Author’s response to reviews

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

Authors:

Andres Ledesma (andres.ledesma@tut.fi)
Niranjan Bidargaddi (niranjan.bidargaddi@flinders.edu.au)
Jörg Strobel (jorg.strobel@sa.gov.au)
Geoffrey Schrader (geoffrey.schrader@adelaide.edu.au)
Hannu Nieminen (hannu.o.nieminen@tut.fi)
Ilkka Korhonen (ilkka.korhonen@tut.fi)
Miikka Ermes (miikka.ermes@vtt.fi)

Version: 1 Date: 29 Dec 2018

Author’s response to reviews:

Dirk Krüger, PhD

Editor-in-Chief
BioMed Central
BMC Medical Informatics and Decision Making
ISSN: 1472-6947
The Campus, 4 Crinan Street
London N1 9XW
United Kingdom

Dear Dr. Krüger,

Please find enclosed our manuscript with the corrections suggested by the reviewers. We deeply appreciate the remarks and criticism since they increased the quality and clarity of the manuscript. This letter provides a point-by-point list of the remarks and observations suggested by the reviewers and a reference to the relevant sections of the manuscript.

Please refer also to the file named “Difference PDF with Highlighted Changes – diff.pdf” for a comparison with the previous submission. This file contains all the highlighted changes made to the manuscript since the last submission.
The comments and their response are now listed in the rest of this letter.

Editor Comments:

Thank you for your submission. Both reviewers agreed that this paper presented an interesting application of the insight methodology to a health data timeline display, providing a potentially interesting and generalizable methodology. However, there were multiple concerns regarding the methodology and validity of the paper. Perhaps most importantly, concerns about the rigor of the counting of insights were raised - without measures designed to assess the reliability of the insights measured, it's hard to know how to interpret the data. There were also concerns regarding the rationale for the proposed design, a lack of appropriate contextualization in prior literature, methods of data collection, appropriateness of the data used, the theoretical basis for the hypotheses, and the description of the methods.

We thank the Editor for these comments and guidelines. We have revised the manuscript as described below to address these concerns and we think that after these revisions the manuscript has improved and the key concerns have been satisfied.

Thomas Reese (Reviewer 1):

1. Goal 1 states, "By applying the methodology, we can determine the degree to which the visualization assists clinicians to understand the clinical data." From the perspective of applying the Insight methodology to a realistic novel use case, there are limitations to the performed analysis.

The limitations of the study are now detailed in the manuscript on page 13. The key limitations are the size of the sample and the criteria used to classify the insights (value coding). A section has been added to explain the recruitment process (sections “Sample Size” and “Participants” on page 11) and the rationale for the design of the study (sections “Time Constraints” and “Visualization Presentation Order” on page 10). We have also explained how the criteria was formulated to assess the insights with the expertise of two subject matter experts (section “Agreed Criteria and Value Coding” on page 8).

* It is suspect that the primary significant finding is, perhaps, the most susceptible to bias (insight score). The manuscript lacks information on how coding objectivity was maintained. Suggest adding information on insight coding (e.g., independently conducted, validated by subject matter experts).

This is an excellent comment and a justified concern. We targeted to maintain objectivity by inviting two subject matter experts to assess the sample of the insights and to revise the criteria used in the evaluation process. The two experts were professional psychiatrists that have previously worked with the clinical data used in the. We believe that hence the results are generalizable but admit that this is a shortcoming in the present study. We have added detailed
Suggest adding information on how insight classification relates to clinical impact. For example, one of the highest rated insights states, "and developed comorbid diabetes." Why does this statement reflect clinician understanding and an impactful insight?

We agree with this observation, the clinical impact is not evident in the insight classification. In page 8, we added a section “Agreed Criteria and Value Coding. This section describes the relationship between insight classifications (value 1 to 5) and the clinical impact. We detailed the criteria, reviewed by subject matter experts, but we also justify that a high value insight demonstrates a deep and comprehensive understanding of the data. With this understanding, the clinician can formulate an accurate diagnosis which in turn allows for a better decision-making process for treatment and prognosis.

The following example has now been used in the manuscript (p8, Agreed Criteria and Value Coding, line 28): “the patient has experienced depression and anxiety, that would explain the prescription and regular use of the drug treatment and also the 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 has improved towards the summer and it seems that changing the drug treatment improved the outlook”.

* The authors mention a couple well-known timeline visualizations (e.g., LifeLines, LifeFlow); however, it is difficult to understand how Health Timeline is unique or amenable to this type of evaluation. Suggest adding design rationale for Health Timeline features. Suggest adding information on how Health Timeline builds from these other displays (e.g., LifeLines).

The manuscript now details the rationale for the design of Health Timeline (page 6, Design Rationale and Requirements). It can be said that Health Timeline does build up from these other visualizations in that it simplifies the user interface, uses a time-based visualization and allows some interactivity (zoom and pan), and the key contribution comes from the assessment of this visualization in a real-life scenario. Hence, we do not contend that Health Timeline is as such unique. Bui and colleagues developed Timeline which contains multiple features for screening and sorting the data, as well as a causal model. By contrast, Health Timeline was designed and built to be used with a simple user interface without complex features. This was a key requirement from Australia's "My Health Record" initiative as well as from interviews with clinicians and Information Technology experts with a background in data visualization. The goal of the study was to conduct the assessments without any introduction to Health Timeline to test out how clinicians would find their way around the visualization and formulate the insights by looking at the data.

Suggest adding justification on how timeline features and insights gained are analogous to those observed in the Bioinformatics context.

We consider this an important and valid observation. Page 4 of the manuscript has a new section named “From Bioinformatics to Clinical Data”. This section explains why the insights in both
studies are analogous. Mainly, subject matter experts reviewed the criteria used for evaluation, thus the validity of an insight was verified by experts who were acquainted with the data used in the experiments.

2. Goal 2 states, "To document the process of applying the methodology in the context of clinical data and assess the effectiveness of the visualization in the decision-making process." From the perspective of analyzing and documenting the impact of the visualization on clinical decision-making, the methods have limitations.

We agree with the reviewer. Additional information has been added to the manuscript to detail the limitations of the study on page 13 under the Discussion section. The main limitations are the scalability of the study (repeat the experiment with more participants or with a longer time for the assessments) and the assessment criteria used to classify the insights. We tried to minimize the bias in the assessment criteria by inviting subject matter experts.

* The title states "Clinical Data," but on Page 6, Lines 18-28, it appears the data source administrative (i.e., Pharmaceutical Benefit claims and Medicare Benefits Scheme claims). Do psychiatrists routinely sort through administrative data for insights? How does this mimic a "real-world" environment?

This is an important observation and to address this comment we explained further the data used in the study. The clinical data used in the study is not exactly the same as the clinical data used in real life because we removed the clinical diagnoses and annotations made by clinicians since the purpose of the study is to allow the participants to reach their own conclusions. The Medical Benefit and Pharmaceutical Benefit data is not limited to medications, in this Australian initiative, the data comprises he visits to General Practitioners, specialists, laboratory tests, emergency hospital admissions, drug treatment, among other categories. This information is now detailed in sub section named “Clinical Data Included in the Study” on page 9.

* How do the insights impact decision-making? Suggest adding information on how the task context is clinically relevant. Suggest adding limitations to the Discussion section.

We agree with the reviewer, a section detailing the limitations of the study has been added to the manuscript on page 13. To address the question on the impact of the insights in the decision-making process, we have now detailed the criteria used to classify the insights. The criteria is described in the section “Agreed Criteria and Value Coding” on page 8. The criteria was revised by subject matter experts and it validates the insights formulated by the participants (the insights are valid and comprise a total or partial understanding of the data). The purpose is to take informed decisions regarding the patient care with a comprehensive understanding of the clinical data. High value insights demonstrate a greater understanding of the data and therefore the clinician is better prepared to take an informed decision.

* The authors and other researchers have mentioned a key aspect of visualizations is interaction with the display to gain insights. Page 6, Lines 44 and 45 state, "The psychiatrist's observations
The assessment sessions were recorded on an answering machine over a phone connection. How did this procedure document visualization display interaction?

The insight-based methodology suggests the use of the “thinking aloud” process. We explained the process to the participants in the experiment protocol. The process consists of narrating the actions “aloud” when the participants use the visualization software. For example, when the participant is panning or zooming, the participant will then say aloud “I am panning to the year 2013” or “I am zooming in to get more information about the prescribed treatment”. We have added an explanation about the “thinking aloud” process to the manuscript on page 9.

* Rationale for selecting patient cases is lacking. Where the patients selected for profile complexity or representativeness?

The patient selection was based on the complexity of their diagnoses. The five cases were selected because they had mental disorders and had received continuous care throughout their treatment. This explanation has now been added to the manuscript on page 9 in the section named “Patient Data Selection”.

Where the same patient cases used for both displays?

All five patients were presented with both visualizations to account for control in the representation of the clinical data. This clarification has been added to the manuscript on page 10, section named “Visualization Presentation and Order” under the Experiment Protocol.

How was sample size determined?

Based on the availability of staff at the Universities and Institutions in rural South-Australia, a total of 10 psychiatrists were invited to take part in the study. Five of them accepted the invitation. However, this is a small sample size even though each participant conducted 10 assessments. The busy schedules of the clinicians was perhaps the main limitation in recruiting a larger sample size. We have added this information to the manuscript on page 11 section named “Sample Size”.

Was the display presentation order controlled?

The display order was controlled. The experiment started first with the Timeline then the tabular data and after that it would alternate between the two. The order of the clinical data was also controlled to prevent consecutive assessments with the same data. This information has been added to the manuscript on section “Visualization Presentation and Order” page 10.

Wolfgang Aigner (Reviewer 2):
In the section „Key contribution" it remains somewhat unclear, whether the focus is put on the visualization or on the evaluation methodology. E.g., in related work prior to the section this is mixed.

We agree with the observation and decided to clarify further the focus of the article. The main focus is on the assessment of visualization tools in the clinical data domain. For that purpose, the manuscript details the process of testing out the insight-based methodology for the assessment of a time-based clinical data visualization. This has been clarified in “Key contributions” on page 3.

In case a focus is put on evaluation methodologies for visualization in this domain, the authors should take a look at works of Bertini et al. for example (7 scenarios paper and follow ups) as well as work featured in the BELiV workshop series.

We thank the referee for this valuable comment and references. The focus of our study is narrowed down to visualizations of clinical data that have been assessed in the Healthcare context. The literature review revealed that similar visualization tools have been introduced in the Healthcare context but lack a formal evaluation, and none of the examples found follows the insight-based methodology or other structured assessment methodologies. We have complimented the manuscript with the BELiV series and the information they provide regarding the assessment scenarios in the section named “Assessment Methods” on page 3. We found that the scenarios defined by Lam and researchers, the “Evaluating Visual Data Analysis and Reasoning (VDAR)” comprises, among other techniques, the insight-based methodology. We have added this information to the manuscript to complement the context of the study on the section named “Evaluating Visual Data Analysis and Reasoning” on page 4, and also to provide a defined scope in the related work section.

- Related Work

The related work section seems to be very selective or even biased as it selects a very small sample of related systems but does not explain on which basis this selection has been made. E.g., in the cited article [6] a lot more systems are presented and it is not clear why other systems are not included.

We thank the referee for this note. The manuscript was indeed missing the rationale for the selection of the works included in the review. The selection criteria and scope definition have been added to the section named “Scope” on page 2 under the Related Work chapter. We narrowed down the scope to data visualization studies made in the context of clinical data (EHR) with any documented evaluation.

The related work section ends very abruptly and seems to be incomplete. Moreover, a summary of the findings from related work is not presented.

A summary of the findings is now added at the end of the Related Work chapter under the section named “Summary” page 3. The key point is to illustrate that the existing body of research
does not follow a well-defined methodology to assess the degree to which a visualization assists an audience to better understand the data in a clinical context (Healthcare).

- Visualization design

The visualization design itself is rather on the simple side and seems not to provide many of the features other systems from the state of the art do have.

We agree with this observation. Additional information has now been added to the manuscript to explain the motivation, design and reasoning behind Health Timeline.

The visualization was designed to be a simple web application to visualize longitudinal clinical data. The simplicity of the design was a key requirement from the target users of the system. The authors have included similar visualization tools that supersede Health Timeline in terms of features and complexity. However, in this study, the participants did not receive an introduction or explanation on how the Health Timeline software works. This particular reason was a motivation to keep the design simple and enable just a minimal set of features, namely zooming and panning.

The description of the visualization design is rather minimalistic and does not include any design rationales. Moreover, user requirements should be laid out systematically.

The requirements have now been added to the manuscript as well as the design rationale in the section named “Design Rationale and Requirements” on page 6. Health Timeline aimed at providing a web-based (usable with any devices with a web browser) visualization tool for clinicians to assist in the process of understanding the overall condition of a patient. 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.

- Study

What I miss in the study is that hypotheses are posed before the experiment is done and that they are based on theoretic considerations.

The authors agree with this observation. The following hypothesis has been proposed:

“A time-based visualization enable clinicians to obtain valuable insights about the clinical condition of the patients. To determine what a valuable insights comprises, we invited subject matter experts to come up with an evaluation criteria, described in this study. Valuable insights would thus translate to complete and correct understanding of the clinical data, and therefore the patient diagnosis and suggested treatment”.
- Discussion & Conclusion

Unfortunately, the findings of the study are not generalized. It is not made clear, what the broader implications of these findings are.

This is a valid observation. We have added to the manuscript a section to explain the implications of the findings and the possible generalizations in the section “Possible Generalization of the Results” on page 13.

For instance, a possible generalization of the findings of the study is that in the context of longitudinal clinical data, a time-based visualization of chronological events is better than textual information. In this context, a “better” visualization means that the data can be understood to a greater extent. Thus, the clinicians will have deeper understanding of the clinical history, diagnosis and treatment of the patients with the assistance of a time-based visualization.

Additionally, it is possible to generalize that to determine the extent to which a visualization can assist its audience to better understand the data, a formal study needs to be thoroughly conducted. Without an objective assessment it becomes subjective to state with confidence that one visualization is “useful” or “better” than another. This generalization confirms the observations made by Lam and Bertini.

Limitations of the study and setup should be discussed.

We have added to the manuscript a section detailing the limitations of the study in section “Limitations and Considerations” on page 13.

The number of participants in the experiment is one important limitation. Another limitation was the time used for the experiment.

In this study, we used the thinking aloud process to facilitate the participation in the experiment. However, this method is time-consuming since it requires a manual transcription of the phone call recordings. If we were to conduct a similar study with more participants, the scalability is serious limitation.

Two subject matter experts took part in the definition of the criteria used to assess the value of the insight. However, there is no guarantee that the criteria is completely free from any bias. The experts provided their own knowledge and judgment as subject matter experts but it is possible that other experts might disagree with the established criteria.

To what extent, the external validity of the study is given, should be mentioned. I.e., how realistic is the task and dataset for the user group in their daily work?

We agree with the reviewers and consider this a valid observation. It was not clear in the manuscript that the study used real life patient data and that it is meant to simulate a real-life scenario. In many cases, clinicians are expected to get acquainted with the clinical history of the patients before the consultation. However, the clinical data used in this study did not include
diagnoses and annotations from previous visits to the clinicians. The reason was to encourage the participants to come up with their own conclusions and then validate their findings with the criteria established with the subject matter experts. This information has now been added to the manuscript on section “Clinical Data in a Real Healthcare Context”, page 14.

What remains unclear, is the comparability of the selected data. Are the selected patient records on the same level from a point of view of their complexity, number of data elements, etc.?

The selected data is comparable to the reality of the healthcare context because the data for this study was extracted directly from real patients. In terms of complexity, we have added a section to the manuscript that explains the selection criteria used. This information can be found on the section named “Patient Data Selection”, page 9.

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 a substantial amount of changes in the overall mental health (schizophrenia, depression and mood disorder).

What was not exactly clear to me is whether each subject used both, visualization and table and whether the study design was counter-balanced or how the order of representation was selected otherwise?

The article now clarifies this information and the rationale behind the design of the study on a section named “Visualization Presentation and Order” on page 10.

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 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.

Moreover, I wonder whether the order of datasets was the same for all participants?

The order of the assessments was the same for all the participants. This information has now been added and explained further in the manuscript in the aforementioned section on page 10.

- Typos & References:
Typo in Background section: Goetze —> Goetz

Typo in Experiment Protocol: „give“ —> „five“

Reference Lesselroth and Pieczkiewicz (2011) is incomplete (contains „???”).

URL references are incomplete (e.g., creation/access dates missing)

These corrections have now been added to the manuscript.

All authors have approved the manuscript and agree with its submission to BMC Medical Informatics and Decision Making.

Please address all correspondence to: Andres Ledesma (Tampere University of Technology) at andres.ledesma@tut.fi and tel. +358 50 30 14 237.

We look forward to hearing from you at your earliest convenience.

Andres Ledesma, MSc

Corresponding Author
andres.ledesma@tut.fi
Tel. +358 50 30 14 237
Project Researcher
Tampere University of Technology
BioMediTech Institute
Faculty of Biomedical Sciences and Engineering
Tampere University of Technology
Korkeakoulunkatu 10
Tampere 33720
Finland