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: 2 Date: 04 Jun 2019

Author’s response to reviews:

Reviewer 1 comments
Author response
While suggesting a thoughtful reorganization of the manuscript, authors did a very little on that, especially in trying to precisely define their research question and outcomes and to efficiently linking experimental design and results. We have now tried to define the RQ p 6 and link this to the experimental design and results by adding (research question 1/2/3) at relevant points.

P6 now says:

The research questions for the DCE were:
1. how much are individuals willing-to-pay for each of the features of the online dashboard
2. which features are monetarily the most desirable, as a proxy for practical desirability in use by health services managers within cancer care
3. how likely they are to purchase or not purchase a dashboard given its features.

While they provided some additional literature about DCE, they failed in doing the same for actual dashboard-based systems implementation in healthcare.

On p 5 we have now added:

2. Prior work and context
UK healthcare policy foregrounds the patient’s perspective on quality of care, and we used as our test case an example of a successful survey that aims to capture this perspective, the National Cancer Patient Experience Survey (CPES). Since 2010 this has been sent over a 3-month period each year to all patients treated for cancer as in-patients or day cases in NHS Trusts in England. But there has been no way of systematically, efficiently and usefully analysing and reporting the free-text responses in this and other similar surveys, despite this being a recognised need. The PRESENT project therefore used rule-based information retrieval to structure CPES free-text comments and Python programming language
(https://www.python.org/) to display results and text analytics in a summary visual format in a digital dashboard display that could be drilled down to the original free-text.

A dashboard approach was chosen because of its popularity in healthcare. In 2009, the NHS Connecting for Health initiative, now incorporated within The Health and Social Care Information Centre (http://webarchive.nationalarchives.gov.uk/20130502102046/http://www.hscic.gov.uk/aboutus), began the Clinical Dashboards Programme. This was an attempt to make dashboard development more accessible to healthcare professionals as a useful way of summarising healthcare-relevant data. But its recommendations do not appear to have been based on domain-specific considerations that incorporated patient views and nor do they appear to have been informed by empirical work on healthcare dashboard usability and design considerations. Moreover this programme was so flexible that it led to very divergent design approaches that often breached Gestalt design principles [7]. Good dashboards should tailor their designs to specific end user requirements without breaking these principles.

While a PRESNET dashboard description is not available via literature, I would expected authors to improve its description in the manuscript rather than providing a link to that. We have now added some dashboard part-screenshots to make this clear as Fig 2. Table 1 should also report the distribution of participants in the stakeholder categories.

We do not understand this point, we believe we have done this. Does the reviewer mean on a group by group basis? Is not clear where is shown that "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 now added the text. I'm still convinced that a graphical representation of the experimental design is necessary. Also if you state "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." I don't see why you still include a paragraph titled "Experimental design"? As for the PRESNET dashboard, only providing a link to the JMP software (which of course should be included) is not useful to explain how you exploit it within your project.

We have included a graphical representation. With regard to the quote, we meant we would not delve into theoretical statistical aspects, but of course we had to explain the design we used, we are sorry for the confusion our miscommunication caused.

JMP was the software that was used to create three D-efficient fractional factorial partial-profile designs with eight choice sets each and three alternative per choice set. This is now clarified in the manuscript. I agree on the reply to comment number 8 and author response clarifies the point, although while re-reading the revised paper from the beginning, my concerns arose again, and I think (again) this is due to a poorly organized/clear method section. I'm still suggestion to rethink the manuscript organization also in order to better link methods and results. We hope we have now managed to deal with this. Figure 2 is missing. Apologies this is now added (it is Fig 3 now)
My comment about user experience was meant to suggest an addition in the discussion regarding for example holistic, evidence based frameworks (e.g. CeHRes) for end-to-end design and development processes in design and implement health informatics technologies.

Thank you for this helpful comment. We have now added the following on p19:

The holistic approach that we used in PRESENT to co-create our dashboard, incorporating human, organisational, economic and technology dimensions, is increasingly recognised as necessary. This is shown for example by the development of holistic, evidence based frameworks such as CeHRes [19] for end-to-end design, development implementation of health informatics technologies.

There is nothing particular about a healthcare dashboard that would make it a specific candidate for the use of DCE in the way we have described per se, but healthcare dashboards have types of data, uses and historical contexts that make their attributes different to those of other dashboards. Thus our use of DCE can be generalised beyond healthcare but the data from our study are only potentially generalizable to other healthcare management dashboards.

Reviewer 3 comments

Main contribution of the study is not presented clearly. The following is what the paper claims to do: "This paper sets out to propose, define and implement a customization and evaluation framework for an online software dashboard that satisfies these needs, using a Discrete Choice Experiment." I found this aim confusing, somehow limited, and not situated within the gaps of the relevant literature. Is the aim to propose a customisation and evaluation framework for software dashboards only? Or for IT systems in general? Or dashboards/IT systems in healthcare or other settings? If the focus is upon healthcare settings, are there any particular factors that separate it from other domains? If yes, what are they? If not, what would be the unique contribution of the paper in using DCE to develop a healthcare IT system? Is the novelty of the current study related to use DCE as part of this framework? Has DCE been used for developing such IT systems in other domains and/or in healthcare domain? If it has been used, what are the outcomes and/or gaps in this line of work? Has DCE been found effective? If so, in what ways? Is this the first study to use DCE in developing a healthcare system or is it more specifically a novel use of DCE in dashboard design & development?

If the contribution is not limited to the use of DCE but is related more to the whole process, then the entire process and framework need to be presented in a much more structured way. Proposing a framework ideally requires a structure preferably presented together with a diagram with various stages, factors and activities comprising the framework. Currently, the series of activities performed to obtain the desired results seem arbitrary and hard to situate within a structured framework.

The framework/process appears to involve the following activities:
- Identifying attributes and levels:
  - A systematic review (unpublished and limitedly reported)
  - A scoping survey with 35 respondents (limited information on its design and demographics)
  - Co-design workshops (6 workshops with 34 participants)
Identification of 21 features and then discussion and 'further co-design work' were performed to reduce them to 10 features. (no information on the process to reduce the no of features from 21 to 10, and why 10? Mentioning of 'further co-design work' is too vague.)

At some point willingness-to-pay amount was gathered from the potential users.

Cognitive interviews with two stakeholders were performed to get feedback on features and potential prices (no details on this).

- Experiment Design

Some experimental design methods (that are not clearly presented) were performed to determine number of attributes and their levels. In the literature, it seems the aim is to generate Fractional Factorial Design, but it is not clear whether this has been the case in this study or not.

Some more descriptions of ad-hoc-looking decisions on attribute and level choices, ending with 'using one of the feature (unspecified) of [their] experimental design software' to generate three different versions of the same design with the various combinations of attributes and levels.

- Conducting the Survey
- Performing the DCE analysis
- Estimation of product valuation

Which part(s) of the work, process or outcome presented in the paper are proposed as the contribution is not clear to me. If the entire set of activities is the contribution, the paper needs to spend more time to re-organise these as a structured process with clearly explained and evidence-based rationales to support its activities.

We hope that we have clarified this with changes including those made in response to reviewer one.
We have not meant to present this as a framework but rather an example of how DCE can be used in the development of health technologies. We have added a diagram of the study and research questions.

We have also added to p 3:

In the PRESENT study we had sought to determine these factors from a literature review, survey of relevant potential end uses of the dashboard, a workshop and interviews. But the review findings were limited, due to a lack of previous work in the field of healthcare dashboard feature preferences, the survey produced comments on what respondents had personal experience of, rather than what they desired, and the qualitative data from the workshops and interviews was a retrospective co-construction between researchers and participants [2]. We confirmed a gap in methodology in conversations with other researchers in a ‘learning set’ set up to foster collaboration between research teams developing similar systems and funded, as we were, by the UK National Institute for Health Research (NIHR).
We have added to p 7:

Through discussion within the research team, and further co-design work – exploring the dashboard and attributes list in a workshop with our advisory group and through their networks - the number of attributes was reduced down to 10 key features of a healthcare dashboard.

In addition we have specified on p19:

There is nothing particular about a healthcare dashboard that would make it a specific candidate for the use of DCE in the way we have described per se, but healthcare dashboards have types of data, uses and historical contexts that make their attributes different to those of other dashboards. Thus our use of DCE can be generalised beyond healthcare but the data from our study are only potentially generalizable to other healthcare management dashboards. The contribution is in the use of a structured way in creating and evaluating dashboards that is mainly guided by the steps of the DCE which implies the identification of attributes and levels, as well as the data collection and analysis of respondents’ preferences.

We have added a detailed description of the other work in the study and placed the DCE within this more clearly to address most of these points. We have also added fig 1 to show the stages. We were worried this would confuse rather than clarify as it would not be appropriate to go into depth about all aspects.

WTP amount were not gathered from respondents but elicited from their responses. WTP represent the marginal rate of substitution between levels of attributes. If one of the attributes is in monetary terms this translate to value that a respondent assigns to the level of an attribute. This is described at the beginning of section 2.4.1.

The exposition and detailed presentation of experimental design methods (as well estimation/modelling techniques) is outside the scope of the paper and as such was kept intentionally brief. Key references for the literature are given so that any interested reader can follow through. However, the reviewer is right that some key information is missing from the manuscript (e.g. fractional factorial designs are created etc.) that has now been added at the bottom of paragraph 3 in Section 2.2.

As previously some of these choices relate to technical aspects of experimental design that are beyond the scope of this paper. Nevertheless, none of our choices are ad-hoc but supported by past evidence and good practice in DCEs. First, we increase the number of alternatives in a choice set, second, we allow only some attributes to vary within each choice set, third, we create three surveys to allow for more coverage of the full-factorial and improve design efficiency and fourth, we use a two-stage question format. All of these choices result in an experiment that is cognitively manageable by respondents and allows the researchers for the elicitation of all necessary information in a statistically efficient way.

This might be the result of the lead author coming from a different discipline. What we propose is a way to develop and evaluate dashboards or online toolkits through DCEs. We have modified the descriptions and re-organised and added text throughout that we hope satisfies this comment.
I do not see any evidence to support the following claim (Page 17, Line 52): "Discrete Choice Experiment can be successfully used to inform development of an online dashboard by determining preferences for particular features and customization options and how this affects individuals' purchasing behaviours." How has the study arrived at the conclusion that DCE successfully informed the development process? Earlier in the discussion, the paper states that "it is not clear what the uptake would be among the overall stakeholder population" (Page 17, Line 6). How was the success measured in this case? There is no way to judge whether using DCE has been effective in determining the features desired as there is no product and associated demand information available. Perhaps, a qualitative study to discuss the findings derived from DCE study with potential users could improve the validity of the results.

We have filled in the gaps – we had omitted this detail to keep the paper brief but this clearly has not been effective.

We have included two new sections, one called Prior work and context

And one called Synthesising the DCE results with those from previous and subsequent stages of the PRESENT project

We believe that these and the new diagrams and table we have added should satisfy these questions.

Finally, the specific healthcare setting has never been mentioned. Healthcare is a very large domain and a statement like generic key features of 'a healthcare dashboard' (Page 5, Line 35) is very vague and unspecific. A dashboard in an emergency room has different kind of requirements than the one in a primary care setting. Was there a particular healthcare setting targeted? Were the participants told about a target domain/setting? Or were the participants knowledgeable/relevant health professionals/stakeholders in the case of a specific target setting?

We have now described at the start (context section) how CPES is collected in primary and secondary care. We have also partly addressed this through our new section on context and prior work in the study. However we purposely designed the dashboard to suit a range of settings, the common theme was patient experience. We agree with the reviewer, but focusing on the patient experience cutting across different healthcare settings, which the CPES itself does, means that we did not need to consider the different features of different settings. This was clear from the feedback we got in early workshops, as we had originally believed we had to produce several different versions of the dashboard. We have now clarified this on p 6.

In early workshops we proposed that we would produce more than one dashboard, to represent different user settings, but it rapidly became clear that potential end-users wanted one system with different layers of access depending on setting. This is in keeping with CPES, the comments in this survey cutting across primary and secondary care of all types.

The focal feature was the patient experience and different end-uses could be satisfied simply by including filters and search functions.