**Author’s response to reviews**

**Title:** A proof-of-concept framework for the preference elicitation and evaluation of health informatics technologies: the online PRESENT patient experience dashboard as a case example

**Authors:**

Emmanouil Mentzakis (E.Mentzakis@soton.ac.uk)

Daria Tkacz (darmal87@googlemail.com)

Carol Rivas (c.rivas@ucl.ac.uk)

**Version:** 1  **Date:** 31 Jan 2019

**Author’s response to reviews:**

Dear Editors and reviewers

Thank you for your helpful comments and the opportunity to revise the paper. We have responded to each comment and summarised responses below.

Yours faithfully

The Authors

Arianna Dagliati (Reviewer 1): The manuscript deal with a very important issue in decision system, i.e. user evaluation of dashboard in healthcare settings on the basis of specific needs. This underline the importance of delivering "desirable" tools in this context. However, I have some doubtful concerns about the paper.

First of all, while the experiment is applied in the context of informatics and decision making, the applied approach and methods are more related to the health economics context. While the two fields are related, authors should better emphasise the aspect of their research that fits the scope of the journal. I recommend the author to rewrite the work carefully in order to clarify specific research questions and outcome through each manuscript section.

1. In the background section 1. I would recommend to show literature evidence of: dashboard-based system in clinical context, DCEs application in healthcare

   In page 4, we add specific mentions on the DCE application in health and evaluation of health care technologies with relevant references.

2. Please check reference [6] to PRESENT dashboard description [6], while is not clear is possible to get an exhaustive description of the project from the provided reference.

   NIHR policy is to delay putting the report on its website until papers are published. Hence this report will be released after this paper is published. To get around this problem I have put the NIHR reports page link on the reference, so if someone chooses to look after this paper is published it will be clear to them that all they need do is type in the report authors or title. This report will have extensive and complete detail on the project.
3. In section 2.1 I would describe the research team for the dashboard design more thoroughly, have you used focus group? In that case please describe the participants in term of relevant attributes (like type of profession, level of expertise etc.)

A table has been added and the text slightly amended.

4. Page 5 line 24/25 "healthcare professional who fitted the profile" this sentence is vague, please define the profile characteristics.

We have now said healthcare managers and decision-makers. It is hard to be more precise because their specific roles only have local meaning, for example local CCG (clinical commissioning group board member), and each person had a different role. Moreover, our aim was to choose a wide range. As this was piloting, more detail should not be necessary as research shows that whatever their identity, if they are in the right education/literacy range, they should be able to pick up any issues. We have amended the wording slightly to make their purpose clearer.

5. I've found section 2.2 very difficult to understand. Some bit are very generic (e.g. "statistical theory" per se is a very vague definition), I think the entire paragraph would greatly benefit and improve its clarity through a graphical illustration of the design setting.

We have removed the phrase in question and have reworded and added text to streamline the section and make it easier to follow. We also offer references for anyone interested in choice experimental designs methods and properties. As experimental design is a large research-active statistical field we opt not to delve into the actual procedures of experimental design but let the interested reader follow through in the references.

6. JMP is a visualization SW for data analysis, how did you employ it in this context? Can you provide some examples?

The software used is JMP Pro 13 (of the SAS family) which on top of its analytics and visual tools has an extensive experimental designs (choice and other) suite. For more see https://www.jmp.com/en_dk/applications/design-of-experiments.html

7. Section 2.4 has to be improved in term of clarity. First of all I think you should very clearly define the analysis outcomes (multiple outcomes). In this context I would expected the use of logit to estimate participants as function of observations, and the use the predicted value to compute (for example) propensity scores, but while you're basically explaining logit and nested models, so I would suggest to reformulate this section, highlighting only the innovative contribution of your research in term of methods and providing relevant references for DCE and RUT.

The reviewer’s suggestion in shortening this section and focusing on providing references is intuitive and improves the flow. Nevertheless, the models for discrete choice are rather different in the formulation from standard logits. The outcome of analysis is again binary but the unit of analysis is the choice set and not the individual. In fact, for the base model, individual characteristics cannot explain the choice between dashboard A and dashboard B. Further propensity scores have little intuitive interpretation and as such we opt for predicted probabilities and WTP values that are also of interest.

All of this is now explicit in the revised sections 2.4 and 2.4.1.

8. I think you should discuss the fact that your model don't deal with confounding effects and how this might impact the analysis results.

It is not clear what the reviewer implies by confounding effects. In experimental choices confounding is mitigated due to the experimental design where the choice sets
presented are created in an almost orthogonal way (i.e. low correlation) and as such estimated effects are causal. At the same time, the fact that individual characteristics cannot directly influence choices mitigates traditional confounding or omitted variables problems. However, possible biases can arise if important dashboard attributes have been omitted. Given the background qualitative work and piloting we are confident that we capture most important feature of the dashboard without omitting any relevant major features. Another possible source of problems is the non-random nature of the sample (i.e. sample selection) which can hinder generalisability. The limitation is common in most DCEs and is now acknowledged in the limitations section.

9. Paragraph 2.4.2, Figure 1 is more a list than a figure.
   We have now made this a table.

10. Results in section 3 are presented in a confused way, but I think this is mainly related to the lack of clarity of the design study. I would suggest to rethink the structure of the manuscript in order to efficiently link the first sections to results, clearly explain how different methods are applied to address different research questions.
    Following the reviewer’s comment we have revised the design and analysis sections. We believe and hope that the changes made in these sections now clarify things. Overall we follow standard choice experimental methodology in analysing, interpreting and presenting our results. We have discussed at length the exact interpretation of each of the presented tables and the intuition behind them and hence, we are unsure how restructuring Section 3 could improve comprehension.

11. Please introduce all the acronyms (e.g. WTP)
    We believe we have done so. WTP is introduced at the top of page 5.

12. Table 6 is actually the only point where you clearly illustrate the design experiment, but this explanation came too late in the manuscript.
    The reviewer is right and indeed this is confusing. We have added a figure (Figure 2) with an example choice set that will help clarify the structure of our experimental survey questions.

13. If you discussed the dashboard design from the UX point of view, this should be included in the discussion.
    Apologies but what is “the UX point of view”? It could be “user experience” but even if this is the case we still don’t understand the comment. The whole set up of the dashboard and the experiment is to elicit potential users’ preferences.

Saif Khairat (Reviewer 2): This study aims to serve as the introduction of such a framework in the product development and evaluation of any health care technology.

The authors mentioned conducting a systematic review however, there are no citation to follow nor the results of the review are not discussed in the paper. (Page 5, line 17).
We plan to publish the review results (updated) but have not completed that paper so cannot give a reference. We have however added a summary of points as a Table.
Participants who were involved in the study "had previously shown interest in the study, or had attended our launch event, were invited and encouraged to share the link with colleagues". There are study design concerns, first this is a convenience sample and participants had attended a lunch event which introduces bias to the findings.

These were not our only participants; as stated in the text we invited 200 altogether, many through relevant organisations and charities. However we did also deploy a form of snowball sampling. Thus respondents were a select group but only inasmuch as we only selected potential stakeholders. Given that we sent invitations to people through the main relevant professional organisations as well as to people in our more immediate networks it is likely that the sample was not as biased as the reviewer believes. Moreover, given the very specific nature of the Dashboard, this selective sampling was necessary and indeed the funders required it to be very targeted - respondents had to understand and appreciate the features of the proposed dashboard as intended future clients. We have added a statement about this in the limitations section however.

Response rates to the survey are very low 33/152 (21%), and the sample size is too small. Males were underrepresented in this study.

The response rate to the survey was actually high. Of the 152 individuals invited, 148 logged into the survey (i.e. response rate 97%). Of those 148, 116 responded to at least one choice set but did not finish the survey rendering the answers usable. This is now stated explicitly in page 11. Overall, online experimental surveys commonly achieve low response rates of 10% to 30%, hence we are comparable to past work. However, we agree with the reviewer that sample size is indeed small and males potentially under-represented. While identification of parameters in choice experiments does not rest on sample size, small sample size probably affects the generalisability of findings. This is now explicitly acknowledged in the limitations section.

At the same time we have altered the paper’s title to reflect the exploratory nature of the paper to: A proof-of-concept framework for the preference elicitation and evaluation of health informatics technologies: an application to the online PRESENT dashboard.

The authors argue state: "we assume the dashboard is offered to 100, 500 and 1000 potential interested clients and the product". What is the assumption based on? Literature? It is unclear.

The size of the potential client base is only hypothetical to allow the cost-benefit calculations. The range of values has been chosen to highlight the possible profitability of the dashboard in a hypothetical market. We have added this to the text.

Table 7 is interesting. One inquiry is the extrapolations to 500 potential clients validated? (This is now Table 9). It is not clear what the reviewer means by validated. This is a suggested hypothetical simplistic cost-benefit exercise to examine the profitability of the dashboard. Following the exercise results were not returned to the actual respondents or potential clients for comment or a further purchasing intention exercise.