Tampere University

ANDRES LEDESMA

Assessment of Data Visualizations for Clinical Decision Support

Tampere University Dissertations 277

Tampere University Dissertations 277

ANDRES LEDESMA

Assessment of Data Visualizations for Clinical Decision Support

ACADEMIC DISSERTATION
To be presented, with the permission of
the Faculty of Medicine and Health Technology
of Tampere University,
for public discussion in the TB104
of the Tietotalo, Korkeakoulunkatu 1, Tampere,
on 18 September 2020, at 12 o'clock.

ACADEMIC DISSERTATION
Tampere University, Faculty of Medicine and Health Technology
Finland

Responsible supervisor and Custos	Professor Ilkka Korhonen Tampere University Finland	
Supervisor	Professor Ilkka Korhonen Tampere University Finland	
Pre-examiners	Professor Mark Van Gils VTT Technical Research Centre of Finland Finland	Professor Minna Isomursu University of Oulu Finland
Opponent	Professor Gearóid ÓLaighin National University of Ireland Galway Ireland	

The originality of this thesis has been checked using the Turnitin OriginalityCheck service.

Copyright ©2020 author

Cover design: Roihu Inc.

ISBN 978-952-03-1622-8 (print)
ISBN 978-952-03-1623-5 (pdf)
ISSN 2489-9860 (print)
ISSN 2490-0028 (pdf)
http://urn.fi/URN:ISBN:978-952-03-1623-5

PunaMusta Oy – Yliopistopaino
Vantaa 2020

PREFACE

The research presented in this thesis has been conducted in Tampere University (formerly known as Tampere University of Technology) from 2015 to 2020. The research was financially supported via research projects funded by the Finnish Funding Agency for Technology and Innovation (TEKES), the European Commission (program for Research and Innovation Horizon 2020 under the Grant Agreement number 689260) and VTT Technical Research Centre of Finland. The support from TEKES was provided as part of the Digital Health Revolution and Finland Distinguished Professor Programme (FiDiPro) research projects. The European Commission and TEKES supported this research under the ARTEMIS-JU WithMe project. The support from the European Union came from the program for Research and Innovation Horizon 2020 under the Grant Agreement number 689260 as part of the project Digi-NewB.

First, I would like to express my gratitude to my supervisor, Adjunct Professor Ilkka Korhonen, for giving me the opportunity to begin my research career as a doctoral student in the Personal Health Informatics group. He has provided guidance, inspiration and advise that motivated me to see through my research work and studies. His invaluable advice in research methodology and scientific writing were crucial to complete my studies.

I would also like to thank my instructor D.Sc. (Tech.) Hannu Nieminen for his constant guidance, ingenuity and invaluable assistance during my research work. With his years of experience, he served as a source of inspiration and advise throughout my studies

My gratitude to the pre-examiners Docent, Research Professor Mark Van Gils and Professor Minna Isomursu for providing valuable comments and objective criticism of this thesis.

I want to thank D.Sc. (Tech.) Miikka Ermes from VTT Technical Research Centre of Finland for his insightful comments and valuable guidance during our research studies.

My gratitude also goes to Professor Misha Pavel and Professor Holly Jimmison from Northeastern University Boston, for providing guidance and invaluable advice in our research studies.

I wish to express my gratitude to Professor Niranjan Bidargaddi from the College of Medicine & Public Health at Flinders University for advise on research methodology. I would also like to thank M.D. Jörg Strobel from the Queen Elizabeth Hospital University of Adelaide and M.D. Doctor Geofrey Schrader from the Country Health Local Health Network and Flinders University for their medial expertise and guidance.

Special thanks to M.D. Päivi Valve for her medical knowledge, which was crucial for our studies. Also my gratitude goes to my colleague Mohammed Al-Musawi and his diligent work that helped validate part of the research presented on this thesis.

I would also like to thank Alpo Värri for inviting me to take part in the research project Digi-NewB which supported my research and the publication of my final article. I would also like to thank VTT for funding the research study made in Australia, which provided the data needed to complete my last publication.

My gratitude also to Movendos, the company where I have been working for the past 5 years. I wish to thank all my great colleagues for providing a welcoming atmosphere and for accommodating several research hours that were much needed to complete my studies.

Last but not least, I would like to thank my family, specially my father for his unconditional support and his shining example of dedication, discipline and academic excellence. I wish to extend my gratitude to all my family for their constant support during my studies. I would also like to thank my friends, many of which live abroad but have nevertheless been an essential company and a constant source of moral support throughout my life.

ABSTRACT

Since the wide adoption of electronic health records, the amount of clinical data has increased dramatically. It has been estimated that in 2013 there was a total of 153 exabytes of clinical data, and by 2020 this number will increase to 2 314 worldwide. Another estimate has calculated that the average patient generates 80 megabytes of data per year. Clinicians rely on clinical data to make informed decisions at the point of care. However, the volume and complexity of clinical data along with time constraints, make the diagnosis process challenging, time-consuming and prone to errors. It has been estimated that the number of deaths due to clinical misdiagnosed are between 44 000 and 98 000 per year in the United States. By using computerized data visualization techniques, clinicians can extract valuable insights, reducing cognitive overload. Consistent and structured methodologies that assess clinical data visualizations and their effect on the decision making process, are still missing. The gap that the thesis aims to bridge is to develop a methodology that allows the assessment of clinical data visualizations in terms of their efficacy in supporting clinical decision making. The purpose of this thesis is to develop such methodology, which studies the reasoning derived from the visualization and how this affects the clinical decision making process at an individual level. The first experiment compared five different visualization techniques. The study measured the quantity and quality of insights obtained by the users of the visualizations. This assessment technique has not been used before in the context of clinical data. By evaluating the visualizations in this way, it was objectively determined that from the visualization techniques used in the study, the radar plots were the most effective in enabling the generation of hypotheses and in acquiring accurate understanding of the data. The second experiment studied a dashboard representing the evolution of health and wellness of a modelled patient. The dashboard included an improved version of the radar plots used in the previous study. By using methods such as heuristics, cognitive walk-through, analytic tasks, and usability questionnaires, it was objectively determined that the

dashboard was effective in assisting users to find critical information and gain accurate understanding of the clinical data. The study was able to quantify and demonstrate the degree to which the dashboard proved useful for its intended audience by scoring an average of 6.02 out of 7 points in the usability studies and a completion rate of analytical tasks of 96 percent. The third and last experiment compared an existing tabular interface for clinical data against an interactive visualized timeline. The methodology used in this study was the same as in the first experiment. However, the chronological and longitudinal nature of the clinical data required adaptations to the methodology. The use of this methodology to evaluate longitudinal data visualizations has not been reported in previous studies. By applying this novel approach, it was objectively determined that the timeline enabled clinicians to deduce the underlying conditions of the patients, reflecting a deep understanding of the data by connecting information scattered over a period of time. These three assessments followed state-of-the-art methodologies in the discipline of data visualization that have not been used before for the purpose of clinical decision making. The objectives of the thesis were met by applying novel assessment techniques. By applying quantitative and qualitative research, it was possible to compare visualizations in a clinical context and provided better understanding on what makes a good visualization. The publications in this dissertation document experiments conducted to study the reasoning derived from the visualization and how this affects the clinical decision making process at an individual level. By utilizing consistent methodologies such as the insight-based, usability testing and cognitive walkthrough, different visualizations were objectively compared and assessed. These documented experiments can serve as blueprints for future studies. With a deeper understanding on the impact of visualization tools in the clinical decision making process, researchers can develop better visualizations to ease the cognitive burden of making sense of complex data. With better visualizations, clinicians can gain deeper understanding of the data, making better decisions, resulting in better patient outcome.

CONTENTS

List of Figures

List of Tables

ABBREVIATIONS

ASQ	After Scenario Questionnaire
CSUQ	Computer System Usability Questionnaire
EHR	Electronic Health Record
T2D	Type II Diabetes
UP	User Performance
VDAR	Visual Data Analysis and Reasoning

ORIGINAL PUBLICATIONS

Publication I	A. Ledesma, H. Nieminen, P. Valve, M. Ermes, H. Jimison and M. Pavel. The shape of health: A comparison of five alternative ways of visualizing personal health and wellbeing. *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. 2015, 7638–7641.
Publication II	M. Al-Musawi, A. Ledesma, H. Nieminen and I. Korhonen. Implementation and user testing of a system for visualizing continuous health data and events. *2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*. IEEE. 2016, 156–159.
Publication III	A. Ledesma, M. Al-Musawi and H. Nieminen. Health figures: an open source JavaScript library for health data visualization. *BMC Medical Informatics and Decision Making* 16.38 (2016), 1.
Publication IV	A. Ledesma, N. Bidargaddi, J. Strobel, G. Schrader, H. Nieminen, I. Korhonen and M. Ermes. Health Timeline: An Insight-based Study of a Timeline Visualization of Clinical Data. *BMC Medical Informatics and Decision Making* 19.170 (2019), 1.

AUTHOR'S CONTRIBUTION

Publication I — The author was the main contributor and had the responsibility of writing the manuscript. He designed the experiment protocol and programmed the visualizations. He also shared the responsibility of transcribing the recordings and conducting the experiment with P. Vale. The author had the responsibility of establishing the assessment criteria with P. Valve and H. Nieminen. He recruited the participants, evaluated the insights and analyzed the data.

Publication II — The author wrote the sections corresponding to the Health Figures and the data used in the experiment. He programmed the longitudinal and combined polar graphs. He assisted M. Al-Musawi in the programming of the coaching timeline. He designing the study with M. Al-Musawi and H. Nieminen.

Publication III — The author was the main contributor and was responsible for writing the manuscript. He programmed the polar coordinate visualization system and documented the algorithms used. He shared designed the study and experiment protocol with H.Nieminen and M. Al-Musawi.

Publication IV — The author was the main contributor and had the responsibility of transcribing the recordings and writing the manuscript. He shared the responsibility of designing the study, experiment protocol and assessment criteria with M. Ermes and N. Bidargaddi. He evaluated the insights and analyzed the data.

1 INTRODUCTION

The amount of clinical data is expected to continue its growth. One estimation projects that clinical data will grow from 153 exabytes in 2013 to 2 314 in 2020 worldwide [1] (one exabyte equals one million terabytes). The chief information officer from Beth Israel Deaconess Medical Center has calculated that, per year, a total of 20 terabytes clinical data are collected for 250 000 active patients. This means an average of 80 megabytes of data per patient per year [2].

It has been estimated that medical consultations last from "48 seconds in Bangladesh to 22.5 minutes in Sweden", while in Finland the average is 17 minutes [3]. The large volume of complex clinical data along with time constraints may overwhelm health practitioners. Failure to acquire adequate understanding of clinical data may increase the risk of misdiagnosis. Estimates on the number of deaths due to clinical misdiagnosis are between 44 000 and 98 000 per year in the United States [4]. About 30% of the annual healthcare spending ($750 billion) has been reported to be lost due to misdiagnosis [5].

Access to clinical information and the ability to gain knowledge based on clinical data are crucial factors in providing better patient care [6]. Clinical data persist across multiple platforms [7] in a wide variety of formats, posing a challenge for health practitioners [8, 9]. Rind and colleagues stated in their study that "information visualization has the potential" to "provide cognitive support to healthcare providers, patients, and families" based on clinical data [10].

Electronic health records (EHRs) are the purest type of electronic clinical data obtained at the point of care [12]. EHRs are representations of longitudinal data collected "during routine delivery of health care" [13, 14]. Recently, EHRs have been studied beyond their initial administrative purpose as mere records of visits to the healthcare providers. Miotto and colleagues [7] stated that researchers have studied their secondary use to "enable data-driven prediction of drug effects and interactions [15], identification of type 2 diabetes subgroups [16], discovery of comorbidity clus-

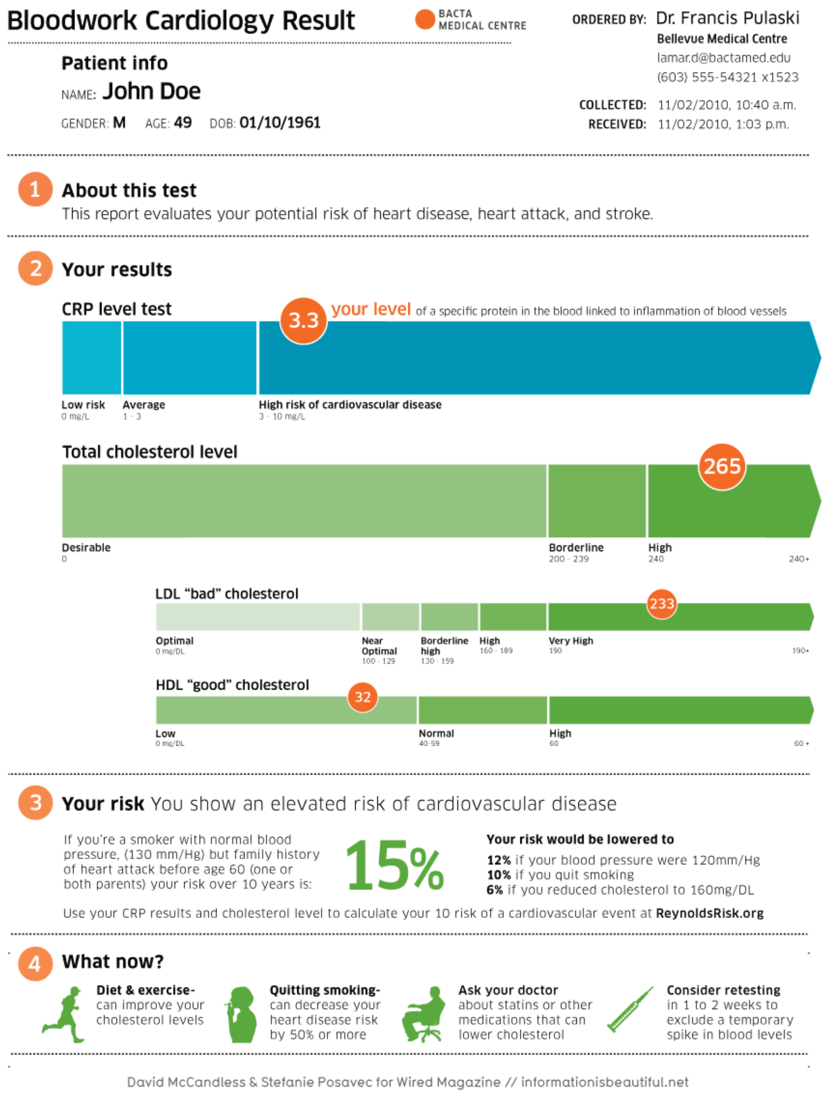


Figure 1.1 An example of laboratory bloodwork results for cardiology as redesigned by Goetz [11]. Creative Commons [2015]
`https://informationisbeautiful.net/`

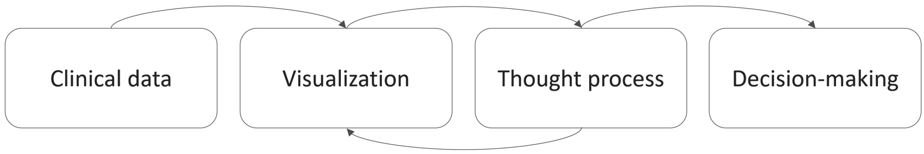


Figure 1.2 A flowchart diagram on the process of understanding clinical data and decision making. The thought-process can feed back into the visualizations stage by interactive data explorations such as zooming in and filtering out elements in the visualization.

ters in autism spectrum disorders [17], and improvements in recruiting patients for clinical trials [18]".

As the amount of clinical data increases, so does the potential value that healthcare providers can extract from the data. To help clinicians and patients make sense of the data, innovative techniques are needed. Researchers are trying to find ways to present complex clinical data in forms that are easier to understand [10, 19].

The goal is to be able to understand heterogeneous, complex, and longitudinal clinical data with the objective of making more informed decisions at the point of care, within the limited time available for each patient. Data visualizations aim to assist the understanding of data [10, 19]. Numerous research efforts have addressed the need to understand clinical data by designing and building a variety of visualizations [10, 19]. Several studies have been published detailing the implementation and design of such visualization tools.

As an example, TimeLine [20] is a software tool that organizes medical records and provides a "problem-centric temporal visualization". Lesselroth and Pieczkiewicz [19] conducted an extensive literature survey on strategies for the visualization of personal health data. They concluded that "smart dashboards" combining different data sources are needed to improve the understanding of our health.

Data visualizations that enable the identification of significant connections over time have been shown to increase the user's self-understanding [21]. Goetz [11] prooved that graphical presentation of data greatly affects the understanding of our health. Figure 1.1 shows an example of the redesign of laboratory test results using a graphical presentation.

With a greater understanding of the data, clinicians can make better decisions regarding patient care. Figure 1.2 shows the cycle that comprises the reasoning from the presentation of clinical data to the decision making process.

However, this notion leads to an open question: how to establish if a visualization is helpful in the decision making process? Assessment methods can measure the effectiveness of a visualization and the degree to which it can assist the decision making process.

The majority of studies assume that a visualization technique effectively reduces the cognitive load required to understand complex data [22, 23]. However, after a literature survey was conducted (chapter 5), it was revealed that a small number of articles contain assessments that compare or study the properties of data visualizations to understand how it helps the intended audience to better understand the data and thus assist in the clinical decision making process.

The assessments of the visualization tools often focus on usability testing and computational performance. Usability testing is of the utmost important to determine the effectiveness of a computerized system that aims to assist professionals in their daily activities. However, to further study the effectiveness of a clinical data visualization and its impact on the decision making process, additional assessment methods are required [22, 23]. It must be therefore understood that the purpose of a visualization is to assist in the comprehension of the data, then the question follows: how can visualizations assist the clinician in the decision making process? An approach that considers visual reasoning is therefore needed[23].

North suggests that "the purpose of visualization is to gain insights on the data it represents [24]. This notion provides the answer to the previous question, as it addresses the user's understanding of the data. North proposes the use of an "insight-based methodology", which focuses on the recognition and quantification of insights gained from exploring the data visualization [25, 26]. With the knowledge obtained from these insights, users can then make better informed decisions regarding patient care. This methodology has been previously applied in visualization tools for genetic data but not in the context of clinical decision making [25, 26, 27].

A literature survey, detailed on chapter 5 of this dissertation, revealed that methodologies that address the visual reasoning of clinical data visualizations and their impact on the clinical decision making process are uncommon. From the published literature, only two studies focused on the impact of the data visualization on the clinical decision making process. These studies did not apply a previously tested methodology such as the insight-based, instead custom made questionnaires were applied. The purpose of this thesis is to develop such methodology, which studies the reasoning derived from the visualization and how this affects the decision making process at an individual level.

The publications compiled for this dissertation deal with clinical data, as either EHR or physiological measurements (heterogeneous). The visualizations were evaluated with methodologies that focus on visual analysis and reasoning and how this affects the decision making process at an individual level. The assessment methods are further explained and analysed in chapter 4.

The visualizations in the studies were instruments to help in the clinical decision making process. The publications compiled for this dissertation feature custom data visualization libraries and implementations in JavaScript, designed for the purpose of representing clinical data.

2 OBJECTIVES

The gap that the thesis aims to bridge is to develop a methodology that allows the assessment of clinical data visualizations in terms of their efficacy in supporting clinical decision making. The purpose of this thesis is to develop such methodology, which studies the reasoning derived from the visualization and how this affects the clinical decision making process at an individual level.

The specific objectives are:

1. to study and compare how different clinical data visualizations affect visual reasoning (Publications I — IV).
2. to apply a methodology that studies visual reasoning and the decision making process in the assessment of clinical data visualizations (Publications I — IV).
3. to develop and objectively measure the scalability and usability of software that visualizes holistic clinical data (Publications II and III).
4. to study how visualizations affect the decision making process (Publications I and IV).
5. to apply a methodology that studies visual reasoning and the decision making process to assess visualization software for clinical data with the participation of domain experts (Publication IV).

The publications in this thesis applied the experiment research method to study the reasoning derived from the visualization and its effect in the clinical decision making process. The research collected data based on the experience of visual reasoning in the decision making process (qualitative) and also collected numerical data for ranking and categorization for statistical analysis (quantitative).

3 BACKGROUND ON DATA VISUALIZATION AND CLINICAL DECISION SUPPORT

3.1 Data visualization principles

Otten and colleagues [28] stated that the origins of data visualization can be traced back to the eighteenth century, when William Playfair developed the first charts to convey information on historical data [29]. Lesselroth and Pieczkiewicz [19] cited Friendly [30] in reference to the beginnings of the scientific investigation of data visualization. Coll and researchers [31] conducted the first studies comparing bar graphs and pie charts to other data representations. Studies shifted from trying to find the best visualization to trying to identify the most suitable graphical representation in specific circumstances.

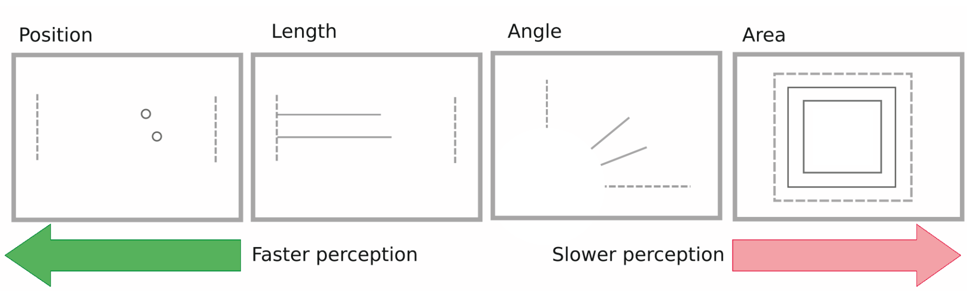


Figure 3.1 Figure adapted from Lesselroth and Pieczkiewicz [19], summarizing the findings of Cleveland and McGill [32] in their study comparing the speed of graphical perception across different visualization techniques.

Lesselroth and Pieczkiewicz [19] focused their studies on two aspects: perception and cognition. Perception is the "low-level acquisition and organization of sensory information" and cognition is the "higher-level interpretation of this information". Cleveland and McGill [32] conducted a study focusing on the speed at which users can perceive and process data based on different representations. These findings were supported by a study conducted by Carswell [33] on “graph reading and comprehension”. Figure 3.1 is illustrates Cleveland and McGill’s findings summarized by Lesselroth and Pieczkiewicz. Cleveland and McGill focused on the speed and accuracy of “data processing”.

Lesselroth and Pieczkiewicz [19] referred to Carswell’s [33] conclusion to the studies, stating that the appropriate visualization is highly dependent on context. The researchers referred to studies on decision theory, demonstrating that users relied on visualizations that “minimized the cognitive burden” [34], and these visualizations varied depending on the context [35, 36, 37].

Visualization task	Suggested modality
Value extraction	Numeric tables
Value comparison	Bar charts, line graphs, scattered plots
Proportions	Pie charts, stacked bar charts
Trend detection	Line graphs

Table 3.1 Suggested modality by visualization task as summarized by Lesselroth and Pieczkiewicz [19].

Attempts have been made to compile “best practices” regarding data visualization. Otten and colleagues [28] referred to Tufte’s work [38, 39, 40] on “best practices for communicating quantitative and qualitative information”. Lesselroth and Pieczkiewicz concluded that ultimately, there is no “grand unified theory”, as it is highly dependent on context. They identified “suggested modalities” after reviewing articles by Kosslyn [41], Tufte [38], Few [42, 43], Schriger and Cooper [44], Gillian et al. [45], Robbins [46] and Cleveland and McGill [47]. Table 3.1 shows their summarized findings [19].

3.2 Clinical decision support systems

In the context of healthcare, clinicians need to gather large clinical data sets with the goal of decreasing uncertainty, patient “risks and costs”. The process of “deciding what information to gather, which tests to order, how to interpret and integrate this information to draw diagnostic conclusions, and which treatments to give is known as clinical decision making” [48]. Typically, clinicians attempt to answer the following questions: "What disease does this patient have? Should this patient be treated? Should testing be done?"

In some cases, clinicians make decisions based on their own experience. Common practice involves recognizing patterns in a disease and following a trial and error approach for patient treatment. For instance, if there is an on-going epidemic of a certain disease, and a patient presents the symptoms of that disease, the clinician might just prescribe the recommended treatment for that disease. However, these practices are prone to mistakes because other diseases might exhibit similar symptoms and might have been misdiagnosed [48].

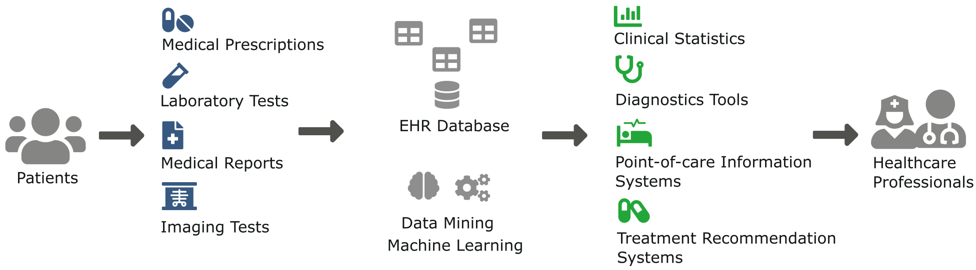


Figure 3.2 Illustration showing a diagram of data acquisition as EHR entries. The patient is the source of multiple data sources. The analysed data can then be used to assist clinicians in the decision making process. The figure is a visual representation of the ideas presented by Miotto and colleagues [7]

A methodical approach is preferred when following the condition of a patient before attempting to make a decision. For example, this means following the practices of "evidence-based medicine, use of clinical guidelines, and use of various specific quantitative techniques (e.g., Bayes theorem)" [48].

Typically, a large portion of the information that a clinician requires in the decision making process resides in the EHR. An EHR is "a mechanism for integrating health care information currently collected in both paper and electronic medical records (EMR) for the purpose of improving quality of care" [49]. Figure 3.2 shows the patient as the source of information for multiple data-driven activities, as proposed by . Miotto and researchers [7]. As an example, the figure shows how EHRs comprising of medical prescriptions, reports and tests can then be stored in databases for later use in machine learning and data mining. The discovered knowledge can then be utilized by information systems for the purpose of assisting healthcare professionals in the decision making process. As an example, the figure shows treatment recommendation and diagnostics systems, as well as point-of-care and clinical statistics tools. The literature review conducted in chapter 5 of this thesis details additional cases in which the clinical data served as the basis for the decision support process.

3.3 Data visualization in clinical decision support systems

Shneiderman and researchers have stated that interactive information visualization "will bring profound changes to personal health programs, clinical healthcare delivery, and public health policy making" [50]. They cite the work of Rind and researchers [10] as examples of the importance of data visualization.

Rind and researchers conducted a literature survey of clinical data visualizations, in their work they reviewed examples that illustrate the application of visualization techniques in clinical decision making process. These examples are: LifeLines [51], MIVA [52], WBIVS [53], Midgaard [54], VisuExplore [55], VIE-VISU [56], LifeLines2 [57, 58], Similan [59], PatternFinder [60], VISITORS [61], Caregiver [62], IPBC [63], Gravi++ [64] and TimeRider [65].

Lesselroth and colleagues have stated that data visualization techniques can significantly improve the quality of health care [19]. They have also provided examples of clinical data visualization tools, these examples are: a glycemic control monitoring software named Glucotron 5000 [66], an interactive graphical modality [53] for home monitoring of lung transplant patients [53] and the TimeLine software that represents longitudinal data for diagnosis and therapy [20].

The potential and realization of data visualization in supporting the clinical decision making process is described in further detail in chapter 5 of this thesis. The chapter presents ample evidence of the effects of visualization techniques and how they can assist the intended audience in the process of making sense of information in a clinical setting.

3.4 Data visualization in commercial solutions for health and wellness

Data visualization plays a key role in commercial solutions for self-monitoring devices that aim at collecting personal data for health and wellness. One example is

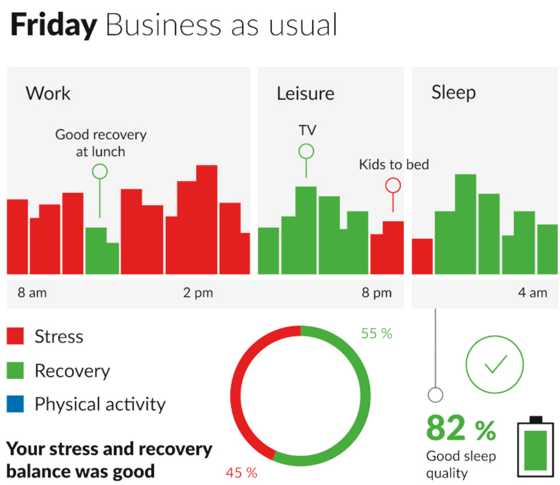


Figure 3.3 An illustration of the Firstbeat Lifestyle Assessment showing the plots that illustrate the day distribution on activity, recovery and stress [67], Firstbeat Technologies Oy ® [2020] `https://perma.cc/7DVW-LEPX`

the dashboard for stress recovery developed by the company Firstbeat [67]. The dashboard shows the stress and recovery balance using radar plots to illustrate the completion of the recovery cycle. The stress and recovery analysis is also plotted using colour-coded bar charts to give an overall picture of the cycles during the day. Figure 3.3 shows an example of the Firstbeat Lifestyle Assessment®.

Withings is another company with a wide-range of health and wellness tracking products. To present the data to its users, the company employs several visualization dashboards. For instance, the Pulse HR® product utilizes an application in the mobile phone to display graphs on the activity of the user such as a step counter and sleep monitoring [68].

The company Oura Cloud also uses a combination of linear and bar plots to show the data collected by ring sensors. The dashboard is a longitudinal visualization of the collected data using an interactive timeline [69].

Similar to Withings, Fitbit offers an alternative for step counting, tracking and sleep assessment. The Fitbit App [70] also offers a dashboard that aggregates the data using visualization dashboards. The dashboard uses mostly bar charts and longitudinal graphs to summarize the data in different categories. Additionally, an overview dashboard is offered via the desktop version of the application [71]. This dashboard

Figure 3.4 The integrated desktop dashboard offered by Fitbit showing and overall view of the collected data from several categories summarized with different visualization techniques [70] Fitbit, Inc.® [2020]
`https://www.fitbit.com/de/app`

aggregates data from all categories collected by the tracking device into one view. Figure 3.4 demonstrated the integrated desktop dashboard offered by Fitbit.

4 ASSESSMENT METHODS FOR DATA VISUALIZATION

4.1 Evaluation scenarios

Lam and researchers conducted an extensive literature review on evaluations of data visualizations [23]. They have proposed seven scenarios that address specific assessment objectives. Researchers are advised to choose from these scenarios the assessment methods to perform depending on the research objectives.

The evaluations listed in the review are independent from the application context, therefore they are not exclusive to clinical data. However, these scenarios still provide valuable guidelines on how to approach the assessment process of data visualizations.

The evaluations are classified into seven categories or "scenarios" depending on the aim of the study. Table 4.1 summarizes the objectives that the evaluation scenarios aim to accomplish.

4.1.1 Evaluation and assessment

Evaluation as defined by the Cambridge dictionary is "the process of judging or calculating the quality, importance, amount, or value of something". Assessment is defined as "the act of judging or deciding the amount, value, quality, or importance of something, or the judgment or decision that is made". Assessment is thus suitable to the objective of studying the impact of visualization in the clinical decision making process, since it is an exercise in judgement to take a clinical decision. The seven scenarios developed by Lam et al. are a classification of different evaluation methods not assessments. However, some of these scenarios study the decision making process, analytical operation and cognition involved in this process. To this extent, the

Scenario	Object of study	Evaluation
Understanding environments and work practices (UWP)	People's workflow Workflow practices	Field observations Interviews
Evaluating visual data analysis and reasoning (VDAR)	Data analysis Decision making Knowledge management Knowledge discovery	(Multidimensional In-depth) Case Studies (insight-based) Laboratory observations and interviews Controlled experiment
Evaluating communication through visualization (CTV)	Communication, learning Teaching, publishing Casual information acquisition Controlled experiments	Field observations and interviews
Evaluating collaborative data analysis (CDA)	Collaboration	Heuristic evaluation Log analysis Field or laboratory observation
Evaluating user performance (UP)	Visual-analytical operation Perception and cognition Usability/effectiveness	Controlled experiments Field logs
Evaluating user experience (UE)	Potential usage Adoption Usability/effectiveness	Informal evaluation Usability test Field observation
Evaluating visualization algorithms (VA)	Algorithm performance Algorithm quality	Visualization quality assessment Algorithmic performance

Table 4.1 The table summarizes the seven scenarios described by Lam and colleagues [23], the objective of the study, and examples of assessment methods.

scenarios provide methods that are also relevant to the assessment process.

4.1.2 Criteria for the selection of the scenarios

All the scenarios are relevant to the assessment of tools that support the clinical decision making process. However, to address the research objectives, two scenarios are directly applicable: Evaluating Visual Data Analysis and Reasoning (VDAR) and Evaluation of User Performance (UP).

VDAR aims to study the decision making process and knowledge discovery [23].

A visualization of clinical data requires assessment of how it facilitates knowledge discovery on a collection of EHRs, and eventually how this affects the clinical decision making process at an individual level. UP focuses on visual-analytic operations, perception, cognition and effectiveness of the visualization [23]. Ideally, an effective visualization of clinical data will reduce the cognitive burden, increase the understanding of relevant information and assist in further data analysis. The other scenarios described by Lam and colleagues do not address the decision making process nor the visual-analytical operation.

Evaluating the Communication Through Visualization (CTV) is of the utmost importance when it comes to collective clinical decision making. Healthcare professionals consult with each other on daily basis, and often times involve the patient in the same process [23]. CTV studies communication, teaching, publishing and information acquisition. These are relevant and necessary evaluations for computerized clinical systems. However, the focus of this thesis is to develop a methodology that studies the reasoning derived from the visualization and how this affects the decision making process at an individual level. To study the collaborative analysis, a different set of tools would be required to conduct experiments involving multiple participants at a time. Such research and experimentation would provide scientific contributions in the collaborative data analysis domain. As stated in the objectives, the gap that the thesis aims to bridge is to develop a methodology that allows the assessment of clinical data visualizations in terms of their efficacy in supporting clinical decision making at an individual level. As such this thesis does not explore collaboration, communication and information acquisition.

Evaluating the User Experience (UE) is also relevant for decision support systems in the clinical domain. UE evaluation methods study systems that are part of the workflow of professionals. These evaluation methods address challenges such as identifying missing features, improving existing ones and prioritizing those that facilitate the work processes [23]. The studies conducted and published for this dissertation were performed on systems that are not used on the work processes of healthcare professionals. The visualization tools were developed for the purpose of experimentation and research. These was done intentionally to investigate the research gap stated in the objectives of this dissertation. UE evaluation methods are therefore suitable for systems that are used in the daily work of healthcare professionals. These methods are thus not suitable to address the research gap of this dissertation. Nevertheless, it would have been advantageous to conduct studies on systems that are used on daily basis in a clinical setting and apply UE evaluation methods to further understand how these systems could improve the daily work of healthcare professionals.

4.1.3 Evaluating visual data analysis and reasoning (VDAR)

VDAR evaluations "study if and how a visualization tool supports the generation of actionable and relevant knowledge in a domain" [23]. In the context of clinical data, the goal is "to support visual analysis and reasoning about" EHRs. VDAR is also suitable due to the output it produces when the evaluation is performed, in the form of "quantifiable metrics". Depending on the evaluation, these could be insights [25, 26].

Evaluations in the VDAR scenario look at how the visualization "as a whole" assists in the "analytical process", whereas UP evaluations focus on interactive aspects of the visualization in "isolation" [23]. These two scenarios complement each other in the overall assessment of the analytical assistive capabilities of a visualization.

The VDAR evaluation scenario was formulated by Lam and colleagues based on the intelligence analysis process developed by Pirolli and Card [22, 23]. This model identifies aspects of data exploration that are studied by VDAR methods. The aspects are a perfect match to what this dissertation aims to research.

Pirolli and Card formulate the aspects of the intelligence analysis process as the exploration of data and how it assists the filtering, searching, reading and extraction of information; discovery of knowledge, schematization of information, and support for analysis of theories; hypothesis generation and examination; and the decision making process [22].

A large variety of approaches have been followed when evaluating a visualization with VDAR. The output of the evaluation tends to be highly specific to the context, and no standardization has been suggested [23]. Instead, Lam and colleagues have recommended use case studies. They presented two examples of VDAR: the Multi-dimensional In-depth Long-term Case (MILC) studies and insight-based evaluations [23].

In MILC case studies, a long-term study takes place with users. Researchers typically explain the features and functions of the visualization system. Users are required to spend the necessary time to get acquainted with the visualization tool.

In the insight-based evaluations, researchers capture the reasoning process of the users as insights, which are "individual observation about the data by the participant, a unit of discovery" [25]. The goals of insight-based evaluations are threefold: "to deepen understanding of the visual analytics process, to understand how existing tools were used in analysis, and to test out an evaluation methodology" [23]. Contrary to the approach in MILC evaluations, insight-based evaluation does not provide users with guidelines or assistance regarding the visualization, to prevent interfering with the "normal data analysis process".

4.1.4 Evaluating user performance (UP)

The evaluations of UP compare visualizations by means of task completion time and accuracy [23]. Lam and researchers describe these assessments as objective studies on certain aspects of the visualizations [23]. These assessments address what are the "limits of the human perception" and how "visualization or interaction techniques compare" to one another [23].

Human perception and cognitive limitations are studied with the objective of deriving models on the use of space for plotting graphical representations and the use of interactions that facilitate exploration [23]. Lam and colleagues refer to the work done by Heer and Robertson, which consisted of a study on how animations can be used to present statistical data [72]. UP evaluations also aim to measure how users perform when given different visualization tools by studying the cognitive impact and the underlying limitations. Lam and researchers have also referred to studies on the "effects of image transformations", such as scaling, rotation and fisheye visual memory [73].

UP evaluations are also used to compare two or more alternatives "head-to-head" [23]. The comparisons typically take place in a controlled environment and include a list of tasks the user is requested to perform. Examples include a study to compare file exploration for investigation using a regular file explorer and a new solution named SpaceTree [23, 74]. In these evaluations, a use case is provided to simulate a real case scenario. Researchers typically conduct the experiments in laboratories, where the participants go through a set of predefined tasks. Measurements such as accuracy and time to completion are objective metrics used to compare different visualization techniques.

4.2 Usability testing

Computerized systems that store, process, and visualize EHRs are subject to usability testing. In the context of software engineering, usability is defined in ISO 9241-11 by Bevan and colleagues as "the degree to which software can be used by specified consumers to achieve quantified objectives with effectiveness, efficiency, and satisfaction in a quantified context of use" [75].

Johnson and colleagues from Westat developed a toolkit for usability testing of computerized EHR systems. The work was commissioned by the Agency for Healthcare Research and Quality from the U.S. Department of Health and Human Services [76]. The researchers identified deficiencies in the usability of existing EHR systems. The studies reviewed in the toolkit recommend establishing usability tests as part of the EHR certification process. These studies also recommend the development of "objective criteria that reflect best practices in EHR usability" [76, 77, 78]. The toolkit aimed to accomplish three objectives: the development of a usability toolkit for "primary care providers", drawing attention to EHR usability issues by "disseminating the toolkit" encouraging "evidence-based usability evaluation methods", and informing about the state of EHR usability testing in the accreditation process [76]. The toolkit also contains a comprehensive literature review of published studies on usability testing of EHRs [76]. The review identified some of the most prevalent usability tests conducted on EHR systems. These usability tests include heuristic evaluation, cognitive walkthrough, remote evaluation, laboratory testing and usability questionnaires [76].

Johnson and researchers have outlined two important aspects in the criteria used to select the methods included in the toolkit. The first aspect is the efficiency and convenience of the method. The practical approach is to select methods that are easy to implement and that measure the required aspects of usability in a prompt manner. The second aspect is the application of the usability method by primary caregivers without the need for usability experts. For practical reasons, healthcare professionals might not always have an available usability expert at hand when providing feedback on EHR systems. For this reason, healthcare professionals should be able to administer the usability tests. The usability toolkit report concludes that with the given criteria and the opinion of subject experts, the most suitable and practical method to evaluate EHR systems is the usability questionnaire [76].

The studies compiled in this dissertation include the usability testing methods recommended by Johnson and colleagues. These methods include heuristic evaluation, cognitive walkthrough, laboratory testing and usability questionnaires. The next section provides further details about the methods used in these studies.

4.3 Assessment methodologies used in the studies

The aim of this dissertation is to develop such methodology, which studies the reasoning derived from the visualization and how this affects the clinical decision making process at an individual level. As described in the intelligence analysis process developed by Pirolli and Card [22, 23], the VDAR evaluation methods study the discovery of knowledge, schematization of information, and support for analysis of theories, hypothesis generation and examination as well as the decision making process. Therefore, as in previous studies [25, 26, 27], the insight-based methodology provides greater opportunities to understand this phenomenon. The analysis of the insights could potentially bring a deeper understanding of how data visualization interacts with the clinical decision making process. As described above, VDAR aims to study "hypothesis generation" and examination, and the decision making process, which fits with the objectives of this thesis.

In addition to the insight-based studies, the UP assessments evaluate the interaction features of a visualization and how these affect data analysis and reasoning. The EHR usability toolkit [76] recommends usability methods that fall in the scenario of UP evaluations [23]. Publications II and III follow UP evaluations to test a computerized visualization system that supports the clinical decision making process. The computerized visualization system consisted of a health and wellness dashboard that visualized clinical data and provided interactive features for data exploration. Lam and colleagues have recommend that interactive features are to be evaluated using techniques described in the UP scenario [23]. Following these recommendations in addition to those from Johsnon and colleagues [76], the evaluation of the computerized visualization system included the heuristic evaluation, cognitive walkthrough, laboratory testing and usability questionnaires.

4.3.1 Insight-based methodology

The insight-based methodology proposed by North [24], focuses on the insights generated by the users of a data visualization. An insight is defined as "the capacity to gain an accurate and deep understanding" [79]. According to the literature survey of this dissertation (chapter 5), the most prevalent approach to evaluating clinical data visualization is to conduct briefing interviews or to utilize UP evaluations, typically with a set of predefined tasks. By contrast, insight-based methodology focuses on recognition and quantification of insights gained from exploratory use of the data visualization. An insight is a unit of discovery based on observation [25, 26, 27].

Insights have a quantifiable value based on the assessment criteria. The criteria should take into consideration the following characteristics of an insight [24, 25, 26, 27, 80]:

- **Observation**: The observation or finding provided by the participant during the process of analysing the data via a representation.
- **Time**: The amount of time taken to reach the insight.
- **Domain Value**: The value, importance, or significance of the insight.
- **Hypotheses**: Some insights enable users to identify a new relevant hypothesis.
- **Directed versus Unexpected**: Directed insights are those that answer specific questions. Unexpected insights are those that were not considered in the design of the study.
- **Correctness**: Insights can be correct or incorrect depending on the data represented in the visualization. Some insights are incorrect conclusions that result from misinterpreting the data visualization. For our study, the insights formulated by the participants need to be clinically valid assessments on the patient's condition.

Researchers stipulate two mechanisms to record the insights, the "thinking aloud process" and the use of a written diary to record the steps taken during the data analysis [24, 25, 26, 27, 80]. Out of these two methods, the "thinking aloud process" required less work from the participants of the experiment and thus allowed the participant to focus solely on deriving insights based on the data visualization. The "thinking aloud process" is one of the most common techniques in usability studies

[81]. It consists of asking the participant to verbally express the thought process while using the system under testing. The insights, comments and other statements can be then captured via audio recording. The recordings need to be transcribed so that they can be assessed based on a defined and established criteria. In this way, the work is shifted from the participant (since there is no longer a need to write the insights in a diary) to the researchers performing the study (transcribe the recordings).

4.3.2 Usability testing methods

Nielsen suggests that "'usability has multiple components and is traditionally associated with the five usability attributes, which are learnability, efficiency, memorability, errors, and satisfaction" [81]. In order to assess the usability of computerized systems, multiple alternatives exist in industry and research.

Usability questionnaires are useful for assessing clinical data visualizations, since they have a high appropriateness ranking [76]. Usability experts can conduct the heuristic evaluation and the cognitive walkthrough; both are recommended techniques to complement the evaluation of a system.

Heuristic evaluation

Heuristic evaluation requires at least one expert in the area of human-computer interaction [76, 81]. Experts in usability conduct the testing using Nielsen's heuristics [81]. The evaluation has 11 metrics that are evaluated using a seven-point Likert scale, value 1 indicates "strongly disagree" and 7 "strongly agree"

Heuristics are "rules of thumb" comprising 10 principles that are meant to assist the human-computer interaction specialist in the usability testing process [76, 82]. The heuristic evaluation principles are described according to Nielsen [82] as follows:

1. *Visibility of the system status*: Refers to continuous feedback on the status of the system "within reasonable time" (Feedback).
2. *Match between the system and the real world*: The use of language should be familiar to the user so that conversations follow a "natural and logical order" avoiding technical terminology unfamiliar to the intended user audience (Speak the User's Language).

3. *User control and freedom* : Allow the user to recover from erroneous navigational options with "clearly marked" access options (Clearly Marked Exits).
4. *Consistency and standards* : Follow the same language and terminology to avoid the user from guessing the meaning of "words, situations, or actions" (Consistency).
5. *Error prevention* : Avoid "error-prone" options in the system whenever possible, and for those cases when the problematic options cannot be avoided, present the user with confirmation dialogues (Prevent Errors).
6. *Recognition rather than recall* : Present visible options to the user at all times so as to avoid the effort of remembering previously stated instructions. Whenever options cannot be visible, make them "easily retrievable whenever appropriate" (Minimize User Memory Load).
7. *Flexibility and efficiency of use* : The interface should accommodate the novice and advanced user by providing "tailored frequent actions" (Shortcuts).
8. *Aesthetic and minimalist design* : The dialogues should only contain relevant and clear information that is needed in a timely manner at that particular state of the interface (Simple and Natural Dialogue).
9. *Help users recognize, diagnose, and recover from errors* : Plain language should be used in error messages, and whenever possible they should provide helpful information so that the users can take constructive actions (Good Error Messages).
10. *Help and documentation* : Some systems require documentation and guidelines to explain briefly how to accomplish specific tasks in concrete steps.

Cognitive walkthrough

Wharton *et al.* developed the cognitive walkthrough for usability testing [83]. Johnson *et al.* summarize this method as a "usability inspection method that compares the users' and designers' conceptual model and can identify numerous problems within an interface" [76, 83].

The cognitive walkthrough has been used successfully to evaluate usability of healthcare information systems [76, 84, 85, 86, 87] and Web Information Systems [88].

Since cognitive walkthroughs "tend to find more severe problems" [76, 89] but "fewer problems than a heuristic evaluation", [76, 90] both methods should be considered for evaluation.

Laboratory testing

Laboratory testing is regarded as the "gold standard" for usability testing [91] when it comes to performing studies in a controlled environment. Laboratory testing collects "qualitative and quantitative" data "since it collects both objective data such as performance metrics (e.g., time to accomplish the task, number of key strokes, errors, and severity of errors) and subjective data such as the vocalizations of users thinking aloud as they work through representative tasks or scenarios" [76].

Controlled user testing comprises "a series of commonly used task scenarios" in which users are asked to conduct these tasks using the "thinking aloud" process [76, 81, 92]. This requires "users to talk aloud about what they are doing and thinking" while they complete the tasks using the system [76, 81, 92].

Usability studies "in the wild" provide higher accuracy since they monitor how users perform their daily activities in reality and not in simulation. However, such studies are difficult to find in literature and are often times costly to implement since they require a usability expert present in the day-to-day activities without interfering with professionals. Johnson and colleagues have stated this as a limitation of the usability testing in a real clinical setting [76].

As the "gold standard" in usability testing, this method has been widely used in evaluating health information systems [76, 93, 94, 95, 96].

Usability questionnaires

Usability questionnaires are "the most common" method to "collect self-reported data" from the "users' experience and perceptions after using the system in question" [76]. Although the data collected is self-reported, some questionnaires have reliability in measuring several usability metrics such as "satisfaction, efficiency, effectiveness, learnability, perceived usefulness, ease of use, information quality, and interface quality" [76].

The Computer System Usability Questionnaire (CSUQ) and After Scenario Questionnaire (ASQ) [97] are recommended for evaluating systems similar to the those

Table 4.2 Standard Questionnaires Table. The table lists the metrics, reliability and length of the Computer System Usability Questionnaire (CSUQ) and the After Scenario Questionnaire (ASQ) used for system evaluation.

Questionnaire	Items	Reliability	Metrics
CSUQ	19	0.93	Usefulness
		0.91	Information Quality
		0.89	Interface Quality
		0.95	Overall Usability
ASQ	3	0.93	Ease of Task Completion
			Time Required to Complete the Task
			Satisfaction

reviewed in this thesis. Table 4.2 shows the length, reliability, and metrics of the questionnaires. These questionnaires use a seven-point Likert scale, where value 1 indicates "strongly disagree" and 7 "strongly agree".

The CSUQ was developed by IBM, and it is a modification of the Post-Study System Usability Questionnaire (PSSUQ) [98]. Table 4.2 shows the reliability of this questionnaire. The questionnaire has a high coefficient alpha with a reliability of 0.95 in total, with 0.93 for system usefulness, 0.91 for informational quality, and 0.89 for interface quality [76, 97, 98]. The questionnaire has been successfully used in the healthcare domain [76, 99] and in the evaluation "of a guideline-based decision support system" [76, 100].

ASQ is an additional questionnaire developed by IBM [76, 97, 101] designed to measure user satisfaction after other usability tests have been completed [76, 98, 102]. This questionnaire measures the "ease of task completion, time required to complete the tasks, and satisfaction with support information" [76]. According to literature review, this questionnaire has not been used for EHR evaluation [76], but researchers recommended it given its properties and appropriateness for use cases related to clinical data visualization where tasks are required to be completed with the assistance of a visualization system.

5 IMPACT OF DATA VISUALIZATION ON THE EFFECTIVENESS OF CLINICAL DECISION-MAKING

From the literature, a total of three review articles have been found that address the prevalence and importance of visualization systems for clinical data, with an emphasis on EHRs.

A survey article published in 2011 identified 14 different articles detailing data visualization tools for EHRs [10]. The survey classifies these articles using two dimensions: representation of single or multiple EHRs, and the type of data represented. The data type can be: categorical, numerical, number of instances, and single or multiple patient representation. The article emphasises the importance of data visualization to enhance the clinical decision-making process and highlights that this is an active and very much needed area of research. The assessment of the articles was not discussed.

Lesselroth and Pieczkiewicz [19] reviewed the existing challenges in utilising existing clinical data to provide better care to patients. The study indicates that data visualizations should help improve the clinical decision making process. The authors conclude that the potential of EHRs has not been realized. In order to do so, multidisciplinary research must address the existing barriers in health informatics. These barriers are the heterogeneous nature of the data, dispersed storage, and the inability to combine the data to better assist clinicians. The review also highlights the need for an objective assessment of clinical visualization tools.

A systematic review conducted by West and colleagues reports on “innovative information visualization” of EHRs [103]. The review focuses on the visualization technique used to deal with heterogeneous data. The methodology of the review is the same as used in this one. The review reports an increasing trend in “innovative” visualization as a natural consequence of the increase in clinical data.

These reviews are not conclusive for the purpose of this thesis, given that they do no study the reasoning derived from the visualization and how this affects the clinical decision making process at an individual level. Therefore, a systematic review is needed to address this problem.

5.1 Preferred reporting items for systematic review and meta-analysis

This review reports on the use of assessment methods in clinical decision support systems that rely on data visualizations. This review follows the preferred reporting items for systematic reviews and meta-analysis (PRISMA) [104]. This review is not constrained by publication year, all articles that fit the search criteria are considered.

The literature search was conducted in July 2019 using PubMed, Web of Science and Scopus. PubMed is a free search engine maintained by the United States National Library of Medicine (NLM) at the National Institutes of Health, and it provides access to MEDLINE, Science, and British Medical Journal, among others. MEDLINE is a bibliographical source for medical articles. Web of Science, previously named Web of Knowledge, has access to seven different databases covering several disciplines and conference proceedings. Elsevier's alternative is called Scopus, which covers more than 34,346 peer-reviewed journals from multiple disciplines.

Additional sources were obtained from the bibliographical references of articles found in this review. This was done to complement the review process and also to include other research work previously covered by other surveys, thus extending this review with additional literature. Table 5.1 lists the criteria and keywords used to perform the search queries in the databases.

5.2 Inclusion and exclusion criteria

The articles must utilise clinical data, that is to say, data used in the context of healthcare. The most common variety is EHR or personal health records (PHR), as it is also referred to in the literature. The emphasis of the article has to be on the visualization of clinical data and how that affects the decision-making process. The

Keywords	Criteria
"health record" OR "clinical data"	Include all articles containing the terms "electronic health record" or "personal health record". In the event that these terms are not used, search for the term "clinical data" to cover all other cases.
visual*	Include all articles discussing any form of visual representation.
usability OR test* OR compar* OR assess* OR study	Include terms such as assessment, study, testing, comparison and usability studies.
"decision support"	articles must make mention of this term as it provides the context in which the clinical data visualization is being used.

Table 5.1 The table shows the four keyword terms used to compose the queries to search for articles. The Boolean operator "AND" was used to combine these terms. Within these terms, the Boolean operator "OR" accounts for similar terms representing roughly the same concept.

visualization should be at the centre of the system that aims to assist in the understanding of the data and the decision making process. The articles should explain the decision making process they aim to support, or provide arguments to substantiate that the visualization is in fact a supportive tool for the decision making process.

Articles must contain figures of the visualization under study. Some studies deal with genetic data, primarily used for classification, exploration and analysis, however they were excluded because they did not support the decision-making process nor did they use clinical data in the context of healthcare. They rather used genetic data to discover answers to research questions not directly related to patient care.

Articles detailing the implementation and computational performance of the data visualization tool were not included because they did use the visualization to assist in the clinical decision making process. Articles that underline current research trends in visualization and decision making in the healthcare context were also excluded.

Surveys, position papers, and articles discussing the importance of clinical data visualization or decision support systems were excluded, as they did not discuss a specific tool and how it supports the clinical decision making process. Articles that conducted assessments on the computational performance or algorithmic complexity of the rendering process of the data visualization were also excluded, as these fall outside of the objectives of this thesis.

5.3 Selection and analysis

The output of the search queries from the databases were exported into a spreadsheet. The information collected included the authors, DOI, year of publication, journal or conference, title and abstract. The first classification of the findings was to determine whether the articles reported any form of assessment of the data visualization. These criteria served as the first step to filter the articles relevant to this review process.

After reading the abstracts, the articles were included or excluded from the review process. The remaining selection was later screened by reading the full article. The main goal was to study the nature and prevalence of assessment methods of clinical data visualizations in the decision making process.

5.4 Search results

A combined total of 228 studies were identified from the electronic databases of PubMed, Scopus, and Web of Science. An additional 28 studies were added to the catalogue by manually reviewing references in the studies. After removing 82 duplicate entries, a total of 174 abstracts were screened manually.

A total of 136 studies were excluded because they were position papers or recommendations on visualizations, did not provide decision support, did not deal with clinical data, did not use data visualization, dealt with policies in healthcare, dealt with genomics data, discussed the importance of visualization in healthcare information systems, described the need for investigation of better visualization techniques, or discussed the potential of clinical data in diagnostics.

The remaining 38 studies were fully read. From these, a total of 19 did not report any form of assessment, 3 did not present the visualization under study, 5 did not explain how the visualization can be used in the decision support process, and 1 was a set of recommendations on experimental design to assess visualization outside the clinical domain. The result of the filtering process for the review is shown in figure 5.1 as a flowchart following the PRISMA approach.

A study of computerized physician order entry (CPOE) shows that a dashboard implementation received positive feedback from nurses at a hospital in Singapore [105]. The dashboard implementation is a software that operates on top of the existing CPOE system. The visualization is rather simple and shows only visual cues for

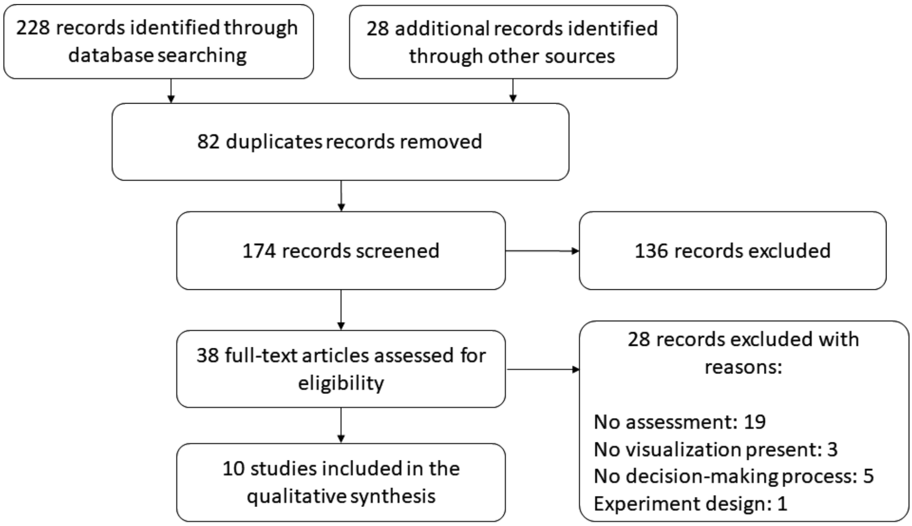


Figure 5.1 The flowchart representing the steps taken for the literature selection, as recommended by PRISMA [104].

alerts, as illustrated in figure 5.2. A total of 106 nurses responded to a questionnaire. The metrics obtained from the questionnaire were overall user satisfaction, usage frequency, system quality, system information quality, impact on work efficiency, and impact on care quality. Mean user satisfaction resulted in 3.6 out of 5 points on a Likert scale. Overall, the nurses expressed a favourable response to the dashboard reducing the cognitive overload.

Rayo and researchers [106] compared two mechanisms to alert physicians to potential problems when prescribing diagnostic imaging studies. They compared the prevalent alert system using textual alerts in the form of pop-ups in the medical software against dynamically annotated visualizations (DAVs). They invited 11 clinicians to take part in the study. The study was designed by the researchers, and it consisted of a series of questions on whether the patient should continue, cancel, or change the recommended treatment. The researchers found a reduction from 34% to 18% of "inappropriate diagnostic imaging tests" when the clinicians used DAVs. The visualization used in the study consisted of visual cues [105]. These visualizations were rather simplistic, as they aimed only to show potential risks as graphical cues.

Franklin and colleagues [108] implemented a software that collects EHRs and presents patients with a personalized view of risks for treatments. The software uses

Figure 5.2 The dashboard implementation used in the study by Tan and colleagues[105], Open Access [2013] IOS Press.

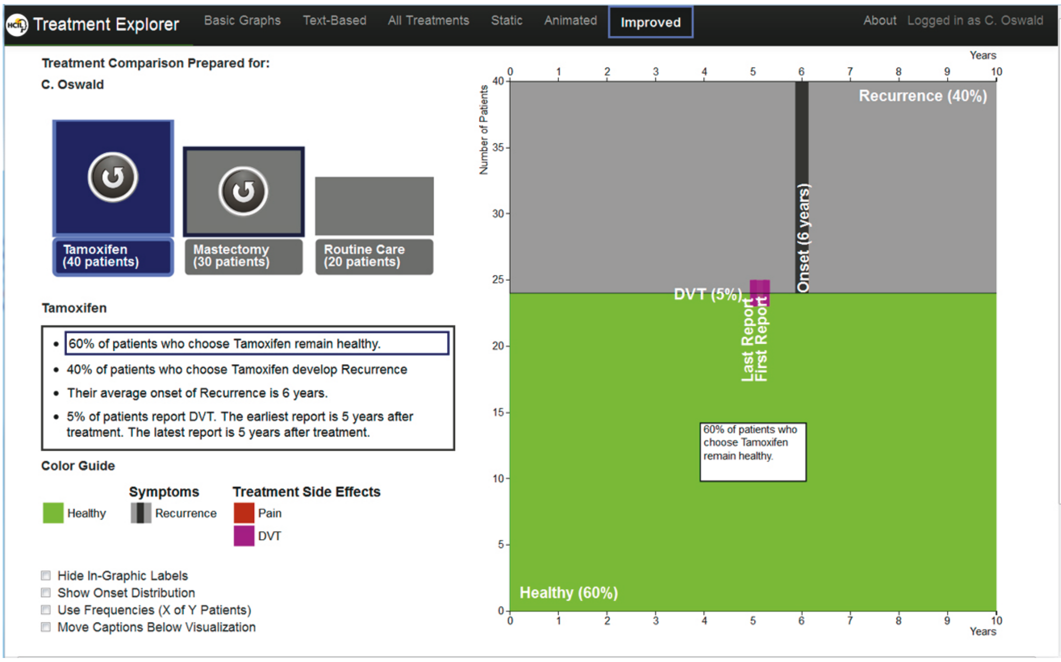


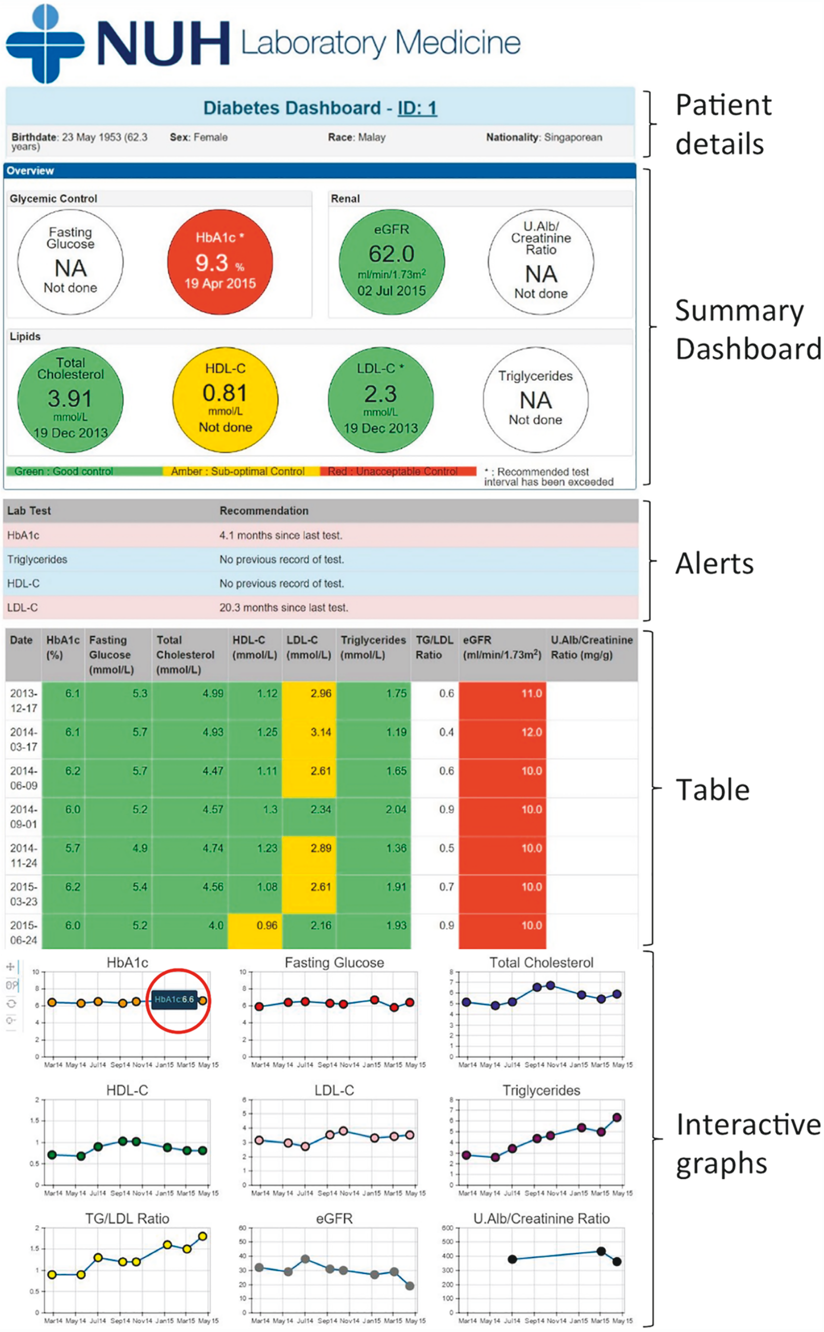
Figure 5.3 The overview of Treatment Explorer showing the distribution of patients that underwent a specific treatment as well as the summary of outcomes for the treatment [107]. Treatment Explorer Interactive Decision Aid for Medical Information [2012] `https://perma.cc/H7PW-E35D`

narration, animations, and visualizations to give patients the best possible information regarding treatment risks. The software was named Treatment Explorer. Figure 5.3 shows the view of a recommended treatment for a patient, summarizing patient outcomes graphically and textually. The software was deployed and tested in three different stages. In the first stage, domain experts from risk analysis and healthcare reviewed the software and provided feedback. The second stage was a user study with 24 participants. The participants used a text-based alternative for risk assessment, a prototype and a full-featured version of the software that included animations. The users were asked 10 questions to which the responses were either correct or incorrect. The third stage consisted of 42 participants answering a similar 10-question test. The results show that the number of incorrect responses was fewer with the full-featured software (from 1.21 to 0.31). The researchers close the study with a set of recommendations for future systems that deal with a similar goal. These include the use of narration, and animations, and allowing the user free control over the time flow of the risk assessment.

A Diabetes dashboard software was implemented by Sim and colleagues [109]. The dashboard featured combined data from EHRs to assist the user in diabetes management. The dashboard featured a visualization of recent test values for lipids, renal function, and laboratory tests. Figure 5.4 shows the dashboard overview. This visualization used colour-coded cues to stress critical values that require attention. A study was designed to evaluate the dashboard compared to the existing system. The study featured a scoring mechanism to determine if the users were able to correctly assess and take adequate action in the management of the disease. The questionnaire was distributed via an online system that reported the results back to the researchers (findsurveys.com). The researchers report that the dashboard "significantly improved the identification of abnormal laboratory results, of the long-term trend of the laboratory tests, and tests due for repeat testing". The dashboard was not substantially better than the existing solution in identifying patients that needed "treatment adjustment".

Forsman and researchers [110] report a software that visualizes clinical data from multiple sources, assisting with intensive-care antibiotic treatment. The visualization aims to provide a "holistic overview of integrated information" required to select the appropriate antibiotic. Graphical cues in the form of "color patterns" were used to indicate "intervals of chemical values and antibiotic treatment". Figure 5.5 shows the holistic overview for antibiotic prescriptions. A study was conducted with 10 physicians and 2 usability experts. The study consisted of 15 tasks that measured "performance time, navigation paths and accuracy". The completion rate for the tasks was 79.4%. A system usability scale (SUS) questionnaire was also used for the evaluation [111]. The score for the SUS questionnaire was 79.5%, which is considered favourable. The software tested in this study was not compared to other solutions.

A patient portal was implemented by Fraccaro and colleagues [112]. The objective was to assist patients in understanding laboratory results so that they can eventually self-manage chronic kidney disease. The patient portal was evaluated using three presentation formats. The first was a numerical and textual presentation of the results, the second presented contextualized results (reference values) and the third presentation grouped similar values into categories. Each presentation used single and reference graphs. A total of 20 participants took part in the study. The researchers asked participants to self-evaluate their health literacy [113]. Laboratory



Date	HbA1c (%)	Fasting Glucose (mmol/L)	Total Cholesterol (mmol/L)	HDL-C (mmol/L)	LDL-C (mmol/L)	Triglycerides (mmol/L)	TG/LDL Ratio	eGFR (ml/min/1.73m^2)	U.Alb/Creatinine Ratio (mg/g)
2013-12-17	6.1	5.3	4.99	1.12	2.96	1.75	0.6	11.0	
2014-03-17	6.1	5.7	4.93	1.25	3.14	1.19	0.4	12.0	
2014-06-09	6.2	5.7	4.47	1.11	2.61	1.65	0.6	10.0	
2014-09-01	6.0	5.2	4.57	1.3	2.34	2.04	0.9	10.0	
2014-11-24	5.7	4.9	4.74	1.23	2.89	1.36	0.5	10.0	
2015-03-23	6.2	5.4	4.56	1.08	2.61	1.91	0.7	10.0	
2015-06-24	6.0	5.2	4.0	0.96	2.16	1.93	0.9	10.0	

Figure 5.4 An overview of the diabetes dashboard use in the study by Sim and researchers [109]. Creative Commons [2017] `doi.org/10.1371/journal.pone.0173021`

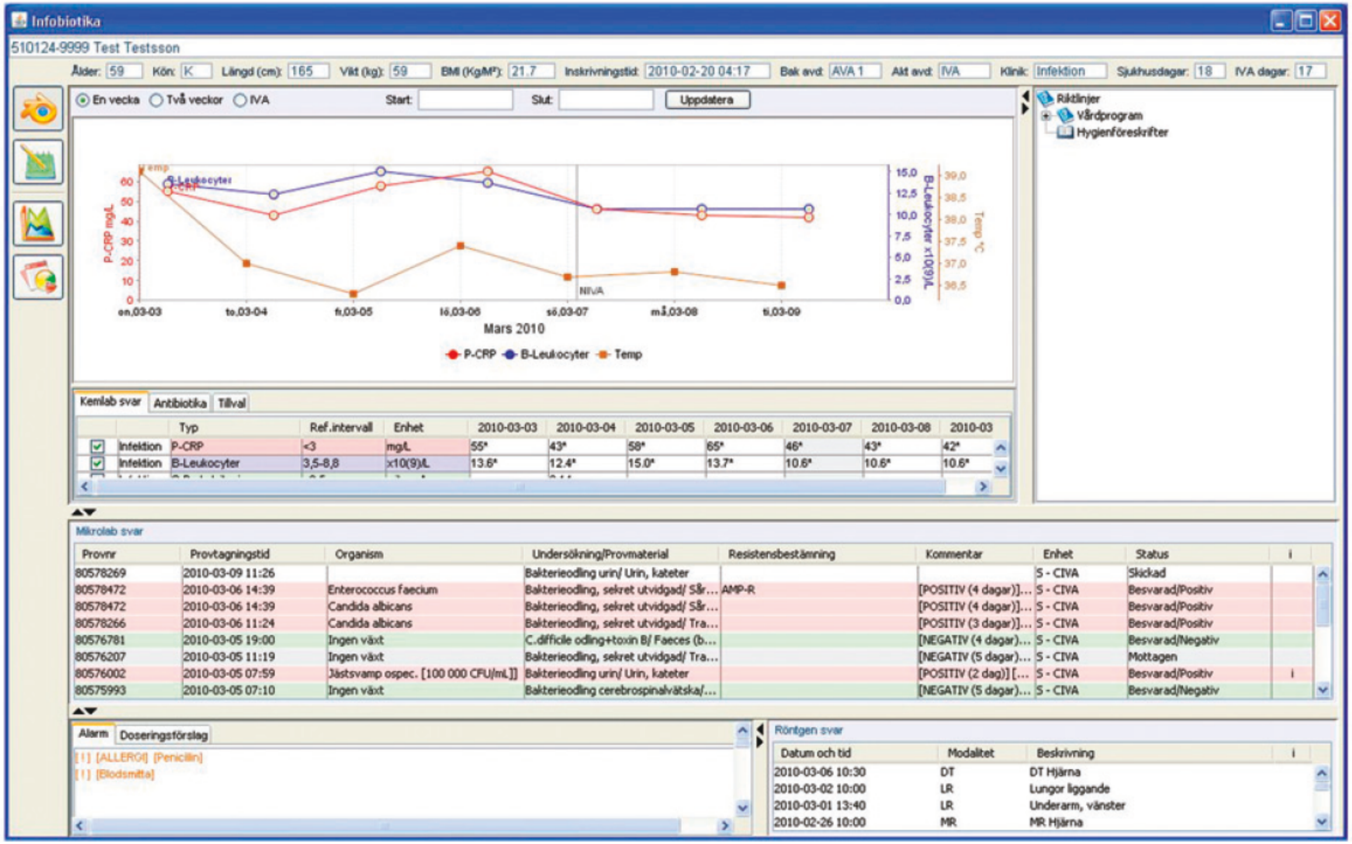


Figure 5.5 The holistic overview that integrates the information required to prescribe antibiotics [110] © [2012] Taylor & Francis.

results were shown to the participants in one of the three formats. The participants were asked to assess how serious the condition of the patient was. They had to select one of three actions, call a doctor immediately, schedule an appointment within the next 4 weeks, or wait for the next check-up in 3 months. A scoring system was used to assess the correctness of the answers. In the results, despite the visualization enhancements, 65% of the participants underestimated the need to take appropriate action. Medium-risk cases were "particularly difficult" to understand. The study concludes that further work is needed to facilitate risk analysis for patients with chronic conditions.

Sim•TwentyFive was reported by Stubbs and researchers as clinical software to compare and analyse patients with similar physiological conditions [114]. The software employs web technologies to run comparisons and visualizations across multiple platforms. The visualizations consist of scattered plots to identify similar patients, and longitudinal physiological data. Computational performance and opinion surveys were used to evaluate the software. The survey was answered by physicians via email invitations. The physicians found the software useful for their clinical practice. No additional assessment was conducted to test the performance, correctness, or analysis and reasoning of the software.

Faiola and colleagues [115] developed and tested a software ICU monitoring of biometrical data. The software was designed using a human-centred approach. The software, named Medical Information Visualization Assistant version 2 (MIVA 2), aimed to reduce the cognitive load on clinicians and assist in better decision-making. The software visualizes various biometrical and longitudinal data arranged as a dashboard. A total of 12 participants took part in the study, comprising nurses and physicians. Fictitious clinical data was used to evaluate MIVA 2 against a baseline representation. The study consisted of a questionnaire and briefing interviews that measured performance, accuracy, context of use, and usability. The first part was a performance test with a set of questions that the participants had to answer. The second part assessed the accuracy of the decisions made by the participants. The results show that users performed faster with MIVA 2 (3.11 seconds) than the baseline (3.65 seconds). They also had increased accuracy from 0.58 to 0.63. Overall, the users were satisfied with the use of MIVA 2.

Gotz and Stavropoulos addressed the challenge of multi-variable temporal data with their software implementation, DecisionFlow [116]. The software aims to assist users in analysing "high-dimensional temporal event sequence data", such as EHRs. The researchers defined a structured approach to treat data as "milestones" organised in sequences, similar to a graph visualization. Figure 5.6 shows the summarized dashboard view of a patient rendered by the software. A study with 12 users was conducted to evaluate DecisionFlow. The study consisted of a questionnaire to evaluate understanding of the events depicted in the graph visualization. The data used was a collection of medical events of anonymized patients. The focus of the study was on measuring accuracy and speed. There was an accuracy of 98% among the participants. The speed at which the participants completed the task varied from 1.7 to 23.6 seconds. The researchers state that these times are considered good values.

Lin and researchers evaluated a paediatric intensive care software called T3 [117]. The software comprises a central visualization and three secondary ones. The central visualization is a combined set of line graphs representing longitudinal physiological signals. The secondary visualizations consist of visual cues for "out-of-target values", sparklines to indicate trends, and a bar graph that uses a 16-parameter proprietary algorithm to estimate "the risk of inadequate oxygen delivery". A human factors usability study was conducted with seven physicians, eight nurses, and seven respiratory therapists. The data used for the study came from a clinical database of newborn cases of post-cardiac surgery. The data was anonymised for the study. The participants were invited to a usability laboratory, and the sessions were recorded. Participants were asked to use the "thinking aloud process" to record their thought process while using the software. Participants were asked to conduct 20 tasks with the software. The tasks involved typical actions and steps required to analyse the data. Usability issues were discovered with the software. The completion rate for the tasks was 88%. The user error rating was 1.3 for clinicians and nurses, and 1.2 for therapists. The researchers state that despite the usability problems, the software was able to assist users in understanding and interpreting the results.

5.5 Summative analysis

A total of four studies used longitudinal data visualizations to help in the clinical decision making process. Three studies focused on patient self-care. Two studies addressed health monitoring, and one the risk analysis of prescribed treatments. Seven studies addressed the decision making process from the perspective of clinicians and

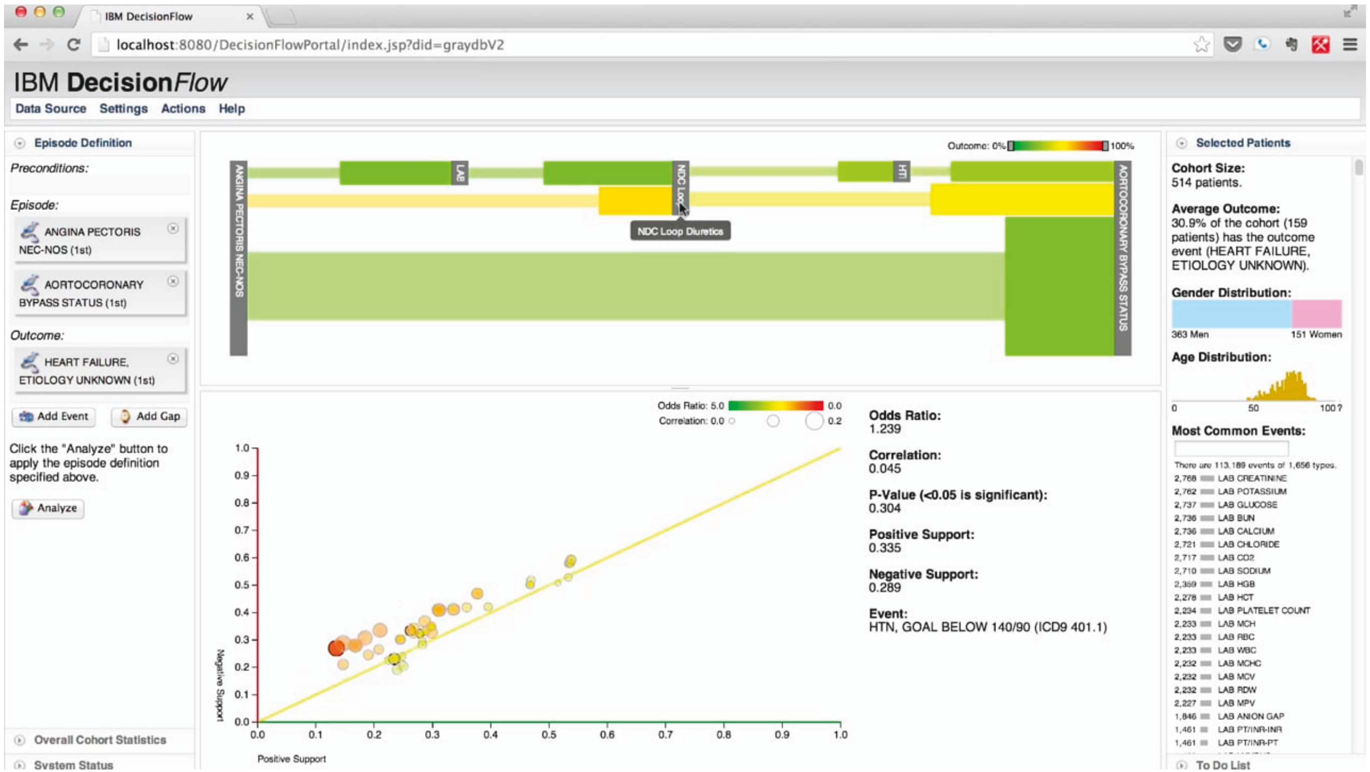


Figure 5.6 The DecisionFlow software summarizing medical data into a dashboard view [116] © [2014] IEEE.

nurses. Two studies focused on ICU monitoring, one on patient monitoring, two on the overall patient condition, one on recommending antibiotic treatment, and one on prescriptions for imaging studies. Four studies dealt with data obtained from laboratory tests as well as physiological measurements.

Half of the assessments made a comparison study between the described system and a baseline representation. All these comparative assessments used questionnaires with a scoring system. Only one of them conducted briefing interviews with the participants. The second half of the assessments did not compare the system to any alternative. From these studies, two relied on usability techniques, including the use of system usability scale (used to collect user feedback). The remaining three, used a set of tasks that the participants performed during the assessments. Only one study used the "thinking aloud process" to document the reasoning behind the actions of the users.

5.6 Discussion and conclusion

The majority of the assessments reported in the articles, fall under the UP scenario classification. Table 5.2 shows the number of studies classified according to the scenarios by Lam and colleagues [23]. This shows an emphasis on performance and user experience among the reviewed studies. The studies aimed to establish that the proposed software increased the accuracy of decisions. A lower number of studies paid attention to the analysis and reasoning obtained from the data visualization and how this in turn affected the clinical decision making process. Some of the studies conducted more than one assessment.

Scenario	Number of studies
Evaluating visual data analysis and reasoning (VDAR)	2
Evaluating user performance (UP)	9
Evaluating user experience (UE)	4
Evaluating visualization algorithms (VA)	1

Table 5.2 The number of studies found in the literature review classified by the scenarios proposed by Lam and colleagues[23]

The articles acknowledge the importance of meaningful visualization techniques to enhance the decision making process. Only a small portion of them conducted assessments to determine how the visualization, as part of the system, can assist users in making better clinical decisions. Figure 5.7 shows the distribution of publications found in this literature review. Before 2012, the studies did not conduct any assessments. After that year, it became increasingly common to publish studies that included any type of assessment.

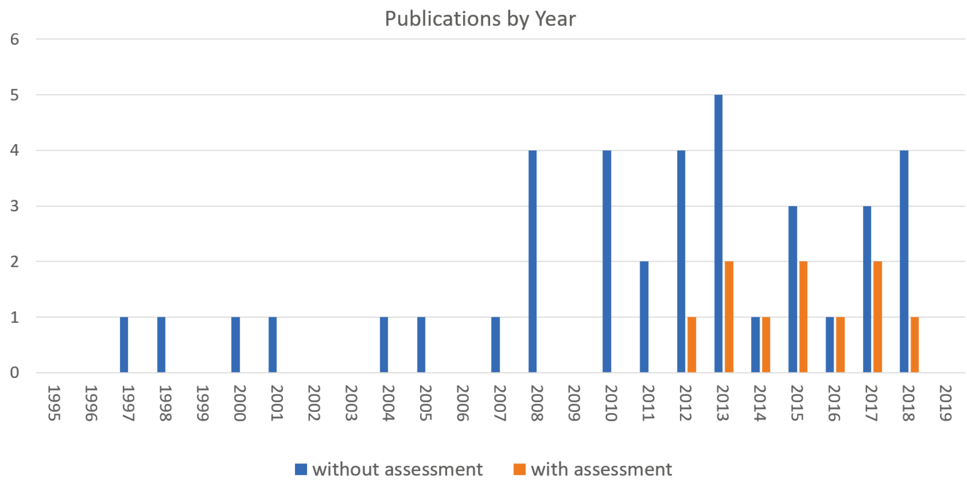


Figure 5.7 The distribution of articles by year, with and without assessments

The articles identified in this survey did not compare assessment methods. No rationale was presented to justify the selection of an assessment methodology. Therefore, challenges remain in existing studies on how to select the assessment method based on the goals of the study. As a first step, the relevant metrics and goals of the study should be defined in detail.

The use of a visualization technique was necessary for an article to be included in this literature review. It is possible that some articles utilized a data visualization but it was not stated in the manuscript. As a result, these articles were excluded from the literature survey.

Although numerous studies have been done that use visualization techniques for clinical decision support, those that did not report any assessment in the abstract were excluded from this review. It might be possible that some studies did not emphasise the assessment, and thus it was not described in the abstract.

The purpose of this review was to determine the prevalence and nature of assessments in decision support systems that rely on clinical data visualizations. The review shows an increase in the use of assessments in reported studies. Table 5.2 shows that only two articles conducted assessments that fall under the VDAR scenario, these are the contributions made by Fraccaro and colleagues [112] as well as Sim and researchers [109]. The gap that the thesis aims to bridge is to develop a methodology that allows the assessment of clinical data visualizations in terms of their efficacy in supporting clinical decision making. Looking further into these two publications, the VDAR assessments consisted of questionnaires designed by the researchers. The questionnaires were made with the intention to capture the ability of the participants to "assess and take adequate action" given the data visualization. The intended design of the experiment was to study the visual reasoning however, the methodology followed in these two publications was defined by the authors. Ideally, the researchers could have used an existing methodology, such as the insight-based, to further study the visual reasoning.

After screening 174 records and identifying 2 instances in which a VDAR evaluation was done, it can be concluded that the evaluation of visual reasoning in the clinical decision making process is uncommon. Even considering these two instances, the experiment design did not follow a previously tested methodology, nor was the experiment design justified or compared to other alternatives.

The focus of the assessments in the rest of the articles is on user performance, meaning the number of correct answers in a given questionnaire. It remains increasingly rare to focus on the reasoning process that leads to the decisions being made. Future studies could aim to provide examples on how to study the reasoning behind the clinical decision making process. The usage of suitable assessment methodology could provide guidance for future studies. With demonstrable evidence that a certain visualization improved the clinical decision making process, better patient outcomes could be expected.

5.7 Clinical data visualization systems in industry

Prominent companies like GE Healthcare and Philips have developed their own commercial solutions for data visualization that support the clinical decision making process.

GE Healthcare offers the Advanced Visualization AW solution for healthcare professionals which focuses primarily on imaging and processing of CT and MRI

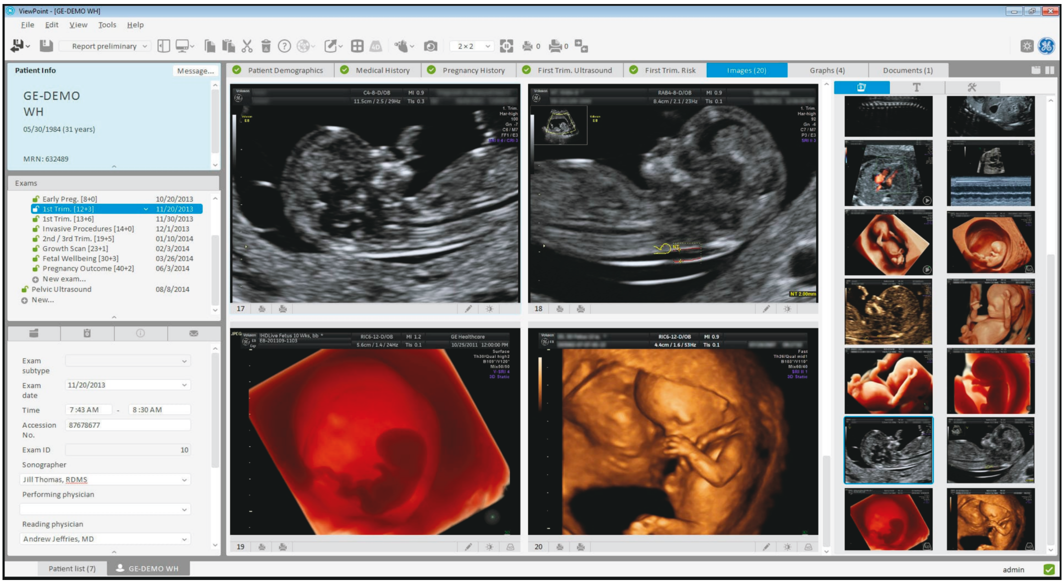


Figure 5.8 A screenshot of the ViewPoint 6 software developed by GE Healthcare to visualize and organize ultrasound imaging. Image obtained from GE Healthcare ViewPoint 6 specification website [118]. `https://perma.cc/P7GT-GXCS`

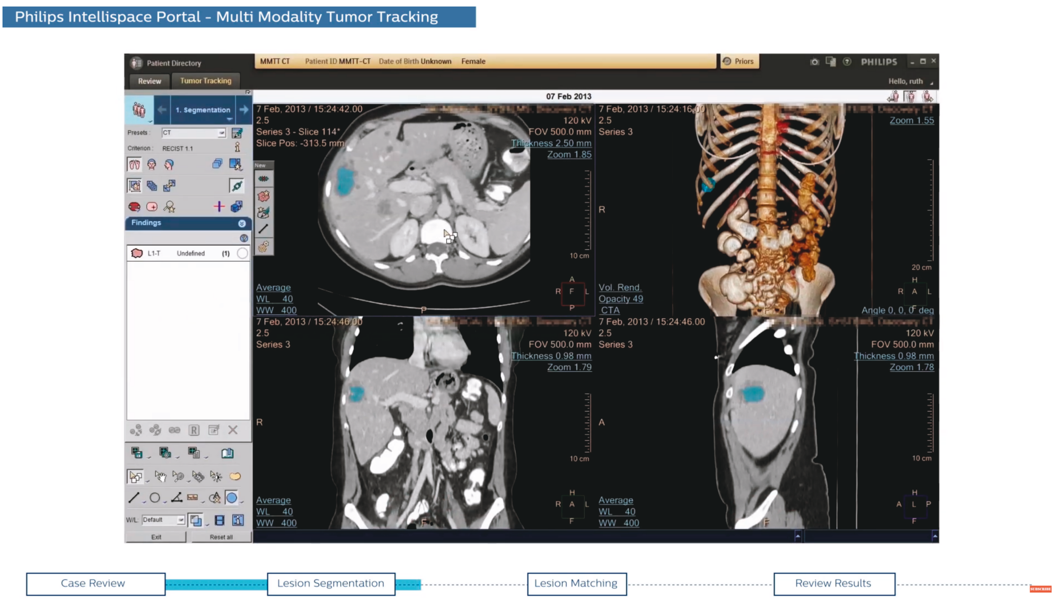


Figure 5.9 The Philips IntelliSpace Portal showing imaging studies in multiple modalities. Image obtained from Philips Healthcare YouTube presentation. `https://perma.cc/8JUD-D8H9`

scans. The solution was designed to assist clinicians in the decision-making process by reconstructing 3D models of the patient imaging and highlighting areas in the studies that are of particular interest for professionals [119]. Another visualization solution developed by GE Healthcare is ViewPoint 6 which is an ultrasound visualization software that also handles the reporting aspect [118]. Figure 5.8 shows a screenshot of the ViewPoint 6 software from GE Healthcare specification website.

Philips developed the IntelliSpace Portal which also aims to provide a software for managing and exploring imaging studies from patients. The software also enables collaborations of image exploration and imaging data centralization [120]. The software features image evaluation, access to different modalities, cardiovascular imaging analysis, quick browsing through patient cases and pulmonary care detection. Figure 5.9 shows a screenshot of the IntelliSpace Portal.

6 ASSESSMENTS OF DATA VISUALIZATIONS FOR CLINICAL DECISION SUPPORT

6.1 Research methodology of the studies

The research conducted in the studies is mixed because it aims to gather data based on the experience of making sense of clinical data to make decisions (qualitative), and to gather numerical data for ranking and categorization for statistical analysis (quantitative). The goal of these experiments is to gain knowledge by observation, recording and analysis of the gathered data to further understand how visual reasoning is affected by visualizations and how that in turn affects the decision making process.

The collection and ranking of insights falls under quantitative research, as well as the comparison between visualization systems via the metrics defined in the insight-based methodology and the scoring system of the usability studies. The formulation of the criteria for the assessment of insights and their classification into hypotheses is qualitative because it studies the previous insights and behaviour of the participant during the data exploration and reasoning via the "thinking aloud process". To capture the thinking process, it was essential to understand the evolution of the data analysis of the participants during the assessment sessions. A mixed research approach enables the study and comparison of different visualizations and how they affect the visual reasoning process mentioned by Pirolli and Card [22] and studied by VDAR evaluations [23].

6.2 Comparison of five visualizations (Publication I)

A study was conducted to compare five visualizations using insight-based methodology. At the time of publication, no other article had applied this methodology in the context of clinical data. The purpose was to understand how different visualizations have an impact on the way people understand health and how this affects in turn, the clinical decision making process. The visualizations used in the study were based on the framework developed by Cleveland and McGill [32]. Additionally, a radar plot visualization technique called "hGraph" was also included in this study. Figure 6.1 illustrates the figures used in the study.

6.2.1 Data and assessment criteria

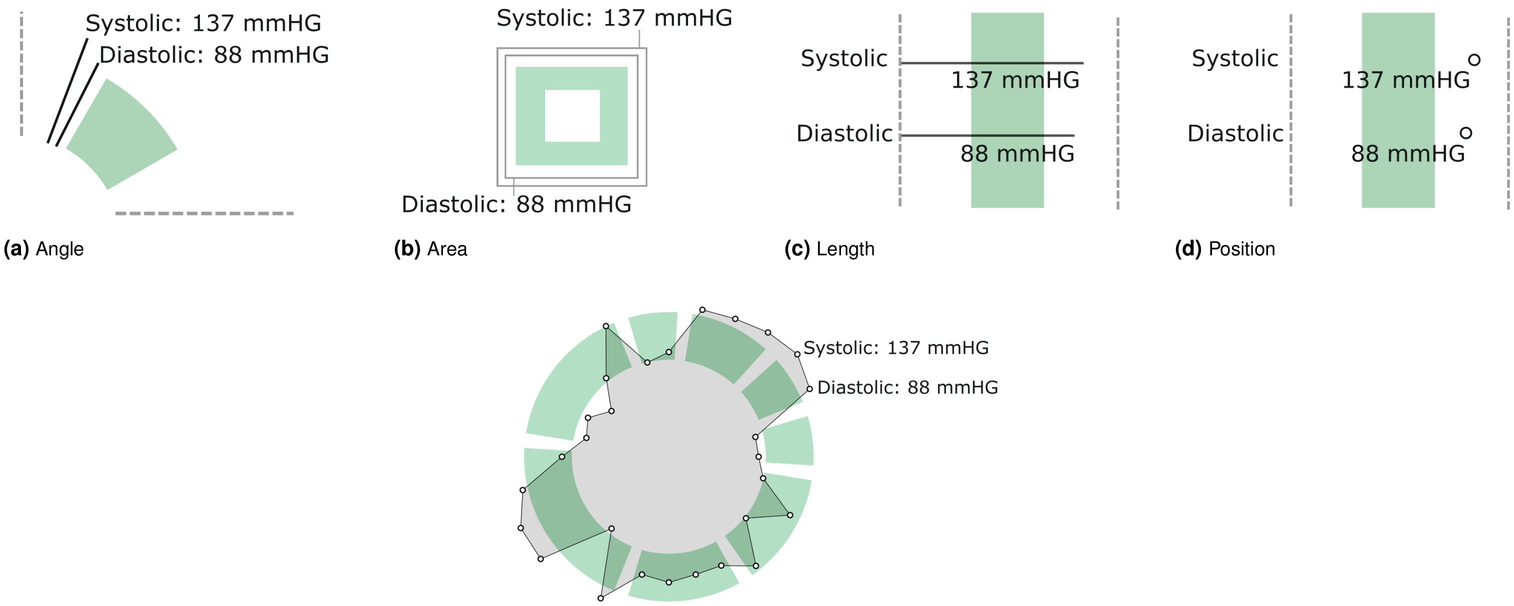
Two data sets were modelled to describe different health situations. The first data set modelled a person with a poor overall health condition and a clear indication of a metabolic syndrome. The modelled data included a high risk of developing Type 2 Diabetes (T2D), based on the calculator developed by Lindström and Tuomilehto [121]. The second set modelled a person with a good health condition and a healthy lifestyle.

Healthy ranges for each parameter, derived based on the national clinical health recommendations [122], were also included in the data sets for comparison purposes.

Depth and complexity [24] were used to establish the criteria to evaluate the insights. Similar to previous insight-based studies [25], a five-point Likert scale was used to determine the value of the insights. Clinically incorrect insights did not receive any points. The insight value was given by a physician following the above-mentioned criteria.

6.2.2 Experiment protocol

A total of 15 female and 15 male participants from 9 different countries took part in the experiment. A self-assessment of the participants' knowledge on health and wellbeing was conducted. The participants were also asked how familiar they are with the risks of developing T2D. Block randomization was used to assign one data representation to each participant. The participants were asked to "see beyond the



(a) Angle **(b)** Area **(c)** Length **(d)** Position

(e) hGraph

Figure 6.1 The blood pressure measurements represented using the angle, area, length and position. The hGraph representation of the data shows only the labels for blood pressure. The green areas represent the recommended range of values for the blood pressure measurements (acceptable minimum and maximum).

figures" and explain the relationships between the measurements using the "thinking aloud" process for a total of ten minutes. The participants were engineering students and members of the staff from a technical university without a degree in the field of medicine.

6.2.3 Results

The *area* visualization had the least correct assessments. The *angle* visualization had all assessments completed correctly. The *hGraph* had four correct and one incorrect assessment. These two visualizations supported participants with low understanding of health literacy in assessing the dataset correctly, as compared with *length* and *position*, which did not support participants with low health literacy.

Table 6.1 The results of the study showing the average time to the first insight of value three or more, average health literacy, and number of hypothesis in total, and per participant.

Visualization	Time(s)	Health Literacy	Hypothesis	
			Total	Per participant
Angle	98	3.2	4	(0,0,2,1,1)
Area	144	3.0	3	(0,0,2,1,0)
hGraph	128	3.2	6	(2,0,1,2,1)
Length	139	2.0	3	(2,0,1,0,0)
Position	57	2.6	4	(0,1,1,1,1)
Table	141	3.2	4	(0,0,1,0,3)

The longest time to the first insight of value three occurred using the *area* and in the *table*. The *position* and *hGraph* enabled the generation of a hypothesis in four out of five participants. An insight of value four or five is considered a hypothesis. Table 6.1 summarizes the insights recorded for each visualization.

6.3 Wellness dashboard and health figures (Publications II and III)

The encouraging results obtained with the hGraph in the previous study, provided the motivation to conduct a second study. A new web library was developed, inspired by the hGraph. The library was released as open source software under the name of Health Figures or hFigures. This library took into account the feedback received by hGraph users and extended its capabilities.

To measure the effectiveness of hFigures, a wellness dashboard was implemented, and a usability study was conducted. Usability methods were used in this study to try to understand how they can measure the effectiveness of a visualization. The study was divided into two publications. The first was a conference publication on the usability study of the dashboard as a tool to support understanding of data representing health and wellness. The second was a journal article on the design and implementation of the hFigures library and its evaluation as a key component of the dashboard.

6.3.1 hFigures

The implementation of hFigures followed the Extreme Programming methodology [123, 124]. During the implementation of the library, the author's research group provided the continuous review process for the software. The research group has expertise in health sciences, signal processing, user design, software engineering and machine learning.

In contrast with hGraph, hFigures does not calculate aggregated measurements that could mislead users into ignoring values outside the recommended ranges that could be skewed. Instead, hFigures hides the labels of individual measurements and shows grouped sectors as categories to avoid clutter and maximize scalability. hFigures is illustrated in Figure 6.2.

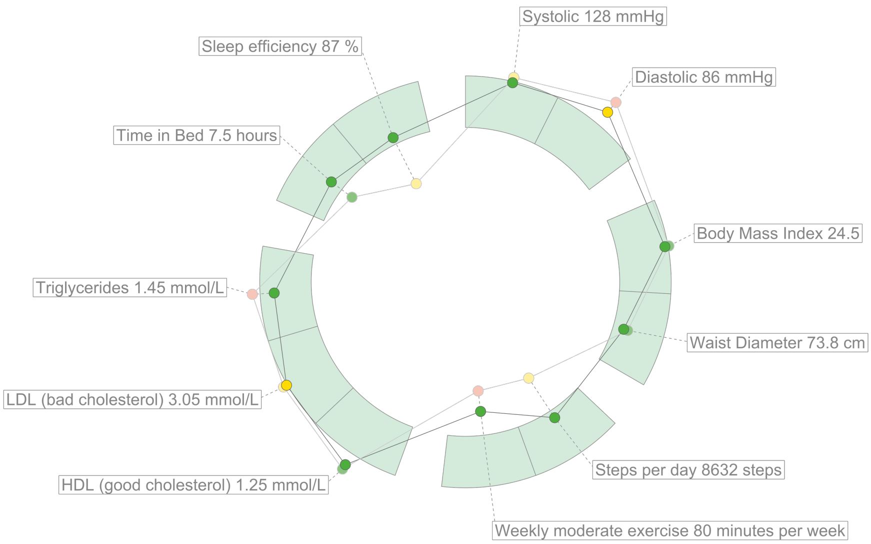


Figure 6.2 The hFigures representation of two different states of the same patient. hFigures does not aggregate the measurements; instead they are separated by groups in sectors.

6.3.2 Wellness dashboard

A wellness dashboard was built to depict the evolution of the previously modelled patient at risk of developing T2D. The patient went through a health coaching programme and as a result, the overall health condition improved.

A timeline shows a series of activities that continuously improved the overall health situation. Users could delineate two points in time to compare the two hFigures as means to understand the evolution of the patient.

6.3.3 Experiment protocol

Previous studies have invited a total of 14 participants to take part in the usability testing [76, 81, 125, 126, 127, 128, 129, 130]. Using that number as a reference, the same number of participants were invited to take part in this study. Out of these, a total of three were usability experts.

The usability questionnaires were filled out using the web portal developed by Perlman [131] and available at the following address `http://garyperlman.com/quest/`.

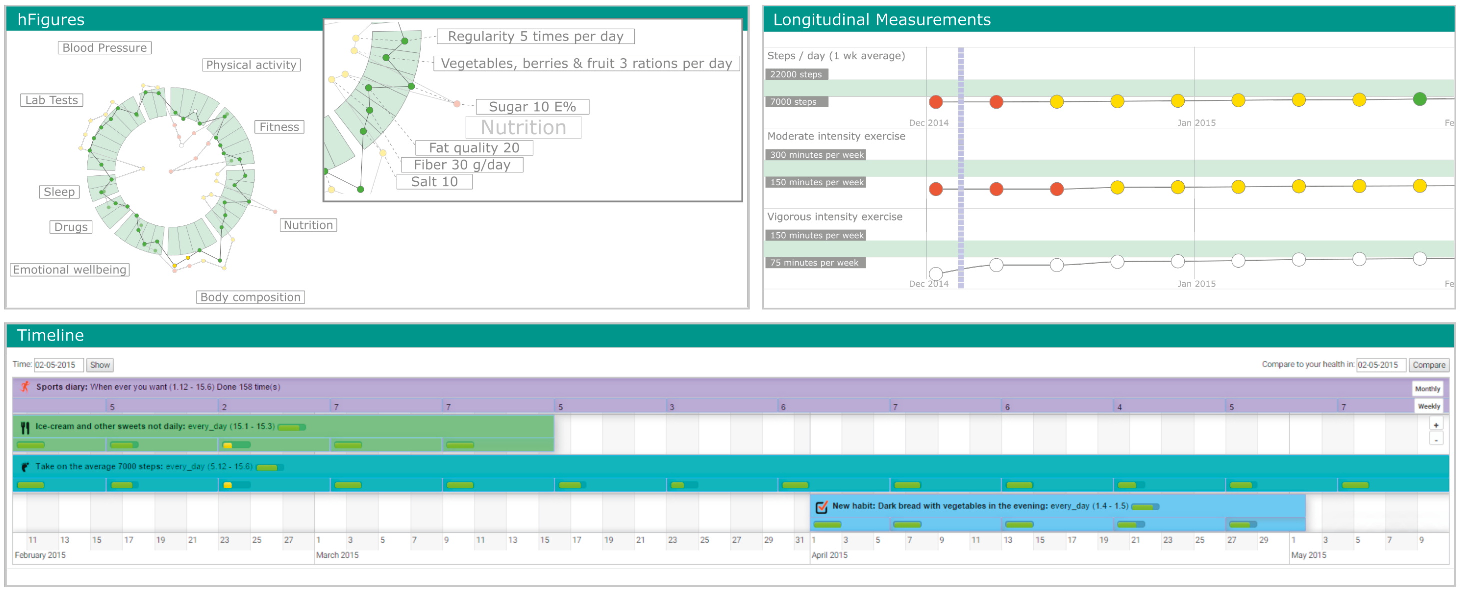


Figure 6.3 A screenshot showing the implemented wellness dashboard with the hFigures, a timeline of coaching activities, and the longitudinal values of the measurements.

The usability experts conducted a heuristic evaluation and a cognitive walkthrough. All the participants conducted 11 data analysis tasks that required the use and exploration of the wellness dashboard. All the participants answered the CSUQ and ASQ.

6.3.4 Results

The average response to the heuristic evaluation was 6.3 out of 7 points. The questions comprising the walkthrough, as described by Wharton *et al* [83]., were correctly answered by the expert users, and thus no design or mismatch errors were found.

All the participants completed 7 of the 9 tasks. The most problematic was to identify the different categories of measurements in the segmented areas of the graph. Steps were taken after the assessment to improve the library using the feedback obtained from the study.

The results for the questionnaires were computed to obtain the overall usability rating of the system as recommended by Lewis [132]. Table 6.2 summarizes the CSUQ results. The dashboard obtained an average of 6.13 out of 7 points in system usefulness, 5.66 in information quality, and 6.24 in interface quality. The score for overall usability is 6.02 with a standard deviation of 1.04.

Table 6.2 The CSUQ results for Overall Usability, System Usefulness, Information and Interface Quality.

Metric	Questions	Average Response	Standard Deviation
Overall Usability	1-19	6.02	1.04
System Usefulness	1-8	6.13	0.93
Information Quality	9-15	5.66	1.20
Interface Quality	16-18	6.24	0.99

Table 6.3 summarizes the results of the ASQ. The average response in the ASQ for the ease of task completion was 6.64 with a standard deviation of 0.842, and for the time required to complete the task it was 6.64 and a standard deviation of 0.497. The overall satisfaction was 6.46 with a standard deviation of 0.53.

Overall, the wellness dashboard proved to be a valuable tool to understand the patient's condition and evolution over a period of time.

Table 6.3 The ASQ results with the average response value and standard deviation.

Question	Average Response	Standard Deviation (seconds)
Overall, I am satisfied with the ease of completing the tasks in this scenario	6.64	0.84
Overall, I am satisfied with the amount of time it took to complete the tasks in this scenario	6.64	0.49
Overall Satisfaction of the system	6.46	0.53

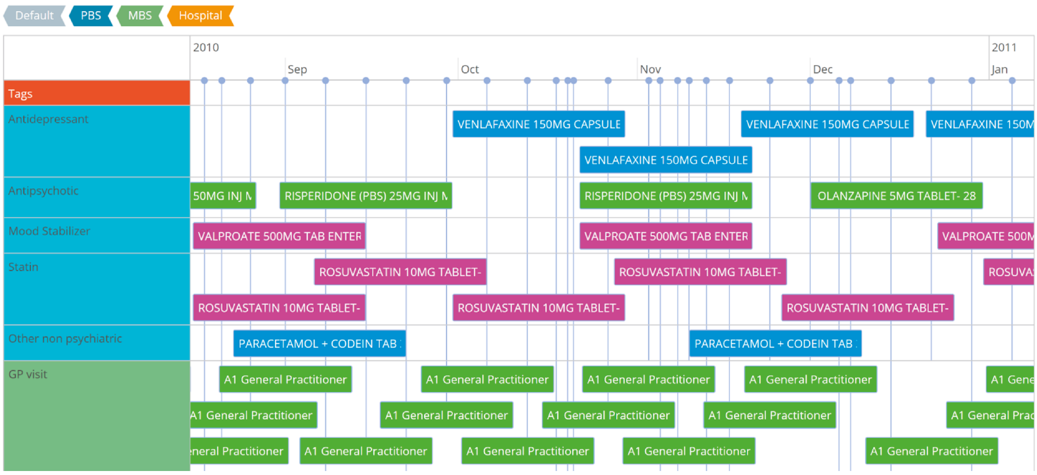
6.4 Health timeline (Publication IV)

The health timeline was designed to provide a simple and interactive way to review a patient's medical history using a software visualization of time-based clinical data. The visualization software aims to assist in the decision making process when assessing the overall status of the patient.

The authors of this study conducted an insight-based evaluation of the visualization software. Five psychiatrists took part in the study by reviewing the clinical data of five anonymised patients. The psychiatrists were instructed to use the "thinking aloud process". Their findings were recorded using voice recording software. These assessments were later transcribed and analysed following insight-based methodology. Figure 6.4a shows an example of the health timeline portraying the clinical history of a patient.

The health timeline tries to address the challenge of representing data clearly and in a meaningful way without overwhelming the user with excessive details and complex interfaces. During the assessment, participants became acquainted with health timeline. Other visualizations, such as LifeFlow [133] and EventFlow [134], required participants to get acquainted with the tool before conducting the experiment.

The current interface to the EHR system consists of an interactive table, shown in Figure 6.4b. This system is considered the baseline representation of the study.



(a) The health timeline visualization showing the collection of EHRs of a patient.

Document	Category	Item	Start Date	End Date	Media
PBS	Antipsychotic	OLANZAPINE 5MG TABLET- 28	28/06/2013	28/07/2013	
PBS	Antipsychotic	OLANZAPINE 5MG TABLET- 28	14/06/2013	14/07/2013	
PBS	Statin	ATORVASTATIN 80MG TABLET- 30	14/06/2013	14/07/2013	
MBS	GP visit	A1 General Practitioner	07/06/2013	11/06/2013	
PBS	Other non psychiatric	RABEPRAZOLE 20MG TAB ENTERIC SO	07/06/2013	07/07/2013	
PBS	Antidepressant	CITALOPRAM 20MG TABLET- 28	07/06/2013	07/07/2013	
PBS	Statin	ATORVASTATIN 80MG TABLET- 30	20/05/2013	19/06/2013	
PBS	Antipsychotic	OLANZAPINE 5MG TABLET- 28	20/05/2013	19/06/2013	

(b) The baseline visualization showing a set of EHRs of a patient as textual information in a table that can be sorted.

Figure 6.4 The health timeline and baseline visualizations used in the study. Both techniques contain the same data but with different graphical representations and interactive elements.

6.4.1 Insight assessment criteria for clinicians

Insight-based methodology has not been applied to clinical longitudinal data. For this reason, new insight assessment criteria were formulated. Professional psychiatrists took part in the process of formulating the insight value criteria.

The criteria are based on insights derived from time patterns and correlation with events along the clinical histories of the patients.

6.4.2 Clinical data

The studies used anonymised patient data obtained with consent from mental health patients with various psychiatric disorders [135].

The selected data included Pharmaceutical Benefit claims (type of medication, amount of medication supplied, and the date of supply), Medicare Benefits claims (tests and services by Medicare eligible practitioners and the associated date of the service), and hospitalization dates over 3 years between 2012 and 2014. The data also included visits to general practitioners, specialists, laboratory tests, emergency hospital admissions, and drug treatment, among other categories.

6.4.3 Experiment protocol

A total of five professional psychiatrists took part in this study. The participants received written instructions on how to conduct the assessments via email. The instructions contained 10 web links used to access a web portal that displayed the patient data. As in previous studies, the participants were instructed to use the "thinking aloud process". The assessments were recorded using a phone communication software.

As in previous studies [25, 26, 27, 80, 136], the insights were transcribed and evaluated. The resulting data was further processed and analysed using the following metrics: insight value per assessment (average and cumulative), number of insights and time to first insight.

6.4.4 Results

A total of 50 assessments were collected, comprising 78,783 seconds of recordings. Table 6.4 summarizes the characteristics of the distributions such as the minimum, maximum, mean, median, and standard deviation.

The health timeline had an average of 22.32 insights compared to the baseline with 23.04 ($p = 0.7047$). The mean value for the timeline (1.70) was higher than the baseline (1.26; $p = 0.01$). The cumulative value for the timeline (34.68) was significantly higher than the baseline (24.96; $p = 0.01$).

The timeline (558) had a lower count of insights than the baseline (576; $p = 0.7$).

Table 6.4 The results comprising the total count of insights, cumulative, and mean value per assessment. Statistical significance is shown in the p column using Mann-Whitney U tests.

Metric	Baseline	Health Timeline	p-values
Number of insights	$min = 6$ $max = 42$ $\mu = 23.04$ $Md = 23$ $\sigma = \pm 10.43$	$min = 5$ $max = 42$ $\mu = 22.32$ $Md = 21$ $\sigma = \pm 9.12$	0.70
Cumultive value	$min = 8$ $max = 45$ $\mu = 24.96$ $Md = 24$ $\sigma = \pm 10.96$	$min = 10$ $max = 56$ $\mu = 34.68$ $Md = 38$ $\sigma = \pm 11.87$	0.01
Mean value	$min = 0.21$ $max = 2.17$ $\mu = 1.26$ $Md = 1.21$ $\sigma = \pm 0.52$	$min = 0.78$ $max = 2.71$ $\mu = 1.70$ $Md = 1.60$ $\sigma = \pm 0.57$	0.01

The average and median insight value for the timeline ($\mu = 1.0833$; $Md = 1$; $\sigma = \pm 1.10$) were higher than the baseline ($\mu = 1.55$; $Md = 2$; $\sigma = \pm 1.25$; p< 0.01)

For insights of any given value, the timeline was slower (13.16 seconds) compared to the baseline (7 seconds; $p < 0.01$). Insights with a value greater than one or two were also slower with the timeline (22.24 seconds and 51.21 seconds) compared to the baseline (20.36 seconds; $p = 0.14$ and 40.01 seconds; $p < 0.01$). For insights with a value greater than 3, the timeline was faster (68.44 seconds) compared to the baseline (92.83 seconds; $p < 0.01$). Only the timeline was able to generate insights of value greater than 4 with an average time of 63.50 seconds.

6.5 Research analysis

Publication I provided results that were similar to those presented by the visual perception framework developed by Cleveland and McGill [32]. The study however, was conducted using a different methodology. By applying the insight-based methodology in a clinical context, the necessary data was obtained from the experiments to gain better understanding on how the graphical representation affects the clinical decision making process. By ranking and analysing the data obtained from the experiments, it was concluded that the visual presentation of the data has an impact similar to the graphical perception that Cleveland and McGill reported. This in turn, affected the visual reasoning and the decision making.

Publication II and III demonstrated the effectiveness of the wellness dashboard and the hFigures library by conducting usability tests. More importantly, the usability testing applied in conjunction with the cognitive walkthrough, revealed the thinking patterns of the participants. Their interactions with the system were recorded and studied. These revealed usability problems that arouse from cognitive overload caused by the graphical interface and lack of clarity in the dashboard system. However, the usability test and cognitive walkthrough also revealed that the users were able to understand multimodal data presented via the hFigures library. The results show that usability testing coupled with cognitive walkthrough can reveal the visual-analytical operation, perception and cognition as reported by Lam and researchers [23]. This was the case when the users were asked to asses the health status and use that knowledge to complete the tasks during the experiment sessions.

In publication IV the insight-based method was applied to study the visual reasoning and decision making process of clinicians via an interactive visualization system. By capturing and analysing the insights, the thinking patterns of the clinicians were recorded. The pattern reveal a deeper understanding of the data when the interactive capabilities of the system were used. The streams of insights and their increasing value show that the clinicians were able to grasp the clinical history of the patients as they became more comfortable with the software. The formulation of hypotheses and the visual reasoning process were documented in the captured insights since these were recorded with the time at which the clinicians verbally expressed their thoughts (thinking aloud process). To this extent, the process described by Pirolli and Card [22] was captured via the experiments and studied via the insight-based methodology [24]. The aforementioned reasoning process involved the searching, reading and extraction of information, as well as the discovery of knowledge, schematization of information, hypothesis generation and the decision making process.

7 DISCUSSION

7.1 Results versus objectives

The experiments documented in the publications compiled for this dissertation were conducted to study the reasoning derived from the visualization and how this affects the clinical decision making process at an individual level. By applying quantitative and qualitative research, it was possible to compare visualizations in a clinical context and provided better understanding on what makes a good visualization. The studies provide documentation on the processes and activities involved in the visual reasoning and knowledge discovery that took place in the experiment sessions. By utilizing evaluation methods such as the insight-based and usability testing with cognitive walkthrough, the documented experiments can serve as blueprints for future studies.

To study and compare how different clinical data visualizations affect visual reasoning (Publications I — IV).

The objective was achieved in all the publications. Publication I compared five visualizations following insight-based methodology. The study collected the insights of non-medical experts and compared their values. In the study, participants evaluated the overall health of the modelled patient based on the data visualization. This exercise in judgement reflected the overall understanding that the participants were able to obtained from the visualization. The study demonstrates that the visualization affects visual reasoning. Visualizations of the data represented by area size had a negative effect (poor performance and data understanding), while angular, tabular, and hGraph representations had a positive effect (deeper understanding of the data and the underlying conditions of the modelled patient).

Publications II and III studied the performance of the execution of tasks by participants in a laboratory. The objective was to evaluate the level of understanding the participants gained from looking at the clinical data using the visualization software. Additionally, usability experts conducted a heuristic evaluation of the system, and the results were positive. The visualization of the modelled clinical data affected the reasoning in a positive way when the tasks were performed correctly. The high completion rate across participation indicates an overall positive result and demonstrates the assistive nature of the visualization in understanding complex clinical data.

Publication IV studied the insights formulated by clinicians using the timeline visualization and the baseline representation. The values of the insights differed greatly depending on the visualization. It can be concluded that the timeline visualization affected visual reasoning positively by enabling clinicians to formulate hypotheses on the patient's health condition.

To apply a methodology that studies visual reasoning and the decision making process in the assessment of clinical data visualizations (Publications I — IV).

All the publications achieved this objective. Publications I and IV applied insight-based methodology to clinical data visualizations. The methodology falls under the VDAR scenario [23]. It comprises a series of steps to collect insights formulated by participants during the experimentation. As described by North [24] and previously used in bioinformatics [25, 26, 27], it has a structure followed by a series of metrics to determine the degree to which the visualization supports hypothesis generation. The studies detail the adjustments required to contextualize the methodology for clinical data.

Publications II and III detail the use of UP evaluations to study clinical visualization software. These evaluation methods are also recommended by Johnson and colleagues when studying EHR visualizations [76]. The cognitive walkthrough [83] and heuristic evaluation [82] constitute structured usability methodologies. Additionally, usability questionnaires measured the UP of the application. These questionnaires "collect self-reported data" and analyse the data to produce usability metrics. These questionnaires have been studied and analysed [76].

To develop and objectively measure the scalability and usability of software that visualizes holistic clinical data (Publications II and III).

The objective was achieved. The software utilized the library hFigures (available in open source from the GitHub repository `https://github.com/ledancs/hFigures`). The wellness dashboard used the library to present the modelled data of a patient at risk of developing T2D, who went through a wellness coaching program. Publications II and III detail the usability methods employed to test the system. The methodologies fall under the UP scenario. The results indicate a positive outcome; users were able to perform the majority of the tasks in a reasonable time. The cognitive walkthrough and heuristic evaluation were conducted with usability experts. Usability questionnaires were also used, and the results indicate a positive outcome.

To study how visualizations affect the decision-making process (Publications I and IV).

The objective was achieved by comparing visualizations based on the correctness of the decisions made by the participants. In Publication I, non-medical users were asked to assess the overall health and wellness of the modelled patient. The visualization technique greatly affected the correctness of the decisions. For instance, the area-based visualization was difficult to grasp and as a result the participants were unable to formulate correct assessments to the same degree as with the other visualizations.

In Publication IV, clinical experts formulated hypotheses and suggested treatments based on the data presented to them during the experiments. The publication details the different outcomes in terms of insight metrics and hypothesis generation. These outcomes are the result of the participants' assessments based on either the baseline representation or the timeline EHR visualization system.

To apply a methodology that studies visual reasoning and the decision making process to assess visualization software for clinical data with the participation of domain experts (Publication IV).

The objective was achieved by recruiting five professional psychiatrists, who took part in the analysis of real patient data. The study followed insight-based methodology, and to date, is the only instance of this methodology applied to study the visualization of real clinical data with domain experts. The publication is as a use case of insight-based assessment which studied the impact of the visualization on the clinical decision making process.

7.2 Impacts of the studies in their research fields

The studies found in the literature survey (chapter 5) did not focus on the assessment of clinical data visualizations in terms of their efficacy in supporting clinical decision making process at an individual level. All the publications (I—IV) compiled in this thesis are practical examples on the use of structured methods that focus on the reasoning derived from the visualization and how this affects the clinical decision making process. The use of structured methods that focus on this goal, could increase the reliability of results obtained from studies. With the provision of a reliable method, the outcome can be expected to be of higher quality and thus produce a better impact in the clinical decision making process. Publication I was the first instance, at its time of publication, in which insight-based methodology was used to evaluate a clinical data visualization that supports patient assessment.

Publications II and III resulted in the implementation of the hFigures library, which has the potential to be used in both clinical and personal wellness applications. The visualization was perceived by the users to portray large datasets effectively while preventing saturation of graphical elements on the screen. Keeping the visualization devoid of clutter decreases the cognitive load required to make sense of the data. Figure 7.1 shows an example of a large visualization of a modelled patient. The comparison of multiple graphs can provide a meaningful visualization for individuals and clinicians alike. The implementation of hFigures follows an extensible approach, and even though it was designed to be used for health data visualization, any dataset that has a target range of values as a reference can be visualized.

Publication II presented a dynamic visualization of the health status of the modelled patient. This was achieved by rendering graphs that respond to the user's command when selecting the time span that is being analysed. Most clinical data sources contain a timestamp that can be used in this way, to present users with a picture of the overall condition of the patient at certain time intervals. The hFigures library draws this essential attribute from EHR data.

In Publications II and III, users demonstrated high satisfaction with the usability of the system, which indicates that the information visualization techniques used in the system were successful. Moreover, this research strongly indicates that the use of different methods and standardized questionnaires in the evaluation phase increases the reliability of the results. In addition, the different tools produced similar

Figure 7.1 A heterogeneous hFigures example. An overview of a modelled person comprising several measurements with two time snapshots showing its evolution over time

results, which suggests that the tools perform in a similar way. The results also show that successful visualization can assist individuals in better understanding health and wellness data.

In Publication IV, a time-based visualization of longitudinal clinical data enabled clinicians to better understand patient histories. The data visualization assisted the participants in obtaining a greater understanding of the data (complete and accurate), compared to the baseline representation. In some cases, clinicians were able to understand the clinical history of the patients, formulate a diagnosis, and suggest treatments. The study shows the need to use a structured assessment methodology in the context of healthcare, to determine the extent to which a visualization can assist clinicians in understanding data. Without an objective assessment, it becomes subjective to state with confidence that one visualization is "useful" or "better". Lam and colleagues [137], as well as Bertini and researchers, [138] also emphasise the importance of structured assessment methods to evaluate data visualizations. This study also serves as a documented example of a contextualized assessment of visualizations in order to provide a use-case evaluation that is based on a real-life scenario. Even though the results of this study are encouraging, a larger study is warranted to examine objective outcomes and the impact that clinical data visualization may have on patient outcomes over longer periods of time.

7.3 Limitations of the studies

A small sample size reduces confidence in the conclusions that can be made based on the results. However, many of the results from Publication I are in line with previous studies, such as the graphical perception study by Cleveland and McGill [32]. The data used was modelled by a clinician based on known patterns that tend to develop T2D. In real-life data often times has a certain degree of noise. Dealing with missing data entries, failure in sensors and borderline parameter values are real problems that the study did not address. Visualizations should account for such cases and make it clear either via annotations or visual cues so that clinicians are aware of these issues.

In the case of Publications I, II, and III, the participants were all recruited from a technical university, and were all academically oriented. This reduces the ability to generalise the conclusions to other user groups, including the general public.

The hFigures library used in Publications I and II does not retrieve data remotely by itself, so data provision is the responsibility of the developer. Other libraries provide interface through HTTP(S) communication to an endpoint in order to retrieve data. The information comes entirely from a data source file, which means that additional information about the measurements cannot be supplied otherwise. The extraction of the SVG file requires the export of the code embedded in the HTML file. Currently, no automatic export function is implemented.

The clinical data used in Publication IV did not include diagnoses and notes taken by the practitioners during previous consultations. This was deliberate, to allow the participants to come up with their own conclusions and assessments.

In a real-life scenario, practitioners would have access to the diagnoses, notes, and observations made by other clinicians. For the purpose of the study, this information was excluded to test the effectiveness of the visualization itself. The insights formulated in the assessments were measurements that reflect the degree to which the participants were able to understand the clinical data without any previous information. This is not an entirely realistic scenario, but for the purpose of the experiment, it allowed the study to emphasise the usefulness of the visualization.

The clinical data was comparable to the reality of the healthcare context because it was extracted directly from real patients. In terms of complexity, the selected data had a high degree of complexity in treatment, in-patient and out-patient events. This was done to base the study on complex real-life scenarios.

To facilitate participation in the study of Publication IV, insight-based methodology was applied with the "thinking aloud process". All the audio recordings were transcribed for evaluation. The sample size and the clinical data used in this study required a large number of working hours for the transcription and evaluation of the insights. Increasing the sample size or number of participants would improve the representation and statistical reliability of the results but at the cost of a greater number of transcribing hours, which was beyond our resources. Therefore, the generalisation of this study may be limited, and further studies would be desirable to confirm the findings.

Another limitation was the time available to conduct the assessments. Even with the relatively brief time window of 30 minutes, it was not possible to recruit more participants. This could probably be explained by the busy schedules of the psychiatrists.

The evaluation criteria, even though revised and peer-reviewed by domain experts, could also be subject to bias. A larger group of domain experts could provide more objective criteria. It is possible that experts that were not acquainted with the data might have provided better evaluation criteria, as these experts would have a fresh look at the clinical data without preconceived notions.

Given the results obtained in Publication IV, further studies should be conducted to test whether the timeline visualization can assist a larger group of experts in providing a higher quality of care to patients. Such a study would require a larger number of participants and patients, with the required ethical approval, as well as randomized controlled trials.

7.4 Directions for future research

Publications I and IV applied insight-based methodology to studies the reasoning derived from the visualization and how this affects the clinical decision making process. To the date of publication, these are the only instances in which a VDAR evaluation has been used. As the literature survey shows in chapter 5, the prevalence of evaluation in clinical data visualization studies is higher in recent articles. However, the publications that do conduct assessments, do not compare the methods nor do they justify the method selection.

Additional studies are needed to increase the understanding and importance of clinical data visualization. Studies have suggested that it can aid in the process of understanding large multidimensional data sets [10, 19]. This is a key component of an effective decision support system, particularly in the case of clinical data, where information gathering is often a challenge for time-constrained clinicians and health practitioners. Additional studies are required to determine if insight-based methodology is the best fit from the VDAR scenario to properly study the effectiveness of these visualizations.

Studies similar to publication IV provide a use case scenario close to practice in healthcare. Due to the increasing demand for healthcare services and the limited resources, specialists are likely to receive and treat patients they are not familiar with. The study demonstrates that clinicians can gain the necessary knowledge to make accurate assessments on patient status. Additional studies with a similar setup should be conducted in the future to refine the evaluation methodology and establish good practices when validating visualization software that supports the clinical decision making process.

Publications II and III consist of the usability study on a wellness dashboard, visualizing clinical and coaching activity data. The evaluation toolkit [76], as well as the evaluations in the UP scenario [137], are well-documented frameworks for selecting the appropriate assessment methods. Online usability questionnaires are useful for practical reasons, because a usability expert is not required to be present during the assessment. [76]. Questionnaires such as CSUQ and ASQ proved useful in the studies. However, additional evidence is needed to establish the benefits of using ASQ for visualizing EHRs that assist in the clinical decision making process. Future studies should see an increase in online usability questionnaires as it is an efficient and convenient evaluation mechanism. Johnson and researchers already present substantial cases that have relied on this method [76], and so similar evaluations of clinical decision support tools can be expected.

The approach to visualizing clinical data looks increasingly like a combination of several techniques, connected by interactive exploration. Rather than focusing on a "silver bullet" to tackle the multidimensional nature of clinical data, future solutions can consist of several visualizations suitable for different exploration tasks. DecisionFlow is an example of combining visualization techniques [116]. New applications could enable users to switch between modes to find the appropriate visualization to explore the data and obtain the necessary insights to make better clinical decisions.

8 CONCLUSIONS

The objective of this thesis is to develop a methodology that studies the reasoning derived from the visualization and how this affects the decision making process at an individual level. The literature survey in chapter 5 revealed that these studies are uncommon. The gap that the thesis aims to bridge is to develop a methodology that allows the assessment of clinical data visualizations in terms of their efficacy in supporting clinical decision making. By applying a mixed research method of qualitative and quantitative studies, this thesis provides a blueprint on the assessment of data visualizations that support clinical decision making.

The experiments conducted required the use of methodologies that fall under the category of VDAR and UP (including the cognitive walkthrough) [23]. These experiments studied the data reasoning activities defined by Pirolli and Card [22]. The scientific contribution of this thesis is the application of VDAR and UP evaluation methodology in the context of visualizations that support the clinical decision making process. By better understanding the reasoning derived from the visualization and how this affects the clinical decision making process, researchers can improve existing techniques to facilitate the process of making sense of data for clinicians. It can be expected that better visualizations that are scientifically tested to ease cognitive burden on clinicians, may lead to better diagnosis and better patient outcomes. The experiments documented in the studies compiled for this thesis, required the implementation of computerized visualization systems.

The key findings of this theses can be summarized as follows:

- Visual analytical and reasoning evaluation methods are mostly missing from current studies that aim to study the reasoning derived from visualization and how this affects the clinical decision making process.
- Insight-based methodology identified visualizations that positively affect the decision-making process for non-medical experts. The results obtained from

the comparison of five visualizations corroborate the previously established framework of graphical perception [32].

- Radar plots have the highest potential for representing clinical data, which by nature are large and multidimensional. As an example the hFigures library was developed to build on the interactivity of web software to enhance the capabilities of radar plots.
- UP evaluation methods with cognitive walkthrough can be used to assess the interaction features of a visualization and how these affect data analysis and reasoning. A software for continuous monitoring of health and wellness proved to be satisfactory in accommodating complex data, according to UP and cognitive walkthrough evaluations.
- VDAR evaluation methods are suitable to assess the efficacy of data visualisation techniques on clinical decision making.
- Timeline representation of clinical data proved to be effective. Insight-based methodology was used to measure its impact on the clinical decision making process. The software allowed for interactive actions that support data analysis while effectively representing the clinical history of a patient, encouraging data exploration and hypothesis generation.

REFERENCES

[1] L. Minor. Harnessing the Power of Data in Health. *Stanford Med. Heal. Trends Rep.* (2017).

[2] J. D. Halamka. *The Cost of Storing Patient Records*. 2011. URL: `http://geekdoctor.blogspot.com/2011/04/cost-of-storing-patient-records.html`.

[3] G. Irving, A. L. Neves, H. Dambha-Miller, A. Oishi, H. Tagashira, A. Verho and J. Holden. International variations in primary care physician consultation time: a systematic review of 67 countries. *BMJ open* 7.10 (2017), e017902.

[4] M. S. Donaldson, J. M. Corrigan, L. T. Kohn et al. *To err is human: building a safer health system*. Vol. 6. National Academies Press, 2000.

[5] PinnacleCare. *The Human Cost and Financial Impact of Misdiagnosis*. Tech. rep. PinnacleCare, 2016. URL: `https://www.pinnaclecare.com/forms/download/Human-Cost-Financial-Impact-Whitepaper.pdf`.

[6] J. A. Osheroff, J. M. Teich, B. Middleton, E. B. Steen, A. Wright and D. E. Detmer. A roadmap for national action on clinical decision support. *Journal of the American medical informatics association* 14.2 (2007), 141–145.

[7] R. Miotto, L. Li, B. A. Kidd and J. T. Dudley. Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. *Scientific reports* 6 (2016), 26094.

[8] P. B. Jensen, L. J. Jensen and S. Brunak. Mining electronic health records: towards better research applications and clinical care. *Nature Reviews Genetics* 13.6 (2012), 395.

[9] N. G. Weiskopf, G. Hripcsak, S. Swaminathan and C. Weng. Defining and measuring completeness of electronic health records for secondary use. *Journal of biomedical informatics* 46.5 (2013), 830–836.

[10] A. Rind, T. D. Wang, W. Aigner, S. Miksch, K. Wongsuphasawat, C. Plaisant, B. Shneiderman et al. Interactive information visualization to explore and query electronic health records. *Foundations and Trends®in Human–Computer Interaction* 5.3 (2013), 207–298.

[11] T. Goetz. *It's time to redesign medical data*. 2015. URL: `http://www.ted.com/talks/thomas_goetz_it_s_time_to_redesign_medical_data?language=en`.

[12] Health Science Library, University of Washington. *Data Resources in the Health Sciences*. Updated: 2018-11-20. Accessed: 2018-12-25. URL: `http://guides.lib.uw.edu/c.php?g=99209%5C&p=642709%5C#12207216`.

[13] M. R. Cowie, J. I. Blomster, L. H. Curtis, S. Duclaux, I. Ford, F. Fritz, S. Goldman, S. Janmohamed, J. Kreuzer, M. Leenay et al. Electronic health records to facilitate clinical research. *Clinical Research in Cardiology* 106.1 (2017), 1–9.

[14] S. C. Denaxas and K. I. Morley. Big biomedical data and cardiovascular disease research: opportunities and challenges. *European Heart Journal-Quality of Care and Clinical Outcomes* 1.1 (2015), 9–16.

[15] N. P. Tatonetti, P. Y. Patrick, R. Daneshjou and R. B. Altman. Data-driven prediction of drug effects and interactions. *Science translational medicine* 4.125 (2012), 125ra31–125ra31.

[16] L. Li, W.-Y. Cheng, B. S. Glicksberg, O. Gottesman, R. Tamler, R. Chen, E. P. Bottinger and J. T. Dudley. Identification of type 2 diabetes subgroups through topological analysis of patient similarity. *Science translational medicine* 7.311 (2015), 311ra174–311ra174.

[17] F. Doshi-Velez, Y. Ge and I. Kohane. Comorbidity clusters in autism spectrum disorders: an electronic health record time-series analysis. *Pediatrics* 133.1 (2014), e54–e63.

[18] R. Miotto and C. Weng. Case-based reasoning using electronic health records efficiently identifies eligible patients for clinical trials. *Journal of the American Medical Informatics Association* 22.e1 (2015), e141–e150.

[19] B. J. Lesselroth and D. S. Pieczkiewicz. *Data visualization strategies for the electronic health record*. Nova Science Publishers, Inc., 2011.

[20] A. Bui, D. Aberle and H. Kangarloo. TimeLine: Visualizing Integrated Patient Records. *Information Technology in Biomedicine, IEEE Transactions on* 11.4 (2007), 462–473. ISSN: 1089-7771. DOI: 10.1109/TITB.2006.884365.

[21] F. Bentley, K. Tollmar, P. Stephenson, L. Levy, B. Jones, S. Robertson, E. Price, R. Catrambone and J. Wilson. Health Mashups: Presenting statistical patterns between wellbeing data and context in natural language to promote behavior change. *ACM Transactions on Computer-Human Interaction (TOCHI)* 20.5 (2013), 30.

[22] P. Pirolli and S. Card. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. *Proceedings of international conference on intelligence analysis*. Vol. 5. McLean, VA, USA. 2005, 2–4.

[23] H. Lam, E. Bertini, P. Isenberg, C. Plaisant and S. Carpendale. Empirical studies in information visualization: Seven scenarios. *IEEE transactions on visualization and computer graphics* 18.9 (2011), 1520–1536.

[24] C. North. Toward measuring visualization insight. *IEEE computer graphics and applications* 26.3 (2006), 6–9.

[25] P. Saraiya, C. North and K. Duca. An evaluation of microarray visualization tools for biological insight. *IEEE Symposium on Information Visualization*. IEEE. 2004, 1–8.

[26] P. Saraiya, C. North and K. Duca. An insight-based methodology for evaluating bioinformatics visualizations. *IEEE transactions on visualization and computer graphics* 11.4 (2005), 443–456.

[27] P. Saraiya, C. North, V. Lam and K. A. Duca. An insight-based longitudinal study of visual analytics. *IEEE Transactions on Visualization and Computer Graphics* 12.6 (2006), 1511–1522.

[28] J. J. Otten, K. Cheng and A. Drewnowski. Infographics and public policy: using data visualization to convey complex information. *Health Affairs* 34.11 (2015), 1901–1907.

[29] W. Playfair. *The commercial and political atlas: representing, by means of stained copper-plate charts, the progress of the commerce, revenues, expenditure and debts of england during the whole of the eighteenth century*. T. Burton, 1801.

[30] M. Friendly. A brief history of data visualization. *Handbook of data visualization*. Springer, 2008, 15–56.

[31] J. H. Coll and R. Coll. Tables and graphs: A classification scheme for display presentation variables and a framework for research in this area. *Information processing & management* 29.6 (1993), 745–750.

[32] W. S. Cleveland and R. McGill. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American statistical association* 79.387 (1984), 531–554.

[33] C. M. Carswell. Choosing specifiers: An evaluation of the basic tasks model of graphical perception. *Human factors* 34.5 (1992), 535–554.

[34] L. R. Beach and T. R. Mitchell. A contingency model for the selection of decision strategies. *Academy of management review* 3.3 (1978), 439–449.

[35] M. Kennedy, D. Te'eni and J. B. Treleaven. Impacts of decision task, data and display on strategies for extracting information. *International Journal of Human-Computer Studies* 48.2 (1998), 159–180.

[36] S. M. Glover, D. F. Prawitt and B. C. Spilker. The influence of decision aids on user behavior: Implications for knowledge acquisition and inappropriate reliance. *Organizational Behavior and Human Decision Processes* 72.2 (1997), 232–255.

[37] P. Todd and I. Benbasat. The influence of decision aids on choice strategies: an experimental analysis of the role of cognitive effort. *Organizational behavior and human decision processes* 60.1 (1994), 36–74.

[38] E. R. Tufte, N. H. Goeler and R. Benson. *Envisioning information*. Vol. 126. Graphics press Cheshire, CT, 1990.

[39] E. R. Tufte. *The visual display of quantitative information*. Vol. 2. Graphics press Cheshire, CT, 2001.

[40] E. R. Tufte. *Beautiful evidence*. Vol. 1. Graphics Press Cheshire, CT, 2006.

[41] S. M. Kosslyn and S. M. Kosslyn. *Graph design for the eye and mind*. OUP USA, 2006.

[42] S. Few. Show me the numbers. *Analytics Pres* (2004).

[43] S. Few. *Now you see it: simple visualization techniques for quantitative analysis*. Analytics Press, 2009.

[44] D. L. Schriger and R. J. Cooper. Achieving graphical excellence: suggestions and methods for creating high-quality visual displays of experimental data. *Annals of emergency medicine* 37.1 (2001), 75–87.

[45] D. J. Gillan, C. D. Wickens, J. G. Hollands and C. M. Carswell. Guidelines for presenting quantitative data in HFES publications. *Human Factors* 40.1 (1998), 28–41.

[46] N. B. Robbins. *Creating more effective graphs*. Wiley, 2012.

[47] W. S. Cleveland and W. S. Cleveland. *The elements of graphing data*. Vol. 2. Wadsworth advanced books and software Monterey, CA, 1985.

[48] D. L. McGee. *MSD and the MSD Manuals: Introduction to Clinical Decision Making*. 2018. URL: `https://www.msdmanuals.com/professional/special-subjects/clinical-decision-making/introduction-to-clinical-decision-making`.

[49] T. D. Gunter and N. P. Terry. The emergence of national electronic health record architectures in the United States and Australia: models, costs, and questions. *Journal of medical Internet research* 7.1 (2005), e3.

[50] B. Shneiderman, C. Plaisant and B. Hesse. *Improving health and healthcare with interactive visualization methods*. Tech. rep. HCIL Technical Report, 2013.

[51] C. Plaisant, B. Milash, A. Rose, S. Widoff and B. Shneiderman. LifeLines: visualizing personal histories. *Proceedings of the SIGCHI conference on Human factors in computing systems*. 1996, 221–227.

[52] A. Faiola and C. Newlon. Medical Information Visualization Assistant (MIVA). (2011).

[53] D. S. Pieczkiewicz, S. M. Finkelstein and M. I. Hertz. Design and evaluation of a web-based interactive visualization system for lung transplant home monitoring data. *AMIA annual symposium proceedings*. Vol. 2007. American Medical Informatics Association. 2007, 598.

[54] R. Bade, S. Schlechtweg and S. Miksch. Connecting time-oriented data and information to a coherent interactive visualization. *Proceedings of the SIGCHI conference on Human factors in computing systems*. 2004, 105–112.

[55] A. Rind, W. Aigner, S. Miksch, S. Wiltner, M. Pohl, T. Turic and F. Drexler. Visual exploration of time-oriented patient data for chronic diseases: Design study and evaluation. *Symposium of the Austrian HCI and Usability Engineering Group*. Springer. 2011, 301–320.

[56] W. Horn, C. Popow and L. Unterasinger. Support for fast comprehension of ICU data: Visualization using metaphor graphics. *Methods of information in medicine* 40.05 (2001), 421–424.

[57] T. D. Wang, C. Plaisant, A. J. Quinn, R. Stanchak, S. Murphy and B. Shneiderman. Aligning temporal data by sentinel events: discovering patterns in electronic health records. *Proceedings of the SIGCHI conference on Human factors in computing systems*. 2008, 457–466.

[58] T. D. Wang, C. Plaisant, B. Shneiderman, N. Spring, D. Roseman, G. Marchand, V. Mukherjee and M. Smith. Temporal summaries: Supporting temporal categorical searching, aggregation and comparison. *IEEE transactions on visualization and computer graphics* 15.6 (2009), 1049–1056.

[59] K. Wongsuphasawat and B. Shneiderman. Finding comparable temporal categorical records: A similarity measure with an interactive visualization. *2009 IEEE Symposium on Visual Analytics Science and Technology*. IEEE. 2009, 27–34.

[60] J. A. Fails, A. Karlson, L. Shahamat and B. Shneiderman. A visual interface for multivariate temporal data: Finding patterns of events across multiple histories. *2006 IEEE Symposium On Visual Analytics Science And Technology*. IEEE. 2006, 167–174.

[61] D. Klimov, Y. Shahar and M. Taieb-Maimon. Intelligent selection and retrieval of multiple time-oriented records. *Journal of Intelligent Information Systems* 35.2 (2010), 261–300.

[62] D. Brodbeck, R. Gasser, M. Degen, S. Reichlin and J. Luthiger. Enabling large-scale telemedical disease management through interactive visualization. *European Notes in Medical Informatics* 1.1 (2005), 1172–1177.

[63] L. Chittaro, C. Combi and G. Trapasso. Data mining on temporal data: a visual approach and its clinical application to hemodialysis. *Journal of Visual Languages & Computing* 14.6 (2003), 591–620.

[64] K. Hinum, S. Miksch, W. Aigner, S. Ohmann, C. Popow, M. Pohl and M. Rester. Gravi++: Interactive Information Visualization to Explore Highly Structured Temporal Data. *J. UCS* 11.11 (2005), 1792–1805.

[65] A. Rind, W. Aigner, S. Miksch, S. Wiltner, M. Pohl, F. Drexler, B. Neubauer and N. Suchy. Visually exploring multivariate trends in patient cohorts using animated scatter plots. *International Conference on Ergonomics and Health Aspects of Work with Computers*. Springer. 2011, 139–148.

[66] B. Lesselroth, R. Felder, C. Locke, R. Klein and V. Church. *A glycemic monitoring window to improve inpatient diabetic management*. Tech. rep. 2006.

[67] A. Davidson. *Firstbeat Sports New Stress and Recovery Feature Allows True Big Picture Player Monitoring*. 2020. URL: `https://www.firstbeat.com/en/news/firstbeat-sports-stress-and-recovery-feature-allows-true-big-picture-player-monitoring`.

[68] Withings. *Pulse HR Health & fitness tracker*. 2020. URL: `https://www.withings.com/us/en/pulse-hr`.

[69] *Oura Cloud – How to make the most of your trends view*. 2020. URL: `https://ouraring.com/how-to-make-the-most-out-of-oura-cloud-and-the-trends-view`.

[70] Fitbit. *Fitbit App*. 2020. URL: `https://www.fitbit.com/fi/app`.

[71] T. M. Team. *Fitbit announces their new Universal Windows App for Windows 10, designed to motivate with enhanced features*. 2015. URL: `https://blogs.windows.com/windowsexperience/2015/08/13`.

[72] J. Heer and G. Robertson. Animated transitions in statistical data graphics. *IEEE transactions on visualization and computer graphics* 13.6 (2007), 1240–1247.

[73] H. Lam, R. A. Rensink and T. Munzner. Effects of 2D geometric transformations on visual memory. *Proceedings of the 3rd symposium on Applied perception in graphics and visualization*. ACM. 2006, 119–126.

[74] C. Plaisant, J. Grosjean and B. B. Bederson. Spacetree: Supporting exploration in large node link tree, design evolution and empirical evaluation. *IEEE Symposium on Information Visualization, 2002. INFOVIS 2002*. IEEE. 2002, 57–64.

[75] N. Bevan, J. Carter and S. Harker. ISO 9241-11 revised: What have we learnt about usability since 1998?: *International Conference on Human-Computer Interaction*. Springer. 2015, 143–151.

[76] C. Johnson, D. Johnston, P. Crowle et al. EHR Usability Toolkit: A Background Report on Usability and Electronic Health Records. *Rockville, MD: Agency for Healthcare Research and Quality* (2011).

[77] D. Armijo, C. McDonnell and K. Werner. Electronic health record usability. *Interface Design Considerations. James Bell Associates. The Altarum Institute* (2009).

[78] D. Armijo, C. McDonnell and K. Werner. Electronic health record usability: evaluation and use case framework. *AHRQ Publication* 09 (2009), 10.

[79] O. E. Dictionary. Oxford english dictionary. *Simpson, JA & Weiner, ESC* (1989).

[80] C. Plaisant, J.-D. Fekete and G. Grinstein. Promoting insight-based evaluation of visualizations: From contest to benchmark repository. *IEEE transactions on visualization and computer graphics* 14.1 (2008), 120–134.

[81] J. Nielsen. *Usability engineering*. New York: Elsevier, 1994.

[82] J. Nielsen. *10 Usability Heuristics for User Interface Design*. 1995. URL: `http://www.nngroup.com/articles/ten-usability-heuristics/`.

[83] C. Wharton, J. Rieman, C. Lewis and P. Polson. The cognitive walkthrough method: A practitioner's guide. *Usability inspection methods*. John Wiley & Sons, Inc. 1994, 105–140.

[84] L. W. Peute and M. W. Jaspers. The significance of a usability evaluation of an emerging laboratory order entry system. *International journal of medical informatics* 76.2 (2007), 157–168.

[85] A. Karahoca, E. Bayraktar, E. Tatoglu and D. Karahoca. Information system design for a hospital emergency department: A usability analysis of software prototypes. *Journal of biomedical informatics* 43.2 (2010), 224–232.

[86] T. Cohen, D. Kaufman, T. White, G. Segal, A. B. Staub, V. Patel and M. Finnerty. Cognitive evaluation of an innovative psychiatric clinical knowledge enhancement system. *Medinfo* 11.Pt 2 (2004), 1295–9.

[87] R. Khajouei, D. de Jongh and M. W. Jaspers. Usability evaluation of a computerized physician order entry for medication ordering. *MIE*. 2009, 532–536.

[88] M. H. Blackmon, P. G. Polson, M. Kitajima and C. Lewis. Cognitive walkthrough for the web. *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM. 2002, 463–470.

[89] A. Sears. Heuristic walkthroughs: Finding the problems without the noise. *International Journal of Human-Computer Interaction* 9.3 (1997), 213–234.

[90] M.-C. Beuscart-Zéphir, P. Elkin, S. Pelayo, R. Beuscart et al. The human factors engineering approach to biomedical informatics projects: state of the art, results, benefits and challenges. *IMIA Yearbook* 2 (2007), 109–127.

[91] W. M. Newman. 10. On Simulation, Measurement, and Pfeeewiaie Usability Evaluation. *Commentary on " Damaged Merchandise?* 13 (1998), 316.

[92] K. A. Ericsson and H. A. Simon. Verbal reports as data. *Psychological review* 87.3 (1980), 215.

[93] E. M. Borycki, L. Lemieux-Charles, L. Nagle and G. Eysenbach. Novice nurse information needs in paper and hybrid electronic-paper environments: a qualitative analysis. *MIE*. 2009, 913–917.

[94] L. M. Currie, M. Graham, M. Allen, S. Bakken, V. Patel and J. J. Cimino. Clinical information needs in context: an observational study of clinicians while using a clinical information system. *AMIA Annual Symposium proceedings*. Vol. 2003. American Medical Informatics Association. 2003, 190.

[95] A. Hasman et al. Development of methods for usability evaluations of EHR systems. *Ubiquity: Technologies for Better Health in Aging Societies: Proceedings of MIE2006*. Vol. 124. IOS Press. 2006, 341.

[96] R. C. Wu, M. S. Orr, M. Chignell and S. E. Straus. Usability of a mobile electronic medical record prototype: a verbal protocol analysis. *Informatics for Health and Social Care* 33.2 (2008), 139–149.

[97] J. R. Lewis. IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. *International Journal of Human-Computer Interaction* 7.1 (1995), 57–78.

[98] J. R. Lewis. Psychometric evaluation of the post-study system usability questionnaire: The PSSUQ. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Vol. 36. 16. SAGE Publications. 1992, 1259–1260.

[99] M. W. Jaspers, L. W. Peute, A. Lauteslager and P. J. Bakker. Pre-post evaluation of physicians' satisfaction with a redesigned electronic medical record system. *Studies in health technology and informatics* 136 (2008), 303.

[100] R. Goud, M. W. Jaspers, A. Hasman and N. Peek. Subjective usability of the CARDSS guideline-based decision support system. *Studies in health technology and informatics* 136 (2008), 193.

[101] J. R. Lewis. Psychometric evaluation of an after-scenario questionnaire for computer usability studies: the ASQ. *ACM SIGCHI Bulletin* 23.1 (1991), 78–81.

[102] A. Bangor, P. T. Kortum and J. T. Miller. An empirical evaluation of the system usability scale. *Intl. Journal of Human–Computer Interaction* 24.6 (2008), 574–594.

[103] V. L. West, D. Borland and W. E. Hammond. Innovative information visualization of electronic health record data: a systematic review. *Journal of the American Medical Informatics Association* 22.2 (2014), 330–339.

[104] D. Moher, A. Liberati, J. Tetzlaff, D. G. Altman, P. Group et al. *Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement.* 2010.

[105] Y.-M. Tan, J. Hii, K. Chan, R. Sardual and B. Mah. An electronic dashboard to improve nursing care. *MedInfo*. 2013, 190–194.

[106] M. F. Rayo, N. Kowalczyk, B. W. Liston, E. B.-N. Sanders, S. White and E. S. Patterson. Comparing the effectiveness of alerts and dynamically annotated visualizations (DAVs) in improving clinical decision making. *Human factors* 57.6 (2015), 1002–1014.

[107] L. Franklin, C. Plaisant, K. Minhazur Rahman and B. Shneiderman. *Treatment Explorer Interactive Decision Aid for Medical Information*. 2012. URL: `http://www.cs.umd.edu/hcil/treatmentexplorer/`.

[108] L. Franklin, C. Plaisant, K. Minhazur Rahman and B. Shneiderman. TreatmentExplorer: An interactive decision aid for medical risk communication and treatment exploration. *Interacting with Computers* 28.3 (2014), 238–252.

[109] L. L. W. Sim, K. H. K. Ban, T. W. Tan, S. K. Sethi and T. P. Loh. Development of a clinical decision support system for diabetes care: A pilot study. *PloS one* 12.2 (2017), e0173021.

[110] J. Forsman, N. Anani, A. Eghdam, M. Falkenhav and S. Koch. Integrated information visualization to support decision making for use of antibiotics in intensive care: design and usability evaluation. *Informatics for Health and Social Care* 38.4 (2013), 330–353.

[111] J. Brooke et al. SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189.194 (1996), 4–7.

[112] P. Fraccaro, M. Vigo, P. Balatsoukas, S. N. van der Veer, L. Hassan, R. Williams, G. Wood, S. Sinha, I. Buchan and N. Peek. Presentation of laboratory test results in patient portals: influence of interface design on risk interpretation and visual search behaviour. *BMC medical informatics and decision making* 18.1 (2018), 11.

[113] L. D. Chew, J. M. Griffin, M. R. Partin, S. Noorbaloochi, J. P. Grill, A. Snyder, K. A. Bradley, S. M. Nugent, A. D. Baines and M. VanRyn. Validation of screening questions for limited health literacy in a large VA outpatient population. *Journal of general internal medicine* 23.5 (2008), 561–566.

[114] B. Stubbs, D. C. Kale and A. Das. Sim• TwentyFive: an interactive visualization system for data-driven decision support. *AMIA Annual Symposium Proceedings*. Vol. 2012. American Medical Informatics Association. 2012, 891.

[115] A. Faiola, P. Srinivas and J. Duke. Supporting clinical cognition: a human-centered approach to a novel ICU information visualization dashboard. *AMIA Annual Symposium Proceedings*. Vol. 2015. American Medical Informatics Association. 2015, 560.

[116] D. Gotz and H. Stavropoulos. Decisionflow: Visual analytics for high-dimensional temporal event sequence data. *IEEE transactions on visualization and computer graphics* 20.12 (2014), 1783–1792.

[117] Y. L. Lin, A.-M. Guerguerian, J. Tomasi, P. Laussen and P. Trbovich. Usability of data integration and visualization software for multidisciplinary pediatric intensive care: a human factors approach to assessing technology. *BMC medical informatics and decision making* 17.1 (2017), 122.

[118] *Advanced Visualization*. 2020. URL: `https://www.gehealthcare.com/products/advanced-visualization`.

[119] *Advanced Visualization*. 2020. URL: `https://www.gehealthcare.com/products/advanced-visualization`.

[120] *One work environment, one advanced visualization solution*. 2020. URL: `https://www.philips.fi/healthcare/solutions/clinical-informatics/advanced-visualization`.

[121] J. Lindström and J. Tuomilehto. The Diabetes Risk Score A practical tool to predict type 2 diabetes risk. *Diabetes care* 26.3 (2003), 725–731.

[122] The Finnish Medical Society Duodecim. *Current Care Guidelines*. URL: `http://www.kaypahoito.fi/web/english`.

[123] K. Roebuck. *Agile Software Development: High-impact Strategies-What You Need to Know: Definitions, Adoptions, Impact, Benefits, Maturity, Vendors*. Brisbane, Australia: Emereo Publishing, 2012.

[124] K. Beck. *Extreme programming explained: embrace change*. Boston, Massachusetts: Addison-Wesley Professional, 2000.

[125] L. Faulkner. Beyond the five-user assumption: Benefits of increased sample sizes in usability testing. *Behavior Research Methods, Instruments, & Computers* 35.3 (2003), 379–383.

[126] D. L. Berry, B. Halpenny, J. L. Bosco, J. Bruyere and M. G. Sanda. Usability evaluation and adaptation of the e-health Personal Patient Profile-Prostate decision aid for Spanish-speaking Latino men. *BMC medical informatics and decision making* 15.1 (2015), 56.

[127] J. Zhang, T. R. Johnson, V. L. Patel, D. L. Paige and T. Kubose. Using usability heuristics to evaluate patient safety of medical devices. *Journal of biomedical informatics* 36.1 (2003), 23–30.

[128] B. Shneiderman. *Designing the user interface: strategies for effective human-computer interaction*. Vol. 3. Boston, Massachusetts: Addison-Wesley Reading, MA, 1992.

[129] R. Molich and J. Nielsen. Improving a human-computer dialogue. *Communications of the ACM* 33.3 (1990), 338–348.

[130] B. Tognazzini. *First principles of information design*. 2014. URL: http://www.asktog.com/basics/firstPrinciples.html.

[131] G. Perlman. Electronic surveys. *Behavior Research Methods, Instruments, & Computers* 17.2 (1985), 203–205.

[132] J. R. Lewis. Psychometric evaluation of the PSSUQ using data from five years of usability studies. *International Journal of Human-Computer Interaction* 14.3-4 (2002), 463–488.

[133] K. Wongsuphasawat, J. A. Guerra Gómez, C. Plaisant, T. D. Wang, M. Taieb-Maimon and B. Shneiderman. LifeFlow: visualizing an overview of event sequences. *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM. 2011, 1747–1756.

[134] M. Monroe, T. E. Meyer, C. Plaisant, R. Lan, K. Wongsuphasawat, T. S. Coster, S. Gold, J. Millstein and B. Shneiderman. *Visualizing patterns of drug prescriptions with EventFlow: A pilot study of asthma medications in the military health system*. Tech. rep. OFFICE OF THE SURGEON GENERAL (ARMY) FALLS CHURCH VA, 2013.

[135] N. Bidargaddi, G. Schrader, C. Piantadosi, R. Adams, B. Ryan, J. Williams and J. Strobel. Embedding research into clinical practice: The Services, Pathways Access–Research and Knowledge (SPARK) study. *Australian Journal of Rural Health* 21.3 (2013), 192–193.

[136] A. Ledesma, H. Nieminen, P. Valve, M. Ermes, H. Jimison and M. Pavel. The shape of health: A comparison of five alternative ways of visualizing personal health and wellbeing. *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. 2015, 7638–7641.

[137] H. Lam, E. Bertini, P. Isenberg, C. Plaisant and S. Carpendale. Empirical studies in information visualization: Seven scenarios. *IEEE transactions on visualization and computer graphics* 18.9 (2012), 1520–1536.

[138] E. Bertini, A. Perer, C. Plaisant and G. Santucci. BELIV'08: Beyond Time and Errors: Novel Evaluation Methods for Information Visualization. *CHI '08 Extended Abstracts on Human Factors in Computing Systems*. CHI EA '08. Florence, Italy: ACM, 2008, 3913–3916. ISBN: 978-1-60558-012-8. DOI: 10.1145/1358628.1358955.

[139] A. Ledesma, H. Nieminen, P. Valve, M. Ermes, H. Jimison and M. Pavel. The shape of health: A comparison of five alternative ways of visualizing personal health and wellbeing. *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. 2015, 7638–7641.

[140] M. Al-Musawi, A. Ledesma, H. Nieminen and I. Korhonen. Implementation and user testing of a system for visualizing continuous health data and events. *2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*. IEEE. 2016, 156–159.

[141] A. Ledesma, M. Al-Musawi and H. Nieminen. Health figures: an open source JavaScript library for health data visualization. *BMC Medical Informatics and Decision Making* 16.38 (2016), 1.

[142] A. Ledesma, N. Bidargaddi, J. Strobel, G. Schrader, H. Nieminen, I. Korhonen and M. Ermes. Health Timeline: An Insight-based Study of a Timeline Visualization of Clinical Data. *BMC Medical Informatics and Decision Making* 19.170 (2019), 1.

PUBLICATIONS

PUBLICATION

I

The shape of health: A comparison of five alternative ways of visualizing personal health and wellbeing

A. Ledesma, H. Nieminen, P. Valve, M. Ermes, H. Jimison and M. Pavel

*2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*2015, 7638–7641

Publication reprinted with the permission of the copyright holders

The Shape of Health: A Comparison of Five Alternative Ways of Visualizing Personal Health and Wellbeing*.

Andres Ledesma[1], *Student Member IEEE*, Hannu Nieminen[1], Päivi Valve[1,2], Miikka Ermes[3], Holly Jimison[4], *Member IEEE* and Misha Pavel[4], *Senior Member IEEE*

***Abstract*— The combination of clinical and personal health and wellbeing data can tell us much about our behaviors, risks and overall status. The way this data is visualized may affect our understanding of our own health. To study this effect, we conducted a small experiment with 30 participants in which we presented a holistic overview of the health and wellbeing of two modeled individuals, one of them with metabolic syndrome. We used an insight-based methodology to assess the effectiveness of the visualizations. The results show that adequate visualization of holistic health data helps users without medical background to better understand the overall health situation and possible health risks related to lifestyles. Furthermore, we found that the application of insight-based methodology in the health and wellbeing domain remains unexplored and additional research and methodology development are needed.**

I. INTRODUCTION

Modern medicine is constantly moving towards a preventive approach. As stated by Rose [1], "common diseases have their roots in life-style, social factors and the environment, and successful health promotion depends upon a population-based strategy of prevention". Self-management of health and wellness plays a key role in making this happen.

A holistic description of health should ideally combine together several measures of physical, mental and social wellbeing [2]. Individuals already start to have access to a wealth of multimodal health-related data e.g. from personal wellness devices, health records, genetic tests and mobile wellness applications. This information, when combined together, can be utilized both to empower people to monitor their evolving health [3] and to enhance the healthcare process. However, understanding the status of a person's health and the underlying factors behind the data is not easy, especially for the non-experts. Presenting this information in a comprehensive and meaningful way is an ongoing effort and an open challenge [3].

As an example, TimeLine [4] is a software tool that organizes medical records and provides a "problem-centric temporal visualization". Lesselroth and Pieczkiewicz [5] conducted an extensive literature survey on strategies for the visualization of personal health data. They argue that "smart dashboards" combining different data sources are needed to improve the understanding of our health. Although various graphs, stylized presentation and textual feedback have been addressed in studies [6], visualizations combining multiple data sources that enable inferences are mostly lacking [3]. Health applications enabling to identify connections that are significant over time have been shown to increase the user's self-understanding [7]. Goetze [8] has shown that the style of the presentation of data greatly affects the understanding of our health. The hGraph [9] is an example of a visual representation of a patient's health status, designed to increase awareness of the factors that can affect one's overall health.

Graphical perception, defined as "the visual decoding of information encoded on graphs" [10], has been widely studied and researched [11]. This paper focuses on evaluating how the participants identify the underlying reasons of the health situation of a person represented as a set of measurements. We study how the visual representation of holistic health data affects the capability of non-medical experts to derive insights. Insights aim to see "beyond the figures" by identifying relationships between the measurements and understanding the underlying reasons for the values. We also consider the effect of health literacy in this study, which is defined as "the degree to which individuals have the capacity to obtain, process, and understand basic health information and services needed to make appropriate health decisions" [12].

II. METHODS

We recruited 15 female and 15 male participants from 9 different countries. We used the local bulletin board to advertise the study. Participants having earlier expertise in health data analysis and visualization were excluded from the study. The students signed up using a Doodle poll. The experimental procedures described in this paper complied with the principles of Helsinki Declaration of 1975, as revised in 2000. All subjects gave informed consent to participate and they had a right to withdraw from the study at any time. Their information was anonymized prior the

*This research was supported jointly by TEKES (the Finnish Funding Agency for Technology and Innovation) as part of the Digital Health Revolution and FiDiPro (Finland Distinguished Professor Programme) projects and the European Commission and TEKES under the ARTEMIS-JU WithMe project.

[1]Andres Ledesma, Päivi Valve and Hannu Nieminen are with the Personal Health Informatics research group as part of the Department of Signal Processing, Tampere University of Technology, Tampere, Finland
andres.ledesma@tut.fi,
hannu.o.nieminen@tut.fi, paivi.valve@tut.fi
[2]Päivi Valve is also with the School of Health Sciences, University of Tampere, Tampere, Finland
[3]Miikka Ermes is with VTT Technical Research Centre of Finland, Tampere, Finland
miikka.ermes@vtt.fi
[4]Misha Pavel and Holly Jimison are with the College of Computer & Information Science and the Bouvé College of Health Sciences, Northeastern University, Boston MA, USA
m.pavel@neu.edu, h.jimison@neu.edu

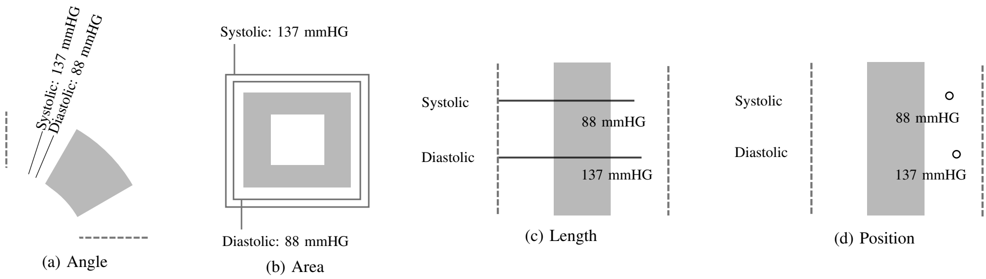


Fig. 1: The blood pressure measurements represented using the angle, area, length and position.

analysis. The participants received a movie ticket at the end of the session.

A. Datasets Used in the Study

We modeled two datasets of measurements to describe two different health situations. The first dataset modeled a person with a poor overall health condition with a clear indication of a metabolic syndrome and therefore a high to very high risk of developing Type 2 Diabetes (T2D), based on the calculator in [13]. The second set modeled a person with a good health condition and with a healthy lifestyle. The participants focused on the modeled person with metabolic syndrome to derive their insights, and the other person was used as a comparison point.

The following health parameters (average values of past month) were chosen to describe their health and wellbeing:

- Blood pressure: systolic and diastolic blood pressure
- Physical activity: weekly active days, steps per day
- Body composition: Body Mass Index (BMI), waist diameter and fat percentage
- Sleep: time in bed, time asleep
- Fitness: resting heart rate, fitness index, muscular force, muscular endurance and balance [14]
- Lab Tests: hemoglobin, fB-Gluc, cholesterol, HDL, LDL, triglycerides
- Drugs: tobacco (cigarettes per day), alcohol abuse, drug abuse (narcotics), medication abuse
- Emotional wellbeing: depression level (DEPS), stress level and stress recovery [15] as well as optimism [16].

Healthy ranges for each parameter, derived based on the national clinical health recommendations [17], were visualized with a light green background. This concept was also explained to the study participants.

B. Visualizations

We utilized the generic visualization framework proposed by Cleveland and McGill [10] to design four different visualizations (Fig. 1). In addition, we included the polar coordinate style "hGraph" [9] visualization, to assess it against the framework of Cleveland and McGill.

1) Cleveland and McGill Framework:

a) Angle: The degree of the angle represents the value of the measurements. Figure 1a shows the lines with an angle as a representation of the blood pressure.

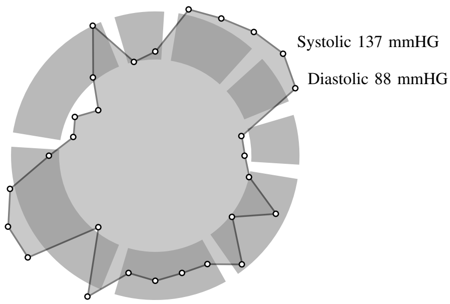


Fig. 2: The hGraph representation of the data with two example labels.

b) Area: The size of a shape represents the value of the measurements. Figure 1b shows the blood pressure represented as squares where the size represents the value.

c) Length: The length of the line represents the value of the measurements. Figure 1c shows the length of the lines representing the values of the blood pressure.

d) Position Along Aligned Scales: The value of a measurement is represented by the position. The position of the circles in figure 1d represent the values of the blood pressure.

2) hGraph: The design emphasizes the presentation of a large number of parameters and how they conform a holistic overview of the health of a person. Figure 2 shows the hGraph of the modeled data with metabolic syndrome. Multiple parameters are arranged as polar coordinates and a figure connecting the coordinates shows how well the measurements are with respect to the recommended values (green area).

3) Control Group: The control group received a table with numeric values. These participants relied only on numeric values and recommended ranges.

C. Experiment Protocol

The experiment started with a self-assessment of participants' knowledge on health and wellbeing, i.e. their familiarity with health and risk behaviors (diet, sleep, physical activity, etc) and how they affect one's health and wellbeing. Additionally, we asked the participants how familiar they think they are with the risks of developing T2D.

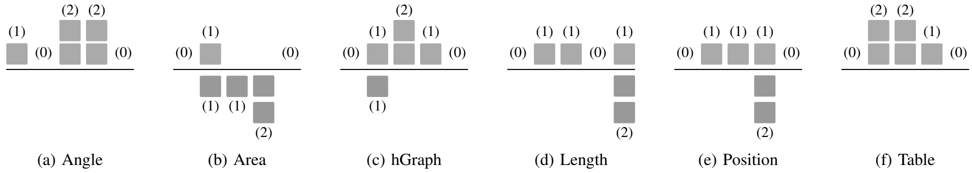


Fig. 3: Assessments of overall health per visualization, ordered left from right according to the subjects' Health Literacy, i.e. familiarity on the risks of T2D (from "Very familiar" to "Not familiar at all"). Green refers to a correct and red to an incorrect assessment.

Block randomization, where only one of the visualizations (or the table) was shown to each subject, was utilized. We used a web browser to render the graphics (JavaScript and SVG). The whole session was recorded with a video camera.

We applied the insight-based methodology proposed by North [18]. We asked the participants to "see beyond the figures" and explain the relationships between the measurements in order to gain understanding of the health situation as well as to identify the possible underlying reasons. We asked the subject to tell us as many insights as possible within ten minutes. Participants were encouraged to use a think aloud process.

Afterwards, we asked the participants to perform an investigative analysis [19] by assessing the overall health and wellbeing as well as the risks of developing T2D of the two datasets. We used a scale of five for both tasks. The options for the overall health and wellbeing assessment ranged from *"Very Poor Health"* to *"Very Good Health"* and for the T2D risk from *"Very High Risk"* to *"Very Low Risk"*.

We used depth and complexity [18] to evaluate the insights. Similar to [20], we used a five-point scale. Clinically incorrect insights did not give any points. Superficial insights that stated e.g. that a value is low or high were rated as one point. Assumptions and relationships between measurements were rated as a value of two. For instance, we assigned two points if the participant mentioned that some measurements are related or that "the person does not exercise" with no explanation. Explanations of the insights were rated as three points. The explanations had complexity, depth or both. For our case, complexity refers to the relationships between measurements and depth to the underlying reasons of the values. When a participant formulated a hypothesis that described the relationship of measurements and the possible reasons behind the values, we evaluated the insight with four points. In order to obtain five points, the participants had to relate all the measurement groups in a single hypothesis explaining the underlying reasons for the values. The insights were evaluated blindly by the authors of this article following the before-mentioned criteria.

TABLE I: Proportion of correct assessments of the T2D risk and overall health at different levels of self-assessed familiarity with T2D risk factors (Health Literacy).

Familiarity	Risk (%)	Health (%)
Very familiar (1)	100%	100%
Familiar (8)	100%	75%
Somewhat familiar (10)	90%	80%
Slightly familiar (8)	75%	50%
Not familiar at all (3)	66%	33%

III. RESULTS AND DISCUSSION

Correct assessment of T2D risk and overall health depended on the participants' familiarity with T2D risk factors (see table I). The correct assessment of overall health and the general knowledge on health and wellbeing did not have such dependency (data not shown). Therefore we selected familiarity with T2D risk factors as our measure of Health Literacy.

Figure 3 shows the assessments of the overall health with different visualization techniques. The *area* had the least correct assessments as shown in figure 3b. Figure 3a and 3c show that the *angle* and *hGraph* supported participants with low Health Literacy to assess the dataset correctly, as compared with *lenght* and *position*. Additionally, figure 3f shows that the *table* also supported correct assessments.

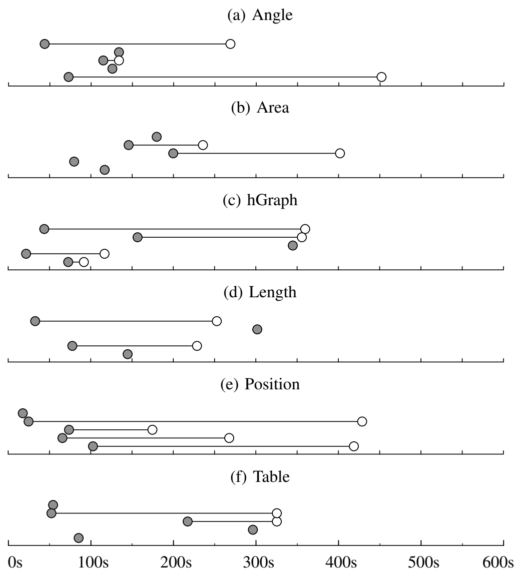


Fig. 4: Time to first insight value three and time to first hypothesis are shown as blue and white circles respectively.

As in Saraiya et al. [20], we computed the time to the first insight. We used the time to first insight of value three, namely when a participant was able to give explanations substantiating the observation. Figure 4 shows the time to first insight of value three as a dot and the time to the first generated hypothesis (insight values as 4 or 5) as a circle. In some cases participants were not able to generate hypothesis or formulate insights of value three at all (as in figure 4d).

Four out of five participants that used the *position* and *hGraph* generated hypothesis, figure 4c and 4e. Participants using the *hGraph* generated six hypotheses in total, while participants with the *angle* and *position* generated a four each, see table II. The participants using the *table* generated a total of four hypotheses, however one participant generated 3 of them.

The average time to insights value three or more is also shown on table II. The longest average time occurred using the *area*, *table* and *length*. The speed is similar to Cleveland and McGill's results [10] except for the *length*. However, the lowest Health Literacy was observed in this group.

TABLE II: The table shows the average time to first insight of value three or more, the average Health Literacy of the participants, the number of hypothesis (insights value four or five) in total and per participant.

Visualization	Time (s)	Health Literacy	Hypothesis	
			Total	Per participant
Angle	98	3.2	4	(0,0,2,1,1)
Area	144	3.0	3	(0,0,2,1,1)
hGraph	128	3.2	6	(2,0,1,2,1)
Length	139	2.0	3	(2,0,1,0,0)
Position	57	2.6	4	(0,1,1,1,1)
Table	141	3.2	4	(0,0,1,0,3)

The evaluation methodology gives low points to obvious insights, such as "physical activity seems to be low, the patient should start exercising more". However, such insight is clinically valuable. Therefore, future work is needed to further develop the methodology.

IV. LIMITATIONS

All the participants in our experiment had an academic degree and in some cases an engineering background. The sample of 30 participants cannot be representative of a wider population.

V. CONCLUSIONS AND FUTURE WORK

In this paper we applied the insight-based methodology for the health data visualizations. Adequate visualization of holistic health data was shown to help users without medical background to better understand the overall health situation and possible health risks related to lifestyles. Visualizations could also be valuable tools for the medical experts to enhance the healthcare process.

At the end of the experiment session, participants selected three visualizations that would have helped them better perform in the previous tasks. We also administered a short form of Raven Advanced Progressive Matrices test [21]. Further study will be addressed in future research.

We evaluated the insights using an informed but subjective assessment based on the principles from North [18]. For future work, we aim to measure the insight value by modeling prior and derived knowledge. This model may provide a more accurate and reliable estimate of the performance of different health data visualizations. The goal is to provide customizable visualizations to support better understanding of personal health and wellbeing.

REFERENCES

[1] G. Rose *et al.*, "The strategy of preventive medicine." *The strategy of preventive medicine.*, 1992.

[2] World Health Organization. (1948) WHO definition of health. [Online]. Available: http://www.who.int/about/definition/en/print.html

[3] B. Shneiderman, C. Plaisant, and B. Hesse, "Improving health and healthcare with interactive visualization methods," HCIL Technical Report, Tech. Rep., 2013.

[4] A. Bui, D. Aberle, and H. Kangarloo, "Timeline: Visualizing integrated patient records," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 11, no. 4, pp. 462–473, July 2007.

[5] B. J. Lesselroth and D. S. Pieczkiewicz, *Data visualization strategies for the electronic health record*. Nova Science Publisher's, Incorporated, 2011.

[6] S. Consolvo, D. W. McDonald, T. Toscos, M. Y. Chen, J. Froehlich, B. Harrison, P. Klasnja, A. LaMarca, L. LeGrand, R. Libby, I. Smith, and J. A. Landay, "Activity sensing in the wild: A field trial of ubifit garden," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '08. New York, NY, USA: ACM, 2008, pp. 1797–1806. [Online]. Available: http://doi.acm.org/10.1145/1357054.1357335

[7] F. Bentley, K. Tollmar, P. Stephenson, L. Levy, B. Jones, S. Robertson, E. Price, R. Catrambone, and J. Wilson, "Health mashups: Presenting statistical patterns between wellbeing data and context in natural language to promote behavior change," *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 20, no. 5, p. 30, 2013.

[8] T. Goetz, "It's time to redesign medical data," 2015. [Online]. Available: http://www.ted.com/talks/thomas_goetz_it_s_time_to_redesign_medical_data?language=en

[9] F. Jonathan and J. Sonin, "hgraph: An open system for visualizing personal health metrics," http://hgraph.org/docs/hGraph_Whitepaper.pdf, involution studios, Arlington, Massachusetts, Tech. Rep., April 2012, [Online; accessed 25-March-2015].

[10] W. S. Cleveland and R. McGill, "Graphical perception: Theory, experimentation, and application to the development of graphical methods," *Journal of the American statistical association*, vol. 79, no. 387, pp. 531–554, 1984.

[11] J. Baird and E. Noma, *Fundamentals of scaling and psychophysics*, ser. A Wiley-Interscience publication. John Wiley & Sons Canada, Limited, 1978. [Online]. Available: https://books.google.fi/books?id=F3V-AAAAMAAJ

[12] C. Baur, *National Action Plan to Improve Health Literacy*. US Department of Health and Human Services, Office of Disease Prevention and Health Promotion, 2010.

[13] J. Lindström and J. Tuomilehto, "The diabetes risk score a practical tool to predict type 2 diabetes risk," *Diabetes care*, vol. 26, no. 3, pp. 725–731, 2003.

[14] R. Laukkanen, P. Oja, M. Pasanen, and I. Vuori, "Validity of a two kilometre walking test for estimating maximal aerobic power in overweight adults." *International journal of obesity and related metabolic disorders: journal of the International Association for the Study of Obesity*, vol. 16, no. 4, pp. 263–268, 1992.

[15] Firstbeat Technologies Ltd. (2014) Stress and recovery analysis method based on 24-hour heart rate variability. white paper. [Online]. Available: http://www.firstbeat.com/userData/firstbeat/research-publications/Stress-and-recovery_white-paper_2014.pdf

[16] M. F. Scheier, C. S. Carver, and M. W. Bridges, "Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): a reevaluation of the life orientation test." *Journal of personality and social psychology*, vol. 67, no. 6, p. 1063, 1994.

[17] The Finnish Medical Society Duodecim. Current care guidelines. [Online]. Available: http://www.kaypahoito.fi/web/english

[18] C. North, "Toward measuring visualization insight," *Computer Graphics and Applications, IEEE*, vol. 26, no. 3, pp. 6–9, 2006.

[19] Y. ah Kang, C. Gorg, and J. Stasko, "Evaluating visual analytics systems for investigative analysis: Deriving design principles from a case study," in *Visual Analytics Science and Technology, 2009. VAST 2009. IEEE Symposium on*, Oct 2009, pp. 139–146.

[20] P. Saraiya, C. North, and K. Duca, "An evaluation of microarray visualization tools for biological insight," in *Information Visualization, 2004. INFOVIS 2004. IEEE Symposium on*. IEEE, 2004, pp. 1–8.

[21] W. Arthur and D. V. Day, "Development of a short form for the raven advanced progressive matrices test," *Educational and Psychological Measurement*, no. 54, pp. 394–403, 1994.

PUBLICATION

II

Implementation and user testing of a system for visualizing continuous health data and events

M. Al-Musawi, A. Ledesma, H. Nieminen and I. Korhonen

*2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*2016, 156–159

Publication reprinted with the permission of the copyright holders

Implementation and User Testing of a System for Visualizing Continuous Health Data and Events*.

Mohammed Al-Musawi[1], Andres Ledesma[1], *Student Member IEEE*, Hannu Nieminen[1], *Member IEEE*, Ilkka Korhonen[1], *Senior Member IEEE*

Abstract— **Efficient ways are needed to visualize the health status of a person and how the lifestyle, daily choices and health care actions are affecting it. Current systems lack a comprehensive interface for interaction and exploration of large and complex data and events affecting the data. Based on state-of-the-art data visualization techniques, we implemented and user tested a system that visualizes health data holistically over time. The system focuses on the dynamic changes by using a timeline of events affecting the overall health status. We conducted an extensive user testing process involving surveys, heuristics and observations in order to evaluate our system. The results show that our system has a high level of User Satisfaction while providing an adequate understanding, interaction and navigation of the data.**

I. INTRODUCTION

The World Health Organization estimates that 77% of diseases in Europe relate to a "chronic diseases" [1]. The "majority of type II diabetes, coronary heart disease and strokes" can be avoided by putting "more effort on preventative health care" by "supporting the behavior change towards healthier" lifestyles reducing "significant costs" [2].

To be able to follow up the health progress, we need efficient ways to visualize the holistic health status of a person and how lifestyles, daily choices and health care actions affect it. Health care professionals also need this information, including for example health coaches, nurses and physiotherapists. A holistic description of health and its progress should combine together different measures of physical, mental and social wellbeing [3][4]. This description should also link the measures to the daily choices and interventions utilized to improve the health.

Good usability is essential to ensure that the information can be understood and effectively utilized by the different users. In the 21st century, the growing use of information technology in health-related fields has resulted in increased significance of evaluation studies in usability [5].

Many "studies suggest that healthcare information systems suffer from numerous usability problems" [6]. Usability issues are especially critical for data visualization, which is a vital building block for ease of use, user satisfaction, and efficiency [7].

Effective visualization enables both, health care providers and patients, to gain better insight into the holistic health data, enabling better management of the patient [8]. Graphical visualizations can significantly assist in decision-making processes reducing the cognitive load on the caregiver by presenting holistic overviews of a patient's data [9][10][11].

We built an information system for the continuous visualization of health status and events potentially affecting it. The system visualizes the data sets comprising a holistic overview of the patient state. The system represents events that affect the health status as timeline elements. These events are e.g. daily actions the user does with a goal to improve his health (health interventions), laboratory tests, visits to a doctor, or medication. Health interventions are typically linked to participation in health coaching, selfcare or homecare programs. The integration of the system resulted in a full scale application that visualizes holistic health and related events. As an example use case, we modeled a health coaching program targeted for patients with an increased risk for cardiovascular diseases.

We conducted an evaluation of the system with general and expert users, using three different methods, namely heuristic based expert evaluation, survey, and observation. Standardized usability questionnaires [12][13][14] were used as a technique to measure different usability metrics and overall user satisfaction. Additionally we compared the results between them in order to obtain higher reliability of the data and to assess the performance of these questionnaires.

II. DESIGN AND IMPLEMENTATION

We designed a system comprised of three parts representing health data over time. Two components represent a snapshot of the status of a modeled patient's health and wellbeing at a given time. The third component represents the dynamic behavior of events affecting the status of the modeled patient.

a) Health Figures: A previous study [15] showed that the Health Graph (hGraph) developed by Sonin and Follet [16] encouraged the greatest number of insights about health. This data visualization is based on polar coordinates where each point represents a measurement positioned between an area representing the optimal value range. Inspired by the hGraph, we integrated the Health Figures (hFigures) library [17] which allows multiple graphs, animations and color coded points.

b) Curves as Time Series: This component represents the time series values of the measurements with a minimum and maximum recommended range. This component is a sequential longitudinal representation of the measurements.

*This research was supported by the European Commission and TEKES under the ARTEMIS-JU WithMe project.

[1] Personal Health Informatics research group, Department of Signal Processing, Tampere University of Technology, Tampere, Finland
{mohammed.al-musawi, andres.ledesma, hannu.o.nieminen}@tut.fi

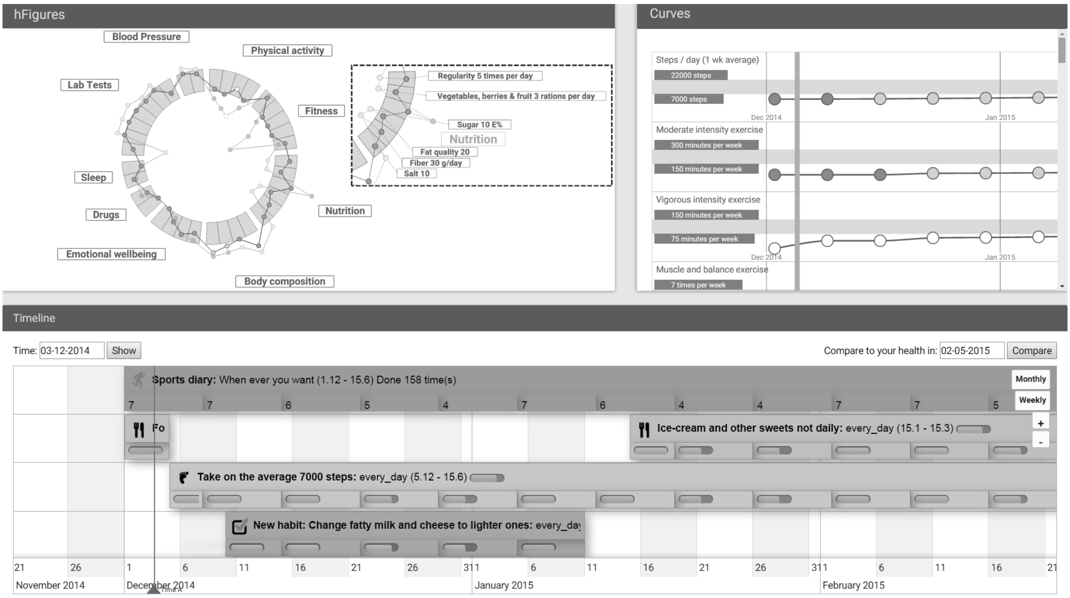


Fig. 1: The implemented system. The hFigures (upper left corner) displays measurement labels when zooming into figure.

c) Timeline: We used a timeline to represent a series of events that affect the values of the measurements. In the case of our study, the health coaching program has a series of tasks with a different degree of completion affecting the overall health situation.

A. Data Model

We modeled the visualized data as measurements and health coaching interventions of a patient at risk of developing type II diabetes. The following parameters were used to represent the patient's health:

- Blood pressure: systolic and diastolic blood pressure
- Physical activity: weekly active days, steps per day
- Body composition: Body Mass Index (BMI), waist diameter and fat percentage
- Sleep: time in bed, time asleep
- Fitness: resting heart rate, fitness index, muscular force, muscular endurance and balance [18]
- Lab Tests: hemoglobin, fB-Gluc, cholesterol, HDL, LDL, triglycerides
- Nutrition: meal regularity, type of meals (vegetables and fruits), sugar intake, fat quality, fiber and salt intake.
- Drugs: tobacco (cigarettes per day), alcohol abuse, drug abuse (narcotics), medication abuse
- Emotional wellbeing: depression level (DEPS), stress level and stress recovery [19] as well as optimism [20].

The recommended values were derived from the Finnish Medical Society Duodecim [21].

III. METHODS

We designed a user evaluation study and recruited 14 participants. Three of them were experts in usability. All participants had university level education. The purpose of the evaluation study was to obtain insights from expert and non-expert users regarding the usability of the data visualization of the system. The aim was to obtain better understanding of the effectiveness of the system as a tool to provide users a graphical understanding of the health data and the events that affect it. Our user evaluation had the following goals:

1) Identify any usability issues such as System Usefulness, Information Quality, Interface Quality, User Satisfaction, Ease of Use, Ease of Learning, Effectiveness, and Efficiency.
2) Assess the user performance.
3) Identify the tasks and functions which present challenges to the user.
4) Identify aspects that need further development.
5) Assess and compare the overall user satisfaction towards the solution between different self-supported metrics techniques, and assess the performance of the standard questionnaires used in the study.

We used three methods in the user evaluation study, namely usability heuristics [22], survey, and observation. The usability heuristics method was used only with individuals experienced in human cognition and interface design rules. The aim was to identify usability problems in the user interface and to obtain suggestions for corrective action. The Walk-through technique and the Heuristic questionnaire were used as techniques for this method.

We used the Survey method with all participants to measure user satisfaction in different usability metrics. We used two techniques for this method, namely interview and standard usability questionnaires - After-Scenario Questionnaire (ASQ), Computer System Usability Questionnaire (CSUQ), and USE Questionnaire - and compared the overall user satisfaction results between the different questionnaires. The questionnaires use a 7-point Likert scale, where 1 means 'Strongly disagree', and 7 means 'Strongly agree'.

We used the Observation method with all participants to collect information about the interaction between the participants and the system. The techniques we used in this method were direct observations, note keeping and video recording. Table I illustrates the techniques and types and number of participants for the different methods.

TABLE I: Test methods, techniques, types and number of participants.

Method	Techniques	Participant Type	# Participants
Heuristic	Walk-through Questionnaire	Expert	3
Survey	Interview Questionnaires	Expert/ General users	14
Observation	Direct Observation Notes and video recording	Expert/ General users	14

We developed three scenarios comprised of tasks organized in order of difficulty, with the easiest task first. The tasks consisted of a set of questions that the participants answered by using the system (interactive exploration of the data set). Two of the three scenarios contained questions related to the understanding of the hFigures data visualization that represents the overall health status at a certain moment. The third scenario contained questions related to the Timeline component which represents the series of events that affect the overall health status continuously.

We conducted the evaluation in a controlled environment, where one participant at a time conducted the series of tasks and questionnaires accompanied by one researcher. Each participant was sitting in front of a computer with a video camera recording the session. In the beginning of the session, an introduction of the study was given, and the participant read and signed an informed consent form. We also collected information about the participants, such as age, gender, and possible prior experience in using health and wellness software. The participants were able to ask questions at any time during the session.

The participant spent five minutes exploring the solution using a thinking aloud process. Afterwards, the scenarios were explained and the tasks given. After finishing the tasks of each scenario, the participant was requested to answer the ASQ and to provide feedback regarding the scenario and tasks. At the end of the session, the participant answered the CSUQ and USE Questionnaire. Participants also provided feedback regarding the system. Expert participants conducted a walk-through process of the system and answer the ten-question heuristic questionnaire in order to evaluate the user interface and to spot possible problems not necessarily discovered by the non-expert participants.

System Usefulness, Information Quality and Interface Quality are the main three metrics that can be obtained by examining the CSUQ. The four internal metrics or subscales of the USE Questionnaire refer to System Usefulness, Ease of Use, Ease of Learning, and User Satisfaction. User satisfaction was also measured after each scenario by using an ASQ. This questionnaire addressed the issues of ease of completing the tasks, i.e. effectiveness, and amount of time it took to complete the tasks, i.e. efficiency.

IV. RESULTS AND DISCUSSION

Based on the evaluation data analysis, using averaging of the values from the standard questionnaires, the evaluation study revealed that users were generally satisfied with the solution. From the standard questionnaires used in the study, it can be seen that satisfaction with the usability of this solution was positive in all the metrics measured.

As Table 2 below shows, the satisfaction rate was highest, above 6 out of 7 in the Likert scale, with efficiency (6.60) and effectiveness (6.55), ease of learning (6.50), interface quality (6.24), and system usefulness (6.13). Further modifications and development of the system are needed in the spheres of ease of use and information quality, which scored 5.94 and 5.66 respectively, by addressing the issues which the participants disliked in the system.

TABLE II: Usability metrics summary for the solution.

Metric	Average response	Standard deviation
System Usefulness	6.13	.930
Information Quality	5.66	1.20
Interface Quality	6.24	.999
Ease of Use	5.94	1.05
Ease of Learning	6.50	.620
Effectiveness	6.55	.739
Efficiency	6.60	.587

When comparing the overall user satisfaction results obtained by utilizing different self-reported metrics tools, it can be seen that the results are relatively uniform, with little dispersion between them, and a relatively low standard deviation rate, as shown in Table 3. This indicates high reliability of the results obtained from the participants.

The Nielsen heuristic questionnaire was answered by the three expert participants. All experts agreed or strongly agreed on all but one of the indicators. One expert found the system to require extensive use of his memory. This expert needed to spend time to distinguish which lines in the hFigures corresponded to the specific times in the Timeline.

TABLE III: The overall user satisfaction result of the questionnaires.

Questionnaire Type	Average response	Standard deviation
After-Scenario Questionnaire	6.46	.531
Computer System Usability	6.02	1.04
USE Questionnaire	6.11	.816
Nielsen heuristic evaluation	6.30	.562

User performance was measured utilizing the method of observation, and we analyzed the data and obtained the following results. With regard to scenarios and tasks that focused on the Progress data component, all 14 participants completed all the tasks successfully, with only a few non-crucial errors in total, and with most of the tasks completed within benchmark time. Benchmark times were selected based on estimated average completion time, and the required level of interaction with the system. The results for the scenario and the related tasks focusing on the components of hFigures and Curves showed that 75% of the tasks were successfully completed by all 14 participants. Finding the measurements in the area of health in the hFigures component proved the most difficult one for the participants, with 3 out of the 14 not completing the task. Only 3 out of 8 tasks were completed without non-crucial errors, but the number of errors was small in the remaining five tasks and did not significantly hamper the accomplishment of the task. The average benchmark time was exceeded only in 3 tasks, and even in two of them only slightly.

After analyzing all the data and feedback from the participants, the following areas of improvement were identified. The three most urgent issues need to be solved to improve user interaction with the system and to increase the users ability to complete their tasks expediently using the system without excessive load on the memory. These issues were: 1. In hFigures, when people click on the measurement, the title of the measurement is not visible due to insufficient zooming; 2. People face difficulty to distinguish between the two specific moments of time between which they choose to compare the health status in hFigures; 3. People need a tutorial or help document.

These issues could be addressed by displaying titles of the measurements when users click on them, adding legends or labels in the hFigures to distinguish between the two moments in time, and implementing a help document to explain the purpose of each part of the solution.

V. CONCLUSIONS AND FUTURE WORK

Users demonstrated high satisfaction with the usability of the system, which indicates that the information visualization techniques used in the system were successful. Moreover, this research strongly indicates that the use of different methods and standardized questionnaires in the evaluation phase increases the reliability of the results. In addition, the different tools produced similar results, which suggests that the tools perform in a similar way. The results also show that successful visualization can assist individuals to understand their holistic health and wellness data.

The evaluation study was conducted on people with a certain level of education, which may present certain bias. It would be beneficial to conduct an evaluation study with small groups of individuals representing different levels of education and computer literacy, and potential real users, such as patients, nurses, and trainers. Other further work could include evaluating the design of the interface on devices with different screen sizes, such as tablets and smart phone. The application's interaction through touch screen technology should be carefully assessed. For smart phones, a new interface, designed for smaller screens, should be designed, implemented, and evaluated, putting emphasis on the context of use.

REFERENCES

[1] World Health Organization. (2015) Noncommunicable diseases. [Online]. Available: http://www.euro.who.int/en/health-topics/noncommunicable-diseases

[2] H. Nieminen, S. Baggström, H. Lehtonen, H. Mikkola, A. Leppisaari, and I. Korhonen, "Mobile solution for increasing the efficiency of health coaching for preventative health care and chronic disease management," *35th Annual International IEEE EMBS Conference*, vol. 51, no. Supplement, pp. M–153, 2013.

[3] World Health Organization. (1948) WHO definition of health. [Online]. Available: http://www.who.int/about/definition/en/print.html

[4] B. Shneiderman, C. Plaisant, and B. Hesse, "Improving health and healthcare with interactive visualization methods," HCIL Technical Report, Tech. Rep., 2013.

[5] J. Viitanen and M. Nieminen, *Usability evaluation of digital dictation procedure-an interaction analysis approach*. Springer, 2011.

[6] J. Viitanen, H. Hyppönen, T. Lääveri, J. Vänskä, J. Reponen, and I. Winblad, "National questionnaire study on clinical ict systems proofs: physicians suffer from poor usability," *International journal of medical informatics*, vol. 80, no. 10, pp. 708–725, 2011.

[7] M. Khan and S. S. Khan, "Data and information visualization methods, and interactive mechanisms: A survey," *International Journal of Computer Applications*, vol. 34, no. 1, pp. 1–14, 2011.

[8] M. de Ridder, L. Constantinescu, L. Bi, Y. H. Jung, A. Kumar, J. Kim, D. D. Feng, and M. Fulham, "A web-based medical multimedia visualisation interface for personal health records," in *Computer-Based Medical Systems (CBMS), 2013 IEEE 26th International Symposium on*. IEEE, 2013, pp. 191–196.

[9] A. Cabarcos, T. Sanchez, J. A. Seoane, V. Aguiar-Pulido, A. Freire, J. Dorado, and A. Pazos, "Retrieval and management of medical information from heterogeneous sources, for its integration in a medical record visualisation tool," *International journal of electronic healthcare*, vol. 5, no. 4, pp. 371–385, 2010.

[10] L. Chittaro, "Information visualization and its application to medicine," *Artificial intelligence in medicine*, vol. 22, no. 2, pp. 81–88, 2001.

[11] T. Le, K. Wilamowska, G. Demiris, and H. Thompson, "Integrated data visualisation: an approach to capture older adults' wellness," *International journal of electronic healthcare*, vol. 7, no. 2, pp. 89–104, 2012.

[12] J. R. Lewis, "Psychometric evaluation of an after-scenario questionnaire for computer usability studies: the asq," *ACM SIGCHI Bulletin*, vol. 23, no. 1, pp. 78–81, 1991.

[13] ——, "Ibm computer usability satisfaction questionnaires: psychometric evaluation and instructions for use," *International Journal of Human-Computer Interaction*, vol. 7, no. 1, pp. 57–78, 1995.

[14] A. M. Lund, "Measuring usability with the use questionnaire," *Usability interface*, vol. 8, no. 2, pp. 3–6, 2001.

[15] A. Ledesma, H. Nieminen, P. Valve, M. Ermes, H. Jimison, and M. Pavel, "The shape of health: A comparison of five alternative ways of visualizing personal health and wellbeing," in *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, Aug 2015.

[16] F. Jonathan and J. Sonin, "hgraph: An open system for visualizing personal health metrics," http://hgraph.org/docs/hGraph_Whitepaper.pdf, involution studios, Arlington, Massachusetts, Tech. Rep., April 2012, [Online; accessed 25-March-2015].

[17] A. Ledesma and H. Nieminen. (2015) Health Figures. [Online]. Available: https://gitlab.com/ledancs/hFigures

[18] R. Laukkanen, P. Oja, M. Pasanen, and I. Vuori, "Validity of a two kilometre walking test for estimating maximal aerobic power in overweight adults." *International journal of obesity and related metabolic disorders: journal of the International Association for the Study of Obesity*, vol. 16, no. 4, pp. 263–268, 1992.

[19] Firstbeat Technologies Ltd. (2014) Stress and recovery analysis method based on 24-hour heart rate variability. white paper. [Online]. Available: http://www.firstbeat.com/userData/firstbeat/research-publications/Stress-and-recovery_white-paper_2014.pdf

[20] M. F. Scheier, C. S. Carver, and M. W. Bridges, "Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): a reevaluation of the life orientation test." *Journal of personality and social psychology*, vol. 67, no. 6, p. 1063, 1994.

[21] The Finnish Medical Society Duodecim. Current care guidelines. [Online]. Available: http://www.kaypahoito.fi/web/english

[22] J. Nielsen, *Usability engineering*. Elsevier, 1994.

PUBLICATION
III

Health figures: an open source JavaScript library for health data visualization

A. Ledesma, M. Al-Musawi and H. Nieminen

BMC Medical Informatics and Decision Making 16.38 (2016), 1

Publication reprinted with the permission of the copyright holders

Ledesma *et al. BMC Medical Informatics and Decision Making* (2016) 16:38
DOI 10.1186/s12911-016-0275-6

BMC Medical Informatics and Decision Making

SOFTWARE Open Access

Health figures: an open source JavaScript library for health data visualization

Andres Ledesma*, Mohammed Al-Musawi and Hannu Nieminen

Abstract

Background: The way we look at data has a great impact on how we can understand it, particularly when the data is related to health and wellness. Due to the increased use of self-tracking devices and the ongoing shift towards preventive medicine, better understanding of our health data is an important part of improving the general welfare of the citizens. Electronic Health Records, self-tracking devices and mobile applications provide a rich variety of data but it often becomes difficult to understand. We implemented the hFigures library inspired on the hGraph visualization with additional improvements. The purpose of the library is to provide a visual representation of the evolution of health measurements in a complete and useful manner.

Results: We researched the usefulness and usability of the library by building an application for health data visualization in a health coaching program. We performed a user evaluation with Heuristic Evaluation, Controlled User Testing and Usability Questionnaires. In the Heuristics Evaluation the average response was 6.3 out of 7 points and the Cognitive Walkthrough done by usability experts indicated no design or mismatch errors. In the CSUQ usability test the system obtained an average score of 6.13 out of 7, and in the ASQ usability test the overall satisfaction score was 6.64 out of 7.

Conclusions: We developed hFigures, an open source library for visualizing a complete, accurate and normalized graphical representation of health data. The idea is based on the concept of the hGraph but it provides additional key features, including a comparison of multiple health measurements over time. We conducted a usability evaluation of the library as a key component of an application for health and wellness monitoring. The results indicate that the data visualization library was helpful in assisting users in understanding health data and its evolution over time.

Keywords: Data visualization, Health data, Health informatics, Javascript

Introduction

The ongoing shift from reactive to preventive medicine requires that citizens have the skills and means to take an active role in developing and maintaining their wellness. Use of self-tracking devices and personal wellness applications is more and more common, ranging from sports tracking applications to personal genome sequence analysis services. These services and devices produce large amounts of data. In addition, Electronic Health Records are increasingly replacing paper records in hospitals and clinics around the world. The combination of these large and heterogeneous data sources is expected to provide a "predictive, preventive, personalized and participatory" ecosystem for the benefit of the general welfare [1].

To better understand our health, we need to combine heterogeneous data sources and present the information to the user in a complete and accurate manner. In order to accomplish this, health information technologies and visualization design need to be integrated [2]. The goal is to provide tools for individuals to take better decisions regarding their health. Similarly, doctors and other medical experts need tools and solutions for getting a complete and accurate view of the patients health, combining together patient's own measurements and clinical data.

An innovative approach for health data visualization is the Health Graph (hGraph), released publicly by MITRE corporation [3]. In an earlier study [4], it was found that the hGraph-type radial plot can enhance deep

*Correspondence: andres.ledesma@tut.fi
Department of Signal Processing, Tampere University of Technology, Korkeakoulunkatu 10, 33720 Tampere, Finland

© 2016 Ledesma et al. **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

understanding of health data and enable the user to create meaningful health insights based on the interrelationships between the measurements. However, the hGraph shows a static overview of a persons wellness. Disease and wellness are processes that change over time. It is also essentially important to be able to follow up the trajectories in the different parameters, what is the rate of the change and how they respond to events such as medical care actions and interventions. Thus, in addition to a static snapshot such as in hGraph, a temporal way of presenting the data is also needed. While hGraph visualization is useful for the purpose of understanding with a quick glance the overall situation, it lacks features such as a clear division of measurements according to their category, a distribution of the labels to avoid clutter and the notion of time or *evolution* of the data [4].

In this article we present a visualization library based on extending the core ideas of the hGraph. Aim of the library is to provide tools to assist experts and non-experts in the decision making process of assessing the situation of a patient and its evolution over time. To test the user satisfaction, ease-of-use and usefulness of the solution, we built an application for health monitoring using the visualization library. The application is a visualization tool that shows a health coaching program and its effects on the evolution of health and wellness of a modeled patient. We evaluated the software solution using usability tests (Heuristic Evaluation, Controlled User Testing and Usability Questionnaires).

First in this article we review the state-of-the-art on health data visualizations and describe the hGraph and results related to its usability and usefulness, which motivated us to develop hGraph further. We named the new library built on top of the hGraph core ideas as hFigures. The article details the design and implementation of the hFigures library, the features we implemented and how they address the users' needs. We also describe the usability test process we utilized to assess the library in the context of an application for health monitoring. We present the results and discuss the further improvements of the library and its pitfalls. We also discuss how this library can be used in Personal Health Informatics and in the Health Care processes.

Background

Visualization tools have mostly focused on Healthcare Information Systems and Electronic Health Records (EHR) [2]. For instance, TimeLine is a software developed to retrieve data from several sources and presented in a hierarchical and timeline based structure where clinicians can browse chronologically through existing EHRs including MRI [5].

Additionally, the growing market for mobile health applications (mHealth) have drawn the attention of researchers, developers and investors [6]. These applications provide large volumes of personal health data. While the market and demand are expected to grow, the use of the data has the potential to contribute to a better understanding of our health.

Goetze [7] demonstrated the impact of data visualization as means to represent health data in a complete and accurate manner. He conducted a project that redesigned laboratory test results from numerical tables into colored graphics. He demonstrated that the patients were able to understand better their health situation when presented with the new designs.

Data integration for health monitoring as a Big Data process for personalized medicine has been approached by Idris et al. [8]. The visualization of this information uses traditional bar and pie charts to report to the user a historical view of a variety of data including mental, social, physical aspects. The novelty of this work is the integration of heterogeneous data sources while the presentation of the information was done following existing graphical representations.

An extensive choice of graphical representation is listed and explained by S. Few [9]. These techniques have been studied and used widely among researchers and individuals alike. Examples include: bar, stacked bar, line and bullet graphs. These visualizations can be combined to provide a personalized wellness indicator system, as proposed by Soomlek and Benedicenti [10].

DeRidder et al. developed a combined approach that retrieves data from Personal Health Records (PHRs), and presents them to individual patients using a "3D medical graphical avatar" [11]. The solution is built using HTML5 and WebGL to render 3D graphics using the web browser. Patients can browse "regions of interests" on their avatar and explore further the information contained in EHRs as well as in PHRs.

As stated by Shneiderman et al. [12], new visual representations are needed for "systematic yet flexible visual analytics processes". We present an existing tentative solution known as the hGraph, released publicly by MITRE corporation. We describe its main advantages and how they can address these challenges. In a previous study [4] we identified possible improvements and based on our own implementation we addressed these issues and extended the features of the hGraph to better address large data sources.

However, a data visualization is only as good as the ability of the intended audience to decode graphical objects into numerical values which conveys a clear message. Therefore, we have to consider the graphical perception of the users who will benefit from the data visualization. Graphical perception is the ability of an individual to decode the information displayed as graphical objects [13], it has been a widely researched field [14]. Graphical perception affect how we understand visualized information. In the context of health data, it remains a challenge to design graphical representations for non-medical experts. Graphical representations in this context should enhance the users' ability to understand their health situation and take informed decisions. With this "deep" understanding on the health situation, individuals can move towards healthier behaviors.

The graphical representation of health data requires a complete and accurate overview of a patient often including large amounts of measurements, which in turn translates to large datasets. Therefore, health data visualizations need to scale to accommodate large datasets.

Radar visualization

Radar visualization scales to large amounts of data entries because the points are distributed among the circumference. Bar charts, lines or scattered plots, pie charts and similar visualizations can quickly grow to a large scale with large datasets. In the case of radar visualizations, the graphical representation scales with the number of points plotted causing the circumference to grow in order to accommodate all the plotted values.

Hoffman et al. seem to have coined the term of "radial visualization" [15, 16]. The term was the foundation of " pie chart, starplot, and radar plot" which are the basis of "virtually all the radial visualization methods found in the state-of-the-art research" [16].

Draper et al. propose a classification of radial visualizations into "three main divisions" each comprised of "design patterns" [16].

Polar plots

These are radar visualizations where the center is the starting point from which "line segments" originate [16]. According to this classification, polar plots are divided into **Tree** and **Star** patterns.

Trees have segments that "branch off" and are mostly used for visualizing hierarchical data [16]. Examples include *Moiregraph* [16, 17] and the *Hyperbolic Browser* [16, 18].

Stars do not have branches but rather straight segments originating from the center, their common uses include "ranking of search results" and "viewing relationships among disparate entities" [16]. Examples of star patterns include *Starstruck* [16, 19] and *Neighbourhood Explorer* [16, 20].

Space filling

Also referred to as *Radial Space Filling* (RSF) [16, 21], this category comprises the visualizations that fill "most or all of the area of a circle" [16]. The classification identifies three patterns: **Concentric**, **Spiral** and **Euler**. These patterns are mostly used for visualizing hierarchical data and "viewing relationships among disparate entities", except for the Spiral pattern which is used to visualize "serial periodic data" [16].

Concentric pattern have an "origin at or near center of canvas" from where rings are plotted outwards and "each ring divided into multiple sectors" [16]. Filelight is an example of concentric pattern and it is a "filesystem browser based in part on the Polar TreeMap metaphor" [16, 22].

Spiral pattern consists of a "spiral-shaped glyph" that starts from the center of the canvas [16]. Certain patterns can emerge when arranging the data according to its periodicity, as observed by Carlis and Konstan [16, 23]. RankSpiral is an "interface for search engines" developed using Spiral patterns [16, 24].

Euler pattern has "multiple circles placed inside (or adjacent to) a larger circle" often linked to represent a nested visualization of a hierarchy [16]. An example is Zoomology, which uses the "outer ring" as the actual root of the hierarchical structure where "each node's children are rendered as inner circles" [16, 25].

Ring

The Ring visualization distributes the nodes "around the circumference" and its common use is to identify relationships between the nodes [16]. The classification divides this group into **Connected** and **Disconnected** Rings.

Connected Rings have the nodes connected by "line segments" and in some cases "additional nodes" are positioned in the "ring's interior" [16]. A popular example is the Circos visualization tool for "identification and analysis of similarities and differences arising from comparisons of genomes" [26].

Disconnected Rings follow the same principle but the nodes have no connections between them, thus representing large datasets without the clutter that Connected Rings have when portraying the relationships between the nodes [16]. SQiRL is an example of this pattern, it is a tool that visualizes the "opinion polls" by breaking down the "respondent's answers to selected questions" placing them "around the circumference" [16, 27].

In the context of health data visualization, radar visualizations have a potential to visualize large amount of datasets due to their clarity in the data representation. However, the potential use of interactivity needs to be addressed by these visualization tools. The ability to represent relevant information should be embedded in the visualization tool leveraging from modern technologies such as Web browsers and Web services as well as with current interfaces such as touchscreens.

The health graph

The health graph, or hGraph was developed by MITRE Corporation and released to the public in 2010 under the

Apache v2.0 license. The design intention of the hGraph is to facilitate the graphical representation and understanding of health data. The data can come from a wide-range of sources such as laboratory tests, physical activity, nutrition, sleep monitors and other sources. The domain of this visualization technique includes Personal Health Informatics, EHR and Personal Health Record (PHR) visualization [3].

Following the classification from Draper et al., the hGraph could be classified as a Polar Plot and Ring, following the design patterns of a Star and a Connected Ring.

The hGraph design consists of a circular space with an area defined by to circumferences. The area represents the minimum and maximum recommended values for a given measurement. For instance, the minimum and maximum recommended fat percentage of a person in a given age. The measurements are represented as circles and their position in the circular space represents how far or close they are from the recommended values. The position is normalized according to the recommended values.

The values are distributed in a circular space. A graph is formed by joining the data points around the circular area. This polygon or graph reveals a pattern and its shape provides a quick overview of the general situation of all the values and how they deviate from the recommendations. The hGraph design highlights values outside of the recommendation by using the red color on the data points and by modifying the shape of the graph. The rationale of the hGraph is that if the same measurements are plotted in the same order for various cases, then the graph patterns can reveal similar shapes associated with certain health conditions.

Web-based solution

Web-based solutions for data visualization provide flexibility, as they can be accessed by any web browser, either from mobile devices or personal computers. The hGraph uses a web approach via HyperText Markup Language (HTML) and Scalable Vector Graphics (SVG). The programming language of the library is JavaScript and is built using the Data-Driven Documents library. Data-Driven Documents (D3.js) library provides free access to the Document Object Model (DOM), which is the substrate that enables the programmer to interface with the graphical representations in a web browser [28].

hGraph as an insightful visualization

Based on the approach proposed by C. North [29], a previous study [4] compared visualizations based on how well users derived meaningful insights. The study compared the hGraph visualization along with four alternatives based on the Graphical Perception Framework proposed by Cleveland McGill [14]. The study compared the same data plotted with these five alternatives plus a control group which had the numerical data with no visualization. The data was comprised by a set of measurements of two modelled patients. The first patient had an elevated at risk of developing Type II Diabetes and the second one had a low risk due to a healthy lifestyle (regular exercise and a balanced diet). The evaluation followed the insight-based methodology similar to other experiments for visualizing genetic expressions [30]. The experiment determined how these visualizations can enable users to understand the overall health situation of the modelled patient with poor health, as well as the possible causes behind that patient's situation. The hGraph was found to be the most effective solutions for creating meaningful insights and to help users to better understand the data.

Figure 1 is an example of the hGraph visualization. The figure was extracted as a snapshot as the library generates an SVG document structure that cannot be exported outside the browser. We address this issue in the next section of the article. The hGraph hides the measurements when the zoom level is low, meaning that the user has zoomed out. The shape is an average of the deviation of each measurement under the same category. When the user zooms in, the details are revealed and the rest of the information becomes visible, that is the numerical values and positions (with respect to the recommendation) of the measurements.

Implementation

In this chapter we detail the implementation methodology and the key features implemented in the library. We named the library Health Figures (hFigures) because it is based on the design principles of the hGraph. hFigures makes an emphasis on multiple graphs, or figures, in order to provide a graphical representation of evolution of the data over time (multiple snapshots of the data at certain points in time).

Methodology

The implementation of the hFigures followed the Extreme Programming methodology [31, 32]. The main key requirement was to provide a visualization which represents the changes in the overall health situation. In addition, the implementation addressed the features that the users requested in a previous study [4].

Extreme Programming focuses on releasing and reviewing functional software continuously [32]. Often these requirements change and the programming practice is to address this changes by prioritizing them at the top of the change list.

During the implementation of the library, our research group provided the continuous review process of the software. The research group has expertise in Health Sciences, Signal Processing, User Design, Software Engineering and

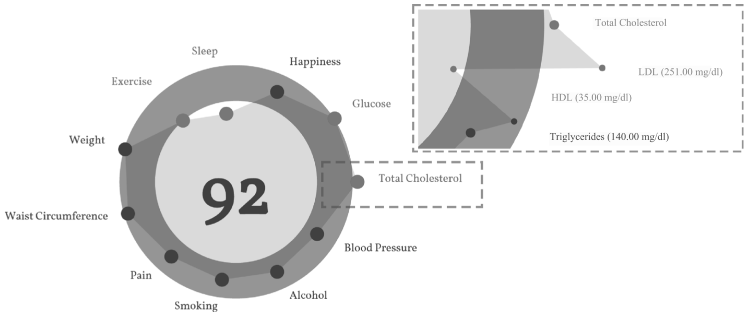


Fig. 1 An example of the hGraph. The hGraph shows an overview at a low zoom level. When the zoom increases, the details of each measurement are revealed

Machine Learning. Requirements often changed and new releases were assessed by the group. The development of the health monitoring application followed the practice of pair programming, as it is often the case in Extreme Programming [31].

Data source

We use a JSON (JavaScript Object Notation) format to read the data, in which the measurements are grouped according to their categories. The groups contain an array of samples, which represent the values obtained from a measurement (steps per day, cholesterol, triglycerides, blood sugar or depression level using [33]). The samples contain a timestamp in Unix Epoch format and the value of the measurement. The Unix Epoch format is the number of seconds since the first of January 1970, Greenwich Meridian Time (GMT). An example of the data source is in Fig. 2, it shows the first measurement of the group "Blood Pressure" which in this case is comprised by Systolic and Diastolic measurements and each of them have two samples taken at two particular times, Friday 9th of January 2015 at 10:10:24 GMT (1420798224 — seconds) and Thursday 12th of February 2015 at 12:05:20 GMT (1423742720 — seconds).

SVG document export

The SVG document structure we designed in our implementation can be exported to a separate file outside the web browser. The short-term objective is to build a toolkit that enables researchers to visualize their data with our implementation so they can use the generated SVG file in articles, posters, presentations or other applications. For instance, Fig. 3 has been exported as an SVG document from the browser into this article. SVG export is possible due to the rendering of our algorithm which does not depend on JavaScript or Cascade Style Sheet (CSS) styling properties to produce a finalized document. The library build the entire image as a stand-alone document. The hGraph library unfortunately does not produce a complete document but instead depends on CSS and JavaScript code to make the image visible.

Constant graph shape

Figure 1 shows an hGraph example and Fig. 3 shows an hFigures example. The hGraph computes the average of the deviation of the measurements in order to show the polygon or graph, as a representation of the overall health assessment. However, some measurement might deviate towards a lower value and while others towards a higher

```
var measurementGroups = [
    {
        label: "Blood Pressure", measurements:[
            {
                min: 90, max: 129,
                yellow_min: 80, yellow_max: 130, red_min: 70, red_max: 140,
                units: "mmHg", label: "Systolic",
                samples: [
                    { timestamp: 1420798224, value: 132 },
                    { timestamp: 1423742720, value: 128 }
                ]
            },
            { /* Additional measurements in the same group */ }
        ]
    },
    { /* Additional groups. */ },
];
```

Fig. 2 JSON data source file. The data source file structured as a JSON file

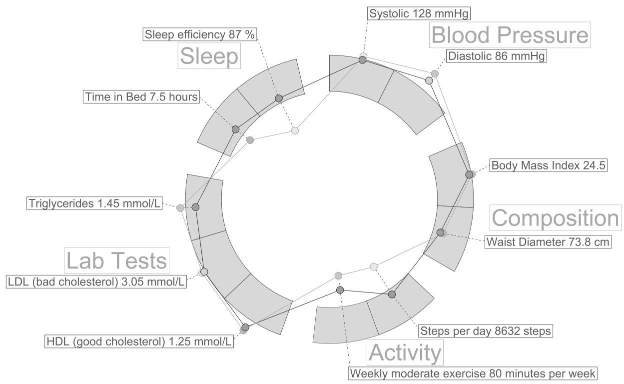


Fig. 3 A simple example of hFigures. The measurements are separated by groups in sectors

one, thus the average position would be roughly the middle recommended area. For this reason, hFigures does not change the shape of the graph if the user zooms in or out. Showing and hiding the measurement labels is the only reaction to the zooming events from the user at the moment. This avoids clutter when the user wished to have a quick glance at the picture but keeps the graph with the same shape.

Layout construction

The measurement groups are represented using a circular layout divided in sectors. The goal is to have a clear division between the groups as they represent the different aspects of the overall health. hFigures uses the d3 pie chart layout and modifies the data source provided to the layout. All the measurements have the same numerical value and at the end of the group, we insert an extra value in order to leave a blank space between the circular area sectors. The result is visible in the hFigures example shown in Fig. 3 and the code that produced this visualization is in Fig. 4. The pie layout constructs the sectors of the circular area based on a data source. When we provide an array of numbers, the layout uses the numbers to calculate the proportions of the area. In order to achieve the layout construction that we have designed, the array has the same constant number multiple times, the number of measurements plus an additional number for each group.

Color-coded entries

The data source can contain additional sets of value ranges. For instance a warning range of values can let the users know when a value has reached a level that requires attention but has not yet reached a critical point. We followed the users' feedback that recommends a traffic light-based approach. The green color means that the values are within the recommended, yellow suggest a warning or follow-up action needed and the red indicated a critical threshold has been passed.

In the implementation, the library verifies if these additional ranges are present in the measurement definition. In order to verify if the property of the object exists, JavaScript provides a qualifier method, `typeof`. The returned value must be compared with the keyword definition for properties that are not present in an object, the keyword `undefined` has been suggested by Mozilla Developer Network [34], a highly reputable source for Web development. The code is shown in Fig. 5.

Multiple graph

The dataset is structured as a set of measurements where each has its own collection of samples. In order to compare the evolution of these measurements, the hFigures library allows the graphical representation of any number of samples. The result is a set of graphs or polygons overlapping or stacking with each other. In order to differentiate them, we use a lighter set of colors so that the users can see the difference between two points in time. As an example, Fig. 3 shows two different samples for each measurement. This example portrays a modeled person that has been active in a health coach program. Some measurements have improved and are closer to the recommendation. Users repeatedly expressed that it would be very helpful to visualize two or more different points in time so as to compare how the person has evolved.

Including multiple graphs has implications in the structure and procedures of the visualization construction. For instance, we structured the SVG document such that each measurement includes one or many plotted circles that map to each sample. The measurement labels need to be positioned considering that a plotted circles can (and

```
function extractMeasurements(groupedMeasurements){
    var group;
    var total = [];
    for(var i = 0; i < groupedMeasurements.length; i ++){
        group = groupedMeasurements[i];
        total = total.concat(group.measurements);
        total.push({ label: "empty" });
    }
    return total;
}
var outerRadius = w * 0.4, innerRadius = w * 0.3;
var arc = d3.svg.arc()
    .innerRadius(innerRadius)
    .outerRadius(outerRadius);
var pie = d3.layout.pie().value(
    function (d) { return d.label === "empty" ? 1: 2;
    }).sort(null); // Avoid sorting to preserve the data source order.
var measurementsArray = extractMeasurements(groups);
var measurementsObjects = pie(measurementsArray);
```

Fig. 4 Layout construction code. The layout for distributing the measurements is build using d3 pie layout leaving spaced between the measurement groups

probably will) overlap. This is challenge that we address in the next section by finding an optimal label space distribution to avoid labels from overlapping and also to reduce the clutter in the visualization space.

Label space distribution

After the measurements are plotted, the labels are added to increase readability. The position of the label needs to be defined within a given range to avoid overlaps and clutter.

Labels need to avoid overlapping with each other and with their measurements. To solve the label overlapping problem we implemented an algorithm that starts by ordering the labels by their angular position, that is the angle at which the measurement is positioned. The next step is to calculate the height of the label and position it over the previous one in the direction that goes from the center of the visualization area upwards or downwards (depending on the angle). The idea is to begin with the center of the area, either to the left or to the right of the circles, then we work our way up or down drawing the labels into the SVG document. We add the labels as SVG elements and the use the `transform` property to position them in the corresponding place. Figure 6 shows the spacing between the labels using the algorithm when drawing the labels from the center to the upper right corner. For each of the four quadrants, the library calls the method shown in Fig. 7 which computed the position of the label as we described.

As mentioned before, labels can also overlap with measurement circles. To avoid this problem we calculate the maximum radius from the center of the visualization area to the highest value of a measurement sample. From that starting point, we place the label in that position. In other words, for each measurement, we find the largest value of the samples. Figure 8 shows a sector of the hFigures where the sugar measurement label has been pushed out for a few pixels in order to avoid overlapping it with the red circle. The rest of the labels adjust to that position by leaving a user-defined margin.

$$r_{label} = max\left(\bigcup_{i}^{n}\{r_i\}\right) + margin \quad (1)$$

```
function getColor(measurement){
    var color = "white", yellowMin = "yellow_min";
    /* Create as many variables as cusom limits. */
    /* Code removed for brevity. */
    sample = measurement.samples[index];
    /* Verify the limits if the exist. */
    if (typeof measurement[yellowMin] != 'undefined'){
        hasAdditionalRanges = true;
        if(sample.value <= measurement[yellowMin]){ color = yellow; }
    }
    /* Follow the same approach for the rest of the limits. */
    /* Code removed for brevity. */
    if (hasAdditionalRanges && color === "white"){
        color = green;
    }
    return color;
}
```

Fig. 5 JavaScript code to determine if additional ranges are provided. In JavaScript the data source could contain additional ranges, these are properties in an object that need to be checked beforehand and if the exists, compare the values accordingly

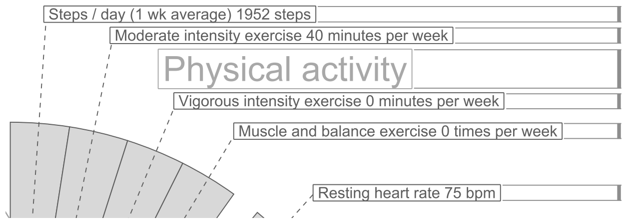


Fig. 6 Label positioning distribution. The labels are positioned according the their height and margin, as a result the labels do not overlap and clutter is avoided

The radius for the label is the maximum value of the samples translated as graphical coordinates plus a margin. Equation 1 obtains the label radius r_{label} given the radii of the samples of a measurement plus the default margin m.

Feature implementation summary

The key improvement of hFigures is the addition of multiple graphs as a mechanism to compare the values of the health measurements over time (Table 1).

The immutable shape of the graph presents the same information (values of the measurements in respect to the recommended target) regardless of zooming. This feature shows the data "as is" without calculating average, mean or deviation. Users stressed the importance of graphically representing the information without any calculations such as mean or accumulated values. The users expressed that showing the measurement values in hFigures helped them to derive valuable insights with just a quick glance at the data, for instance they identified measurement that fall outside of the recommended range easier and without requiring them to zoom in or out.

The hFigures library does not calculate an overall score since the users considered that this task should be the sole responsibility of a health care professional. An overall score also depends on each person under a case by case basis. For instance, the hGraph allows the user to assign weights to each measurements' value, the score is then calculated summing the value of the weights times the measurement's deviation from the recommendation. The users participating in the design process of our library expressed that an overall score would complicate the integration of the library into daily health care processes as specialists would need to review case by case to find the adequate score formula, which means specifying the weights (importance) of each measurement for a given person.

Measurements in the hFigures library are grouped in sectors which represent the category they belong to. Grouped measurements allow a clear division of categories resulting in enhanced understanding on how certain areas of wellness have changed and how, if any, they affect each other. The sectors remain visible regardless of the zoom level, users expressed that this feature provides an informative approach as the categories are always showing to which category the measurements belong to. Hiding the category labels and displaying the measurements without divisions would complicate understanding the status of health categories, such as sleep, nutrition, physical activity and others. Users expressed their confusion when they were unable to determine when a category starts or ends after zooming in and out of the hGraph.

```
var verticalLabelLimit = 0;
var verticalLabelMargin = 5;
function moveLabelVertically(i, angle, height, y){
    var delta = i === 0 ? -1 * verticalLabelMargin: 0;
    var upper = angle >= (3/2 * Math.PI) || (angle <= Math.PI/2);
    var collision = upper ?
    y + height + verticalLabelMargin >= verticalLabelLimit :
    y <= verticalLabelLimit;
    // Negative value moves the object up
    if(collision && i > 0)
        delta = upper ?
            verticalLabelLimit - (y + height + verticalLabelMargin) :
            verticalLabelLimit - y; // positive to move it down
    // Add the height and the margin plus possible delta.
    verticalLabelLimit = upper ?
        y + delta : y + height + verticalLabelMargin + delta;
    return delta;
}
```

Fig. 7 JavaScript code for label positioning. The function in JavaScript distributed the positioning of the label to avoid overlapping and clutter

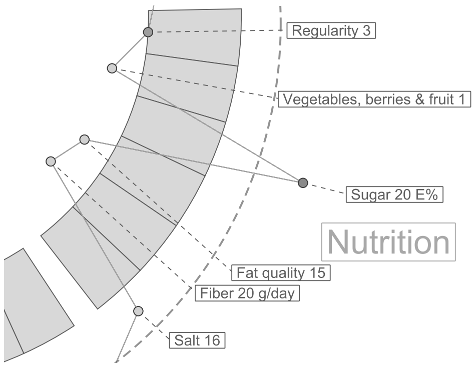


Fig. 8 Label positioning avoiding overlapping with measurements. Some measurements can be positioned outside of the recommended range, the labels are adjusted to avoid overlapping

The possibility to export the generated figure as an SVG file, allows the integration into research articles, presentations, websites, posters and other Software applications to further enhance the utility of the hFigures.

Evaluation

The health data visualization library was placed in the context of a full application. We tested the library in a contextualized scenario where the users conducted a series of tasks and answered usability questionnaires. In this section we present the methods we used for recruiting the participants and for the usability testing of the library. We also explain the metrics measured and the rationale behind the selection of the usability testing methods.

Nielsen suggests that "usability has multiple components and is traditionally associated with the five usability attributes, which are learnability, efficiency, memorability, errors, and satisfaction" [35]. In order to assess the usability of the software solution, multiple alternatives exist in industry and research. Johnson et al. developed a toolkit for usability testing of Electronic Health Records commissioned by the Agency for Healthcare Research and Quality of the U.S. Department of Health and Human Services [36]. The toolkit is built on the basis of the assessment of existing usability methods in the context of Electronic Health Records and Health Information systems. The toolkit is a detailed analysis of the usability methods, their advantages, disadvantages and appropriateness ranking.

Table 1 Comparative functionality table. The table lists the features supported by hFigures and hGraph

Functionality	hFigures	hGraph
Color-coded measurements	✓	Partially
Recommended area	✓	✓
Circular distribution of measurements	✓	✓
Plot values at specific times	✓	—
Time series data source	✓	—
Immutable shape	✓	—
Weighted score	—	✓
Multiple graphs	✓	—
Visible category labels	✓	—
Label space distribution	✓	Partially
Grouped measurements	✓	—
SVG export	✓	—

We selected the Usability Questionnaires since it has a high appropriateness ranking [36]. We were able to recruit three usability experts to conduct the Heuristic Evaluation and the Cognitive Walkthrough, both are recommended techniques to complement the evaluation. We concluded the evaluation using the principles of Controlled User Testing.

Continuous health monitoring application

In order to test the hFigures library, we designed an application for visualizing the health situation of a modeled patient and how this has changed over time within a health coaching program. The objective is to help the users in the decision making process of assessing the overall health situation and whether or not the health program has provided benefits.

The application has three components: activity timeline, the hFigures data visualization library and longitudinal measurements. Figure 9 shows the three components. The figure is not a screenshot of the application but rather an extraction of the SVG documents embedded in the HTML file, except for the timeline.

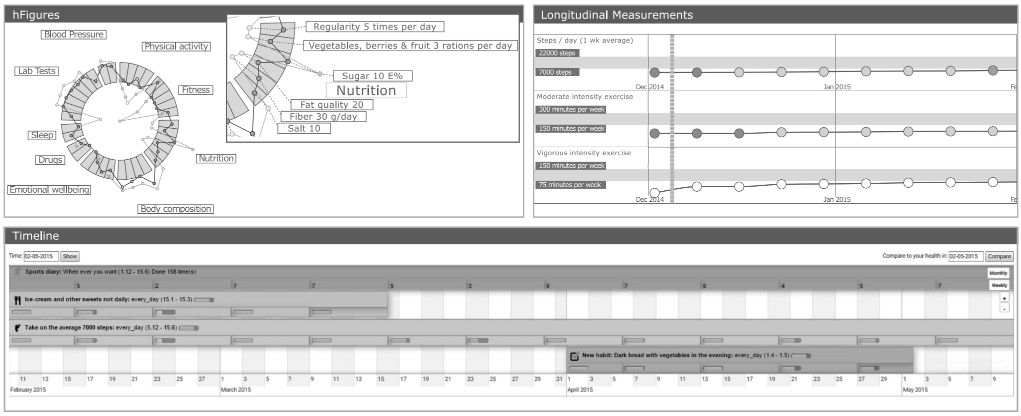


Fig. 9 Health Monitoring Application. The health monitoring application with the three components: hFigures, Activity Timeline and Longitudinal Measurements

Activity timeline

This component represents the health interventions (particular actions) that the modeled patient has done during the health coaching program. During the program, several snapshots of the patient's overall health are taken and visualized using the hFigures library.

hFigures

The developed hFigures library is utilized to display the set of measurements taken during the health coaching program. These measurements describe an overview of the health situation of the modeled patient. Users can change the time at which the snapshot was taken to compare changes over time as a result of the health interventions.

Longitudinal measurements

The application also displays the same set of measurements using longitudinal temporal representation. We included this component to provide additional details and trends on how the measurements have changed over the coaching program.

Heuristic evaluation

Heuristic Evaluation requires at least one expert in the area of human-computer interaction [35, 36]. For our evaluation we recruited three experts, and they assessed the application using Nielsen's heuristics [35]. The evaluation has 11 metrics evaluated using a seven point Likert scale, where the value 1 indicates "strongly disagree" and 7 "strongly agree"

Heuristics are "rules of thumb" comprised of 10 principles meant to assist the Human-Computer Interaction specialist in the usability assessment [36, 37]. We explain the principles of the Heuristic Evaluation according Nielsen [37].

1. *Visibility of the System Status*: Refers to continuous feedback on the status of the system "within reasonable time" (Feedback).
2. *Match between system and the real world*: The use of the language should be familiar to the user so that conversations follow a "natural and logical order" avoiding technical terminology unfamiliar to the intended user audience (Speak the User's Language).
3. *User control and freedom*: Allow the user to recover from erroneous navigational options with "clearly marked" access options (Clearly Marked Exits).
4. *Consistency and standards*: Follow the same language and terminology to avoid the user from guessing the meaning of "words, situations, or actions"(Consistency).
5. *Error prevention*: Avoid "error-prone" options in the system whenever possible and for those cases when the problematic options cannot be avoided, then present the user confirmation dialogues (Prevent Errors).
6. *Recognition rather than recall*: Present visible options to the user at all times so as to avoid the effort of remembering previously stated instructions. Whenever options cannot be visible, make them "easily retrievable whenever appropriate" (Minimize User Memory Load).
7. *Flexibility and efficiency of use*: The interface should accommodate the novice and advance user by providing "tailored frequent actions" (Shortcuts).
8. *Aesthetic and minimalist design*: The dialogues should only contain relevant and clear information

that is timely needed at that particular state of the interface (Simple and Natural Dialogue).

9. *Help Users recognize, diagnose, and recover from errors*: Plain language should be used in error messages, and whenever possible they should provide helpful information so that the users can take constructive actions. (Good Error Messages)
10. *Help and documentation*: Some systems require documentation and guidelines to explain briefly how to accomplish specific tasks in concrete steps.

Cognitive walkthrough

Wharton et al. developed the Cognitive Walkthrough for usability testing [38]. Johnson *et al.* summarize this method as a "usability inspection method that compares the users' and designers' conceptual model and can identify numerous problems within an interface" [36, 38].

Cognitive Walkthrough has successfully been used to evaluate usability of Healthcare Information Systems [36, 39–42] and Web Information Systems [43].

Since Cognitive Walkthroughs "tend to find more severe problems" [36, 44] but "fewer problems than a Heuristic Evaluation" [36, 45] we included both methods in our evaluation.

Laboratory testing

Regarded as the "golden standard" for usability testing [46], Laboratory Testing collects "qualitative and quantitative" data "since it collects both objective data such as performance metrics (e.g., time to accomplish the task, number of key strokes, errors, and severity of errors) and subjective data such as the vocalizations of users thinking aloud as they work through representative tasks or scenarios" [36].

Controlled user testing is comprised of "a series of commonly used task scenarios" where users are asked to conduct these tasks using a "thinking aloud" [35, 36, 47]. This process requires "users to talk aloud about what they are doing and thinking" while they complete the tasks using the system [35, 36, 47].

As the "golden standard" in usability testing, this method has been widely used in evaluating Health Information Systems [36, 48–51]

The data exploration tasks are designed to assist the decision making process on the health situation of the modeled patient. The usability scenario was the main goal of the intended use of the application. We explain the participants the purpose of the application, which is to facilitate the decision making process weather or not the overall health situation of the modeled patient is favourable and weather or not the health coaching program was beneficial for the patient. The tasks are designed to represent the common usage of the application, namely to find the measurements inside and outside of the recommendation and to identify the areas that improved and need even further improvement. The tasks given to the participants are shown in the following list.

1. How many areas of health are displayed in the hFigures?
2. Choose one of these areas and point to its measurements.
3. Identify one measurement inside the recommended values and another one outside.
4. Identify the measurement that is the furthest from the recommended values.
5. What does the green, yellow and red circles mean?
6. Has the overall health improved after coaching?
7. Which area of health has improved the most after health coaching?
8. Which measurements show the biggest improvement?
9. Understand the difference between the points inside and outside the recommended area.

Usability questionnaires

We followed the recommendations from Johnson et al. and used this method in our evaluation. Usability Questionnaires are "the most common" method to "collect self-reported data" from the "users' experience and perceptions after using the system in question" [36]. Although the data collected is self-reported, some questionnaires have reliability in measuring several usability metrics such as "satisfaction, efficiency, effectiveness, learnability, perceived usefulness, ease of use, information quality, and interface quality" [36].

We used two Usability Questionnaires to evaluate the usability of our application, Computer System Usability Questionnaire (CSUQ) and After Scenario Questionnaire (ASQ) [52]. Table 2 shows the length, reliability and the metrics of the questionnaires. These questionnaires use a seven-point Likert scale from "strongly disagree" up to "strongly agree".

Table 2 Standard questionnaires table. The table lists the standard questionnaires we used for the user evaluation of the system with their length, reliability and metrics

	Items	Reliability	Metrics
CSUQ	19	0.93	Usefulness
		0.91	Information quality
		0.89	Interface quality
		0.95	Overall usability
ASQ	3	0.93	Ease of Task Completion
			Time Required to complete the task
			Satisfaction

Computer system usability questionnaire (CSUQ)
The questionnaire was developed by IBM and it is a “slight” modification of the Post-Study System Usability Questionnaire (PSSUQ) [53]. Table 2 shows the reliability of this questionnaire. The questionnaire has high “coefficient alpha” with a reliability 0.95 in total and “0.93 for system usefulness, 0.91 for informational quality, and 0.89 for interface quality” [36, 52, 53]. We selected this questionnaire since it has been successfully used in the Healthcare domain [36, 54] and in the evaluation of “of a guideline-based decision support system” [36, 55].

After scenario questionnaire (ASQ)
An additional questionnaire developed by IBM [36, 52, 56] and designed to measure the user satisfaction after scenario usability studies have been completed [36, 53, 57]. This questionnaire measures the “ease of task completion, time required to complete the tasks, and satisfaction with support information” [36]. Since we already designed the scenario for the evaluation of the system, we included this questionnaire in our study.

Data model
Similar to the study we conducted in the insight-based methodology [4], we modeled a patient using clinical expertise of a physician along with the most common symptoms for developing Type II Diabetes. The modeled patient consisted of a set of measurements over time comprised of the following parameters:

- Blood pressure: systolic and diastolic blood pressure
- Physical activity: weekly active days [58, 59], steps per day [60]
- Body composition: Body Mass Index (BMI), waist diameter and fat percentage
- Sleep: time in bed, time asleep
- Fitness: resting heart rate, fitness index [61, 62], muscular force, muscular endurance and balance [63]
- Lab Tests: hemoglobin, fB-Gluc, cholesterol, HDL, LDL, triglycerides
- Nutrition: meal regularity, type of meals (vegetables and fruits), sugar intake, fat quality, fiber and salt intake.
- Drugs: tobacco (cigarettes per day), alcohol abuse, drug abuse (narcotics), medication abuse
- Emotional wellbeing: depression level [33], stress level and stress recovery [64, 65] and optimism [66].

Recruitment
We recruited a total of 14 participants following similar usability studies and Faulkner’s [67] recommendation of conducting usability tests with 10 to 20 users “in order to find 90 to 95 % of usability problems” [68]. Among the 14 participants we were able to recruit 3 usability experts, following the recommendations from Nielsen and a number of previous studies stating that 3 to 5 experts are needed to conduct the Heuristics Evaluation [35, 36, 69–72].

Participants were recruited through the university’s student email lists, self-study groups, lectures and workshops. After completing the usability tests, the participants received a movie ticket.

Ethics
The study we conducted was a usability evaluation using simulated data not belonging to a real person. The results of the usability tests were kept anonymous and the collected data does not include sensitive information from the participants. According to the ethical principles applied by the Finnish Advisory Board on Research Integrity, our study did not need ethics approval [73].

The experimental procedures described in this paper complied with the principles of Helsinki Declaration of 1975, as revised in 2000. All subjects gave informed consent to participate and they had a right to withdraw from the study at any time. The informed consent also explained that their names and identities will be kept confidential, that the results will not be linked to their identities, the sessions will be recorded using a Web camera and microphone for further study and that the clinical data visualized did not belong to a real person.

Experiment protocol
The testing process started with the signature of an informed consent where we explained the participants the purpose of the test. Afterwards we proceed to explain a usability scenario and the tasks that the participants were asked to complete. The participants were allowed to ask questions at any time. After performing the tasks we asked the participants to fill in the Usability Questionnaires. We close the session with a briefing interview where we asked the users what they liked and disliked about the application as well as what were their recommendations for further improvements. The sessions were recorded for further study and to find the correct timing of the task completion.

Materials and tools
We conducted the usability tests in our laboratory. We used a computer with a local HTTP server running our server application and Google Chrome as the browser running our front-end application. The computer was a laptop with a camera and microphone which were used to record the session for later study. The computer was connected to a 23 inch display and a separate keyboard and mouse. The usability questionnaires were filled out using the Web portal developed by Perlman and available at the following address http://garyperlman.com/quest/.

Results

Heuristic evaluation

The three expert users answered the Heuristic Questionnaire in order to identify problems with the user interface of the health monitoring application. The three experts agreed and in some cases strongly agreed with most of the indicators. One expert found the instructions for adjusting the time of the visualization tool to be demanding. The expert addressed this comment to the integration interface that allows the time to be adjusted and thus visualized. The remark was not addressed to the graphical representation of the data using hFigures. The results of the evaluation are summarized in Table 3. The average response was 6.3 out of 7 points (Additional file 1).

Cognitive walkthrough

During the Cognitive Walkthrough, the concept of the health monitoring application was explained to the usability experts. The purpose of the application was explained in the context of the health situation of the modeled patient and how the application visualizes the changes in the health situation over time. We used the usability scenario and tasks to confirm that the interface supports the intended use of the application. The questions comprising the walkthrough, as described by Wharton et al. [38]., were correctly answered by the expert users thus no design or mismatch errors were found.

Controlled user testing

Table 4 summarizes the results of the completed, number of errors, average time to complete the task and the standard deviation. All participants completed 7 of the 9 tasks. Task 2 was the most problematic, we asked users to "choose one of these areas and point to its measurements", we found that 3 participants were not able to understand the task thus unable to complete it. The second most problematic task was number 9, "understand the difference between the points inside and outside the recommended area", where one participant was unable to complete it successfully incurring in one non-crucial mistake (an error that prevent the task to be completed).

Table 3 Heuristic evaluation results. The table summarizes the results of the Heuristic evaluation conducted by three usability experts

Heuristic	Average response	Standard deviation
Visibility of system status	6.00	1.00
Match between system and the real world	6.33	0.57
User control and freedom	6.33	0.57
Consistency and standards	6.67	0.57
Error prevention	6.33	0.57
Recognition rather than recall	4.67	0.57
Flexibility and efficiency of use	6.67	0.57
Aesthetic and minimalist design	7.00	0.00
Help users recognize, diagnose, and recover from errors	6.33	0.57
Help and documentation	6.67	0.57
Nielsen heuristic evaluation	6.30	0.56

Table 4 Controlled User Testing Results. The table summarizes the results of the 14 users performing the 9 tasks

Task	Successfully completed	Errors	Average time (seconds)	Standard deviation (seconds)
Task 1	14	4	12.21	12.60
Task 2	11	0	10.00	12.38
Task 3	14	1	10.78	5.38
Task 4	14	2	6.78	4.98
Task 5	14	0	17.50	6.60
Task 6	14	2	16.07	16.52
Task 7	14	0	6.21	5.591
Task 8	14	1	7.85	5.882
Task 9	13	1	9.23	4.729

Additional non-crucial errors occurred in tasks 1, 3, 4, 6 and 8. The large number occurred in the first task due to the initial values set in the default zoom level of the hFigures component. After the usability testing, we corrected this problem by adjusting the initial zoom level to include the whole figures inside the container (Additional files 2 and 3).

Usability questionnaires

Computer system usability questionnaire (CSUQ)

We computed the results according to Lewis, obtaining the average of "items 1 through 19" to determine the overall usability rating of the system. System usefulness is the average of items 1 to 8, information quality 9 through 15 and interface quality 16 through 18 [74].

Table 5 shows the results of the system usefulness. The system obtained an average of 6.13 out of 7. Table 6 shows the results of the information quality metric where the application scored a total average of 5.66. The average value is still within the "agree" response of the participants, however the notable low value compare to the other metrics might be due to the amount of information presented in textual format in the application. The information was encoded using graphical representations and even though a help document was included in the system, the text was not likely to fulfil the users' expectations. Table 7 shows the score for the interface quality where the

Table 5 Computer system usability questionnaire results for the system usefulness assessment. The table shows the results of the questions corresponding to the system usefulness with its average and standard deviation

Question	Average response	Standard deviation
Overall, I am satisfied with how easy it is to use this system	6.29	0.99
It was simple to use this system	6.07	1.20
I can effectively complete my work using this system	6.07	1.07
I am able to complete my work quickly using this system	5.86	1.40
I am able to efficiently complete my work using this system	6.21	0.89
I feel comfortable using this system	6.21	0.97
It was easy to learn to use this system	6.43	0.85
I believe I became productive quickly using this system	5.93	1.26
System Usefulness	6.13	0.93

application obtained an average of 6.24 out of 7. The combined results are shown in Table 8. The score of the overall usability is 6.02 with a standard deviation of 1.04. We can determine that all the participants at least "agreed" in the Likert scale that the application was useful for the decision making process of assessing the health situation and evolution of the modeled patient (Additional file 4).

Table 6 Computer system usability questionnaire results for information quality

Question	Average response	Standard deviation
The system gives error messages that clearly tell me how to fix problems	4.50	2.44
Whenever I make a mistake using the system, I recover easily and quickly	5.43	1.95
The information (such as online help, on-screen mes-sages, and other documentation) provided with this system is clear	5.29	1.90
It is easy to find the information I needed	6.07	1.27
The information provided for the system is easy to un-derstand	5.93	1.39
The information is effective in helping me complete the tasks and scenarios	6.14	1.17
The organization of information on the system screens is clear	6.29	1.14
Information Quality	5.66	1.20

Table 7 Computer system usability questionnaire results for interface quality

Question	Average response	Standard deviation
The interface of this system is pleasant	6.36	1.00
I like using the interface of this system	6.36	0.92
This system has all the functions and capabilities I expect it to have	6.00	1.18
Interface Quality	6.24	0.99

After scenario questionnaire (ASQ)

Table 9 the average response for the ease of task completion was 6.64 with a standard deviation of 0.842 and for the time required to complete the task 6.64 and a standard deviation of 0.497. The overall satisfaction was 6.46 and a standard deviation of 0.53. The usability of the system had a high score in the ASQ results meaning that the system was suitable for the scenario in the context of the health data visualization of the modeled patient and its evolution over time (Additional file 5).

Identified issues and suggested improvements

The feedback shows that the main problem was the incomplete visibility of the hFigures in the application component window. Users also requested to show the detailed information as a hovering pop up window in the second figure (measurements before the coaching program). Currently only the latest measurements have the hovering window however users requested that both measurements (the before and after) should contain the same functionality. Additional information was needed in the measurements that contained numerical scales, such as the depression index. A more contextualized approach explaining the meaning of the values can help the user understand the measurements and thus the overall health situation of the patient better.

Discussion

The value of a data visualization depends on the knowledge that it can convey to the public. In this section, we claim that hFigures has the potential to be used both in the

Table 8 Computer system usability questionnaire results for overall usability, system usefulness, information and interface quality

Metric	Questions	Average response	Standard deviation
Overall usability	1–19	6.02	1.04
System usefulness	1–8	6.13	0.93
Information quality	9–15	5.66	1.20
Interface quality	16–18	6.24	0.99

Table 9 After scenario questionnaire results

Question	Average time (seconds)	Standard deviation (seconds)
Overall, I am satisfied with the ease of completing the tasks in this scenario	6.64	0.84
Overall, I am satisfied with the amount of time it took to complete the tasks in this scenario	6.64	0.49
Overall Satisfaction of the system	6.46	0.53

clinical and personal wellness applications. Large amounts of measurements do not clutter the visualization area as a result of our implementation, Fig. 10 shows an example of a large visualization of a modeled patient. The comparison of multiple graphs can provide a meaningful visualization to individuals and clinicians alike. The implementation of hFigures follows an extensible approach and even though it was designed to be used for health data visualization, any dataset that has a target range of values as a reference can be visualized.

Translation to health care

The wide variety of EHR formats and data sources from self monitoring applications comprise a challenge in unifying the data in order to provide an overview of a patient. Currently, most of the data sources contain the date when the sample was extracted, whether it is a tracking device or a blood test. This sample date already provides the timestamp required by the hFigures data source file. The values of the samples are the main object of study in a measurement, for instance the levels of cholesterol or sugar at a given time, the number of steps per day, the percentage of body fat and several others. This information can be transformed in a simple process to build the hFigures data source following the JSON structure.

hFigures is a visualization library based on Web technologies, it uses a Web browser and the rendering of the SVG is compatible with current HTML standards. Thus providing interoperability across multiple devices including tablets, smartphones, workstations or laptops is indeed feasible. The D3.js library that hFigures is built on, enables compatibility with Internet Explorer versions 8 and higher using a component named *Aight* [75]. Internet Explorer 8 is prevalent in hospitals and clinics due to the restrictions in installing custom software.

Patient evolution

Multiple graphs plotted on the layout of recommended values shows the change in the data over time. This could allow clinicians to understand the evolution of certain

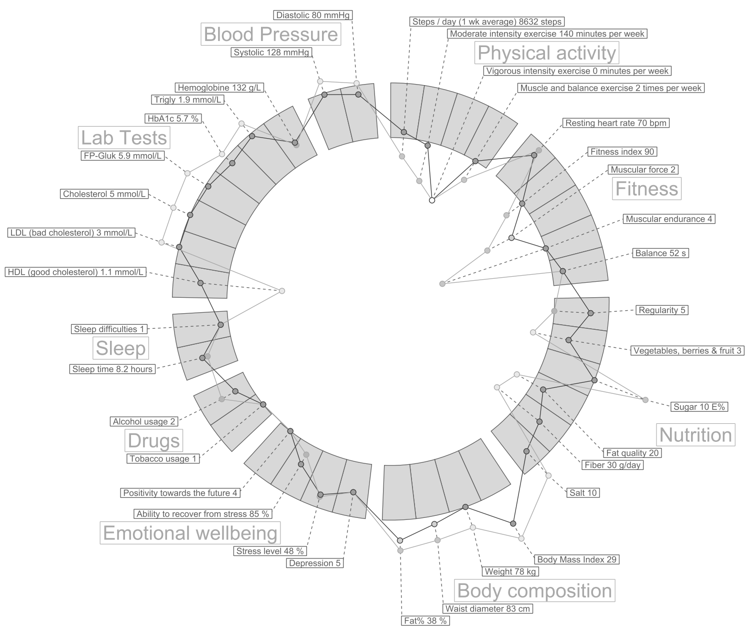


Fig. 10 A heterogeneous hFigures example. An overview of a modeled person comprised of several measurements with two time snapshots showing its evolution over time

aspects in the health of the patient. For instance, health professionals would be able to look at the effect of a trail drug treatment over time. Possibly, the collected samples of a patient before and after starting the trial would be plotted as the multiple graphs portraying the evolution of the patient. Figure 10 shows an example of a more heterogeneous dataset. The labels for each individual measurement are usually hidden when a full zoom out is performed by the user. For the purpose of demonstrating the visualization library we have made all the labels visible.

Personal health monitoring

As an example, Fitbit provides activity trackers and a wide-range of devices. The data collected can be obtained through their API. In most cases, providers such as Fitbit follow an HTTP REST interface. Figure 11 shows an example of the data Fitbit provides through its interface. The data has `"activities"` as properties of a JavaScript object. These entries have in turn a property `"startTime"` which provides the timestamp required for the hFigures data source file.

In this case, a step counter contains the property `"steps"` with the number of steps registered by the device. This and other measurement can be plotted in the hFigures following the data source file structure.

We can also use multiple graphs to show the user-defined goals as a reference in addition to the actual values of the measurement (steps per day). Figure 11 only show steps per day, however the documentation of the API specifies that additional data is available. This data includes distance travelled, sedentary activity, floors climbed, calories burnt and more.

Other device manufacturers such as Withings or Jawbone provide their users the possibility to use their APIs to extract collected data in a similar way. As in the case of Fitbit, all data has at least a timestamp (date, time or both) and a set of values. Recommended values for the health measurements can be obtained from healthcare professionals and public health information sources.

Limitations

The library reads the data in JSON format and does not support XML, which is still used by some Health Information Systems. Data files need to be included in the same HTML file and used to create an instance of the hFigures class. The library does not retrieve the data remotely by itself so the data provision is the responsibility of the developer. Other libraries provide an AJAX interface through HTTP(S) communication to an endpoint in order to retrieve the data. The information in the nodes comes entirely from data source file, which means that additional information about the measurements cannot be supplied otherwise. The extraction of the SVG file requires to export the code embedded in the HTML file. Currently no automatic export functionality is implemented.

Further development

We plan to develop the library further to address the suggestions obtained from the participants of the usability testing. The next release of the library already included the fixes for the automatic adjustment of the initial zoom level to show the complete figures within the given container, usually a `<div>` element in the HTML document. The next item to address is the inclusion of additional information explaining the measurements in the hover pop up window. Additionally we need to develop an algorithm to display the values when hovering on the two figures so that the pop up windows do not overlap.

Further development contemplates a Web Service which consumes a JSON data source file and produces an SVG or a HTML document with the interaction features as a JavaScript file attached. Such service has been already requested in other projects for research purposes in order to provide a better software tool for medical decision making processes.

Conclusion

Complete and accurate visualizations of health data have been thought to empower individuals, citizens and health professionals alike, to better understand situations and take better informed decisions [2, 12, 76]. These decisions can be medical treatment, behaviour change practices, wellness development, health coaching program and more. In this article we detailed the underlying motivation to develop a visualization library inspired by the hGraph.

We tested the visualization library in the context of an application by conducting usability tests comprised of Heuristics Evaluation, Cognitive Walkthrough and Usability Questionnaires. In the Heuristics Evaluation the average response was 6.3 out of 7 points and the Cognitive Walkthrough done by usability experts indicated no design or mismatch errors. In the CSUQ usability test the system obtained an average score of 6.13 out of 7, and in the ASQ test the overall satisfaction score was 6.64 out of 7. The results indicate that the library was helpful in assisting users in understanding health data and its evolution over time.

The library is an open source tool inspired by the hGraph but with additional key improvements. However, additional improvements and fixes are needed to further develop this tool. In this article, we also discussed how this library can be used in wellness and health processes to understand the evolution of a patient's health and wellness.

```
var response = {
    "activities": [
        {
            "calories": value,
            "distance": value,
            "duration": value,
            "startTime": value,
            "steps": value
        }
    ],
    "goals":
        {
            "caloriesOut": value,
            "distance": value,
            "floors": floors,
            "steps": value
        },
    "summary":{
        "activityCalories": value,
        "caloriesBMR": value,
        "caloriesOut": value,
        "distances":[
            {"activity":"total", "distance": value},
            {"activity":"veryActive", "distance": value},
            {"activity":"moderatelyActive", "distance": value},
            {"activity":"sedentaryActive", "distance": value}
        ],
    "elevation": value,
    "fairlyActiveMinutes": value,
    "floors": value,
    "sedentaryMinutes": value,
    "steps": value,
    "veryActiveMinutes": value
    }
};
```


Fig. 11 Fitbit API example response. An example JSON response from a Fitbit activity sensor that can be transformed to a data source to be visualized by hFigures

Open challenges remain in studying alternative features that can help users identify relationships between measurements, visualize patterns and enable deeper exploration of the data with a higher degree of interactivity.

Availability and requirements

- **Project name:** hFigures
- **Project home page and source code repository:** https://github.com/ledancs/hFigures
- **SciCrunch Resource ID** SCR_014201
- **Operating System:** Platform independent.
- **Programming language:** JavaScript.
- **Other requirements:** Developers willing to deploy the application need to serve the files via a Web server. Users require a Web browser to visualize the application.
- **License:** MIT License.
- **Any restrictions to use by non-academics:** No.

Availability of data

The dataset supporting the conclusions of this article is available in the BioSharing repository with the identifier biodbcore-000734 at the following url: https://biosharing.org/biodbcore-000734.

The dataset is also available at Tampere University of Technology Personal Health Informatics website in the following url: http://www.tut.fi/phi/?p=319.

Additional files

Additional file 1: Nielsen's Heuristic evaluation. The data contains the results form Nielsen's Heuristic Evaluation conducted by three usability experts. (CSV 116 kb)

Additional file 2: Results from laboratory testing. The data contains the task identifier, the average time to completion, number of times the task was successfully completed and the total number of errors. (CSV 209 kb)

Additional file 3: Laboratory testing tasks. The data contains the task identifier and the instructions given to the participants to complete the task. (CSV 618 kb)

Additional file 4: Computer system usability questionnaire results. The data contains the results of the Computer System Usability Questionnaire answered by 14 participants. (CSV 639 kb)

Additional file 5: After scenario questionnaire results. The data contains the results of the After Scenario Questionnaire answered by 14 participants. (CSV 149 kb)

Abbreviations

API: application programming interface; D3: data-driven documents; EHR: electronic health record; hFigures: health figures; hGraph: health graph; HTML:

hypertext markup language; JSON: JavaScript object notation; PHR: personal health record; SVG: scalable vector graphics.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
AL developed the hFigures library. AL and HN wrote jointly this article. HN suggested several use cases that helped shaped the design process. HN contributed in the design and conception of the library. AL, HN and MAM designed the application for health coaching. AL and MAM implemented the application. AL and MAM integrated the application into a Web service. MAM conducted the user testing and analysed the results. HN and AL wrote jointly the introduction and background. AL wrote the implementation and discussion. MAM and AL wrote jointly the evaluation and results section. HN and AL wrote jointly the conclusion. All authors read and approved the final manuscript.

Authors' information
AL and HN are part of the Personal Health Informatics research group from the Department of Signal Processing of Tampere University of Technology. AL is a Software Engineer pursuing PhD studies in Health Data Visualization with a strong background in Web Engineering and Health Information Systems. HN is a PhD senior researcher with background in biomedical engineering, signal processing, electrical engineering, user interface design and service design.

Acknowledgements
This research was supported jointly by TEKES (the Finnish Funding Agency for Technology and Innovation) as part of the Digital Health Revolution project, as well as the European Commission and TEKES under the ARTEMIS-JU WithMe project.

Received: 20 July 2015 Accepted: 10 March 2016
Published online: 22 March 2016

References

1. Hood L, Flores M. A personal view on systems medicine and the emergence of proactive {P4} medicine: predictive, preventive, personalized and participatory. New Biotechnol. 2012;29(6):613–624. doi:10.1016/j.nbt.2012.03.004. Molecular Diagnostics and Personalised Medicine.
2. Lesselroth BJ, Pieczkiewicz DS. Data visualization strategies for the electronic health record In: Berhardt LV, editor. Advances in Medicine and Biology, vol. 16. New York: Nova Science Publishers, Inc; 2012. p. 107–40.
3. Follett J, Sonin J. hGraph: An Open System for Visualizing Personal Health Metrics. Technical report, Involution Studios, Arlington, Massachusetts (April 2012). Accessed 25 Mar 2015. http://www.hgraph.org/docs/hGraph_Whitepaper.pdf.
4. Ledesma A, Nieminen H, Valve P, Ermes M, Jimison H, Pavel M. The shape of health: A comparison of five alternative ways of visualizing personal health and wellbeing. In: Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE; 2015. doi:10.13140/RG.2.1.1321.6725.
5. Bui AAT, Aberle DR, Kangarloo H. Timeline: visualizing integrated patient records. IEEE Trans Inf Technol Biomed. 2007;11(4):462–473. doi:10.1109/TITB.2006.884365.
6. mHealth App Developer Economics 2014: The State of the Art of mHealth App Publishing. Berlin, Germany: research2guidance; 2014. http://mhealtheconomics.com/mhealth-developer-economics-report/.
7. Goetz T. It's Time to Redesign Medical Data. Accessed 25 Mar 2015. http://www.ted.com/talks/thomas_goetz_it_s_time_to_redesign_medical_data?language=en.
8. Idris M, Hussain S, Ahmad M, Lee S. Big data service engine (bise): Integration of big data technologies for human centric wellness data. In: Big Data and Smart Computing (BigComp), 2015 International Conference On; 2015. p. 244–248. doi:10.1109/35021BIGCOMP.2015.7072838.
9. Few S. Information Dashboard Design. North Sebastopol: O'Reilly; 2006.
10. Soomlek C, Benedicenti L. An agent-based wellness indicator: experimental results and future directions. J Inf Technol Res. 2013;6(2):1–23. doi:10.4018/jitr.2013040101.
11. de Ridder M, Constantinescu L, Bi L, Jung YH, Kumar A, Kim J, Feng DD, Fulham M. A web-based medical multimedia visualisation interface for personal health records. In: Computer-Based Medical Systems (CBMS), 2013 IEEE 26th International Symposium On. Porto, Portugal: IEEE; 2013. p. 191–196.
12. Shneiderman B, Plaisant C, Hesse BW. Improving health and healthcare with interactive visualization methods. Technical report, Citeseer. 2013.
13. Baird JC, Noma EJ. Fundamentals of Scaling and Psychophysics. New Jersey: John Wiley & Sons Canada, Limited; 1978.
14. Cleveland W, McGill R. Graphical perception: Theory, experimentation, and application to the development of graphical methods. J Am Stat Assoc. 1984;79(387):531–554.
15. Hoffman P, Grinstein G, Marx K, Grosse I, Stanley E. Dna visual and analytic data mining. In: Visualization'97., Proceedings. Phoenix, Arizona (AZ), USA: IEEE; 1997. p. 437–441.
16. Draper GM, Livnat Y, Riesenfeld RF. A survey of radial methods for information visualization. IEEE Trans Vis Comput Graph. 2009;15(5):759–76.
17. Jankun-Kelly T, Ma KL. Moiregraphs: Radial focus+ context visualization and interaction for graphs with visual nodes. In: Information Visualization, 2003. INFOVIS 2003. IEEE Symposium On. Seattle, Washington (WA), USA: IEEE; 2003. p. 59–66.
18. Lamping J, Rao R. The hyperbolic browser: A focus+ context technique for visualizing large hierarchies. J Vis Lang Comput. 1996;7(1):33–55.
19. Hetzler B, Whitney P, Martucci L, Thomas J. Multi-faceted insight through interoperable visual information analysis paradigms. In: Information Visualization, 1998. Proceedings. IEEE Symposium On. Research Triangle, California (CA), USA: IEEE; 1998. p. 137–44.
20. Spence R. Information visualization. New York: ACM Press; 2001, pp. 85–88.
21. Yang J, Ward MO, Rundensteiner E, et al. Interring: An interactive tool for visually navigating and manipulating hierarchical structures. In: Information Visualization, 2002. INFOVIS 2002. IEEE Symposium On. Boston Massachusetts (MA), USA: IEEE; 2002. p. 77–84.
22. Howell M. Filelight. Accessed 26 Nov 2015. http://methylblue.com/filelight/.
23. Carlis JV, Konstan JA. Interactive visualization of serial periodic data. In: Proceedings of the 11th Annual ACM Symposium on User Interface Software and Technology. San Francisco, California (CA), USA: ACM; 1998. p. 29–38.
24. Spoerri A. Rankspiral: Toward enhancing search results visualizations. In: Information Visualization, 2004. INFOVIS 2004. IEEE Symposium On. Austin, Texas (TX), USA: IEEE; 2004. p. 18–18.
25. Hong JY, D'Andries J, Richman M, Westfall M. Zoomology: comparing two large hierarchical trees. In: Poster Compendium of the IEEE Symposium of Information Visualization (InfoVis). Seattle, Washington (WA), USA: IEEE; 2003. p. 120–121.
26. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. Circos: an information aesthetic for comparative genomics. Genome Res. 2009;19(9):1639–1645.
27. Draper GM, Riesenfeld RF. Who votes for what? a visual query language for opinion data. IEEE Trans Vis Comput Graph. 2008;14(6):1197–1204.
28. Bostock M, Ogievetsky V, Heer J. D^3 data-driven documents. IEEE Trans Vis Comput Graph. 2011;17(12):2301–2309.
29. Chris N. Toward measuring visualization insight. IEEE Comput Graph Appl. 2006;26(3):6–9.
30. Saraiya P, North C, Duca K. An evaluation of microarray visualization tools for biological insight. In: Information Visualization, 2004. INFOVIS 2004. IEEE Symposium On. Austin, Texas (TX), USA: IEEE; 2004. p. 1–8.
31. Roebuck K. Agile Software Development: High-impact Strategies-What You Need to Know: Definitions, Adoptions, Impact, Benefits, Maturity, Vendors. Brisbane: Emereo Publishing; 2012.
32. Beck K. Extreme Programming Explained: Embrace Change. Boston: Addison-Wesley Professional; 2000.
33. Poutanen O, Koivisto AM, Kääriä S, Salokangas RK. The validity of the depression scale (deps) to assess the severity of depression in primary care patients. Fam Pract. 2010;27(5):527–534.
34. The Mozilla Developer Network (MDN). The Typeof Operator - JavaScript Reference. Accessed 25 Mar 2015. https://developer.mozilla.org/en-US/docs/Web/JavaScript/Reference/Operators/typeof.
35. Nielsen J. Usability Engineering. Amsterdam: Elsevier; 1994.
36. Johnson C, Johnston D, Crowle P, et al. EHR usability toolkit: A background report on usability and electronic health records. Rockville,

Maryland (MD), United States: Agency for Healthcare Research and Quality; 2011. https://healthit.ahrq.gov/ahrq-funded-projects/electronic-health-record-usability-toolkit/citation/ehr-toolkitbackground.
37. Nielsen J. 10 Usability Heuristics for User Interface Design. Accessed 19 May 2015. http://www.nngroup.com/articles/ten-usability-heuristics/.
38. Wharton C, Rieman J, Lewis C, Polson P. The cognitive walkthrough method: A practitioner's guide. In: Usability Inspection Methods. Hoboken, New Jersey (NJ), USA: Wiley; 1994. p. 105–40.
39. Peute LW, Jaspers MW. The significance of a usability evaluation of an emerging laboratory order entry system. Int J Med Inform. 2007;76(2):157–68.
40. Karahoca A, Bayraktar E, Tatoglu E, Karahoca D. Information system design for a hospital emergency department: A usability analysis of software prototypes. J Biomed Inform. 2010;43(2):224–32.
41. Cohen T, Kaufman D, White T, Segal G, Staub AB, Patel V, Finnerty M. Cognitive evaluation of an innovative psychiatric clinical knowledge enhancement system. Medinfo. 2004;11(Pt 2):1295–9.
42. Khajouei R, de Jongh D, Jaspers MW. Usability evaluation of a computerized physician order entry for medication ordering. In: Medical Informatics Europe (MIE). Sarajevo, Bosnia-Herzegovina; 2009. p. 532–536.
43. Blackmon MH, Polson PG, Kitajima M, Lewis C. Cognitive walkthrough for the web. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM; 2002. p. 463–470.
44. Sears A. Heuristic walkthroughs: Finding the problems without the noise. Int J Hum Comput Interact. 1997;9(3):213–234.
45. Beuscart-Zéphir MC, Elkin P, Pelayo S, Beuscart R, et al. The human factors engineering approach to biomedical informatics projects: state of the art, results, benefits and challenges. IMIA Yearb. 2007;2:109–127.
46. Newman WM. 10. on simulation, measurement, and pfeeewiaie usability evaluation. Comment Damaged Merch. 1998;13:316.
47. Ericsson KA, Simon HA. Verbal reports as data. Psychol Rev. 1980;87(3):215.
48. Borycki EM, Lemieux-Charles L, Nagle L, Eysenbach G. Novice nurse information needs in paper and hybrid electronic-paper environments: a qualitative analysis. In: Medical Informatics Europe (MIE). Sarajevo, Bosnia-Herzegovina; 2009. p. 913–917.
49. Currie LM, Graham M, Allen M, Bakken S, Patel V, Cimino JJ. Clinical information needs in context: an observational study of clinicians while using a clinical information system. In: AMIA Annual Symposium Proceedings, vol. 2003. Washington DC, USA: American Medical Informatics Association; 2003. p. 190.
50. Hasman A, et al. Development of methods for usability evaluations of ehr systems. In: Ubiquity: Technologies for Better Health in Aging Societies: Proceedings of MIE2006, vol. 124 IOS Press. Maastricht, The Netherlands: Medical Informatics Europe; 2006. p. 341.
51. Wu RC, Orr MS, Chignell M, Straus SE. Usability of a mobile electronic medical record prototype: a verbal protocol analysis. Inform Health Soc Care. 2008;33(2):139–49.
52. Lewis JR. Ibm computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. Int J Hum Comput Interact. 1995;7(1):57–78.
53. Lewis JR. Psychometric evaluation of the post-study system usability questionnaire: The pssuq. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 36. Thousand Oaks, California (CA), USA: SAGE Publications; 1992. p. 1259–1260.
54. Jaspers MW, Peute LW, Lauteslager A, Bakker PJ. Pre-post evaluation of physicians' satisfaction with a redesigned electronic medical record system. Stud Health Technol Inform. 2008;136:303.
55. Goud R, Jaspers MW, Hasman A, Peek N. Subjective usability of the cardss guideline-based decision support system. Stud Health Technol Inform. 2008;136:193.
56. Lewis JR. Psychometric evaluation of an after-scenario questionnaire for computer usability studies: the asq. ACM SIGCHI Bull. 1991;23(1):78–81.
57. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. Int J Hum Comput Interact. 2008;24(6):574–594.
58. United States Department of Health and Human Services. Physical Activity Guidelines for Americans: Be Active, Healthy, and Happy! ODPHP publication No. U0036. Washington D. C. October 2008. 2008. http://health.gov/paguidelines/pdf/paguide.pdf.
59. Physical Activity Guidelines Advisory Committee. Physical activity guidelines advisory committee report, 2008. Washington, DC: US Department of Health and Human Services. 2008;2008:5.
60. Tudor-Locke C, Bassett Jr DR. How many steps/day are enough? Sports Med. 2004;34(1):1–8.
61. Oja P, Mänttäri A, Pokki T, Kukkonen-Harjala K, Laukkanen R, Malmberg J. UKK walk test. Tampere, Finland: Centre for Health Promotion Research; 2013. http://www.ukkinstituutti.fi/en/products/ukk_walk_test.
62. Laukkanen R, Oja P, Pasanen M, Vuori I. Validity of a two kilometre walking test for estimating maximal aerobic power in overweight adults. Int J Obes Relat Metab Disord: J Int Assoc Study Obesity. 1992;16(4):263–268.
63. Suni J, Husu P, Rinne M. Fitness for Health: The ALPHA-FIT Test Battery. Tampere, Finland: European Union, DG SANCO, and UKK Institute for Health Promotion Research; 2009. http://www.ukkinstituutti.fi/en/alpha.
64. Firstbeat Technologies Ltd. Stress and Recovery Analysis Method Based on 24-hour Heart Rate Variability. White Paper. Accessed 25 Mar 2015. http://www.firstbeat.com/userData/firstbeat/research-publications/Stress-and-recovery_white-paper_2014.pdf.
65. Teisala T, Mutikainen S, Tolvanen A, Rottensteiner M, Leskinen T, Kaprio J, Kolehmainen M, Rusko H, Kujala UM. Associations of physical activity, fitness, and body composition with heart rate variability-based indicators of stress and recovery on workdays: a cross-sectional study. J Occup Med Toxicol. 2014;34:26–40.
66. Scheier MF, Carver CS, Bridges MW. Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): a reevaluation of the life orientation test. J Pers Soc Psychol. 1994;67(6):1063.
67. Faulkner L. Beyond the five-user assumption: Benefits of increased sample sizes in usability testing. Behav Res Methods Instrum Comput. 2003;35(3):379–383.
68. Berry DL, Halpenny B, Bosco JL, Bruyere J, Sanda MG. Usability evaluation and adaptation of the e-health personal patient profile-prostate decision aid for spanish-speaking latino men. BMC Med Inform Decis Mak. 2015;15(1):56.
69. Zhang J, Johnson TR, Patel VL, Paige DL, Kubose T. Using usability heuristics to evaluate patient safety of medical devices. J Biomed Inform. 2003;36(1):23–30.
70. Shneiderman B. Designing the User Interface: Strategies for Effective Human-computer Interaction, vol. 3. Boston: Addison-Wesley Reading, MA; 1992.
71. Molich R, Nielsen J. Improving a human-computer dialogue. Commun ACM. 1990;33(3):338–348.
72. Tognazzini B. First Principles of Information Design. Accessed 25 Mar 2015. http://www.asktog.com/basics/firstPrinciples.html.
73. Finnish Advisory Board on Research Integrity: Ethical Principles of Research in the Humanities and Social and Behavioural Sciences and Proposals for Ethical Review. Helsinki. 2009. Finnish Advisory Board on Research Integrity. Accessed 8 Mar 2016. http://www.tenk.fi/en/ethical-review-human-sciences.
74. Lewis JR. Psychometric evaluation of the pssuq using data from five years of usability studies. Int J Hum Comput Interact. 2002;14(3-4):463–488.
75. Allen S. Aight – JavaScript Shims and Shams for Making IE8-9 Behave Reasonably.
76. Rind A, Wang TD, Wolfgang A, Miksch S, Wongsuphasawat K, Plaisant C, Shneiderman B. Interactive information visualization to explore and query electronic health records. Found Trends Hum Comput Interact. 2011;5(3):207–298.

PUBLICATION
IV

Health Timeline: An Insight-based Study of a Timeline Visualization of Clinical Data

A. Ledesma, N. Bidargaddi, J. Strobel, G. Schrader, H. Nieminen, I. Korhonen and M. Ermes

BMC Medical Informatics and Decision Making 19.170 (2019), 1

Publication reprinted with the permission of the copyright holders

Ledesma *et al. BMC Medical Informatics and Decision Making* (2019) 19:170
https://doi.org/10.1186/s12911-019-0885-x

BMC Medical Informatics and Decision Making

RESEARCH ARTICLE Open Access

Health timeline: an insight-based study of a timeline visualization of clinical data

Andres Ledesma[1*], Niranjan Bidargaddi[2], Jörg Strobel[3], Geoffrey Schrader[4], Hannu Nieminen[1], Ilkka Korhonen[1] and Miikka Ermes[5]

Abstract

Background: The increasing complexity and volume of clinical data poses a challenge in the decision-making process. Data visualizations can assist in this process by speeding up the time required to analyze and understand clinical data. Even though empirical experiments show that visualizations facilitate clinical data understanding, a consistent method to assess their effectiveness is still missing.

Methods: The insight-based methodology determines the quality of insights a user acquires from the visualization. Insights receive a value from one to five points based on a domain-specific criteria. Five professional psychiatrists took part in the study using real de-identified clinical data spanning 4 years of medical history.

Results: A total of 50 assessments were transcribed and analyzed. Comparing a total of 558 insights using Health Timeline and 576 without, the mean value using the Timeline (1.7) was higher than without (1.26; $p<0.01$), similarly the cumulative value with the Timeline (11.87) was higher than without (10.96: $p<0.01$). The average time required to formulate the first insight with the Timeline was higher (13.16 s) than without (7 s; $p<0.01$). Seven insights achieved the highest possible value using Health Timeline while none were obtained without it.

Conclusions: The Health Timeline effectively improved understanding of clinical data and helped participants recognize complex patterns from the data. By applying the insight-based methodology, the effectiveness of the Health Timeline was quantified, documented and demonstrated. As an outcome of this exercise, we propose the use of such methodologies to measure the effectiveness of visualizations that assist the clinical decision-making process.

Keywords: Clinical data, Data visualization, Health informatics, Electronic health record, Insight-based methodology

Background

Researchers estimate that clinical data will grow [1], from 153 exabytes in 2013 to 2314 in 2020 [2]. Electronic health records (EHRs) constitute most of the clinical data. These records are the "purest type" of electronic clinical data obtained at the point of care [3]. EHRs collected over time represent a patient's clinical history. Healthcare professionals rely on them for diagnosis and treatment.

As clinical data increases in volume, so does the potential value for healthcare providers to benefit from it. IT-based information systems can summarize and visualize clinical data. Health Informatics is the "interdisciplinary study of the design, development, adoption and application of IT-based innovations in healthcare services delivery" [4]. Numerous research efforts have addressed the need to understand clinical data by designing and building a variety of visualizations [5, 6].

North suggests that the purpose of visualizations is to help readers derive insights [7]. A measure of an effective visualization is its ability to generate new insights that go beyond predefined data analysis tasks [8]. The effectiveness of clinical data visualizations is an area of active research [5, 6, 9]. Graphical representation techniques have been addressed in several studies [10]. One significant example was reported by Goetz who demonstrated that data presentation greatly affects the understanding of a person's health [11]. Thus, visualization concepts applied to clinical data should enable healthcare professionals to obtain insights about the condition of the

*Correspondence: andres.ledesma@tut.fi
[1]Tampere University, Faculty of Medicine and Health Technology, Tampere, Finland
Full list of author information is available at the end of the article

© The Author(s). 2019 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

patient more comprehensively and efficiently than traditional textual or tabular presentations.

Researchers have implemented several visualization tools for clinical data [5]. A literature survey identified fourteen computerized tools for visualizing EHRs [6]. The articles describe in detail the design and implementation of these tools. However, they do not use a structured method, such as the insight-based methodology, to perform assessments.

We extracted data from the Australian “My Health Record” initiative [12] and developed a computerized solution to visualize clinical data chronologically. The solution was named Health Timeline to emphasize the significance of time in the clinical data. The software organizes and displays the clinical data in an interactive timeline using visual enhancements to facilitate readability. This article documents the assessment of Health Timeline which applies the insight-based methodology. Five psychiatrists took part in the study by reviewing the clinical data of five de-identified patients. The psychiatrists were instructed to use the “thinking aloud process”. This process consists of verbalizing the thought pattern followed while inspecting the data. Their findings were recorded using a voice recording software. These findings were later transcribed and analyzed following the insight-based methodology.

Related work

The findings of Lesselroth and Pieczkiewicz [5], as well as those of Rind and colleagues [6] are further discussed. The study reported in this article relates to other computerized solutions. The key distinction is the assessment methodology.

Scope

The focus of this study is on time-based visualizations of clinical data (EHR) and the assessment methodology used to validate them. Time-based visualizations are graphical representations of data collected over time. The research literature has a large number of data visualization techniques that vary in their strategies. However, we refined our search to include only visualization tools based on time, and the longitudinal nature of the data. The search was narrowed down further to only include those techniques that were used in the context of clinical data. We were particularly interested in the assessment methodology used to evaluate these visualizations.

Review of similar solutions in the healthcare context

LifeLines is a computerized tool that displays clinical data [13] using dots positioned along horizontal lines [14]. A study showed that participants responded 50% faster to a “post-experimental memory test” ($p < 0.004$) [15]. LifeLines was extended in a second version with support for aggregation of temporal events [16]. The focus was on emphasizing prevalence and temporal order of the clinical data. A study revealed that the clinicians were able to confirm hypotheses on the hospital length of stay of patients. LifeLines [13] and LifeLines2 [16] are Java software applications, thus they must be installed in a Java-capable device. These tools provide data filters to narrow down the data exploration and enable the user to focus on certain aspects of the timeline.

TimeLine is a visualization software that displays EHRs chronologically [17]. The data is grouped by categories and displayed along a visual timeline. No assessment of the software was reported in the article. Timeline [17] is the closest application to Health Timeline. It features web support, EHR interfacing and a timeline representation of data with a focus on oncology. Timeline also has a large number of features such as causal models, imaging files, data search, disease progression visualization and data category toggling. It is probable that using Timeline involves a learning curve as it offers several features that would require the user to become acquainted with.

LifeFlow is a visualization tool that summarizes data in sequences using temporal spacing of events [18]. One physician took part in a briefing interview about the visualization of patient transportation data. EventFlow is a drug prescription pattern visualization [19]. A study on the use of asthma drugs was conducted to identify patterns that complied with regulations. LifeFlow and EventFlow are also software applications that require installation. The data are not visualized in a timeline but instead the representation is chronologically ordered as a series of events and outcomes. These visualizations are optimized for understanding the causes and outcomes of patient admissions to hospital.

Assessment methods

Bertini and colleagues [20] made a strong case for the objective assessment of visualization tools. A literature review on assessment methods for information visualization reports a number of practical cases and proposes a classification of these methods. The Visual Data Analysis and Reasoning (VDAR) classification group is relevant to our study because it emphasizes the decision-making process, knowledge discovery and visual data analysis. We found that no assessments of this kind have been conducted using clinical data and medical experts.

Summary

Time-based visualizations have been found helpful in several use cases. However, without a systematic assessment method, it is difficult to demonstrate how they improve the understanding of the data. To provide a contribution towards the good practice of assessment methods for clinical data visualizations, we conducted and documented the assessment of the Health Timeline using the insight-based methodology. This methodology has been previously used in bioinformatics [8, 21–23] and well-being data analysis [9, 24].

Key contribution

The key contribution is to test out the insight-based methodology for the assessment of a clinical data visualization of EHRs. To our understanding, researchers have not documented an insight-based evaluation in this specific context.

Evaluating visual data analysis and reasoning

The VDAR classification group proposed by Lam and colleagues [25] is relevant to this study due to the nature of the clinical data. The purpose of our study is to measure the degree to which a time-based visualization can assist the data analysis and the reasoning process of a clinician when analyzing the clinical history of a patient. In the context of clinical data, we found in the research literature that studies tend to focus on the usability of the interactions with the visualization system. Other studies conduct briefing interviews and questionnaires to try to assess the participants' understanding of the data. Nevertheless, a consistent evaluation methodology is still missing.

The purpose of our study is to measure the degree to which a time-based visualization can assist the data analysis and reasoning process of a clinician when analyzing the clinical history of a patient. In the context of clinical data, we found in the research literature that studies tend to focus on the usability of the interactions with the visualization system. Other studies conduct briefing interviews and questionnaires to try to assess the participants' understanding of the data. Nevertheless, a consistent evaluation methodology is still missing.

Contextualizing the insight-based methodology

The insight-based methodology provides an assessment method focusing on the insights generated by a visualization [7]. In this approach, the insights formulated by the visualization audience are assessed in a Likert scale from 1 to 5. To contextualize the insight-based methodology to the clinical data domain, we consulted with two subject matter experts to assist us in establishing the assessment criteria of the insights.

From bioinformatics to clinical data

The insights gained in this study are analogous to those observed in the bioinformatics context because they demonstrate different degrees of completeness and accuracy in the understanding of the data represented by the visualization. In the bioinformatics context, predetermined information deemed as correct and complete, was used to compare the insights obtained in the study [8, 21, 22]. The same is true for our study, since two professional psychiatrists reviewed the evaluation criteria. Due to their experience, these psychiatrists were qualified to evaluate the degree of completeness and accuracy of the insights. These experts were also acquainted with the clinical data used in our study.

The insight-based methodology could be used in future studies to quantify the effectiveness of clinical data visualizations. This study provides details on the application of the methodology, including the process required to establish criteria to assess the insight value. The insight value is a key component to analyze and interpret the results of the evaluation. We hope to provide enough documentation about the steps followed to apply the insight-based methodology. The purpose is to provide a case study on evaluating visualizations using the insight-based methodology contextualized to the healthcare domain.

Methods

In this study we documented the application of the insight-based methodology. This process involved the recording of insights, the formulation of a consistent criteria for insight assessment and the application of those criteria to collect and analyze the results. To this extent, we define the goals of the study as follows:

1 To apply the insight-based methodology to assess the Health Timeline visualization using clinical data and Healthcare professionals.
2 To document the assessment process so that it may serve as a reference for future studies on clinical data visualizations.

In this section we present the experiment protocol and the hypothesis that we proposed on the outcome of the study. We also present the Health Timeline and explain its design rationale. The counterpart of the Health Timeline is a table showing the clinical history of the patient in textual format. We refer to this table as the "baseline". This section also provides information about the clinical data used in the study, as well as the order in which the visualizations were displayed to the participants. Finally, we explain the metrics used to analyze and interpret the results of the study in accordance with the insight-based methodology.

Hypothesis

We propose the following hypothesis: a time-based visualization will enable clinicians to obtain valuable insights regarding the clinical condition of patients. A valuable insight demonstrates a complete and accurate understanding of the data. This understanding of the data is corroborated by the evaluation criteria used to assess the insight value. The criteria were designed and reviewed by subject matter experts. Valuable insights will translate to better patient diagnosis. This in turn will result in better decision-making processes thereby improving patient outcomes.

Health timeline

The Health Timeline was designed to provide a simple and interactive way to review a patient's medical history using a time-based visualization. For healthcare professionals, the visualization software aims to assist in the decision-making process when assessing the overall status of the patient.

The Health Timeline is a cloud-based software that acquires data from the Medicare and Pharmaceutical Benefits Scheme (PBS) claims systems as part of Australia's "My Health Record" initiative. These claims include a multitude of information as well as pharmaceutical aspects of patient data, including consultations, laboratory tests, in-patient admissions and other relevant information. The application handles the consent, sign up and compliance processes for using the clinical data. It has received conformance certification from Australia's National E-Health Transition Authority (NEHTA) and the Commonwealth Department of Health for use with production servers and deployment in clinical practice.

Graphical representation and interactivity

The events in the timeline are grouped in adjacent rows. Each event is depicted as a box. The boxes are grouped in rows with labels on the left edge of the timeline. The events are shown at the corresponding start date position displayed on the horizontal axis. The width of the event box is adjusted to the corresponding end date. The group labels correspond to domains that are clinically relevant, for example medical treatments, visits to the general practitioner, psychiatric consultations and laboratory tests. Users can interact with the timeline by adjusting the displayed time frame (start and end dates). This is done by zooming in and out or panning left and right. Figure 1 shows an example of the Health Timeline portraying the clinical history of a patient.

Design rationale and requirements

The rationale behind the design of Health Timeline is to provide a web-based (usable with any device with a web browser) visualization tool for clinicians to assist in the process of understanding the overall condition of a patient. The Health Timeline extracts information from EHR systems and presents the data in an interactive interface.

The design principles followed to develop Health Timeline were based on a simple and comprehensive visualization. In the "Related work" section, we identified similar visualization tools. These tools supersede the Health Timeline in terms of features and complexity. However, in this study, the participants did not receive an introduction or explanation on how Health Timeline software works, this was part of the intended strategy of the study. That is, to design and develop a simple yet comprehensive visualization that would require no introduction to clinicians. In other words, we aim to provide an intuitive visualization of clinical

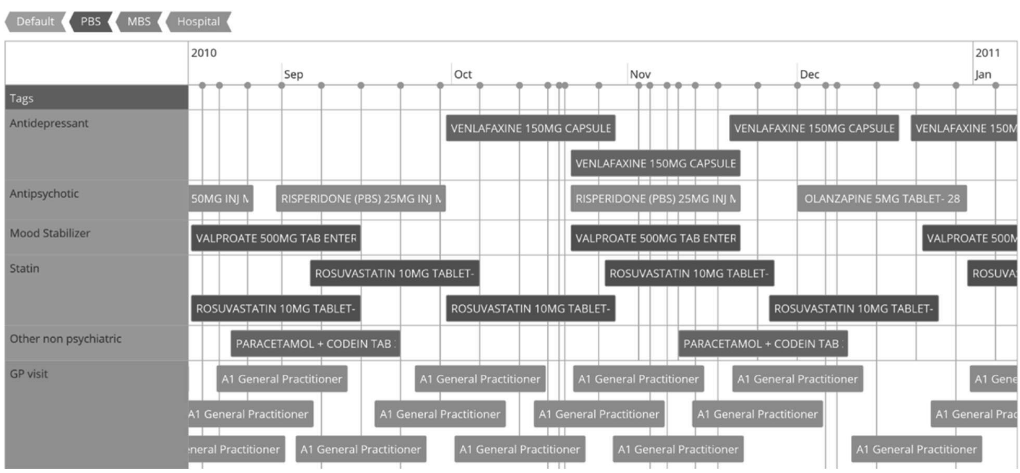


Fig. 1 Health Timeline. Timeline visualization showing the collection of EHRs of a patient used in the study

data that clinicians can understand without requiring training.

The Health Timeline builds on similar time-based visualizations and tries to address the challenge of representing data clearly and in a meaningful way without overwhelming the user with excessive details and complex interfaces. During the first assessment, the participants became acquainted with the Health Timeline visualization. In contrast, more complex visualization software, such as LifeFlow [18] or EventFlow [19] would probably require the participants to get acquainted with the tool before conducting the experiment.

The key requirements of Health Timeline came from Australia's "My Health Record" initiative and were gathered from interviews with clinicians and Information Technology experts with a background in data visualization. These requirements are summarized as follows:

- Web-based visualization software
- Acquisition of data from Medicare and PBS
- Cloud-hosted application software
- A clean and simple user-interface. For the purpose of the study, no tutorials or explanations were given to the participants. This was done to "stress test" the design of the graphical interface.
- Interactivity for data exploration (time adjustment via panning and zooming)

Data representation baseline

The baseline representation of the clinical data is the current interface that clinicians have to the EHRs. The tabular data show in the baseline is the starting point of the study and contains the same information as the Health Timeline. The tabular data representation allows the user to sort the events chronologically, alphabetically, by document type and category. Figure 2 shows an example the tabular format representation of the baseline.

Insight-based methodology

The study documented in this article follows the insight-based methodology as proposed by North [7]. An insight is defined as "the capacity to gain an accurate and deep understanding" (Oxford, 2016). As presented in the literature survey, the most prevalent approach to evaluate clinical data visualization is to conduct briefing interviews or to measure the performance of the user conducting a set of predefined tasks. By contrast, the insight-based methodology focuses on recognition and quantification of insights gained from exploratory use of the data visualization. In the context of this study, an insight is a unit of discovery based on observation [8, 9, 21, 22, 24].

Insights have a quantifiable value based on the assessment criteria. The proposed criteria take into consideration the following characteristics of an insight [7, 8, 21–23]:

- **Observation**: The observation or finding provided by the participant during the process of analyzing the data via a representation.
- **Time**: The amount of time taken to reach the insight.
- **Domain Value**: The value, importance, or significance of the insight.
- **Hypotheses**: Some insights enable users to identify a new relevant hypothesis.
- **Directed versus Unexpected**: Directed insights are those that answer specific questions. Unexpected insights are those that were not considered in the design of the study.
- **Correctness**: Insights can be correct or incorrect depending on the data represented in the visualization. Some insights are incorrect conclusions that result from misinterpreting the data visualization. For our study, the insights formulated by the participants need to be clinically valid assessments on the patient's condition.

Document	Category	Item	Start Date	End Date	Media
PBS	Antipsychotic	OLANZAPINE 5MG TABLET- 28	28/06/2013	28/07/2013	
PBS	Antipsychotic	OLANZAPINE 5MG TABLET- 28	14/06/2013	14/07/2013	
PBS	Statin	ATORVASTATIN 80MG TABLET- 30	14/06/2013	14/07/2013	
MBS	GP visit	A1 General Practitioner	07/06/2013	11/06/2013	
PBS	Other non psychiatric	RABEPRAZOLE 20MG TAB ENTERIC SO	07/06/2013	07/07/2013	
PBS	Antidepressant	CITALOPRAM 20MG TABLET- 28	07/06/2013	07/07/2013	
PBS	Statin	ATORVASTATIN 80MG TABLET- 30	20/05/2013	19/06/2013	
PBS	Antipsychotic	OLANZAPINE 5MG TABLET- 28	20/05/2013	19/06/2013	

Fig. 2 Visualization Baseline. The baseline visualization showing a set of EHRs of a patient used in the study

Time, domain value, hypothesis and correctness are characteristics included in this study. We did not compile a set predefined of insights that the participants were required to formulate. Instead, the focus of the study was on allowing the participants to explore the data and formulate observations freely. This was designed to simulate the typical use case in which a healthcare provider is presented with a patient history and has to become acquainted with the data in a short span of time before the consultation. Therefore, the directed versus unexpected characteristic was not used in our study.

Domain value criteria and insight value coding

General Criteria: The domain value of an insight is determined using a five-point Likert scale as suggested in the methodology [7] and applied in previous studies [8, 9, 21–23]. The value depends on the insight's complexity and depth. An insight is complex if it can associate more than one element in the data by providing a relationship or association between them, for example; "if "a" increases then "b" decreases". The depth is the degree to which the underlying reasons behind the data are explained, that is, a plausible explanation was given that could justify the state of the data. For instance, an explanation for values outside of the recommended range for certain physiological measurements could indicate the presence of a disease. In our study, we recognized that additional refinement was needed for the evaluation criteria used to assess the insight value, due to the complexity of the data.

Longitudinal Nature: Previous studies focused on the visualization of "static" data. The data used for this study was essentially chronological since it was comprised of clinical histories of psychiatric patients. The medications, appointments, inpatient and outpatient treatments were all visualized over time for the duration of the clinical history. Therefore, suitable criteria should evaluate the insights considering the temporal nature of the data.

Criteria Formulation and Review Process: To determine the insight value, we applied the complexity and depth criteria as used in previous studies. However, as the Insight-based Methodology had not been previously applied in this context and lacked a precedent, to overcome this limitation, we asked two professional psychiatrists to assess the usefulness of the insights using a five-point Likert scale.

The consulting psychiatrists analyzed a sample of 100 insights (extracted from the assessments made by the participants) and evaluated them according to what they considered informative and helpful in understanding the patient's condition. The consulting psychiatrists were acquainted with the data and thus were able to understand the underlying condition and overall health of the patients. The insights provided to the consulting psychiatrists included the patient identifier number and the text transcribed manually from the voice recording. This allowed them to have a reference to the original data to objectively determine if an insight was accurate and meaningful in understanding the medical history and clinical condition of the patient.

Agreed Criteria and Value Coding: The result of this process was a set of rules that enabled us to systematically and consistently assign a numeric value (domain value) to the insights. For example, an insight with a value of one, describes an event in the timeline without associations to other events or without possible explanations for the causes behind the event. Examples of these insights are: "the patient has an elevated heart-rate" or "the patient is taking medication for diabetes".

Insights that describe associations of events regarding their frequency, pattern and irregularities were evaluated with two points. These insights connect multiple occurrences of similar events and show that the clinician can track down a pattern. For example, insights like "the patient has regular appointments with a General Practitioner" or "the patient is taking a high dose of the medication with regular frequency". These insights have a time component as the participant identifies a regularity or irregularity in their occurrence.

The insights that include theories that could explain some or all the symptoms and thus the reasoning behind a prescription, or a specialist appointment, were evaluated with three, four or five points depending on how much information they could combine to produce a valid clinical statement. These insights could be considered a hypothesis and could explain valid clinical scenarios that result in drug treatment, laboratory tests, specialists' visits and in-patient treatments. The following insight provides an illustration: "the patient has experienced depression and anxiety, that would explain the prescription and regular use of the drug treatment and also visits to the specialists, this also ties together an emergency episode in January 2013 and another admission in June, overall the patient's mental health improved towards the summer and it seems that changing the drug treatment improved the outlook". Table 1 summarized the criteria for each value with examples extracted from the transcription of the assessments.

Thinking aloud process

The insight-based methodology recommends two mechanisms to record the insights, the "thinking aloud process" and the use of a written diary to record the steps taken during the data analysis. We aimed at making the participation in the study as simple as possible since the participants were professional psychiatrists with busy schedules. For this reason, we opted to use the "thinking

Table 1 Criteria used to determine the insight value

Value	Criteria	Example
1	Describe the data. No patterns or periodicity spotted. Values described as "low" or "high". No awareness of the times an event appears in the dataset.	"this is a patient on injectable antipsychotic medications"
2	Describe periodicity or frequency of an event. Found patterns, irregularities and amount of repetitions. No conjectures or assumptions.	"the patient has quite a number of GP visits in 2011, 2012 and 2013"
3	Requires a conjecture or assumption. Try to explain why an event or value is repeating, missing, following a pattern. Speculation on the treatment, status, follow-up, treatment or behavior of the patient. Prediction on the future status of the patient based on a single event. No correlation with other events. Single conjecture explaining one phenomenon	"patient also has an investigation suggesting that he has some metabolic disease"
4	Conjecture or assumption about two or more events. Explanation of probable cause and effect. Ties two events or phenomena together with a probable reason or explanation. Not all the elements in the dataset are explained some relationships remain unknown to this insight.	"in summary I think it is a patient with psychotic illness"
5	Hypothesis that ties together the discovered elements and events into one possible explanation. Explains relationship between events. Explains probable underlying reasons of the events. Ties together all the events mentioned beforehand.	"overall this patient presents quite of a complex picture of mainly depression probably complicated with psychotic component anxiety"

Examples are provided from the insights we obtained during the study

aloud process" and capture the insights, comments and other statements via phone calls and recording software. We later transcribed the recordings manually to proceed to the analysis stage of the study.

The "thinking aloud process" is one of the most common techniques in usability studies [26]. It consists of asking the participant to verbally express the thought process while using the system under testing. In this study, we asked the participants to follow a link received via email and once they were ready to begin, they called a number and started the "thinking aloud process" of verbalizing their thoughts as they explored the data. This was an effective and convenient way to conduct the study since the participants had the freedom to conduct the assessments at their convenience.

By using the "thinking aloud process" we were able to capture the actions the participants were conducting during the assessments, for example scrolling, panning and zooming to a particular time frame.

Clinical data

De-identified patient data used for the purpose of the study was obtained with consent from mental health patients with various psychiatric disorders [27]. The patients were either outpatients seen in clinics in rural South Australia (77) or patients reviewed in an urban hospital emergency department (63). Ethical approval for the study was obtained from the South Australian Health Ethics Research Ethics Committee and the Medicare Australia Ethics Committee. Each patient gave written consent to use their data for research purposes. From the pool of 144 patients, the de-identified data of five patients was selected for this evaluation based on the complexity of diagnosis and treatment.

Patient data selection

The patient selection was based on the complexity of the diagnosis. All five cases had a diagnosed mental disorder and received continuous care and monitoring throughout their treatment. Three cases had in-patient episodes. A fourth patient was compliant with the treatment and the disorder was managed throughout the recorded period. A fifth patient had substantial changes in overall mental health.

Clinical data included in the study

The selected data included Pharmaceutical Benefit claims, Medicare Benefits claims and hospitalization dates over 3 years between 2012 and 2014. The data also included visits to General Practitioners, specialists, laboratory tests, emergency hospital admissions, drug treatment, among other categories. The claims records were sourced from Medicare Australia and hospitalization events were sourced from SA Health. Australia's Pharmaceutical Benefits Scheme (PBS) contains information about the type of medication, amount of medication supplied, and the date of supply, while the Medicare Benefits Scheme (MBS) contains information on tests and services by Medicare eligible practitioners and the associated date of the service. The PBS and MBS measures were relabeled into clinical terminologies by psychiatrists involved in the project. The data used are considered a clinical history of a patient as part of the EHR.

Experiment protocol

The participating psychiatrists received written instructions on how to conduct the assessments. The instructions contained 10 web links used to access a web portal that displayed the patient data. Five assessments were conducted with the Health Timeline and another five with the tabular format. The web links displayed one patient data set at a time in either representations.

A log-out timer was set after accessing each link. This was done to provide the psychiatrist with three minutes to analyze the available information and formulate insights. Subsequently, the psychiatrist was automatically logged out of the system and the assessment session was concluded. The psychiatrists accessed the links with their personal computers and thus they were familiar with the browser and operating system of their choice. The psychiatrist's observations during the assessment sessions were recorded on an answering machine over a phone connection.

As in previous studies [8, 9, 21–23], the assessments were transcribed from the recorded sessions. The transcriptions were annotated with the times at which the insights were mentioned by the participant. The insights were evaluated using the previously specified criteria.

Visualization presentation and order

All five patients were presented with both visualizations to account for control. The display order was controlled to first show the Health Timeline and direct any bias towards this visualization upon starting the assessments. There was no introduction or instructions given to the participants on how the Timeline works. The idea was to build a visualization that would be intuitive from the start.

The participants conducted the same assessments independently. The participants were asked to make one assessment at a time. The order was the same for all the participants. Each assessment featured the clinical data of a single patient. The patient clinical data was presented to the participant twice, one time in tabular data and another in the Health Timeline visualization. The participants conducted a total of 10 assessments, 5 with the Health Timeline and another 5 with the tabular data. The visualizations were alternated throughout the study. The first assessment used the Health Timeline, the second tabular data, the third Health Timeline, the fourth tabular data and so on. The order of the patients (clinical data) was also controlled so that no consecutive visualization would show the same patient data.

Time constraints

The experiment was intended to be performed by clinicians. As such, time was a major constraint. Empirically we have observed that taking no more than 30 min of their time would be ideal. Keeping each assessment to no more than three minutes was considered to be short enough for them to be willing to take part in the experiment. The 30-min time frame was determined by the three-minute duration of each assessment, a total of 10 assessments which showed the clinical data of five patients with both visualizations, Health Timeline and the textual data. Additionally, we allowed clinicians to take the assessments at their convenience, so if a clinician was able to spare three minutes, then that would be enough to conduct one of the 10 assessments.

Sample size

A total of 10 psychiatrists were invited to take part in the study. Five of them accepted the invitation. Each participant conducted a total of 10 assessments, 5 with the Health Timeline and 5 in textual data. The presentation order was the same for all participants.

Participants

We recruited five professional psychiatrists to participate in the assessments. The participants had experience in the field and had previously worked with similar patients as those used in the study. The participants also expressed their written consent to participate.

Data analysis

A total of 121 min of audio were transcribed. The transcribed text for each insight was organized using a spreadsheet that included the patient identifier, the participant identifier and the time at which the insight was mentioned. The data set produced was saved as a comma separated value file for further analysis and contained the following attributes (Additional file 1):

- The **experimentId** refers to the sequential identifier number of the assessments.
- The **patientId** refers to an identifier created for the sole purpose of this study and is used to link together the assessments with the de-identified patient data. The identifier does not contain clinical or personal information.
- The **assessmentId** refers to the assessment during which the participants used either the Timeline or the baseline to conduct their observations.
- The **visualization** refers to either the Health Timeline (referred to as timeline) or the baseline (referred to as table).
- The **time** refers to the time at which the insight began to be voiced by the participant. In some cases, multiple insights share the same starting time indicating that they were given by the participant during the same sentence (thought process).

- The **insight** refers to the textual insight as transcribed from the audio.
- The **value** refers to the domain value of the insight which follows the criteria detailed in this article.

Metrics

As in previous studies [8, 9, 21–23], we analyzed the data and applied the following metrics to evaluate the data representations:

- **Metrics per assessment**: For each assessment we calculated the average number of insights, as well as their average and cumulative value.
- **Number of insights**: the total count of insights across all the assessments. We also counted the number of insights separated by their value (from one to five points).
- **Time to first insight**: the time required by the participant to formulate the first insight regardless of its value. Additionally, we also calculated separately the insight time for each of the values.

Statistical tests

Mann–Whitney–Wilcoxon (MWW) was used to calculate the statistical significance of the metrics per assessment and the time to first insight since they are not normal distributions. MWW is the recommended method for these distributions [28, 29].

The statistical significance of the total number of insights was calculated using standard Chi squared because the two populations could be treated as categorical data. The insight value corresponds to the categories, the values are discrete from 1 to 5 using a Likert scale. Therefore, the statistical significance can be obtained with Chi squared [29].

Results

The participants completed all 50 assessments per the protocol. The total duration of resulting recordings was 78,783 s, mean and standard deviation was calculated for the distributions ($69.47\mu \pm 51.63\sigma$). The recordings resulted in 1,134 transcribed insights (per participant $22.68\mu \pm 9.70\sigma$).

Metrics per assessment

The collection of insights was analyzed to calculate the average number of occurrences, the mean and cumulative value for each assessment. Figure 3 illustrates, using box plots, the difference in the metrics per assessment and Table 2 summarizes the characteristics of the distributions such as the minimum, maximum, mean, median and standard deviation.

The distribution of the number of insights per assessment was not statistically significant, with the Timeline having an average of 22.32 insights compared to the baseline with 23.04 ($p = 0.7047$). The mean value with the Timeline (1.70) was higher than the baseline (1.26; $p = 0.01$). Similarly, the cumulative value with the Timeline (34.68) was significantly higher than the baseline (24.96; $p = 0.01$).

Number of insights

The value of each insight was collected from the dataset. The distribution is shown in Table 3 as a total count of insights by value. The total distribution of insights was not statistically significant. The Timeline (558) had a lower count that the baseline (576; $p = 0.7$). Table 4 shows the characteristics of the value distribution. The average and median insight value with the Timeline ($\mu = 1.0833$; $Md = 1$; $\sigma = \pm 1.10$) was higher than the baseline ($\mu = 1.55$; $Md = 2$; $\sigma = \pm 1.25$; p< 0.01)

Table 3 shows that the distribution was statistically significant for insights of value 1 ($p = 0.03$), 2 ($p < 0.01$), 4 ($p < 0.01$) and 5 ($p < 0.01$) but not for value 3 ($p = 0.53$). Insights of value 3 or more demonstrate an understanding of the data that can detect patterns. A total of 81 insights were observed using the baseline compared to 118 with the Health Timeline. Additionally, only 7 insights of value five were generated and all occurred with the Timeline.

Time to first insight

The average time at which the participants formulated their first insight was analyzed. Figure 4 shows the distribution and Table 5 shows the calculated values with their statistical significance. For insights of any given value, the Timeline was slower (13.16 s) compared to the baseline (7 s; $p < 0.01$). Insights with a value greater than one and two were also slower with the Timeline (22.24 s and 51.21 s) compared to the baseline (20.36 s; $p = 0.14$ and 40.01 s; $p < 0.01$). For insights with a value greater than 3, the Timeline was faster (68.44 s) compared to the baseline (92.83 s; $p < 0.01$). Only the Timeline was able to generate insights of value greater than 4 with an average time of 63.50 s.

Discussion

The insight-based methodology applied to this case study, provided an approach to quantify the degree to which a visualization tool assisted the process of understanding longitudinal clinical data. The design of the study enabled us to investigate the usefulness of a visualization in a real-world environment with medical doctors as participants of the study (domain experts and the intended audience for the visualization software).

The high number of low value insights from the baseline visualization could be partly explained by difficulty in putting together smaller units of information contained in each row of the tabular format. This, in turn made it difficult to derive deeper and more complex insights leading

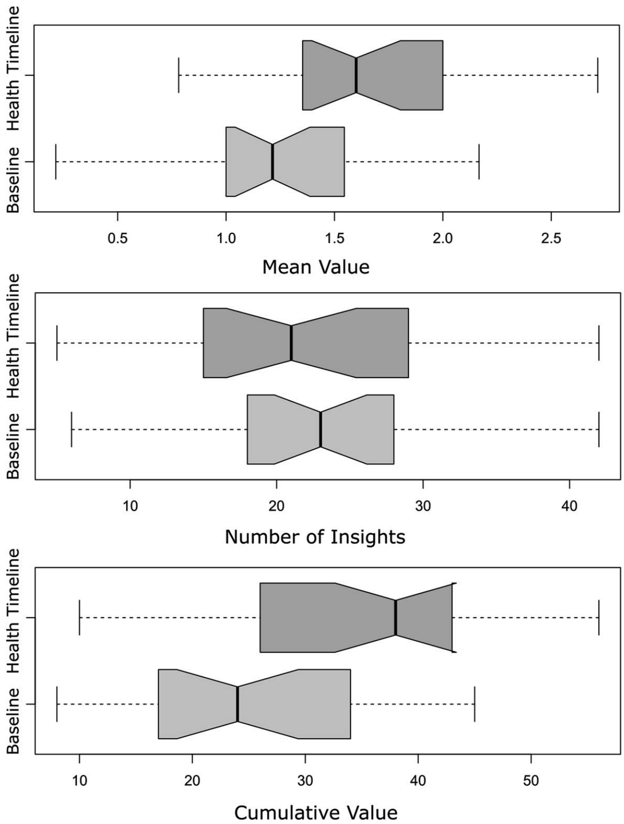


Fig. 3 Metrics per assessment. The box plots represent the number of insights, mean and cumulative value of the insights per assessment

to a hypothesis. In contrast, the Health Timeline assisted participants in generating higher value insights at the cost of a longer time delay. Health Timeline could have encouraged the participants to take a more "deliberate" approach when making decisions about the data, which was in turn reflected in the insights derived during the assessments.

Possible generalization of the results

It can be said that in the context of longitudinal clinical data, a time-based visualization of chronological events assists its audience better than textual information. In this context, the data visualization assisted the participants in gaining a greater understanding of the data (complete and accurate). In some cases, clinicians were able to understand the clinical history of the patients, formulate a diagnosis and suggest treatments.

This study shows the need to use a structured assessment methodology in the context of healthcare to determine the extent to which a visualization can assist clinicians to understand data. Without an objective assessment, it becomes subjective to state with confidence that one visualization is "useful" or "better". Lam and colleagues [25], as well as Bertini and researchers [20] also emphasize the importance of structured assessment methods to evaluate data visualizations. This study also serves as a documented example of a contextualized assessment of visualizations in order to provide a use-case evaluation that is based on a real-life scenario. Even though the results of this study are encouraging, a larger study is warranted to examine objective outcomes and the impact that clinical data visualization may have in patient outcomes over longer periods of time.

Limitations and considerations

To facilitate participation in our study, we applied the insight-based methodology with the "thinking aloud

Table 2 Metrics per Assessment

Metric	Baseline	Health Timeline	p-values
Number of insights	$min = 6$	$min = 5$	0.70
	$max = 42$	$max = 42$	
	$\mu = 23.04$	$\mu = 22.32$	
	$Md = 23$	$Md = 21$	
	$\sigma = \pm 10.43$	$\sigma = \pm 9.12$	
Cumultive value	$min = 8$	$min = 10$	0.01
	$max = 45$	$max = 56$	
	$\mu = 24.96$	$\mu = 34.68$	
	$Md = 24$	$Md = 38$	
	$\sigma = \pm 10.96$	$\sigma = \pm 11.87$	
Mean value	$min = 0.21$	$min = 0.78$	0.01
	$max = 2.17$	$max = 2.71$	
	$\mu = 1.26$	$\mu = 1.70$	
	$Md = 1.21$	$Md = 1.60$	
	$\sigma = \pm 0.52$	$\sigma = \pm 0.57$	

The table shows the insights generated by participants using Health Timeline and the visualization baseline per assessment. Statistical significance is shown in the p column using Mann-Whitney U tests

process". All the audio recordings were transcribed for evaluation. The sample size and the clinical data used in this study required a large number of working hours for the transcription and evaluation of the insights. Increasing sample size or participants would improve representativeness and statistical reliability of the results but at the cost of a greater number of hours, which was outside our resources, unfortunately. Therefore, the generalizability of this study may be limited, and further studies would be desirable to confirm our findings.

Another limitation of the study was the time available to conduct the assessments. Even with the relatively brief time window of 30 min, we failed to recruit more subjects. This could probably be explained by the busy schedules of the psychiatrists.

The evaluation criteria, even though revised and peer-reviewed by domain experts, could also be subject to bias. A larger group of domain experts could provide more objective criteria. It is possible that experts that were not acquainted with the data might have provided better evaluation criteria, as these experts would have a fresh look at the clinical data without preconceived notions.

Table 3 Insight Distribution

Metric by Insight Value	Baseline	Health Timeline	Ratio	p-values
Any	576	558	1.16	0.70
1	175	114	0.65	0.03
2	98	178	1.82	0.01
3	71	78	1.10	0.53
4	10	32	3.20	0.01
5	0	7	—	Non-significant

The table shows the distribution of the insights according to their domain value. The ratio column shows the observations of the Health Timeline divided by the baseline. The p-values are obtained using Wilcoxon Signed-Rank Test

Table 4 The table compares the distributions showing the minimum, maximum, median, average and standard deviation of the two populations, in this case these are the Health Timeline and the baseline representation

Metric	Baseline	Health Timeline	p-values
Insight Value	$min = 0$	$min = 0$	< 0.01
	$max = 4$	$max = 5$	
	$\mu = 1.0833$	$\mu = 1.55$	
	$Md = 1$	$Md = 2$	
	$\sigma = \pm 1.10$	$\sigma = \pm 1.25$	

The statistical tests were conducted with Chi-squared since the populations can be treated as categorical data (value 1 to 5 each comprise their own category)

Given the results obtained in this study, we propose a further study to test if the visualization can assist a larger group of experts in providing higher quality of care to patients. Such a study would require a larger number of participants, a larger number of patients with the required ethical approval and a randomized controlled trial.

Clinical data in a real healthcare context

The clinical data used in this study did not include diagnoses and notes taken by the practitioners during previous consultations. This was deliberate to allow the participants to come up with their own conclusions and assessments.

In a real life scenario, practitioners would have access to the diagnoses, notes and observations made by other clinicians. For the purpose of the study, we did not want to include this information to test the effectiveness of the visualization. The insights formulated in the assessments were our measurement to let us know the degree to which the participants were able to understand the clinical data without any previous information. We acknowledge that this is not an entirely realistic scenario however, for the purpose of the experiment, it allowed us to understand the usefulness of the visualization.

The clinical data was comparable to the reality of the healthcare context because it was extracted directly from real patients. In terms of complexity, we selected patients with a high degree of complexity of their treatment, in-patient events and diagnosis to base the study on complex real-life scenarios.

Conclusion

This study detailed the assessment of a time-based EHR visualization software by applying the insight-based

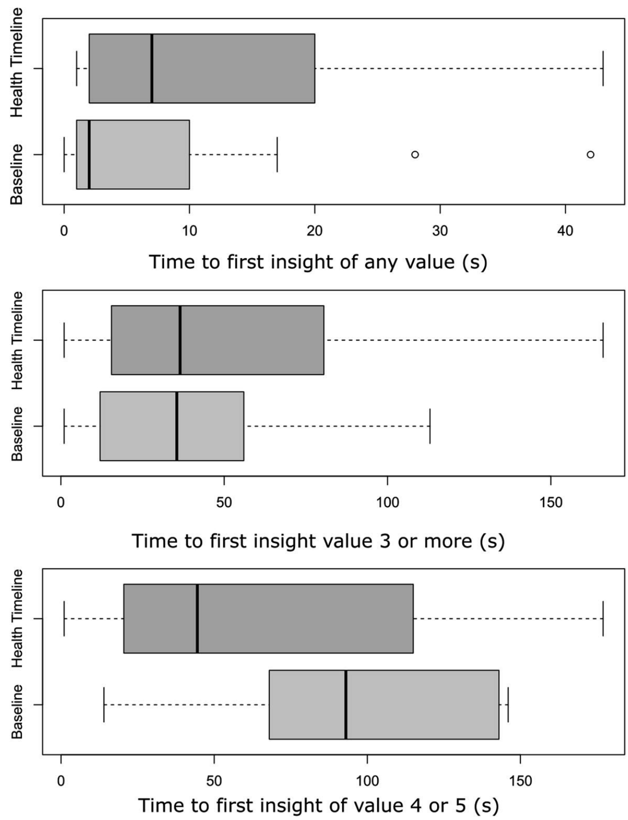


Fig. 4 Time to first insight. The box plots represent the time to first insight of any value, value 3 or higher and value 4 or higher

Table 5 Time to First Insight

Insight Value	Baseline	Health Timeline	Ratio	*p*-values
Any	$\mu = 7$	$\mu = 13.16$	1.88	< 0.01
	$\sigma = \pm 10.02$	$\sigma = \pm 14.24$		
> 1	$\mu = 20.36$	$\mu = 22.24$	1.09	0.17
	$\sigma = \pm 16.66$	$\sigma = \pm 18.48$		
> 2	$\mu = 40.01$	$\mu = 51.21$	1.28	< 0.01
	$\sigma = \pm 29.96$	$\sigma = \pm 45.98$		
> 3	$\mu = 92.83$	$\mu = 68.44$	0.74	< 0.01
	$\sigma = \pm 51.27$	$\sigma = \pm 59.88$		
> 4	$\mu = 0$	$\mu = 63.50$	—	< 0.01
	$\sigma = 0$	$\sigma = \pm 60.06$		

The table shows the mean and standard deviation of the time to first insight from any value to values 1 to 5. The ratio column shows the observations of the Health Timeline divided by the baseline. The *p*-values are obtained using Wilcoxon Signed-Rank Test

methodology. The results show that the Health Timeline data visualization tool supported the generation of valuable insights following the proposed criteria. Furthermore, the use of the assessment method demonstrates the feasibility of applying a consistent methodology to assess visualization techniques and tools in the clinical context. We propose that the insight-based methodology could be used in future studies as a methodological approach to assess the value of a clinical data visualizations.

Additional file

Additional file 1: A comma-separated values data file containing the transcribed insights, with the annotations used for the study. The file is in a standard format and is machine-readable. The annotations of the file are detailed in the "Data analysis" section of this article. (CSV 24 kb)

Abbreviations

EHR: Electronic health record; MBS: Medicare Benefits Scheme; MHR: Medical health records; MWW: Mann–Whitney–Wilcoxon; NEHTA: National E-Health Transition Authority; PBS: Pharmaceutical Benefits Scheme

Acknowledgements
This research work was supported by VTT Technical Research Centre of Finland and the EU program for Research and Innovation Horizon 2020.

Authors' contributions
AL conducted the data analysis including the transcription of the audio into the machine-readable file. AL applied the insight-based methodology to obtain the domain value of the insights. ME, NB, JS, GS and AL designed and agreed on the criteria for the insight domain value. ME, NB and AL conducted the statistical tests. HN and IK provided counseling and feedback on the study methodology and the article writing. NB developed the Health Timeline. JS, GS and NB designed the Health Timeline. ME, NB, JS, GS and AL designed the experiment protocol. JS, GS and NB facilitated the de-identified clinical data. All authors have read and approved the manuscript.

Funding
VTT Technical Research Centre of Finland has funded the design of the study, collection, analysis and interpretation of data. The writing of the manuscript was funded by the EU program for Research and Innovation Horizon 2020 under the Grant Agreement number 689260.

Availability of data and materials
All data generated or analyzed during this study are included in this published article and its supplementary information files.

Ethics approval and consent to participate
Ethical approval was obtained from the South Australian Mental Health Observatory (MHO). Each patient gave a written informed consent to use their data for research purposes. MHO is currently undertaking a study, the "SPARK Study - A project of Mental Health Observatory". The de-identified patient data used in the study was obtained with the informed consent of the patients through the SPARK study [27]. Consent for publication was also obtained from the South Australian MHO under the SPARK study. Additionally, the data published in this study is comprised of clinical observations, or insights, made by psychiatrists as a result of examining de-identified patient information. These insights are also approved for publication by South Australian MHO.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Tampere University, Faculty of Medicine and Health Technology, Tampere, Finland. [2]College of Medicine & Public Health, Flinders University, Medical Sciences Rd, Adelaide, Australia. [3]Discipline of Psychiatry, The Queen Elizabeth Hospital, South Australia, University of Adelaide, 28 Woodville Rd, Woodville South, Adelaide, Australia. [4]Country Health Local Health Network, South Australia, and Flinders University, Medical Sciences Rd, Adelaide, Australia. [5]VTT Technical Research Center Ltd, Tekniikankatu 1, Tampere, Finland.

Received: 16 July 2018 Accepted: 31 July 2019
Published online: 22 August 2019

References

1. Shickel B, Tighe PJ, Bihorac A, Rashidi P. Deep ehr: A survey of recent advances in deep learning techniques for electronic health record (ehr) analysis. IEEE J Biomed Health Inform. 2017;PP(99):1. https://doi.org/10.1109/JBHI.2017.2767063.
2. Corbin K. How to Navigate Big Data in Healthcare. 2015. https://www.cio.com/article/2851986/healthcare/how-to-navigate-big-data-in-healthcare.html. Accessed 25 Dec 2018.
3. Health Science Library UniversityofWashington. Data Resources in the Health Sciences. Updated: 2018-11-20. http://guides.lib.uw.edu/c.php?g=99209&p=642709#12207216. Accessed 25 Dec 2018.
4. Procter R. Definition of health informatics. 2009. https://www.nlm.nih.gov/hsrinfo/informatics.html. Accessed 25 Dec 2018.
5. Lesselroth B, Pieczkiewicz D. Data Visualization Strategies for the Electronic Health Record. Hauppauge, NY: Nova Science Publisher's, Incorporated; 2011.
6. Rind A, Wang TD, Aigner W, Miksch S, Wongsuphasawat K, Plaisant C, Shneiderman B, et al. Interactive information visualization to explore and query electronic health records. Found Trends® Hum Comput Inter. 2013;5(3):207–98.
7. North C. Toward measuring visualization insight. IEEE Comput Graph Appl. 2006;26(3):6–9.
8. Saraiya P, North C, Lam V, Duca KA. An insight-based longitudinal study of visual analytics. IEEE Trans Vis Comput Graph. 2006;12(6):1511–22.
9. Ledesma A, Nieminen H, Valve P, Ermes M, Jimison H, Pavel M. The shape of health: A comparison of five alternative ways of visualizing personal health and wellbeing. In: Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE. IEEE; 2015. p. 7638–41.
10. Cleveland WS, McGill R. Graphical perception: Theory, experimentation, and application to the development of graphical methods. J Am Stat Assoc. 1984;79(387):531–54.
11. Goetz T, Ted.com. It's Time to Redesign Medical Data. Created: 2011-01-25. http://www.ted.com/talks/thomas_goetz_it_s_time_to_redesign_medical_data?language=en. Accessed 25 Dec 2018.
12. Commonwealth of Australia ABN: 83 605 426 759. My Health Record. https://myhealthrecord.gov.au/internet/mhr/publishing.nsf/content/home. Accessed 25 Dec 2018.
13. Plaisant C, Milash B, Rose A, Widoff S, Shneiderman B. Lifelines: visualizing personal histories. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM; 1996. p. 221–7.
14. Plaisant C, Mushlin R, Snyder A, Li J, Heller D, Shneiderman B. Lifelines: using visualization to enhance navigation and analysis of patient records. In: Proceedings of the AMIA Symposium. American Medical Informatics Association; 1998. p. 76.
15. Alonso DL, Rose A, Plaisant C, Norman KL. Viewing personal history records: A comparison of tabular format and graphical presentation using lifelines. Behav Inf Technol. 1998;17(5):249–62.
16. Wang TD, Plaisant C, Shneiderman B, Spring N, Roseman D, Marchand G, Mukherjee V, Smith M. Temporal summaries: Supporting temporal categorical searching, aggregation and comparison. IEEE Trans Vis Comput Graph. 2009;15(6):1049–56.
17. Bui AA, Aberle DR, Kangarloo H. Timeline: visualizing integrated patient records. IEEE Trans Inf Technol Biomed. 2007;11(4):462–73.
18. Wongsuphasawat K, Guerra Gómez JA, Plaisant C, Wang TD, Taieb-Maimon M, Shneiderman B. Lifeflow: visualizing an overview of event sequences. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM; 2011. p. 1747–56.
19. Meyer TE, Monroe M, Plaisant C, Lan R, Wongsuphasawat K, Coster TS, Gold S, Millstein J, Shneiderman B. Visualizing patterns of drug prescriptions with eventflow: a pilot study of asthma medications in the military health system. Tech Rep. 2013. OFFICE OF THE SURGEON GENERAL (ARMY) FALLS CHURCH VA.
20. Bertini E, Perer A, Plaisant C, Santucci G. Beliv'08: Beyond time and errors: Novel evaluation methods for information visualization. In: CHI '08 Extended Abstracts on Human Factors in Computing Systems. CHI EA '08. New York: ACM; 2008. p. 3913–6. https://doi.org/10.1145/1358628.1358955. http://doi.acm.org/10.1145/1358628.1358955.
21. Saraiya P, North C, Duca K. An evaluation of microarray visualization tools for biological insight. In: Information Visualization, 2004. INFOVIS 2004. IEEE Symposium On. IEEE; 2004. p. 1–8.
22. Saraiya P, North C, Duca K. An insight-based methodology for evaluating bioinformatics visualizations. IEEE Trans Vis Comput Graph. 2005;11(4): 443–56.
23. Plaisant C, Fekete J-D, Grinstein G. Promoting insight-based evaluation of visualizations: From contest to benchmark repository. IEEE Trans Vis Comput Graph. 2008;14(1):120–34.
24. Ledesma A, Al-Musawi M, Nieminen H. Health figures: an open source javascript library for health data visualization. BMC Med Inf Decis Mak. 2016;16(1):38.
25. Lam H, Bertini E, Isenberg P, Plaisant C, Carpendale S. Empirical studies in information visualization: Seven scenarios. IEEE Trans Vis Comput Graph. 2012;18(9):1520–36.
26. Nielsen J. Usability Engineering. New York: Elsevier; 1994.
27. Bidargaddi N, Schrader G, Piantadosi C, Adams R, Ryan B, Williams J, Strobel J. Embedding research into clinical practice: The services, pathways access–research and knowledge (spark) study. Aust J Rural Health. 2013;21(3):192–3.

28. Mann HB, Whitney DR. On a test of whether one of two random variables is stochastically larger than the other. Ann Math Stat. 1947;18(1):50–60.
29. Nayak BK, Hazra A. How to choose the right statistical test? Indian J Ophthalmol. 2011;59(2):85.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

BMC