoping review. Int J Med Inform. 2019;132:103978.
[17] Yasavur U, Lisetti C, Rishe N. Let's talk! speaking virtual counselor offers you a brief intervention. J Multimodal User Interfaces. 2014;8:381-398.
[18] Inkster B, Sarda S, Subramanian V. An Empathy-Driven, Conversational Artificial Intelligence Agent (Wysa) for Digital Mental Well-Being: Real-World Data Evaluation Mixed-Methods Study. JMIR Mhealth Uhealth. 2018;6(11):e12106.
[19] Fulmer R, et al. Using Psychological Artificial Intelligence (Tess) to Relieve Symptoms of Depression and Anxiety: Randomized Controlled Trial. JMIR Ment Health. 2018;5(4):e64.
[20] Swartout W, et al. Virtual Humans for Learning. AIMag. 2014;34:13-30.
[21] Nanomi Arachchige IA, Sandanapitchai P, Weerasinghe R. Investigating Machine Learning & Natural Language Processing Techniques Applied for Predicting Depression Disorder from Online Support Forums: A Systematic Literature Review. Information. 2021;12(11):444.
[22] Guntuku SC, et al. Detecting depression and mental illness on social media: an integrative review. Current Opinion in Behavioral Sciences. 2017;18:43-49.
[23] Ophir Y, Asterhan CS, Schwarz BB. The digital footprints of adolescent depression, social rejection and victimization of bullying on Facebook. Computers in Human Behavior. 2019;91:62-71.
[24] Nath S, Kurpicz-Briki M. BurnoutWords - Detecting Burnout for a Clinical Setting. In: Wyld D, Nagamalai D, editors. Proceedings of the 10th International Conference on Soft Computing, Artificial Intelligence and Applications (SCAI 2021); 2021 Nov 20-21; Zurich, Switzerland. Chennai: AIRCC Publishing; 2021. p. 177-191.
[25] Kim JY, Liu N, Tan HX, Chu CH. Unobtrusive Monitoring to Detect Depression for Elderly With Chronic Illnesses. IEEE Sensors Journal. 2017;17:5694-5704.
[26] Matthews M, et al. Development and Evaluation of a Smartphone-Based Measure of Social Rhythms for Bipolar Disorder. Assessment. 2016;23(4):472-483.
[27] Tokuno S. Pathophysiological Voice Analysis for Diagnosis and Monitoring of Depression. In: Kim YK, editor. Understanding Depression. Singapore: Springer; 2018. p. 83-95.

Apps for Patients and Healthcare Professionals Part 1 (Young Researcher)

This page intentionally left blank

Healthcare of the Future 2022
T. Bürkle et al. (Eds.)

doi:10.3233/SHTI220324

Intervention Platform for Action Observation and Motor Imagery Training After Stroke: Usability Test

Katrin FUCHSHOFER [a,b,1], Christian MERZ [a], Kerstin DENECKE [a], Corina SCHUSTER-AMFT [a,c,d]
[a] *Bern University of Applied Sciences, Bern, Switzerland*
[b] *FH Joanneum Graz, University of Applied Sciences, Graz, Austria*
[c] *Research Department, Reha Rheinfelden, Switzerland*
[d] *Department for Sport, Exercise and Health, University of Basel, Switzerland*

Abstract. Action observation (AO) and motor imagery (MI) are considered as promising therapeutic approaches in the rehabilitation of patients after a stroke (PaS). Observing and mentally rehearsing motor movements stimulate the motor system in the brain and result in a positive effect on movement execution. To support patients in the early rehabilitation phase after a stroke, ANIMATE, a digital health intervention platform was developed. The platform guides the user through 6 activities of daily living by observing and imagining the corresponding movements. We conducted a scenario-based usability test with 9 PaS at a rehabilitation centre to identify existing usability issues. PaS found the app easy to use and they could interact with it without problems. Although they judged the app as useful, they stated to be not willing to use the app on a regular basis. Including features for customising ANIMATE regarding the individual rehabilitation goals and needs of PaS, as well as personalisation could help in increasing the motivation to use and the benefits of the platform.

Keywords. neurorehabilitation, stroke, smart device application, motor imagery, action observation, user-centered design, usability

1. Introduction

A stroke is a massively life-changing event [1]. Due to the aging population, the absolute number of strokes per year will increase to 1.5 million in Europe by 2025 [2]. Action observation (AO) and motor imagery (MI) are powerful training techniques in motor learning and motor control, facilitating brain plasticity. Both are easy to learn [3] mental processes, allowing individuals to train a given motor act by observation or rehearsal even if they cannot perform it physically, e.g. due to an injury or paresis. In rehabilitation of patients after a stroke (PaS), the combination of physical practice (PP) and AO or MI can improve motor function more than PP alone [3,4]. Recent developments combined AO and MI with promising results (AO+MI) [3]. To support patients in either AO or MI,

[1] Corresponding Author: Katrin Fuchshofer, Department of Applied Computer Sciences eHealth, FH Joanneum Graz, University of Applied Sciences, Graz, Austria; E-mail: katrin.fuchshofer@edu.fh-joanneum.at

several apps are available. 'Recognise'[2] and 'Orientate'[3] support practicing left/right discrimination of body parts. 'Recovery Exercises'[4] promotes exercising everyday tasks and arm movements by AO or by physical practice. 'ViaTherapy'[5] offers some exercises for AO, some for MI, and some for performing tasks physically. Mirror Therapy VR[6] requires virtual reality glasses; the app simulates mirror therapy. No app is available that provides AO+MI training for subacute PaS. The aim of this work is to develop the ANIMATE app, a prototype addressing this research gap, and to study its usability.

2. Methods

System development and ANIMATE app. An interdisciplinary team of researchers from rehabilitation and medical informatics iteratively developed the intervention platform ANIMATE in a user-centred design process. The current prototype implements 6 activities of daily living (ADL; walking, drinking from a cup, getting up from a chair, washing hands, climbing and descending stairs, drawing a line with ruler and pencil). After having selected an activity, the user watches a video of a person performing it (AO). Next, the user is asked to imagine this activity (MI) for a predefined time. To minimise accessibility issues, user input is kept simple (e.g., large buttons, no control gestures like swiping or pinching). In an onboarding tutorial, the user is introduced to the AO+MI concept and the structure of the app. ANIMATE is realised as a progressive web application (PWA). It is available here: *https://orange-sky-041267503.azurestaticapps.net/*.

Usability test. Following the maxim of user-centred design, we aimed at studying the target group's perception of ANIMATE. Since no qualitative analysis was intended, we aimed at recruiting 10 [5] inpatients at the rehabilitation centre in Rheinfelden after approval by the responsible ethics committee (EKNZ Req-2021-01158). Eligible for participating were PaS aged 18 or older, being able to sit upright without assistance, use at least hand to operate a smart device, and understand spoken and written instructions in German. Recruitment and test took place between 13th and 21th November 2021. Potential participants received an information leaflet at recruitment, and gave informed consent immediately before the test.

The study consisted of a pre-test interview, a scenario-based test and a post-test interview. First, their affinity for IT, their experience with smart devices and (health/exercise) apps, and their expectations regarding our app were inquired. For the test, two scenarios had been designed to sequentially address all features implemented in the app: 1) "Imagine yourself being between two rehabilitation therapy sessions and eager to exercise some more. Start the app, watch the onboarding sequence, and do an exercise of your choice." 2) "You want to watch the onboarding tutorial again. Find and run it." The test supervisor observed the PaSs' interactions with the app, and documented the observations for each view in the app. If patients encountered problems, they were encouraged to find solutions themselves. In the post-session interview, the participants were asked some open questions (about experiences during the test and for ideas for improvement), and were finally guided through a questionnaire of established rating

[2] https://www.noigroup.com/product/recogniseapp/
[3] https://apps.apple.com/de/app/orientate-pain-management/id479540062
[4] https://www.my-therappy.co.uk/app/reps-recovery-exercises
[5] https://www.viatherapy.org
[6] https://play.google.com/store/apps/details?id=com.sixdimensions.mirrortherapy&hl=en&gl=US

items. The System Usability Scale questionnaire (SUS) [6] and the User Version of the Mobile Application Rating Scale (uMARS) [7] were customised to address the peculiarities of our app. We used 6 out of 10 SUS items and 10 out of 20 uMARS items (plus another 2, by slightly rephrasing 2 of those 10). It was not feasible to deploy both questionnaires in their entirety, nor to rely on only one of them. Furthermore, it would have been too strenuous for most candidates to answer all 30 items of SUS and uMARS combined. The complete test concept (German) is available online[7].

3. Results

From 10 volunteering PaS, 1 withdrew his consent before the test. The remaining 9 completed the entire test. Average Extended Barthel Index (EBI) was 44.8 (min: 32 - median: 44 - max: 60), Functional Independence Measure (FIM) 95.6 (62-101-124). Most PaS said they use smartphones, but some almost exclusively as telephones. Only 1 PaS deemed themselves notably interested in technology, and apps, and used a health app (steps counter). Few at least had heard of health apps, like fitness or diabetes apps.

Table 1 provides an overview on all 18 items of the questionnaire. All PaS felt confident in using the app, and they deemed the app easy to use (items 1,2,6). However, 4 PaS claimed they would need help to use the app. Most rated the provided information as useful and of good quality (items 9,10). In contrast, PaS could not imagine to use the app on a regular basis (item 12): 2 PaS indicated they would hardly use the app and 2 estimated they would certainly not use it within the next 7 days. Almost all PaS were not willing to pay for the app (item 13).

Table 1: Results from the customised usability questionnaire.

no.	Question	patient ID								Median	Average
		p1	p2	p3	p4	p5	p7	p8	p9	p10		
1	I managed well and felt confident while using the app.	4	2	5	5	4	4	5	5	4	4	4.2
2	I find the app very cumbersome to use.*	4	4	4	5	4	2	5	5	4	4	4.1
3	I'd need help while using the app.*	5	1	5	5	2	1	5	1	4	4	3.2
4	Most people will quickly learn to use the app.	2	4	4	4	4	5	3	3	3	4	3.6
5	The app is well adapted to the target user's needs.	3	5	5	4	2	5	5	5	3	5	4.1
6	Navigation works fine and is easy.	4	3	5	5	4	1	5	5	4	4	4.0
7	Layout is good (position and size of text, buttons, etc).	5	4	4	5	3	2	5	3	4	4	3.9
8	The app looks good and engaging.	3	NA	4	4	2	5	4	4	4	4	3.8
9	Provided information is important and understandable.	3	NA	5	4	4	5	5	5	4	5	4.4
10	Amount of provided information is good.	3	NA	4	5	4	5	5	5	4	5	4.4
11	Would you recommend the app to other patients?	3	4	2	4	3	5	4	5	4	4	3.8
12	How often would you use the app in the next 7 days?	2	2	4	4	4	1	5	1	3	3	2.9
13	Would you pay for the app?	2	3	4	5	1	1	5	2	3	3	2.9
14	Should the health insurance or the clinic pay for the app?	3	1	5	5	4	5	5	3	1	4	3.6
15	What is your overall opinion of the App? (stars)	4	3	3	4	3	4	4	4	4	4	3.7
16	What is your overall opinion of the App? (statements)	3	4	5	4	3	5	4	4	4	4	4.0
17	I'm better aware of how helpful solo, mental training is.	4	4	5	5	5	2	5	2	4	4	4.0
18	I'm more motivated for solo, mental training.	4	5	4	4	2	4	5	2	5	4	3.9
	SUS (x/60)	43	35	53	55	38	30	55	30	45	43	42.5
	uMARS (x/5)	3.2	3.2	4.2	4.4	3.1	3.7	4.7	3.8	3.5	4	3.7

Legend: * = answers reversed for this analysis (1=worst, 5=best); NA = not available (p02 said he had seen too little of the app to be able to rate items 8 to 10). Questions 5-16 contribute to the uMARS score, the others to the SUS score. For comparison, our total score needs to be adapted for SUS (x/60 instead of originally x/100), as a uMARS score is the average of all valid items anyway. The questions are shortened for depiction here.

[7] https://www.researchgate.net/publication/357543052_

The average SUS and uMARS values (42.5 out of 60 (normalised to the original SUS score maximum: 71.66/100), and 3.7 out of 5) can be interpreted as meaning "above average" to "good" (for interpreting SUS results, see [6]).

From our observations during the scenario-based tests we identified some usability issues: 1) sometimes navigation buttons were difficult to find and 2) PaS found it hard to follow text- or symbols-based directives in the app (in line with perceiving the voice-based instructions in the onboarding tutorial as helpful). While 1 PaS preferred the motor tasks suggested for the AO+MI training to be much simpler, like a basic finger exercise, another PaS considered that more complex AO+MI movements would be more valuable for the therapy process.

4. Discussion and conclusion

The results show that this current prototype is well received by patients. Since AO+MI are known to improve stroke rehabilitation outcome, the ANIMATE app may become a helpful instrument in further meliorating PaS chances of regaining autonomy in their lives. However, our study has some limitations. Variability in the scorings per item is wide. 5 out of 18 items received both extremes of the rating scale (1 and 5). For other items, however, the result is clear (e.g., overall impression, items 16 and 17). Although working with 9 participants is within the range of existing studies on the effectiveness of MI+AO, correlating user perceptions to clinical parameters was out of reach. Reasons for reservations against using the app still have to be assessed in more detail. Participants dismissed the idea of having to pay for the ANIMATE app; it might be that they expect the health insurance to pay for the app (item 14 in Table 1).

Potential for improvement resides mainly in providing customisation options, e.g. provide every activity for deficits on either body side. Further, GUI controls should be fine-tuned, and the range of ADLs widened. Gamification aspects (e.g. personalised avatar) may increase the time spent doing exercises, which is obviously crucial for rehabilitation outcome. A later step might include adding meta features that could help increasing the overall usefulness in daily routine, like user accounts, statistics of use, diary and/or self-assessment options, or superuser access (e.g., for managing and monitoring user access, changing global settings, exchanging components).

References

[1] Virani SS, Alonso A, Benjamin EJ, et al. Heart disease and stroke statistics-2020 update: A report from the American Heart Association. Circulation. 2020;141(9):e139-596.
[2] Béjot Y, Bailly H, Durier J, Giroud M. Epidemiology of stroke in Europe and trends for the 21st century. Presse Med. 2016;45(12, Part 2):e391-8.
[3] Eaves DL, Riach M, Holmes PS, Wright DJ. Motor Imagery during Action Observation: A brief review of evidence, theory and future research opportunities. Front Neurosci. 2016;10:514.
[4] Guerra ZF, Lucchetti AL, Lucchetti G. Motor Imagery training after stroke: a systematic review and meta-analysis of randomized controlled trials. J Neurol Phys Ther. 2017;41(4):205-14.
[5] Nielsen J, Landauer TK. A mathematical model of the finding of usability problems. In: Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems. Amsterdam: ACM Press; 1993;206-13.
[6] Bangor A, Kortum P, Miller J. Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. J Usability Studies. 2009;4(3):114–23.
[7] Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the user version of the Mobile Application Rating Scale (uMARS). JMIR Mhealth Uhealth. 2016;4(2):e72.

Healthcare of the Future 2022
T. Bürkle et al. (Eds.)

doi:10.3233/SHTI220325

A Structured Review and Evaluation of Android Mobile Applications for Yoga Support

Sayantani SARKAR[a], Yong K. CHOI[a], and Katherine K. KIM[a]

[a] *School of Medicine, Department of Public Health Sciences, Division of Health Informatics, University of California Davis, USA*

Abstract. Although there are hundreds of mobile yoga apps in the app market space, the quality and usefulness of these apps have not been systematically tested. We conducted a structured quality evaluation of apps from the Google Play store, applying the validated Mobile Application Rating Scale (MARS) by two independent raters. 18 out of 250 apps were identified for evaluation after applying inclusion/exclusion criteria. The mean MARS score is 4.11 (out of 5) with SD = 0.38. There was high interrater reliability (ICC = .88; 95% CI 0.85-0.91). Apps performed well on functionality and aesthetics. However, there is much room for improvement in information and engagement. Designers and researchers should focus on improving user engagement and building the evidence base for informational content provided in apps.

Keywords. mobile application quality assessment, Android, yoga app

1. Introduction

An age-old eastern healing tradition, yoga is widely accepted and practiced by approximately 36.7 million Americans [1]. Traditionally taught by in-person training, some of the issues such as time and transportation might hinder the continuity of practice. A previous intervention study using a mobile application (app) that offered yoga, meditation, sound, breathing exercise, health advice (e.g., diet, exercise) found significant improvement in mental health status [2]. Apps may offer the convenience of doing yoga when an individual chooses and eliminate the need for transportation and scheduling. To provide a better user experience, evaluation of the functionality and quality of the yoga apps is necessary. Despite the widespread availability of commercial yoga apps, there is a lack of evidence about the potential usability of these tools. In this pilot study, we aimed to rigorously evaluate mobile yoga tools from the Google Play store with a validated tool.

2. Methods

Two distinct processes were applied for this study. First, we selected apps based on our selection criteria established a priori. In the second step, we evaluated our final sample of apps with a validated tool. We used the Python library, Google-Play-Scraper, to retrieve app information from the Google Play store. We used the search keyword 'yoga' on 4/18/2019 for retrieving yoga-related apps.

We selected apps for inclusion if they fulfilled the following criteria based on the description mentioned in the Play store: a) in English; b) for general adult population; c) had a recent update in the year 2017 or later; d) had textual, visual, and audio guidance of the Asana or yoga postures; e) are free to download and offer some yoga support free of cost; and f) had current star rating above 4 with a minimum of 1,824 raters. The number of raters criterion was selected as an indication that the app was commercially available and in active use (rather than being recently published and potentially untested) and represented the mean number of raters among all 142 apps with minimum 4-star rating identified in the search. The median number of raters was 84, and the range was 2 to 86,761. We excluded apps that offered only yoga wallpapers, yoga magazine articles, finger yoga poses, or meditation music as they lacked audio-visual demonstration and guidance for body yoga poses.

We applied the Mobile Application Rating Scale (MARS) developed by Stoyanov et al. because of its widespread use as an instrument for assessing app quality along multiple dimensions (engagement, functionality, aesthetics, information, and subjective quality) and its high internal consistency (Cronbach alpha = .90) [3]. Each of these MARS components consists of a 5-point Likert scale (1-Inadequate, 2-Poor, 3-Acceptable, 4-Good, 5-Excellent) with distinct and identifiable characteristics. The investigators agreed upon the methodology and the interpretation of the MARS components. Two raters reviewed each app independently, spending 30 to 40 minutes with each. After individually rating apps, they met to discuss their ratings and consulted with a third investigator to mediate and resolve any discrepancies in the interpretation of the scale or ratings.

3. Results

The flow chart (Figure 1) describes the step-by-step process of screening and selecting 18 apps for the final analysis from the initial pool of 250 apps. Based on the ratings (shown in Table1), the Track Yoga app received the highest score (4.82), which is 0.32 higher than its published average star rating in the Google play store. The lowest MARS score observed was the Yoga Challenge App1, although this app had a higher published star rating. Most apps performed well on functionality (M=4.65; SD=0.34) and aesthetics (M=4.22; SD=0.41) compared to the engagement (M=3.83; SD=0.62) and information (M=3.75; SD=0.83) domains. There was excellent inter-rater reliability between the two independent raters (two-way mixed ICC = 0.88; 95% CI 0.85-0.91). Similar inter-rater reliability was noted in another study using MARS to evaluate mindfulness apps [4].

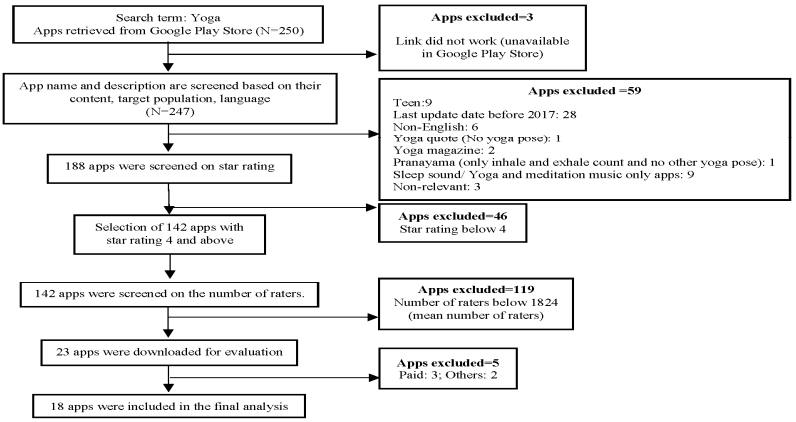


Figure 1. Flow chart describing the app selection process

Table 1. The MARS scores compared to Google Play star ratings for yoga applications

Name of App	Average sub-score of two reviewers				Average total score (a+b+c+d)/4	Google star rating*
	Engagement	Functionality	Aesthetics	Information		
Daily Yoga – Yoga Fitness Plans	4.50	3.50	4.33	3.75	4.02	4.4
Keep Yoga - Yoga &Meditation, Yoga Daily Fitness	4.30	4.62	4.16	4.08	4.29	4.7
5 Minute Yoga	3.30	5.00	4.49	4.08	4.21	4.5
Yoga poses & Classes	3.57	4.75	4.33	3.83	4.12	4.3
Yoga-Track Yoga	4.90	5.00	4.83	4.58	4.82	4.5
Yoga for weight loss -Loss weight in 30 days plan	4.70	4.5	4.33	4.33	4.46	4.6
Simply Yoga - Fitness Trainer for Workouts & Poses	3.70	4.62	4.00	4.33	4.16	4.1
Yoga Challenge App1	3.00	4.37	3.83	1.58	3.19	4.4
Yoga daily fitness - Yoga workout plan	4.30	4.50	4.00	3.91	4.17	4.6
Yoga Studio: Mind & Body	3.40	4.50	4.83	3.74	4.11	4.3
Yoga Workout - Yoga for Beginners - Daily Yoga	4.50	4.87	4.49	3.99	4.46	4.7
Yoga for Beginners	3.40	5.00	4.66	4.24	4.32	4.3
7pranayama: Yoga Daily Breath Fitness Habit – Calm	4.00	4.62	3.83	4.49	4.23	4.6
Yoga Flexibility for Beginners	3.60	4.62	4.49	4.08	4.19	4.1
Yoga for Kids	3.30	4.62	3.66	2.41	3.49	4.4
Yoga for Weight Loss	4.20	4.75	4.66	3.99	4.40	4.2
Complete Yoga Guide	2.60	4.87	3.66	4.08	3.80	4.2
Yoga Challenge App2	3.70	5.00	3.49	2.16	3.58	4.6

Note. Yoga Challenge App1 and Yoga Challenge App2 are two different apps with the same name.
* The scores are based on the app versions from the time of our search (2019) and may not be representative of the most up-to-date versions of the apps.

4. Discussion

Most apps (78%) scored above 4 on MARS (M=4.11; SD=0.38, range 3.19 to 4.82). indicating little variability among apps. This may be due to the selection criterion of 4-star rating or above. The majority of apps (72.2%) also had a lower MARS score than the star rating. Our findings highlight the need for attention to the information and engagement domains compared to the functionality and aesthetics. A low score in the engagement domain suggests that developers should focus on enhancing user engagement strategies (e.g., gamification). Another study [4] with mindfulness apps also noted low engagement scores and recommended design-specific attention in this domain. The lowest MARS score was observed in the information domain, which is consistent with other app review studies [5], indicating the lack of evidence-based content.

We were unable to consider the MARS item 19 (Evidence Base), which asks the rater to assess the evidence from the literature as most of the apps were not represented in the scientific literature, consistent with other studies [4]. There is a possibility that high-quality apps were not included in our review since we excluded apps with low number of raters, apps requiring payment, and those that did not provide full-body demonstration or guidance. In addition, use of single search term'yoga' might eliminate some of the apps that contain yoga support components. For example, an app with yoga content 'Nike Training Club' was missing from our app retrieval. Multiple key search terms may reduce the possibility of such automatic elimination of the target apps.

Our study findings highlight that the MARS is a useful tool for the initial quality evaluation and provides similar scores to Google star ratings. However, the use of the instrument alone cannot be a replacement for other evidence-based research methods involving end-users. There is a possibility that the recent status of the apps after publication may differ from the app status during evaluation. Despite limitations, our findings have given a preliminary idea about the status of the selected free content of the yoga apps and will encourage further research.

Acknowledgment. SS was funded by a doctoral fellowship from Betty Irene Moore School of Nursing. Special thanks to Kougang T Anne Mbe for contributions with app rating. Feedback on an early draft of this manuscript was provided by faculty of the Transatlantic Telehealth Research Network-CITRIS Summer Institute.

References

[1] 2016 Yoga in America Study Conducted by Yoga Journal and Yoga Alliance Reveals Growth and Benefits of the Practice [Internet]. 2016. Available from: https://www.yogaalliance.org/Get_Involved/Media_Inquiries/2016_Yoga_in_America_Study_Conducted_by_Yoga_Journal_and_Yoga_Alliance_Reveals_Growth_and_Benefits_of_the_Practice

[2] Hwang WJ, Jo HH. Evaluation of the effectiveness of mobile app-based stress-management program: A randomized controlled trial. International journal of environmental research and public health. 2019;16(21):4270.

[3] Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani MJJm, et al. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. 2015;3(1).

[4] Mani M, Kavanagh DJ, Hides L, Stoyanov SR. Review and evaluation of mindfulness-based iPhone apps. JMIR mHealth and uHealth. 2015;3(3):e4328.

[5] Bardus M, van Beurden SB, Smith JR, Abraham C. A review and content analysis of engagement, functionality, aesthetics, information quality, and change techniques in the most popular commercial apps for t management. International Journal of Behavioral Nutrition and Physical Activity. 2016;13(1):35.

Workflow Based Support in Patient Care (Young Researcher)

This page intentionally left blank

Healthcare of the Future 2022
T. Bürkle et al. (Eds.)

doi:10.3233/SHTI220327

Digitizing the ECG Workflow
A State-of-the-Art Analysis

Daniel REICHENPFADER[a], Mootaas ABU BAKER[a] and Thomas BÜRKLE [a,1]
[a] *Bern University of Applied Sciences, Biel, Switzerland*

Abstract. ECG is one of the most common examinations in hospitals, e.g. for diagnosing cardiovascular diseases - the most frequent cause of death worldwide. Goal of this paper was to identify and describe the typical digitized workflow, IT systems and data formats for ECG in hospitals: A survey on current ECG-data management practices was conducted with four German speaking hospitals. A generic model of ECG data management was drafted. Today, these hospitals do not use DICOM as exchange format nor do they implement IHE profiles such as REWF for the ECG. Reasons include missing IT infrastructure such as Master Patient Index or electronic archive. ECG data management could be improved at different levels, with the chance to reduce error sources and to improves in patient safety. Storage of ECG raw data promises better diagnoses based on big data and machine learning technologies.

Keywords. Interoperability, IHE, ECG data

1. Introduction

Bern University of Applied Sciences (BFH) provides a Medical Informatics Study of 180 ECTS and maintains a medical informatics living lab [1] with a Schiller AT170 ECG and SEMA ECG Software as well as an Orchestra integration server and a Synedra image archive. In a student project the task was to implement a use case for inpatient ECG ordering and results communication (CPOE) in the lab. For the background, we wanted to know which ECG modalities, IT systems and communication standards are today typical for German speaking hospitals. Any obstacles to implementation were to be identified, alternatives highlighted and suggestions for improvement defined.

2. Methods

A questionnaire with 20 questions regarding IT architecture, existing hard- and software as well as details on how ECG recordings and data are managed was distributed to 9 hospitals in Germany, Austria and Switzerland (DACH region). Four hospitals (2 Swiss, 1 German, 1 Austrian) agreed to either provide written feedback or a semi structured telephone interview. Results were analysed and compared and an architectural concept for ECG integration was drafted. In addition, a functional use case was implemented in the BFH medical informatics lab.

[1] Corresponding Author: Thomas Bürkle, Bern University of Applied Sciences, Quellgasse 21, Biel, Switzerland, E-Mail: thomas.buerkle@bfh.ch

3. Results

Results of the survey have been condensed to an integration architecture (Fig 1). The participating Swiss hospitals A and B had 200 respectively 285 beds. An Austrian hospital with 1500 beds (C) and a German hospital with 1400 beds (D) co-operated. Three hospitals used SAP IS-H, one Nexus Hospis for ADT data <2>. Patients are registered on admission and data is sent using HL7 to the subsystems including the ECG management system (EMS) <1>.

For ECG recording, PC-based systems with LAN connection, ECG printers and mixed functionality systems from Schiller, Custo Med, GE and Siemens are used <3>. Patients are identified by barcode ID, which is either typed or scanned at the modality. ECG printers used in the ED and shock room typically have no connection to the EMS. Instead, recordings are printed to support rapid diagnosing. In one institution, devices with both functionalities are used in the ED. Reports are printed and sent to the EMS.

All hospitals used a proprietary EMS <1> as the "hub" for the entire data exchange between the subsystems of the hospital information system (HIS) and the ECG modalities. Systems in use were Muse (GE), Sema (Schiller) and Customed (Custo Med). Furthermore, all hospitals stored PDF files of the ECG in an archive system <4>. Typically, the EMS converts the proprietary format of modalities into a pdf file and transmits this file, encapsulated in an ORU or MDM message to the archive solution of the institution (Synedra AIM Universal Archive respectively Doxis4/SER group). Depending on the institution, the organizational unit and the EMS, ECG reporting and validation is performed either in the EMS or a clinical workstation (CWS) <5> of the HIS (Kisim/Cistec, Soarian/Cerner, IS-H*Med/SAP, Phoenix/CGM) using front end integration with the archive solution (n=2). One hospital sends ECG PDF files directly to the CWS for temporal storage. One institution transmits ECG measurement data to the CWS via HL7-ORU.

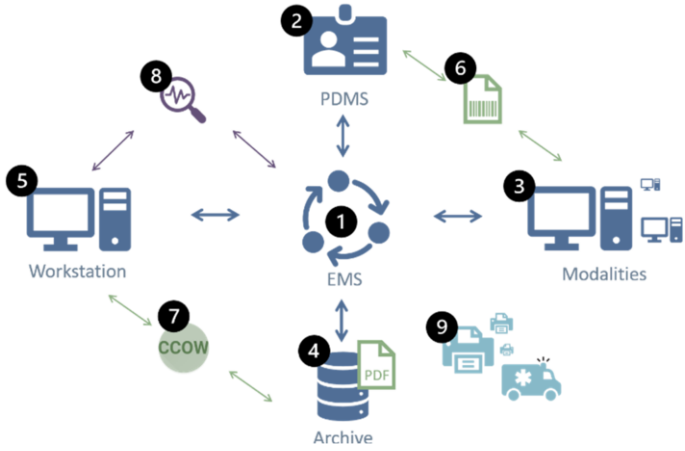


Figure 1. Typical ECG recording architecture within the DACH region. References in brackets

In the two largest institutions, C and D, order communication has been implemented for ECG. These orders are sent to the EMS and a task list can be retrieved on the modality. None of the four hospitals used IHE profiles in the ECG workflow. Reasons comprised

high complexity and implementation effort combined with missing functionality. Neither was DICOM used in any of the hospitals for data exchange or storage.

4. Discussion and conclusion

The IHE initiative provides the profiles Resting ECG Workflow (REWF) and Scheduled Workflow B (SWF.b) for ECG integration [3,4]. REWF, a supplement of the Cardiology Technical Framework (CTF), is still in trial implementation dating back to 2013. It covers three use cases such as scheduled ECG recording using order management, recordings of non-identified patients with subsequent identification and scheduled ECGs recorded offline. Furthermore, the diagnostic process is part of the profile as well. REWF defines specific DICOM tags for transmitting additional ECG data but does not cover long-term or stress ECG. The permanent trial status, missing functionalities, and no update of CTF since 2013 may discourage institutions from implementing REWF.

SWF.b was included in the Radiology Technical Framework (RTF) in its latest revision in 2020. It defines generic use cases including order management and fulfilment of identified and unidentified patients as well as various exceptions and corrections. In contrast to REWF, diagnosis and display of ECG recordings are not part of SWF.b but defined in other RTF profiles and no ECG-specific DICOM tags are defined. In comparison, SWF.b seems to be the better choice: it covers many use cases whose implementation in REWF would not be possible, e.g., update of obsolete /missing patient master data. The generic specification of SWF.b profile should enable long-term and stress ECGs in future.

In our survey, none of the four hospitals used DICOM or IHE for the digital ECG workflow. Instead, data exchange is based upon HL7 and PDF files. Although this is a small sample, other studies suggest that in addition important systems such as MPI or digital archives may be unavailable in many hospitals in the DACH region [2]. Besides, respondents in our survey saw no added value in the use of IHE and DICOM compared to the PDF based solution. Thus, for SWF.b the "Results Distribution" profile would be needed to support export of the examination results as a PDF or CDA document, encapsulated in an HL7 ORU message.

Stamenov et al. identify four common ECG data formats [5]. Plain pdf has the disadvantage to lose most of the acquired numerical data [6]. An effective way to preserve raw data while having minimal impact on existing processes is provided by Sassi et al. with the PDF-ECG format. With only minimal changes to the workflow and a file size increase of ~27% on average, this format enables the preservation of the original measurement data [7-9].

Based on our findings, we would like to present the following recommendations: First, the missing IT infrastructure components that hinder the implementation of IHE profiles should be identified and measures to reduce or remove these barriers defined. In Germany, such investments could be financed based on the Krankenhauszukunftsgesetz, a state-funded investment program to promote digitalization of hospitals [10]. For institutions that want to increase the data quality and availability of ECG recordings or are considering changing their system(s), the following options with increasing complexity are available:

- Implementation of the PDF-ECG format with minimal impact on current processes [9].

- Review of the current infrastructure and realization of missing components such as MPI, electronic archive and order communication [2].
- Parallel implementation of DICOM data exchange based on the SWF.b profile. A good overview of this is provided in [11,12].
- Full implementation of the SWF.b profile at process and semantic level.

The preservation of ECG raw data and measurements provides various future analysis options, e.g., data mining and machine learning, which may generate new medical knowledge and better diagnostic and treatment options [13-15].

References

[1] Holm J, Bürkle T, Gasenzer R, von Kaenel F, Nüssli S, Bignens S, Il Kim S, Lehmann M. A Novel Approach to Teach Medical Informatics. Stud Health Technol Inform. 2015;216:1011.
[2] Hübner U, Esdar M, Hüsers J, et al. IT-Report Gesundheitswesen 2020 [Internet]. Forschungsgruppe Informatik im Gesundheitswesen (IGW); 2020 [cited 2022 Jan 5]. Available from: https://www.hs-osnabrueck.de/fileadmin/HSOS/Homepages/IT-Report_Gesundheitswesen/IT-Report_Gesundheitswesen_2020_final.pdf.
[3] IHE Cardiology Technical Committee. IHE Cardiology Technical Framework Supplement. Resting ECG Workflow (REWF). Trial Implementation. 2013 [cited 2022 Jan 5]. Available from: https://www.ihe.net/uploadedFiles/Documents/Cardiology/IHE_CARD_Suppl_REWF.pdf.
[4] IHE International, Inc. IHE Radiology Technical Framework. Volume ,IHE RAD TF-1, Integration Profiles. Revision 19.0 2020 [cited 2022 Jan 5]. Available from: https://www.ihe.net/uploadedFiles/Documents/Radiology/IHE_RAD_TF_Vol1.pdf.
[5] Stamenov D, Gusev M, Armenski G. Interoperability of ECG standards. 2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO). 2018.
[6] Badilini F, Young B, Brown B, et al. Archiving and exchange of digital ECGs: A review of existing data formats. Journal of Electrocardiology. 2018;51.
[7] Sassi R, Sparagino L, Stockbridge NL, et al. Proof of concept for an international long-time preservation ECG format. Computing in Cardiology 2014. 2014. p. 461–464.
[8] Badilini F, Sassi R. Development of PDF-ECG: Further steps towards the long-term preservation of clinical electrocardiograms. Journal of Electrocardiology. 2016;49:753–754.
[9] Sassi R, Bond RR, Cairns A, et al. PDF-ECG in clinical practice: A model for long-term preservation of digital 12-lead ECG data. J Electrocardiol. 2017;50:776–780.
[10] Bundesministerium für Gesundheit. Krankenhauszukunftsgesetz (KHZG) [Internet]. 2020 [cited 2022 Jan 6]. Available from: https://www.bundesgesundheitsministerium.de/krankenhauszukunftsgesetz.html.
[11] Schwind F, Münch H, Schröter A, et al. Long-term experience with setup and implementation of an IHE-based image management and distribution system in intersectoral clinical routine. International Journal of Computer Assisted Radiology and Surgery. 2018;13.
[12] Wu C, Chiu R, Yeh H, et al. Implementation of a Cloud-Based Electronic Medical Record Exchange System in Compliance with the Integrating Healthcare Enterprise's Cross-Enterprise Document Sharing Integration Profile. International Journal of Medical Informatics. 2017;107.
[13] Denaxas SC, Asselbergs FW, Moore JH. The tip of the iceberg: challenges of accessing hospital electronic health record data for biological data mining. BioData Min. 2016;9:29.
[14] Roopa CK, Harish BS. A Survey on various Machine Learning Approaches for ECG Analysis. International Journal of Computer Applications. 2017;163:25–33.
[15] Navarro-Paredes C, Jing M, Finlay D, et al. P8 Big data platform for cardiovascular healthcare data storage and data viewer. Heart. 2019;105:A8–A8.

Healthcare of the Future 2022
T. Bürkle et al. (Eds.)

doi:10.3233/SHTI220328

An Interoperable Resuscitation Registry for the University Hospital of Bern

Marko MILETIC[a,1], Manuela ITEN[b], Thomas BÜRKLE[a], Alain NIPPEL[a]
[a] *Bern University of Applied Sciences, Biel, Switzerland*
[b] *Department of Intensive Care Medicine, Inselspital, Bern University Hospital, University of Bern, Switzerland*

Abstract. During resuscitation, the patient is the primary focus with the documentation of actions and outcomes being secondary. In most cases, a cardiac event leads to further treatment or hospitalization, in which complex patient pathways, independent documentation systems and information loss represent the key challenges for successful quality management. Hence, the need for a system that takes all these aspects into account. Market research, system analysis and requirements engineering for such a solution were performed and a prototype was created. A complete reference architecture for a web-based electronic data capture system was developed and implemented that enables healthcare professionals to enter resuscitation-relevant data uniformly and store it centrally in compliance with human research legislation. A qualitative evaluation concerning the process flows of the as-is and the to-be situation suggests that there is potential to achieve benefits in the form of improved data quality and quantity.

Keywords. resuscitation registry, electronic data capture, quality management

1. Introduction

With 8,500 cases per year, cardiovascular arrest is a common occurrence in Switzerland. The guiding principle "It takes a system to save a life" shows how important it is for various actors in the rescue chain, i.e., emergency call centers, emergency services (ES), emergency departments (ED) and the intensive care unit (ICU), to operate using coordinated measures [1]. The electronic documentation of all resuscitation activities represents a critical success factor for patient survival, as well as process and outcome quality assurance in the treatment process [2]. Tablet or even app-based systems have been presented previously [3,4,5] and demonstrated positive effects on accuracy [3,4] and completeness of data [4] with the potential to improve cardiopulmonary resuscitation [5]. There is however a lack of interoperability with other IT systems and registries. Jensen et al. report on a Danish electronic pre-clinical medical record which is interoperable with the ED information system [6]. Other commercial systems, e.g., NaProt by pulsation IT and MedicalPad by WEINMANN Emergency Medical Technology, focus mostly on the pre-clinical part of the rescue chain.

As a result, clinicians are often confronted with fragmented data in different IT systems and forced to compile additional comprehensive data sets, e.g., for the follow-

[1] Corresponding Author, Marko Miletic, Bern University of Applied Sciences, Quellgasse 21, 2501 Biel, Switzerland; E-mail: marko.miletic@bfh.ch

up of resuscitation patients (RP) in registries for quality assurance. At the University Hospital of Bern (Inselspital), these sources include several different ES and clinical professionals from various departments, e.g., the ICU, whose data forms are as heterogeneous as their respective medical fields. Pre-clinical data from local ES is collected in the Swiss resuscitation registry SWISSRECA. In addition to digital documentation in the ED system, the Inselspital ED collects internal paper-based documentation. Other patient-relevant information for RPs is collected in the primary systems of the respective departments. The complete data of the resulting rescue chain is collected retrospectively by study assistants and transferred to the German Resuscitation Registry (GRR) of the German Society of Anaesthesiology and Intensive Care Medicine in conformity with the international standards of the Utstein-style protocol [7].

In this situation, the Inselspital ED requested support to implement a mobile tablet based application which should replace parts of the separate documentation and result in an ED registry for RPs with an enriched dataset including preclinical data from the different ES. The registry should include clinical follow-up data and support internal quality benchmarking and export to the GRR.

2. Methods

Initially, the process flows of the rescue chain at the Inselspital were analyzed and modelled in Business Process Model and Notation (BPMN).

Different IT system architectures suitable for real-time operation were examined and discussed with the partner hospital and the Swiss Society for Intensive Care Medicine (SGI), because the future system should be made available to other Swiss EDs as well. As a result, it was decided to use a REDCap database for the ED registry, which could potentially be hosted at the SGI for all Swiss EDs. No patient data is stored locally or outside of REDCap. The resulting final IT architecture is given in Figure 1.

The next step comprised extensive dataset mapping activities. The continuing care dataset WV-CAC version 1.0 from the GRR [8] was used as the future core dataset following in-depth analysis and extensive comparison with its predecessors. Assignable data fields between the WV-CAC and the SWISSRECA dataset were analyzed and mapped to ensure interoperability. Local data fields for adaption to an individual ED were devised.

The IT architecture was implemented using the agile software development method Scrum and consists of several components. These include: a Node.js backend, a NGINX web server, an encrypted PostgreSQL user database for local authorization management, a KeyStore, an initialization script and the web application written with the JavaScript framework Angular and the Angular Material UI design library. Systematic tests were conducted with a focus on plausibility mechanisms, correct data persistence and consistent data import. In addition, an internal safety audit was carried out.

To evaluate the potential benefits, the time spent for gathering the different data items today for the data entry into the GRR was measured and contrasted with the future workflow, which was also modelled in BPMN.

3. Results

The resulting client-server architecture called ReaReg can be seen in Figure 1. It comprises the REDCap database for the clinical data, a ReaReg web server to connect the tablets and PCs in the ED and the import interface for ES data. The use of REDCap as the central data storage ensures, that highly sensitive patient data can be saved securely. The web server and middleware can be deployed in the hospital network, whereas the REDCap instance could be hosted with a neutral third party such as the SGI for use by other institutions.

The architecture supports real-time documentation through a web application and standardized storage compliant to the WV-CAC dataset [8]. An import interface adds pre-clinical data from SWISSRECA, provided by surrounding ES using the secure Swiss medical communication network (HIN). Record linkage is achieved using the unique mission number, hereinafter referred to as the protocol number that is given to each Swiss RP by the ES. During ED and inpatient treatment, clinicians document activities and outcomes on the tablet and receive additional support in the form of assistance functions and plausibility mechanisms. Special data security precautions were taken to prevent unauthorized access, e.g., a KeyStore that can only be accessed by a trusted party.

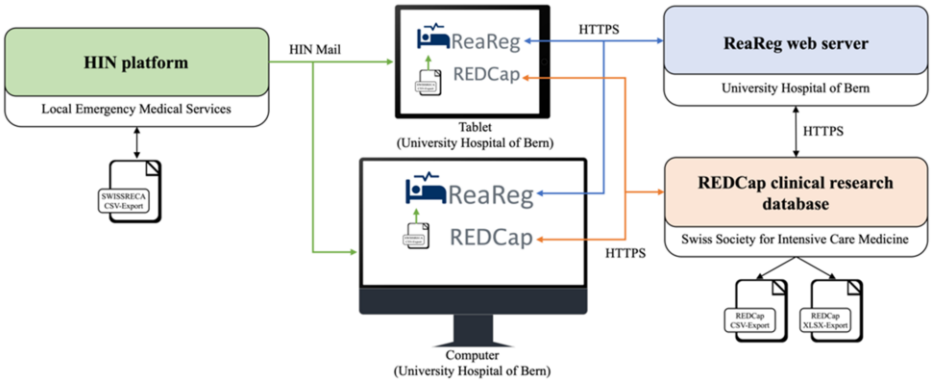


Figure 1. System architecture of ReaReg.

The as-is and to-be BPMN documentation workflows have three parts, “patient admission/patient transfer”, “hospital-internal care” and “dataset completion”. For “dataset completion” the ED staff currently spend an average of 19.6 minutes per RP (n=15). This will be considerably reduced by means of the future automated data export to the GRR. Improvements in documentation quality may be expected due to the automated record linkage of imported data from the external ES using the protocol number, which is also used for export to the GRR. In the workflow “hospital-internal care” we could pinpoint three examples which result in better data quality and quantity: a) the web application enables real-time documentation of resuscitation-relevant measures and outcomes using mobile devices in the ED and ICU, b) data is ubiquitously accessible independent of the primary systems, c) information currently lost for the reanimation registry can be persisted through documentation in ReaReg. The latter concerns, e.g., ECG data, which is currently only available on paper for ICU patients. Furthermore, data quantity within ReaReg will profit from the import of the pre-clinical care protocols from SWISSRECA which are available earlier. In addition, ReaReg can support data analysis and descriptive statistics using REDCap functionalities.

4. Discussion

Today ReaReg has been fully implemented for testing, but not deployed at the Inselspital as the required web server is missing. Therefore, we do not yet have proof for success.

The concept presented here aims for a workflow where the tablet should accompany the patient during his stay, e.g., to the ICU. The decision not to integrate with the different IT systems of the ED and ICU was made deliberately. Fortunately, the number of patients arriving in an ED under resuscitation is limited. Our goal was the reconstruction of patient pathways to support retrospective studies for research and enable better treatment quality. More often than not, the data quality for such research is not contained in the clinical information systems. But before real-time usage, we will not know if the duplicate documentation will be accepted. The plan is to engage senior students with this task. It may be necessary to establish appropriate incentive systems to ensure that data collection is carried out consequently.

Similar work focuses on ensuring data quality and quantity using tablet or web-based solutions that are not or only partially connected to other systems [7], i.e., mostly in the pre-clinical setting, but without connections to registries [4,5]. The ReaReg system provides added value with import and export interfaces and semantic mapping from SWISSRECA and towards GRR. Currently, GRR cannot yet import WV-CAC, but plans its realization. ReaReg has the potential to eliminate the need for error-prone retrospective manual data collection and transfer to registries.

As soon as registries such as the GRR provide import interfaces for continuing care datasets, considerable time savings and improved data quality can be achieved and collaboration across the rescue chain can be enhanced. With the introduction of a web-based data collection system, the objective of improved outcome and process quality for patients under resuscitation may come closer. ReaReg has been published open source under the GPLv3 license. Thus, we would like to stimulate its use and further development in the hope of qualitative and empirical evaluation of the system regarding time savings and error reduction.

References

[1] Burkart R, Regener H. Die erste nationale Überlebensstrategie bei Kreislaufstillstand. Schweiz Ärzteztg. 27. November 2019;100(48):1615–7.

[2] Messelken M, Fischer M, Wnent J, Seewald S, Gräsner J-T, Andresen D, u. a. Ohne Daten kein messbarer Fortschritt. Notfall Rettungsmed. Juni 2014;17(4):327–8.

[3] Grundgeiger T, Albert M, Reinhardt D, Happel O, Steinisch A, Wurmb T. Real-time tablet-based resuscitation documentation by the team leader: evaluating documentation quality and clinical performance. Scand J Trauma Resusc Emerg Med. 2016 Apr 16;24:51.

[4] Peace JM, Yuen TC, Borak MH, Edelson DP. Tablet-based cardiac arrest documentation: a pilot study. Resuscitation. 2014 Feb;85(2):266-9.

[5] Grundgeiger T, Hahn F, Wurmb T, Meybohm P, Happel O. The use of a cognitive aid app supports guideline-conforming cardiopulmonary resuscitations: A randomized study in a high-fidelity simulation. Resusc Plus. 2021 Aug 17;7:100152.

[6] Jensen FB, Ladefoged KT, Lindskou TA, Søvsø MB, Christensen EF, Teli M. Understanding the Effect of Electronic Prehospital Medical Records in Ambulances: A Qualitative Observational Study in a Prehospital Setting. Int J Environ Res Public Health. 2021 Feb 27;18(5):2330.

[7] Otto Q, Nolan JP, Chamberlain DA, Cummins RO, Soar J. Utstein Style for emergency care – the first 30 years. Resuscitation. 1. Juni 2021;163:16–25.

[8] Seewald S, Jakisch B, Gräsner J-T, Brenner S, Jantzen T, Fischer M, Bohn A, Wnent J, Bein B. Strukturierte Datenerfassung nach erfolgter Reanimation: Datensatz Cardiac Arrest Center. Anästh Intensivmed 2020;61:V1–V3.

Apps for Patients and Healthcare Professionals Part 2 (Young Researcher)

This page intentionally left blank

Healthcare of the Future 2022
T. Bürkle et al. (Eds.)

doi:10.3233/SHTI220330

Rehago - A Home-Based Training App Using Virtual Reality to Improve Functional Performance of Stroke Patients with Mirror Therapy and Gamification Concept: A Pilot Study

Chien-Hsi CHEN[a,c], Theresa KREIDLER[b,c], and Anika OCHSENFAHRT[c]
[a] *Institute for Medical Informatics, Statistics and Epidemiology, Medical Faculty, University Leipzig, Germany*
[b] *Faculty of Behavioural and Social Science, Technical University of Chemnitz, Germany*
[c] *ReHub GmbH, Germany*

Abstract. Stroke is one of the prevalent diseases which leads to functional disabilities such as hemiparesis or hemiplegia. It is common practice to treat patients with proper rehabilitation as early as possible for better prognosis after the onset of stroke. One of the effective therapeutic techniques for treating stroke patients is mirror therapy, which can potentially facilitate patients' motor function recovery through repetitive practice. "Rehago" is a software as medical device that implements the concept of mirror therapy in combination with gamified exercises into virtual reality (VR) to provide a home-based rehabilitation environment for stroke patients. In this study, 48 stroke patients completed the full course of intervention with Rehago and their functional performance of pre- and post-intervention was investigated. The intervention with Rehago was predefined as 30 minutes training per day, 5 days per week over a course of 6 weeks. The patient's progress was evaluated by their therapists every 14 days, with a baseline assessment before the intervention began. The results showed an average improvement of 5.54 points in the Functional Independence Measurement score, and an improvement of 7.13 points in the assessed quality-of-life score (EQ5D-5L). An improvement of the FIM score and the quality-of-life score in EQ5D-5L was observed, indicating it is beneficial to the patients using Rehago as a home-based rehabilitation tool.

Keywords. virtual reality, VR, stroke, rehabilitation, home-based, mirror therapy, Rehago

1. Introduction

Cerebrovascular disease, also known as stroke, is one of the leading causes of death worldwide. 80% of stroke survivors suffer from motor impairments, muscle weakness (paresis) and pain [1]. The impairments and pains can dramatically reduce the patients' independence in performing activities of daily living (ADL) and quality of life. To retrieve the functions of the affected side, an intensive rehabilitation is typically introduced in the early stages of stroke. However, therapists often have limited time to treat each patient. Additionally, patients usually have lower motivation without professional guidance to continue rehabilitation at home.

Mirror therapy has been proven to be effective for stroke patients to improve motor function of the upper extremity and improve the performance of activities of daily living [2]. Unlike other conventional interventions relying on somatosensory stimulation, mirror therapy is mainly based on visual input to induce neurosynaptic reconnection via complex neurological mechanisms and commonly included in rehabilitation sessions [2]. As VR technology advances and becomes more cost-effective and accessible to the public, the possibility to offer the traditional rehabilitation in a digital form has been intensively researched and considered as a safe intervention [3]. Meanwhile, the concept of mirror therapy can be implemented in VR to help patients learn to complete simulated ADL tasks [3-5]. Weber et al. found that mirror therapy as a digital application improved upper extremity motor skills in stroke patients [4, 5]. Moreover, a playful and self-directed therapeutic method showed comparable effects to the conventional occupational therapy [5]. However, it is still unknown whether patients' functional performance can be improved with mirror therapy using VR. Therefore, the aim of this study was to preliminarily prove the concept that combining mirror therapy and gamification concepts in VR can improve the functional performance of stroke patients.

2. Methods

2.1. Subjects

Fifty-three stroke patients were initially recruited for this study. Five patients dropped out due to personal reasons, so in total 48 subjects (including 14 females and 34 males) completed the study. The detailed age distribution of the subjects is shown in Table 1. Inclusion criteria were (1) stroke survivors minimum two months post-stroke, (2) older than 18 years old, (3) able to understand and follow training instructions, and (4) able to use and operate a VR headset independently. Exclusion criteria were (1) musculoskeletal disorders such as shoulder pain or limited range of motion, (2) unstable cardiovascular status, (3) any cognitive or emotional disorder, and (4) other systemic disease such as epilepsy, diabetes, or hypertension. Additionally, patients were excluded if they were receiving conventional mirror therapy or other therapy using virtual reality.

Table 1. Subjects composition by age groups

Age group (year)	31-40	41-50	51-60	61-70	71-80	81-90	Total
Female	2	4	5	2	1	0	14
Male	2	6	9	8	8	1	34

2.2. Procedure

Before the study began, a brief introduction of this study was given to the subjects and the subjects signed the consent forms. Next, an explanation regarding the operation of the head-mounted display (HMD) and the Rehago software and a practice session was instructed and supervised by therapists to prevent any biased results due to technical issues. Therapists then suggested a series of training in Rehago to individual subject based on their ability and functional level. Daily training duration of 30 minutes was required for each subject to accomplish [11, 12]. After the subjects were fully informed with the training process and requirement, each subject's baseline data was evaluated by the therapists using Functional Independence Measure (FIM) and the EQ5D-5L, and a follow-up evaluation was conducted every 14 days using the same instruments (Fig. 1).

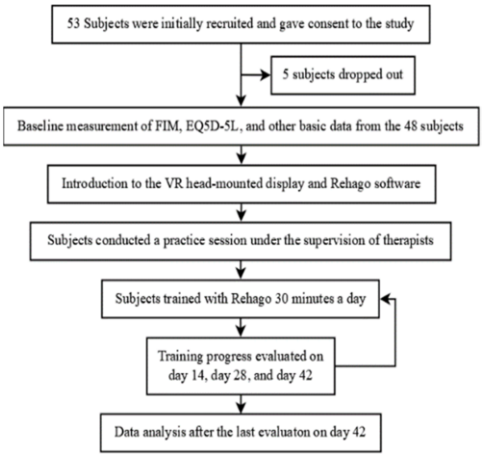


Figure 1. *Flow chart of the subject's journey*

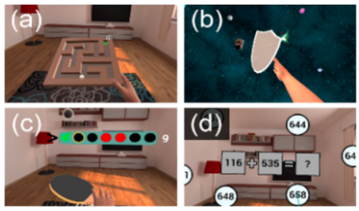


Figure 2. *Screenshots of training examples of Rehago (a) labyrinth (b) asteroids blocking (c) ball bouncing (d) calculations*

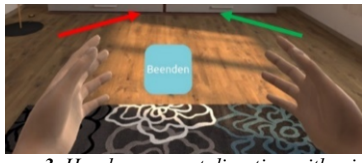


Figure 3. *Hand movement direction with mirror mode activated. If the controller is in right hand and move towards left (green arrow), the left hand will move towards right (red arrow).*

2.3. Hardware and Software

In this study, the standalone HMD Pico Neo 2 (Pico Interactive, China) was chosen for its performance and cost-effectiveness. The headset of Pico Neo 2 has 6 degrees of freedom (DOF) using inside-out tracking, 101 degrees of field of view, and is equipped with two controllers both with 6 DOF tracking. Meanwhile, Pico Neo 2 provides 4K display resolution to ensure the high fidelity of the content. Rehago software (v1.2.0, ReHub GmbH, Germany) was pre-installed on all the HMDs. Rehago is a CE marked software, which has 10 gamified trainings targeting different body parts for the purpose of upper limb rehabilitation for stroke patients. Screenshots of the example training in Rehago are shown in Fig. 2. A simulated mirror (therapy) mode was implemented in Rehago. If a stroke patient has left-side hemiplegia, he needs to hold the controller in the right hand and the movement of the right hand will be “mirrored” to the left hand, as shown in Fig. 3.

2.4. Outcome Measures

2.4.1 Functional Independence Measure (FIM)

This instrument measures the patient's independence of various aspects, such as self-care, including sphincter control, transfers, locomotion, communication, and social cognition. FIM is a reliable tool and is widely used to evaluate the progress of a patient's level of disability after receiving medical treatment or rehabilitation.

2.4.2 EQ5D-5L

The EQ5D-5L is a short questionnaire containing five questions with five possible answers each (5L), which was introduced by the EuroQol Group in 2009. EQ5D-5L contains five aspects, including mobility, self-care, usual activities, pain, or discomfort, as well as anxiety or depression. The “current perception of health”, Quality of Life (QoL), is evaluated on a scale from 1 to 100. This is a subjective score which reflects the

self-perceived health status of the subject. This parameter was recorded for quantitative analysis in terms of changes in QoL after using Rehago.

2.5 Statistics analysis

A paired t-test was performed to check whether there was a significant difference between the initial baseline and the follow-up evaluation (i.e., day 14, day 28, and day 42).

Table 2.: Mean scores (Mean ± SD) of the FIM, QoL, and EQ5D-5L from different time. The p-values were calculated against baseline scores (day 1) of each measure with the significance level set to 0.05.

Measures \ Time	Day 1	Day 14	Day 28	Day 42
FIM	101.48 ± 19.08	103.77 ± 17.79 (p = 0.001*)	105.96 ± 16.69 (p = 0.0002*)	107.02 ± 16.32 (p = 0.0001*)
QoL	69.25 ± 17.63	70.25 ± 17.93 (p = 0.397)	74.81 ± 15.53 (p = 0.003*)	76.38 ± 13.61 (p = 0.00003*)
EQ5D-5L†	12.52 ± 4.19	12.1 ± 3.85 (p = 0.044*)	12.15 ± 3.81 (p = 0.272)	11.62 ± 3.93 (p = 0.026*)

†EQ5D-5L ranges from 1 (more independent) to 5 (more dependent), lower score indicates higher independence.

3. Results

At the end of the study, 48 subjects completed the study, and their data was analyzed. Table 2 summarized the mean scores (Mean ± SD) of the FIM, QoL, and EQ5D-5L from different time. The p-values were calculated against baseline scores (day 1) of each measure with the significance level set to 0.05.

The baseline score of FIM was 101.48±19.08, and it increased to 107.02±16.32 at day 42. The FIM score between baseline and the last evaluation showed a significant difference with paired t-testing ($p < 0.01$).

The mean QoL score describes personal momentary well-being and is measured by the EQ5D-5L questionnaire. The baseline score of QoL was 69.65±17.63 and it increased to 76.38±13.61 at day 42. The QoL between baseline and the last evaluation showed a significant difference with paired t-testing ($p< 0.01$). The mean 5L score from the EQ5D-5L was 12.52±4.19 and it reduced to 11.62±3.93 at day 42. The 5L score between baseline and the last evaluation showed a significant difference with paired t-testing ($p< 0.05$).

4. Discussion

This study aimed to prove the concept that stroke patients can benefit from digital mirror therapy in virtual reality by examining the functional performance with reliable evaluation instruments. The main finding in this study showed that patients' functional performance improved after 42 days of training using Rehago. Because 82.5% of the subjects participating in this study were already in chronic stage (more than 12 months since the onset of stroke), a huge improvement in functional performance or motor skills may not be expected as suggested in the previous study [6]. Besides, the evaluation instruments used in this study may not be sensitive enough to detect subjects' improvement of functional performance due to the design. Nevertheless, an average

improvement of 5.54 points of the FIM and an average improvement of 7.13 points in the QoL during the study were observed, indicating that the subjects acquired improvement of independence in executing ADLs and a higher quality of life after training with Rehago. One thing to note is that the EQ5D-5L score shows a significant difference between the baseline and day 14 ($p=0.044$), baseline and day 42 ($p=0.026$), respectively. However, there is no significant difference between the baseline and day 24 ($p=0.272$), and the reason may result from the design of the instrument, the status of the subjects during the evaluation, and the relatively small sample size of this study. In addition, several publications advocate further research of digital technology in rehabilitation. The use of VR technology can not only increase the expected adherence but relieve the health care system and the therapists and thus counteract subsequent costs through additional home training [3, 5, 7].

5. Conclusion

The results of this pilot study show an observable improvement of patients' functional performance, which supports our hypothesis that the functional performance of stroke patients can be improved by conducting gamified mirror therapy in VR. Our results also align with other relevant publications [8-10]. However, to obtain more evidence of the training effects with Rehago, a comprehensive study with a structured interventional plan and subject grouping may be beneficial. Moreover, it is also meaningful to investigate whether the training effect of performing ADL tasks using VR can be carried over to real world scenarios. Therefore, Rehago as one of the pioneers of VR-based rehabilitation software will continue improving and constantly update new training content to offer better user experience and training effect to users.

References

[1] S. Knecht, S. Hesse, and P. Oster, "Rehabilitation After Stroke," *Dtsch. Aerzteblatt Online*, Sep. 2011.
[2] H. Thieme *et al.*, "Mirror therapy for improving motor function after stroke," *Cochrane Database Syst. Rev.*, vol. 7, p. CD008449, 2018.
[3] R. Ballantyne and P. M. Rea, "A Game Changer: 'The Use of Digital Technologies in the Management of Upper Limb Rehabilitation,'" *Adv. Exp. Med. Biol.*, vol. 1205, pp. 117–147, 2019.
[4] L. M. Weber, D. M. Nilsen, G. Gillen, J. Yoon, and J. Stein, "Immersive Virtual Reality Mirror Therapy for Upper Limb Recovery After Stroke: A Pilot Study," *Am. J. Phys. Med. Rehabil.*, vol. 98, no. 9, pp. 783–788, Sep. 2019.
[5] M. A. Ahmad, D. K. A. Singh, N. A. Mohd Nordin, K. Hooi Nee, and N. Ibrahim, "Virtual Reality Games as an Adjunct in Improving Upper Limb Function and General Health among Stroke Survivors," *Int. J. Environ. Res. Public. Health*, vol. 16, no. 24, 2019.
[6] G. Kwakkel, B. J. Kollen, J. van der Grond, and A. J. H. Prevo, "Probability of Regaining Dexterity in the Flaccid Upper Limb: Impact of Severity of Paresis and Time Since Onset in Acute Stroke," *Stroke*, vol. 34, no. 9, pp. 2181–2186, Sep. 2003.
[7] G. Burdea, "Virtual rehabilitation-benefits and challenges. Methods of information in medicine." 2003.
[8] Y.-W. Hsieh, Y.-H. Lin, J.-D. Zhu, C.-Y. Wu, Y.-P. Lin, and C.-C. Chen, "Treatment Effects of Upper Limb Action Observation Therapy and Mirror Therapy on Rehabilitation Outcomes after Subacute Stroke: A Pilot Study," *Behav. Neurol.*, vol. 2020, p. 6250524, 2020.
[9] J.-Y. Park, M. Chang, K.-M. Kim, and H.-J. Kim, "The effect of mirror therapy on upper-extremity function and activities of daily living in stroke patients," *J. Phys. Ther. Sci.*, vol. 27, no. 6, pp. 1681–1683, 2015.
[10] M. Invernizzi, S. Negrini, S. Carda, L. Lanzotti, C. Cisari, and A. Baricich, "The value of adding mirror therapy for upper limb motor recovery of subacute stroke patients: a randomized controlled trial," *Eur. J. Phys. Rehabil. Med.*, vol. 49, no. 3, pp. 311–317, Jun. 2013.
[11] S. B. O'Sullivan, T. J. Schmitz, and G. D. Fulk, *Physical rehabilitation*, 6th ed. Philadelphia: F.A. Davis Co, 2014.
[12] N. A. Aziz, J. Leonardi-Bee, M. Phillips, J. R. F. Gladman, L. Legg, and M. F. Walker, "Therapy-based rehabilitation services for patients living at home more than one year after stroke," *Cochrane Database Syst. Rev.*, no. 2, p. CD005952, Apr. 2008.

Healthcare of the Future 2022
T. Bürkle et al. (Eds.)

doi:10.3233/SHTI220331

Towards a FHIR-Based Data Model for Coronary Angiography Observations

Florian HOLWEG[a,1], Stephan ACHENBACH[b], Noemi DEPPENWIESE[c], Luise GAEDE[b] and Hans-Ulrich PROKOSCH[a]
[a]*Chair of Medical Informatics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany*
[b]*Cardiology and Angiology Clinic, Universitätsklinikum Erlangen, Germany*
[c]*Center of Medical Information and Communication Technology, Universitätsklinikum Erlangen, Germany*

Abstract. Coronary heart disease is among the most frequent causes of death globally. Thus, our research project aims to develop prognostic models, to predict the risk of spontaneous myocardial infarctions based on a combination of clinical parameters and image data sets (invasive coronary angiograms). To train such models we use data from more than 30,000 coronary angiograms acquired at the cardiology department of Erlangen University Hospital. To linking such proprietary data with additional clinical parameters and to harmonize it for future cross-hospital federated machine learning approaches we defined a mapping for coronary angiography based on the symptom/ clinical phenotype HL7® FHIR® module of the German medical informatics initiative. In this paper we describe the final design of the coronary angiography information model and our mapping approach to ICD-10 and SNOMED CT. From the database we use a subset of 15 required values patient characteristics to create the HL7® FHIR® resource.

Keywords. HL7 FHIR, data model, coronary angiography, SNOMED CT

1. Introduction

Cardiovascular diseases, prominently including myocardial infarction, are the most frequent cause of death globally. Myocardial infarction often occurs in patients without prior symptoms since the most frequent mechanism is the sudden rupture of non-stenotic atherosclerotic coronary artery lesions [1-3]. Algorithms to identify at-risk individuals would therefore be of utmost clinical relevance.

A frequently performed test in patients with suspected coronary artery disease is invasive coronary angiography. In Germany alone, more than 900,000 coronary angiograms are acquired annually. In approximately 60% of these, coronary artery stenoses are identified and revascularization usually ensues. However, no guidance exists for further treatment in patients without stenosis. If, among these, individuals with increased risk for future myocardial infarction could be reliably identified, this would offer substantial opportunity for targeted, intensive risk factor modification with subsequent reduction of the risk for myocardial infarction and death. However, no such

[1] Corresponding Author, Florian Holweg, Friedrich-Alexander University Erlangen-Nürnberg, Chair of Medical Informatics, Am Wetterkreuz 15, D-91058 Erlangen, Germany, E-mail: florian.holweg@fau.de

algorithms exists today. Automated classification of angiograms is challenging due to their complex nature and large data volumes.

It is the long-term goal of our project ER-MIKKI (Determination of myocardial infarction risk from coronary angiography and clinical data with AI) to develop and validate risk prediction tools for future myocardial infarction based on coronary angiography image data and clinical parameters, making use of artificial intelligence-based models. The project is based on a cohort of 30,000 patients whose structured coronary angiography reports are stored in an MS Access® database (MSAd) using a proprietary format, without link to any further clinical data. In order to develop and validate the risk prediction algorithms, integration of the structured angiographic reports with clinical parameters such as laboratory data, procedures, medication and procedures performed, as well as data harmonization to a common data model was required.

This first subproject of ER-MIKKI therefore aims at developing a HL7® FHIR®-(HL7) based data model based on the core dataset definitions of the German medical informatics initiative (MII) [4-6], thus allowing to smoothly integrate the coronary angiography reports with clinical data within the Erlangen University Hospital data integration center (DIC) for subsequent processing.

2. Methods

Based on a comprehensive analysis of MSAd, the following parameters were identified as the basis for record linkage with further clinical data: First and last name, the date of birth and administrative gender. The date and start/end times of the clinical procedure serve as time stamps. Clinical parameters comprise the classified stenosis severity (% value) for the four major coronary arteries: Left Main Stem (LM), Left Anterior Descending (LAD), Ramus Circumflex (RCX) and Right Coronary Artery (RCA). In addition, a short free text describes the findings for each artery. The MSAd proprietory database contains one table with more than hundred attributes for saving the patient data from coronary angiography and 30 reference tables comprising value lists for categorical data items. The database contains more than 100 data points for each patient for a coronary angiography. Unfortunately, the structure of the MSAd is not based on a standard Entity Relationship Modell. One line in the output table corresponds to one observation.

For providing such data in a harmonized data model, the core data set (MII cds) definitions from the German medical informatics initiative were evaluated. The MII cds comprises six basic modules (person, encounter, diagnosis, procedure, laboratory results, medication) and numerous extension modules. After an analysis of all such cds modules we decided to apply the module person to map the identification and demographic data of a patient and the module symptom/clinical phenotype for representing the clinical state or clinical observations/findings of a patient. Even though the implementation guide of the final HL7 profile of this MII module is not yet finalized and balloted, an ART-DÉCOR [7] based data model and UML class diagram were already available, mainly being built upon the HL7 resource Observation [8].

3. Results

By comparing the proprietary data structure of the original MSAd with the HL7 modelling options, two slightly different models would have been possible. Figure 1 shows the ultimately consented HL7 data model (in a reduced form). For transforming the demographic information of the patient, we use the core data set module PERSON. The patient identification field is used for transporting the unique key of the MSAd. The fields birth name, name prefix, name affix and family name are not available in the original system and are omitted in our model in figure 1. Linked with PERSON we apply the core data set module CLINICAL SYMPTOM/CLINICAL PHENOTYPE and reduce this to OBSERVATION only (illustrated in figure 1).

The observation is referencing to the HL7 resource Observation [8]. This resource has as a main component the code of any medical classification system e.g., ICD or SNOMED-CT (SCT). With these classification systems a comprehensive information model for coronary angiography findings can be established [9,10]. In our project, we create a main component with an ICD-10 Code to classify the type of cardiologic disease (e.g., I25.11 single-vessel-disease, I25.12 two-vessel-disease). For mapping the coronary angiography findings to the OBSERVATION resource, the data elements category, state, and methods were set to the fixed values PROCEDURE, FINAL and 3367005 as the SCT code for coronary angiography. To describe the detailed finding related to the respective coronary arteries a hierarchical observation tree is built by linking associated 14 SCT codes with each other, using the "Reference to" connection. In total 35 SCT codes and four ICD-10 Codes can be used for descripting the disease.

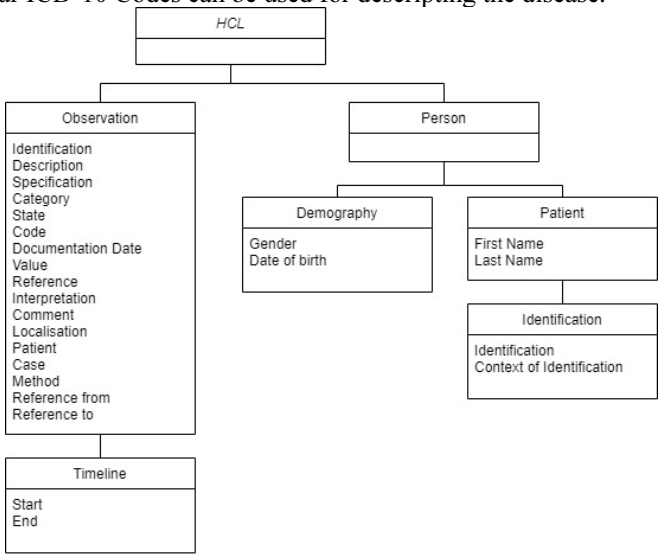


Figure 1. Art-Decor FHIR® Data Model for representing coronary angiography observations from the cardiology heart catheter laboratory (HCL)

Two hierarchically referenced OBSERVATION resources with SCT codes 3227004 and 3711007 for example describe that the left coronary artery main stem is smooth walled, whereas 3227004 referenced with 123641001 describe a left coronary artery occlusion. Documentation date, as well as start/end time are filled with the procedure date and times available in the MSAd. Data elements representing medical information not available within the original MSAd are left blank in the OBSERVATION.

4. Discussion

The ER-MIKKI research team aims at the determination of myocardial infarction risk from coronary angiography and clinical data with AI technologies. Even though in the current project phase we plan to only use data from one cardiology department to train our AI model, future enhancements could build on datasets from multiple hospitals, as it is supported by the German medical informatics initiative. For this goal, data must be provided in an interoperable manner, implying the standardization of data, so it can be easily shared among various institutions, avoiding the creation of silos. Similar to González-Castro in the CASIDE project [11] we responded to this need with the design of the ER-MIKKI-MODEL, a standardized data model based on HL7 which allows structured representation of coronary angiography observations. As already experienced by other researchers [12] we found multiple ways to first create an ART DÉCOR model and then map the clinical findings to SCT. The flexibility that HL7 offers to adapt to specific use cases and the power of SCT to sometimes express clinical meaning in different manners also poses challenges, since there often are multiple ways to map the same concept using available resources and codes. As we also wanted to link our data with clinical data from the university hospital EHR system, which has already been provided according to the MII core data set basic modules within our DIC, we had decided to follow the MII HL7 profile implementation guides and also use the ART DÉCOR tool to describe our model. For defining the mappings of the ICD-10 and SCT Codes to the dataset in the MSAd we can apply the standard MIRACUM mapping service [13] developed in our DIC. For quality assurance a workflow was defined.

References

[1] Little WC. Angiographic assessment of the culprit coronary artery lesion before acute myocardial infarction. Am J Cardiol. 1990;66(16):44G-47G. doi:10.1016/0002-9149(90)90395-h
[2] Nittel D, Seltrecht A. Krankheit: Lernen im Ausnahmezustand?: Springer; 2013.
[3] Nittel D. Krankheit: Brustkrebs und Herzinfarkt aus interdisziplinärer Perspektive. Mit Online-Material für Fachleute. Dordrecht: Springer; 2012. 574 p. ger.
[4] Semler SC, Wissing F, Heyder R. German Medical Informatics Initiative. Methods Inf Med. 2018;57(S 01):e50-e56.
[5] Medizininformatik Initiative. Der Kerndatensatz der Medizininformatik Initiative [Internet] [cited 2021 Dec 21]. Available from: https://www.medizininformatik-initiative.de/de/der-kerndatensatz-der-medizininformatik-initiative
[6] Medizininformatik Initiative. Basismodule des Kerndatensatz der MII [Internet] [cited 2021 Dec 21]. Available from: https://www.medizininformatik-initiative.de/de/basismodule-des-kerndatensatzes-der-mii
[7] Art-Decor. Art-Decor [Internet]. 2021 [updated 2021 Oct 7; cited 2022 Jan 3]. Available from: https://art-decor.org/mediawiki/index.php?title=Main_Page
[8] HL7. Observation [Internet] [cited 2021 Dec 21]. Available from: https://www.hl7.org/fhir/observation.html
[9] Benson T, Grieve G. Principles of Health Interoperability: SNOMED CT, HL7 and FHIR. 3rd ed. Cham, s.l.: Springer International Publishing; 2016. 451 p. (Health Information Technology Standards). eng.
[10] Uysal S. ICD-10-CM Diagnosis Coding for Neuropsychological Assessment. Arch Clin Neuropsychol. 2019;34(5):721–30. doi:10.1093/arclin/acy084 Cited in: PubMed; PMID 30357265.
[11] González-Castro L, Cal-González VM, Del Fiol G, López-Nores M. CASIDE: A data model for interoperable cancer survivorship information based on FHIR. J Biomed Inform. 2021;124103953.
[12] Oeppert L, Wehner K, Schrader T. FHIR data model for data exchange between cancer registries and IQTIG. 2021.
[13] Mate S, Seuchter SA, Ehrenberg K, Deppenwiese N, Zierk J, Prokosch HU, Kraska D, Kapsner LA. A Multi-User Terminology Mapping Toolbox. Stud Health Technol Inform. 2021 May 24;278:217-223.

This page intentionally left blank

Research in Medicine and Medical Informatics (Young Researcher)

This page intentionally left blank

Healthcare of the Future 2022
T. Bürkle et al. (Eds.)

doi:10.3233/SHTI220333

Classifying Numbers from EEG Data - Which Neural Network Architecture Performs Best?

Sugeelan SELVASINGHAM[a], Kerstin DENECKE[a,1]

[a] *Bern University of Applied Sciences, Bern, Switzerland*

Abstract. This paper presents a comparison of deep learning models for classifying P300 events, i.e., event-related potentials of the brain triggered during the human decision-making process. The evaluated models include CNN, (Bi | Deep | CNN-) LSTM, ConvLSTM, LSTM + Attention. The experiments were based on a large publicly available EEG dataset of school-age children conducting the "Guess the number"-experiment. Several hyperparameter choices were experimentally investigated resulting in 30 different models included in the comparison. Ten models with good performance on the validation data set were also automatically optimized with Grid Search. Monte Carlo Cross Validation was used to test all models on test data with 30 iterations. The best performing model was the Deep LSTM with an accuracy of 77.1% followed by the baseline (CNN) 76.1%. The significance test using a 5x2 cross validation paired t-test demonstrated that no model was significantly better than the baseline. We recommend experimenting with other architectures such as Inception, ResNet and Graph Convolutional Network.

Keywords. Convolutional neural networks, EEG, P300, Classification

1. Introduction

Brain-computer interfaces (BCI) enable communication without muscle activity based on brain signals measured with electroencephalography (EEG). The P300 is an event-related potential that is triggered during the decision-making process of a human. P300-based BCIs have gained attention in recent years and are considered one of the most important BCI categories [1]. Compared to other BCI paradigms, P300 BCIs are relatively fast, effective for most users, straightforward and require virtually no training of subjects. The challenge is to classify the P300 events with sufficient accuracy to enable a good communication. Deep learning and neural networks have been applied to this classification task. Vareka created a Convolutional Neural Network (CNN) model and trained it with EEG data [2]. CNN is an artificial intelligence method and is often used in image classification. The author was able to achieve an average classification accuracy of 62.18% using CNN. He also tested Recurrent Neural Networks (RNN) [3]. The RNN architecture is often used for classifying time series, i.e. also for EEG. Since the accuracy for guessing the number using P300 event streams with only three channels is still

[1] Corresponding Author, Kerstin Denecke, Bern University of Applied Sciences, Institute for Medical Informatics, Quellgasse 21, 2502 Biel, Switzerland; E-mail: kerstin.denecke@bfh.ch.

insufficient for reliable BCI, the objective of this work is to develop and test different deep learning architectures to achieve a better accuracy. A practical application of the algorithm could support individuals who are unable to communicate verbally in their interaction with computers. The main contribution of this work is a comprehensive analysis of multiple neural network architectures for the classification of numbers from EEG data (P300 event streams).

2. Material and methods

P300 dataset. We used the P300-dataset described by Mouvcek et al. [4] collected using the "guess the number" experiment. Participants are asked to pick a number between 1 and 9. During the following EEG measurement phase, the individual is stimulated with these numbers. He or she is silently counting the number of occurrences of the selected number. The target number is supposed to trigger the P300 response. After the experiment, this number is revealed and compared with the guess of the experimenters observing averages EEG waveforms [4]. The dataset used in this paper was collected in experiments with 250 participating school-age children which were carried out in elementary and secondary schools in the Czech Republic. Electroencephalographic data from three EEG channels (Fz, Cz, Pz) and stimuli markers were stored. Since we want to compare our results to the work published by Vareka [2], we used the dataset as prepared by the author without additional preprocessing. Before classification, the data were randomly split into training (75 %) and testing (25 %) sets. Training data was additionally split into 75% training and 25% validation set.
Experimental setup. We used two evaluation strategies to test the performance of the models: K-fold cross validation and Monte Carlo Cross Validation (MCCV [5]). In this validation method, the entire data set or the number of epochs is randomly divided into test and training data. Epochs are randomly divided into test and training data. With each iteration, the division of test and training data varies. Due to the random division, the same data components can serve as test data several times in some iterations. K-fold cross validation splits training data into K equal parts. To assess the quality of classification, we calculated the following metrics: AUC, Precision, Recall, Accuracy and Duration for training and classification. Results are tested for statistical significance using the 5x2 Paired t-Test. The following hyperparameter were fix throughout all experiments: Optimizer = Adam, Loss = categorical_crossentropy, epoch = 30, batchsize = 16. While adapting iteratively the model architecture or adjusting the hyperparameters of the model, the performance is tested with the validation data. Using the testing set, we computed results in each cross-validation iteration and averaged scores at the end of the processing. The experiments were run on a DGX station with Tesla V100-DGXS-32GB. Different deep learning models have been tested, among them CNN, (Bi | Stacked)LSTM, CNN-LSTM, ConvLSTM und LSTM + Attention. The models were derived from literature when they considered EEG data for classification or a similar classification problem: Models as tested by Vareka [2,3] have been rebuilt as baseline without any modification (labeled with “gtn_”). Other architectures were taken from Zhang et al. [7] who used the models to classify emotions from EEG signals (labeled with “emotion_”). Anguita et al. [8] applied different architectures to a smartphone dataset for activity recognition. Architectures based on this work are labeled with “mastery_”. Finally, the CNN-BiLSTM architecture suggested by Mansar for classifying sleep stages based on EEG data have been used [9]. All models were optimized (label

"optimized"): We identified good performing hyperparameters using grid search, resulting in optimized architectures.

3. Results

Best performing model was the Deep LSTM *emotion_deep_lstm* with an accuracy of 63.7% in the single trial (see Table 1) and 77.1% in the averaged trial (see Table 2). However, it was not significantly better than the baseline CNN (model *gtn_cnn* shown in the second line in Table 1). We define the model "emotion_deep_lstm" as a Deep LSTM Model with an LSTM layer being the first layer, followed by a Dense layer with 50 units. After that we define another LSTM layer with 6 units, a Dropout layer (dropout rate = 0.9) and an output layer with 2 units (softmax activation). The averaged trial used average values from six epochs as was done by Vareka [3]. Using the average trial, almost all models performed better, but also the standard deviation increased.

modelName	train_acc	dev_acc	test_acc	auc	precision	recall	trainable_param	architecture
emotion_deep_lstm	63.05%	62.71%	63.71%	68.28%	63.71%	63.71%	36428	Deep LSTM
gtn_cnn	**65.13%**	**61.96%**	**63.04%**	**67.43%**	**63.04%**	**63.04%**	89974	**CNN**
optimized_emotion_recog_deep_lstm_model	63.37%	62.31%	63.02%	67.47%	63.02%	63.02%	38450	Deep LSTM
mastery_lstm_attention_stackoverflow	59.79%	59.08%	60.56%	63.94%	60.56%	60.56%	53202	LSTM - Attention
optimized_gtn_cnn	69.69%	60.06%	60.49%	64.00%	60.49%	60.49%	268772	CNN
optimized_emotion_regoc_cnn_lstm_model	60.53%	58.83%	60.05%	63.64%	60.05%	60.05%	590	CNN - LSTM
emotion_cnn_lstm	60.60%	58.79%	59.97%	63.89%	59.97%	59.97%	590	CNN - LSTM
emotion_bidirectional_lstm_attention_stackoverflow	60.15%	58.89%	59.90%	63.34%	59.90%	59.90%	43330	BiLSTM - Attention
emotion_lstm_attention_stackoverflow	58.38%	57.64%	59.46%	62.44%	59.46%	59.46%	1460	LSTM - Attention
emotion_more_layer_cnn_v1	66.66%	58.83%	59.45%	63.18%	59.45%	59.45%	35717	CNN
optimized_gtn_lstm	55.59%	58.52%	59.30%	63.15%	59.30%	59.30%	122132	LSTM
mastery_lstm_cnn	58.71%	58.76%	59.10%	62.23%	59.10%	59.10%	2602622	CNN - LSTM
optimized_sleep_timedistributed_cnn_lstm_model	60.95%	58.16%	58.82%	62.17%	58.82%	58.82%	373202	CNN - BiLSTM
sleep_cnn_lstm	60.22%	57.82%	58.64%	61.94%	58.64%	58.64%	373202	CNN - BiLSTM
gtn_lstm	**56.24%**	**57.74%**	**58.45%**	**62.13%**	**58.45%**	**58.45%**	**120592**	**LSTM**

Table 1. Single trial: Results of the 15 best performing models. Model *gtn_cnn* and gtn_lstm are the baseline models from Vareka [2,3]. Accuracy values for training, validation and test set are shown, as well as values for AUC, precision, and recall.

4. Discussion and future work

Our results show that there are models that can perform slightly better on the considered classification task. Taking Vareka's LSTM (gtn_lstm) as a baseline [3] our best performing model differs from this baseline model in two ways: our model (emotion_deep_lstm) has as first layer an LSTM layer with 64 time steps followed by a dense layer with 50 nodes. In future work, it could be tested to add additional layers since it seems to have an impact on the accuracy. The optimization of the models was not as successful as expected. A reason might be that during Monte Carlo Cross Validation only one iteration was performed. Vareka optimized with 30 iterations. The averaged trial

experiments achieved good performance. Through averaging noise in the data is reduced and only relevant data is considered for classification.

Our work has several limitations: complex architectures were not implemented. Future work is needed to test the performance of such models. We used an existing EEG dataset that only recorded 3 channels. Other researchers use EEG data with 8 recorded channels which might result in more expressive data. Furthermore, we could not test all possible variations of hyperparameter. This also remains open for future work. New deep learning architectures such as Inception, Graph convolutional network or ResNet could be tested on the data.

modelName	train_acc	dev_acc	test_acc	auc	precision	recall	trainable_param	architecture
emotion_deep_lstm	63.10%	62.71%	77.11%	84.97%	77.11%	77.11%	36428	Deep LSTM
gtn_cnn	**65.13%**	**61.96%**	**76.15%**	**83.70%**	**76.15%**	**76.15%**	89974	**CNN**
optimized_emotion_recog_deep_lstm_model	63.41%	62.11%	74.51%	81.75%	74.51%	74.51%	38450	Deep LSTM
optimized_gtn_cnn	69.53%	60.16%	72.64%	79.53%	72.64%	72.64%	268772	CNN
mastery_lstm_attention_stackoverflow	59.79%	59.08%	69.75%	75.78%	69.75%	69.75%	53202	LSTM - Attention
mastery_lstm_cnn	58.23%	58.10%	69.56%	75.16%	69.56%	69.56%	2602622	CNN - LSTM
optimized_gtn_lstm	55.75%	58.68%	69.13%	76.16%	69.13%	69.13%	122132	LSTM
mastery_ConvLSTM	67.54%	57.68%	68.73%	74.85%	68.73%	68.73%	1319214	ConvLSTM
emotion_bidirectional_lstm_attention_stackoverflow	60.31%	58.96%	68.71%	75.12%	68.71%	68.71%	43330	BiLSTM - Attention
optimized_mastery_convLSTM_model	67.50%	57.95%	68.68%	74.73%	68.68%	68.68%	647470	ConvLSTM
optimized_emotion_regoc_cnn_lstm_model	60.44%	58.69%	67.96%	76.03%	67.96%	67.96%	590	CNN - LSTM
emotion_more_layer_cnn_v1	66.59%	58.97%	67.79%	74.52%	67.79%	67.79%	35717	CNN
emotion_cnn_lstm	60.46%	58.86%	67.18%	74.74%	67.18%	67.17%	590	CNN - LSTM
gtn_lstm	**56.25%**	**58.04%**	**67.13%**	**75.07%**	**67.13%**	**67.13%**	**120592**	**LSTM**
optimized_sleep_timedistributed_cnn_lstm_model	60.95%	58.16%	66.90%	72.77%	66.90%	66.90%	373202	CNN - BiLSTM

Table 2. Averaged trial: Results of the 15 best performing models. Model *gtn_cnn* is the baseline. Accuracy values for training, validation and test set are shown, as well as values for AUC, precision, and recall.

References

[1] R. Fazel-Rezai, B. Z. Allison, C. Gugeret al. P300 brain computer interface: current challenges and emerging trends, Frontiers in neuroengineering. 2012;5:14.

[2] L. Vareka, Evaluation of convolutional neural networks using a large multi-subject p300 dataset, Biomedical Signal Processing and Control. 2020;58:101837.

[3] L. Vareka, Comparison of convolutional and recurrent neural networks for the p300 detection, In: Proceedings of the 14th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC) 2021, Vol. 4 - BIOSIGNALS, pp. 186–191.

[4] R. Moucek, L. Vareka, T. Prokop, J. Stebetak, P. Brha, Event-related potential data from a guess the number brain-computer interface experiment on school children, Scientific data. 2017;4(1):1–11.

[5] Q.-S. Xu, Y.-Z. Liang, Monte carlo cross validation, Chemometrics and Intelligent Laboratory Systems . 2001; 56 (1):1–11.

[6] H. Shan, Y. Liu, T. P. Stefanov, A simple convolutional neural network for accurate p300 detection and character spelling in brain computer interface., in: IJCAI, 2018, pp. 1604–1610.

[7] Y. Zhang, J. Chen, J.H. Tan et al. An investigation of deep learning models for EEG-based emotion recognition. Frontiers in Neuroscience. 2020;14.

[8] D. Anguita, A. Ghio, L. Oneto et al. A public domain dataset for human activity recognition using smartphones. In: Esann 2013 Apr 24; 3, p. 3.

[9] Y. Mansar. Sleep Stage Classification from Single Channel EEG using Convolutional Neural Networks [Internet]. Medium. 2020 [last access 13.12.2021]. Accessible at: https://bit.ly/32CgZmA

Healthcare of the Future 2022
T. Bürkle et al. (Eds.)

doi:10.3233/SHTI220334

A Consent Tool for Secondary Use of Biomedical Data

Simon ADAMS[a,1,*], Gausegan UTHAYATHAS[a,*], Murat SARIYAR[a]
[a] *Bern University of Appl. Sciences, Dept. Medical Informatics, Switzerland*

Abstract. To pursue scientific goals with patient data usually requires informed consent from the data subjects. Such a consent constitutes a contract between the research institute and the patient. Several issues must be included in the consent to be valid, for example, how the data is processed and stored as well as specifics of the research questions for which the data is going to be used. Here, we describe the development and the implementation of a user-friendly IT solution that supports the process-oriented obtainment of consents. Current solutions often focus only on the benefits for the researcher. Our solution intends to add value to all participants and to reduce paperwork to a minimum. The consent Tool was evaluated by a usability test using the UEQ Method (User Experience Questionnaire) and received positive feedback – both efficiency and originality were rated above the average UEQ-Benchmark. Nevertheless, the lack of compatibility with the technical infrastructure of the hospital was a significant shortcoming. Hence, although there is a general interest in digitized solutions in the healthcare sector, there are still many hurdles to implement them and roll them out.

Keywords. General consent; informed consent; data protection; clinical trials.

1. Introduction

Research projects in the biomedical domain often require routine health data, i.e., data that is collected in the context of medical treatments. To pursue certain scientific goals with such data usually requires informed consent from the data subjects. Such a consent constitutes a contract between the research institute and the patient. Several issues must be included in the consent to be valid, for example, how the data is processed and stored as well as specifics of the research questions for which the data is going to be used. One consent type that is often used is the general consent, as it allows to formulate very broad research questions [1].

Even if the informed consent contract is not signed, there is still the possibility that using the data for research might be legal. A mechanism that leads to such an unintuitive consequence is the right of objection. In some constellations, it is present as a contract clause that requires explicit objection (see Section Results). In other words, a data subject must either explicitly agree or explicitly object (by a legally valid signature) to the usage of her data [2]. Such a knowledge is often not present in data subjects, as they often have only a vague image of what informed consent implies, and decisions or actions based on such images are sometimes referred to as “empty performative” [3].

[1] Corresponding Author, Simon Kofi Adams, Bern University of Applied Sciences, Quellgasse 21, CH2502 Biel/Bienne, Switzerland; E-mail: simonkadams@protonmail.com.
* Both authors contributed equally to this work and share first authorship.

In Switzerland, the political agenda for future healthcare is called "Strategy eHealth 2.0", which has the vision to provide a better, more secure, and more efficient healthcare system. One ingredient in that strategy is to enhance informational self-determination, which avoids "empty performative" decisions. It is assumed that informed and conscious decisions regarding one's own health data contributes to an increased quality of treatment: patient safety is enhanced, and efficiency is achieved through health literacy [4]. To facilitate such a self-determination, paper-based consent and ad-hoc elicitation of the consent should be replaced by a digitized consent workflow that supports the information needs of the data subjects with a user-friendly and digestible annotation of the digitized consent template [5]. Current solutions often focus only on the benefits for the researcher. Our solution intends to add value to all participants and to reduce paperwork to a minimum.

2. Methods

For obtaining the necessary information, we investigated how the consent is practically completed and collected within medical research projects. For that purpose, we collaborated with the university hospital Bern (Insel), which also took the role of our main stakeholder in this project. We divided our target group into two parts: on the one hand, researchers as processors and administrators of the consent; and on the other hand, the patients, as signatories and consumers of the consent. We performed a literature search for gaining an overview on existing types of informed consent together with their pros and cons. We divided this into two subtasks, a general search in search engines (e.g., Google) and a targeted search in scientific databases (e.g., Google Scholar, PubMed etc.), covering literature from the last 20 years. The following search terms were used: "general consent", "clinical trial consent", "efficiency". In addition to that, we conducted two semi-structured online- interviews with our stakeholder, each lasted about 1.5 hours.

The gathered requirements served as the basis for an AdobeXD-mockup, which allowed us to discuss potential adaptations for implementing the prototype. We gathered feedback on the mockup based on a live-web-demo, taking notes on the recommendations of our stakeholder. It was then decided that the solution should be a user-friendly web-based application (according to the Nielsen-Usability-Heuristics: e.g., "user control and freedom", "flexibility and efficiency of use" amongst others) to ensure platform independency. The implementation was done in five scrum sprints and was based on the following technology stack: vue.js for the frontend and node.js for the backend with mongoDB as a database. For receiving feedback with respect to our solution, we conducted a usability-test with the UEQ Method.

3. Results

Concerning the types of consents, one can first distinguish between general and informed consent. General consent is mostly used in research projects where there is no active involvement of the data subjects, e.g., the research can be carried out solely by relying on previously collected data. Informed consent is used when an active participation of the data subject (e.g., in a clinical trial) is required. Our stakeholder wanted us to focus on the general consent for this project. Regarding the ingredients of

such a consent, we relied on the draft template of the Swiss Academy of Medical Sciences (SAMW ASSM) [6].

Besides the content of the general consent, three further issues had to be solved: the consequences of (not) consenting, data security requirements, and digital signatures. Table 1 summarizes the consequences of consenting, refusing or ignoring the consent contract. The most striking case is ignoring the consent for anonymized non-genetic data and samples, which allows their usage. With respect to data security, we referred again to the SAMW ASSM recommendations, e.g., related to revocations or the history of changes in the consent status. Finally, regarding the digital signature we relied on the Swiss Code of Obligations, which stated that only the "qualified electronical signature" is considered equivalent to a handwritten signature. To ensure this, a certificate-based digital ID issued by accredited trust services is required. In addition to the integrity of the document and the assignment to a person, a time stamp is added here as confirmation of the existence of the document at a specific point in time. To make our prototype more flexible, we relied only on a simple electronic signature and waived the legally valid one.

General Consent signed?	Encrypted genetic data and samples	Encrypted non-genetic data and samples	Anonymized, genetic data and samples	Anonymized, non-genetic data and samples	Anonymously collected data and samples
"Yes"	✓	✓	✓	✓	✓
"No"	✗	✗	✗	✗	✓
"No Answer"	✗	✗	✗	✓	✓

Table 1 Consent matrix for the further use of data/samples

For the implementation of our prototype, we created a BPMN process diagram (not shown here) to consider the conventional consent process involving the data subject, the researcher, and the ethics committee. Once research questions and design are described, the research institute must a request to the local ethics committee. Only after passing critical aspects like the adequacy of the purposes, data security, data protection, study design and added value for medicine in general, is a project allowed to start. After the required study population has been defined, institutes can start to contact suitable candidates. This is mostly done when a patient visits a medical facility in case of a routine examination or due to other medical reasons. They are made aware of the possibility of data usage for research purposes – but since patients came to the facility for another reason, they often don't pay particular attention to it. After their visit, they receive a letter including the form to sign.

Our prototype is designed to annotate all the relevant parts in the digitized consent template to convey their contents and reasons. Thereby, all relevant information is provided in a digestible manner. Further, our prototype allows interactions between patients and the researcher in case something remains unclear or questions beyond the consent are posed. Researchers could draft the documents which are required by the ethical committee digitally and redirect it to them on the same platform. Based on the draft, they could send the document to suitable patients who on the other hand can view it at home at their leisure. Complex terms are explained in a sort of dictionary which can be fed by researchers to avoid "empty performative" actions.

Results from the usability-test show: in comparison to the UEQs Benchmarking dataset (>20'000 persons, >450 studies), our consent tool received above average ratings for attractiveness and transparency and even better ratings for efficiency and originality.

Only stimulation was rated below average. In addition to that, 3 out of 5 researchers rated the digitized version of consenting as highly safe and informative.

4. Discussion

We have found few existing solutions for digitally obtaining consents. There are, for instance, an EU driven research project which tried to formulate a proof-of-concept for consents and the solution E-ConsentPro from the German company Thieme, which we didn't test in depth however [7]. Hence, both economically and politically the matter is still in its infancy and should receive more attention. Interestingly, this is also due to the fact that the digital equivalent of a handwritten signature is not fully clarified. Even though the Swiss law indeed describes one possible variation of that signature, real-life examples are scarce, as there is a lack of microeconomic network effects.

Originally, our tool was implemented to increase efficiency of the consent process, but the lack of compatibility with the technical infrastructure of the hospital was a significant shortcoming. As it was unclear whether the initially increased efficiency would be eaten up by the additional administrative burden and manual transfer of data, the tool has not been used in practice yet. Hence, although there is a general interest in digitized solutions in the healthcare sector, there are still many hurdles to implement them and roll them out. For example, most clinical trials are multicentric studies which are conducted at several sites – each with their own clinical information systems – which lacks a universally applicable data format to avoid data disruptions. HL7 FHIR knows a resource, which could be further developed to address this interoperability issue [8].

References

1. SR 810.30 - Bundesgesetz vom 30. September 2011 über die Forschung am Menschen (Humanforschungsgesetz, HFG) [Internet]. [cited 2021 Dec 13]. Available from: https://www.fedlex.admin.ch/eli/cc/2013/617/de
2. Jörger F, Krebs D, Maurer J, Roth C, Samer C, Sessa C. Recommendations concerning the application of the General Consent version 2019. [cited 2021 Dec 13]. Available from: https://www.fedlex.admin.ch/eli/cc/2013/617/de
3. Ehrlich SL, Eades D, Ainsworth J. Discursive Constructions of Consent in the Legal Process. Oxford University Press; 2016. 345 p.
4. eHealth Suisse, Kompetenz- und Koordinationsstelle von, Bund und Kantonen. Strategie eHealth Schweiz 2.0. [cited 2021 Dec 13]. Available from: https://www.e-health-suisse.ch/fileadmin/user_upload/Dokumente/2018/D/181214_Strategie-eHealth-Suisse-2.0_d.pdf
5. De Nardi G, Ehrensberger-Dow M, Matic I, Steiner F. Verständlichkeitsstudie Generalkonsent: Aufklärung und Einwilligung zur Weiterverwendung von biologischem Material und gesundheitsbezogenen Personendaten für die Forschung. [Internet]. Zürcher Hochschule für Angewandte Wissenschaften, Departement Angewandte Linguistik. [cited 2021 Dec 3]. Available from: https://digitalcollection.zhaw.ch/handle/11475/18831
6. Gruberski T. Generalkonsent [Internet]. SAMWASSM. [cited 2021 Dec 3]. Available from: https://www.samw.ch/de/Ethik/Themen-A-bis-Z/Generalkonsent.html
7. Diez-Domingo, Javier, Fons-Martínez, Jaime, García-Bayarri, Júlia, Calvo Rigual, Fernando, Petrova Dobreva, Dafina, Vázquez-Moreno, Mónica, et al. Guidelines for Tailoring the Informed Consent Process in Clinical Studies. [cited 2021 Dec 3]. https://i-consentproject.eu/wp-content/uploads/2021/03/Guidelines-for-tailoring-the-informed-consent-process-in-clinical-studies-2.pdf
8. HL7.org 2011+. HL7 FHIR Release 4 - 6.2 Resource Consent - Content. [cited 2021 Dec 3]. Available from: https://www.hl7.org/fhir/consent.html

Healthcare of the Future 2022
T. Bürkle et al. (Eds.)

doi:10.3233/SHTI220335

Navigation with Augmented Reality in a Hospital

Joshua DREWLOW [a,1], Michael DÄPPEN [a] and Michael LEHMANN [a]
[a]*Bern University of Applied Sciences, Biel, Switzerland*

Abstract. To improve wayfinding in hospitals or other complex buildings a prototype of an indoor navigation app was implemented that uses Augmented Reality for positioning and guiding users. The iOS and Android app allow the navigation from anywhere inside an area of 690 m². In a usability test 8 from 12 users preferred using the app over a map or verbal directions. Along with the proposed improvements this approach allows covering areas as large as 100'000 m².

Keywords. indoor navigation, wayfinding, positioning, augmented reality, mobile application

1. Introduction

Wayfinding inside large and complex buildings like hospitals is difficult due to multiple buildings, floors and insufficient aids like signage or maps [1, 2]. Complicated wayfinding can lead to patients frequently asking staff for directions, treatments being started late, etc. Most existing indoor navigation apps for smartphones rely on Wi-Fi or Bluetooth for which the acquisition and maintenance often are associated with high costs [3]. These apps mostly show information as a 2D map or text instruction, which some people find difficult to understand. Augmented Reality (AR) promises to present directions in a more understandable way by embedding 3D objects as an overlay in a live video feed of the surroundings captured by the rear camera [4, 5]. Until now, AR on smartphones was limited to showing 3D content on small surfaces. Recent advances in computer graphics algorithms and better calibrated smartphone cameras (vSLAM and viSLAM) enable complex AR experiences in a larger environment such as an entire building [6]. Such AR experiences require point clouds, a set of image data describing unique points in the environment. These functionalities became available for the public in the most known AR-frameworks in September 2020, when this project was started.

2. Methods

2.1. Implementation of the app

Requirements and mockups were defined in cooperation with the Hospital Centre located in Biel, Switzerland. Different AR frameworks were evaluated. Criteria included: the

[1] Corresponding Author: Joshua Drewlow, Bern University of Applied Sciences, Quellgasse 21, 2502 Biel, Switzerland, E-mail: joshua@drewlow.ch

capabilities of the framework (the size of point clouds, the creation methods, the deployment for Android and iOS), easiness of using the functionalities and the ability to use the framework in the game engine Unity.

To create point clouds, the authors used the *Matterport Pro2* camera and the *Area Target Creator* app on an Apple iPad Pro 11 (2020). For route calculation the *NavMesh system* and the extension *NavMeshComponents* of Unity was used.

Due to COVID-19 restrictions the prototype app was only implemented and tested at the Bern University of Applied Sciences in Biel and not at the hospital.

2.2. Usability tests

The first test was carried out during the implementation phase when the app included one point cloud. Following tasks were performed: read and explain the onboarding, navigate to two destinations, and complete a guided tour. The second test was conducted after the main implementation phase and included three point clouds. The probands were randomly assigned to navigate on three different routes, overlapping multiple point clouds, by using three wayfinding aids: the app, a map created by the authors and verbal directions. For both tests a guide observed the proband and observed the behavior and how the tasks were solved. After the tests the probands filled out a questionnaire.

3. Results

3.1. Indoor navigation app

The framework Vuforia (version 9.6.4) was chosen because of its easy-to-use functionality to manage point clouds. The framework is known to support many devices and is able to manage up to 255 *Area Targets* (an optimized point cloud) of which one can have a maximum size of 450 m² when using a 3D capturing camera like *Matterport Pro2* [7]. During the project seven *Area Targets* of the total size of 850 m² were created. In the final app three *Area Targets* with the total size of 690 m² were used. The file size of the *Area Targets* used in the app is 46.8 MB and the automatically generated mesh used for occlusion is 144.3 MB. One *Area Target* that included a long corridor generated with the *Matterport Pro2* had to be discarded because of an error while postprocessing. Three *Area Targets* were used for testing purposes only. For 100 m² it took about one hour for scanning with the *Matterport Pro2* and five minutes with the *Area Target creator* app. For postprocessing it took the *Matterport* service between two and ten hours and less than a minute with the *Area Target creator* app.

The app was deployed for iOS and Android and consists of two main functions: navigation and a guided tour. A navigation or tour can be started from any position inside the area where *Area Targets* previously have been created.

Before starting a navigation or tour the user is asked to select the current floor. This activates the respective *Area Targets* needed for positioning. The user is asked to point the device's rear camera continuously at the environment so the app can find its exact position. As soon as the device position is found a green line and arrows show the fastest route to the destination (Figure 1). At the destination a green pin and an orange board is visible. On the bottom of the screen an estimated arrival time and the remaining distance is shown.

Figure 1. Screenshot of the navigation mode on a tablet. A video is available on YouTube [8].

Figure 2. A test person using the app on a smartphone.

3.2. Usability test results

All the eight probands that participated in the first test were able to navigate with the app while facing some issues: not knowing the ideal way of holding the device for optimal positioning, insufficient instruction when the position was lost and overseeing obstacles.

Table 1 shows the results of the second test were twelve probands participated. The additional app time indicates positioning at the starting point. Encountered issues included: positioning at the starting point took a lot longer than previously (brackets) due to an algorithm trying to automatically detect the floor (was deactivated later) and the green line was partially visible through the wall which distracted some probands from completing the navigation task.

Table 1. Characteristics and results of wayfinding on three routes by using three wayfinding aids.

Distance	Turns	Change of floor	App	Map	Verbal
61.8 m	7	-	2.2 (+ 1.2) min	1.6 min	1.4 min
50.9 m	6	+ 1	2.1 (+ 0.6) min	1.8 min	0.7 min
59.5 m	10	- 1	1.3 (+ 0.6) min	3.5 min	1 min

8 out of 12 probands indicated they would prefer to use the app in a hospital. 5 of 12 probands were over 55 years old. The constant holding of the device in front of the eyes sometimes led to fatigue or obstacles were overlooked. Some probands indicated that an additional 2D map would be helpful.

4. Discussion

This project shows that an AR indoor navigation app is technically realizable for large buildings and was preferred by probands over maps and verbal directions. However, the usefulness and acceptance in a hospital environment could not be conclusively investigated, considering the more stressful circumstances and floors and walls being very look alike. The technology stack has proven to be a good combination. Based on the chosen approach and considering some improvements, a surface of about 100’000 m² can be covered.

4.1. Usability

Navigation with the app was only faster than with the map on one of three routes, nevertheless with ten turns and change of one floor it is the most complex one. To navigate with verbal directions was faster as expected because the test environment was relatively small and simple.

Most of the test persons found it intuitive to follow the displayed path and would prefer using the app over a map or verbal directions. Some of the encountered issues during the tests are software issue which can be solved by updates, and some depend on the digital literacy of the user which can be improved by better instruction. It was observed that the 5 of 12 probands older than 55 years old had most difficulties to concentrate to follow the line or overseeing obstacles. Furthermore, to navigate with the app the user is required to point the smartphone camera constantly at the environment. Stable localization therefore requires active action on the part of the user.

4.2. Recommendations

Scaling the app from 690 m² to 100'000 m² could lead to saturation of device storage or impact WLAN performance at a hospital. To overcome these difficulties different approaches are possible: download data dynamically, provide pre-installed devices and the use of low polygon objects for occlusion instead of the generated mesh. The authors recommend using the *Matterport Pro2* because *Area Targets* can be four times larger than with the *Area Target creator* app.

An inaccurate *Area Target* created with the *Matterport Pro2* revealed that the current software has difficulties to cope with similar looking environments. To improve positioning under such conditions the app can be provided with the approximate position of the user, by using GPS, using QR codes or let the user select the current floor.

To the authors it is unknown how robust *Area Targets* are to changes in the environment. This should be investigated further, but is not considered a concern due to the assumption that areas relevant for navigation won't change the appearance much.

References

[1] Zijlstra E, Hagedoorn M, Krijnen WP, van der Schans CP, Mobach MP. Route complexity and simulated physical ageing negatively influence wayfinding. Appl Ergon. 2016;56:62–7. doi:10.1016/j.apergo.2016.03.009.

[2] Butler DL, Acquino AL, Hissong AA, Scott PA. Wayfinding by Newcomers in a Complex Building. Hum Factors. 1993;35:159–73. doi:10.1177/001872089303500109.

[3] A. Billa, I. Shayea, A. Alhammadi, Q. Abdullah, M. Roslee. An Overview of Indoor Localization Technologies: Toward IoT Navigation Services. 2020;2020 IEEE 5th International Symposium on Telecommunication Technologies (ISTT):76–81. doi:10.1109/ISTT50966.2020.9279369.

[4] Gerstweiler G, Vonach E, Kaufmann H. HyMoTrack: A Mobile AR Navigation System for Complex Indoor Environments. Sensors (Basel, Switzerland) 2015. doi:10.3390/s16010017.

[5] Huang B-C, Hsu J, Chu ET-H, Wu H-M. ARBIN: Augmented Reality Based Indoor Navigation System. Sensors (Basel, Switzerland) 2020. doi:10.3390/s20205890.

[6] Servières M, Renaudin V, Dupuis A, Antigny N. Visual and Visual-Inertial SLAM: State of the Art, Classification, and Experimental Benchmarking. Journal of Sensors. 2021;2021:1–26. doi:10.1155/2021/2054828.

[7] PTC. Area Targets. https://library.vuforia.com/environments/area-targets. Accessed 28 Jan 2022.

[8] Joshua Drewlow. Indoor Navigation with Augmented Reality in a hospital. 30/01/2021. https://www.youtube.com/watch?v=CGphw9r__5k. Accessed 28 Jan 2022.

Healthcare of the Future 2022
T. Bürkle et al. (Eds.)

Subject Index

This page intentionally left blank

Healthcare of the Future 2022
T. Bürkle et al. (Eds.)

Author Index

This page intentionally left blank