Author’s response to reviews

Title: Involving End-Users in the Design of an Audit and Feedback Intervention in the Emergency Department Setting - a Mixed Methods Study

Authors:

Welmoed van Deen (Welmoed.vanDeen@cshs.org)

Edward Cho (Edward.Cho.2018@marshall.usc.edu)

Kathryn Pustolski (pustolskik@gmail.com)

Dennis Wixon (dwixon@cinema.usc.edu)

Shona Lamb (shonalam@usc.edu)

Thomas Valente (Thomas.Valente@med.usc.edu)

Michael Menchine (menchine@usc.edu)

Version: 1 Date: 12 Mar 2019

Author’s response to reviews:

Reviewer reports:

Bridget Koo (Reviewer 1): Thank you for the opportunity to review this paper which aimed to develop design requirements for an ED performance dashboard and to understand the role of culture and social networks in the adoption and associated behaviour change. The paper is very well written and follows a logical flow. The methodological approach is overall sound and the conclusions drawn seem appropriate for the study findings. I have some issues in relation to the overall aim of the project and elements of the methods, however none of these are unsurmountable but I believe do require additional explanation/information. I hope the authors find my feedback useful. NB the manuscript came with two sets of line numbers I have elected to refer to the continual set.

Response: We thank Dr. Kool for reviewing our paper and for providing this useful feedback. We hope we addressed the reviewer’s comments and concerns below.
GENERAL FEEDBACK

I would encourage you to reconsider the use of the term 'provider/s' throughout the manuscript. It would seem when you use this term you are predominately refereeing to ED Doctors/physicians. This journal has an international readership and therefore the use of more generic terms may be advantageous

Response: We agree with the reviewer and changed the term ‘provider/s’ into ‘physician/s’ throughout the manuscript.

It should also be noted that in many countries it is not only Drs who can order laboratory tests, in many places advanced nurse practitioners can. However, for the purpose of this study the focus was doctors

Response: Thank you for this comment. We agree that this is a limitation of the current setting, in which the attending physicians were the only providers who can independently make treatment decisions. We added a sentence in the discussion to address this limitation: “Moreover, our study only included attending physicians at a single site, limiting generalizability to other settings and other healthcare providers such as nurse practitioners.” (line 477-479)

It would be useful to draw in some of the Quality Improvement literature in this section, to highlight that in healthcare we should aim to reduce harm, waste, and variation. In relation to diagnostic tests, the overall goal should not be to reduce the amount of diagnostic tests ordered but to reduce the amount of inappropriate tests ordered. Practice should be evidence-based and as such follow the recommended suit of diagnostic tests recommended by evidence based guidelines for the management of certain conditions, this would help to reduce waste and variation

Response: We agree with the reviewer that the goal should not only be higher efficiency, but primarily improved quality of care. As non-evidence based, non-guideline driven utilization does not only increase costs but also causes harm to patients, this is a key-concern to address in the context of quality. We clarified the importance of quality in the introduction by adding the following: “…overutilization is a common quality concern, which delays the care delivery process and can cause harm to patients.” (line 79-81);
We also added a paragraph in the discussion to discuss quality implications (also see later comment): “Lastly, we would like to acknowledge that caution should be exercised when attempting to implement a performance dashboard such as the one developed in this study. Both intended and unintended consequences should be monitored closely, for example through the inclusion of balancing measures. This is especially important when appropriateness measures are not readily available for metrics such as test utilization and LOS in the ED; reductions in test ordering may not always be appropriate and reductions in LOS might not always improve quality of care even if they may increase patient satisfaction. While we included 72-hour return rates as a balancing measure on the dashboard, several other potential unintended consequences were identified in the interview process, including physician stress levels, dehumanization of the patient experience, effects on teaching, and effects on quality and safety of care.” (line 483-492)

As it currently stands it seems that the goal of your dashboard is to reduce the volume of tests. The limitations of this as a goal need to be made more explicit.

Response: Based on this comment, we wanted to clarify that the initial goal when we initiated the design process was to reduce LOS as LOS was above the desired benchmark in this ED. Only later, in the interviews, test utilization was identified as one of the most important metrics to include on the dashboard as it can be influenced by the physician and is known to affect LOS.

We clarified this goal and process in the introduction: “Here, we describe the development and exploratory evaluation of a performance feedback dashboard in the ED setting with the goal to reduce LOS. We employ a human-centered design approach to identify the most suitable measures to include on the dashboard and to ensure essential design requirements are met.” (line 106-109);

And in the methods: “Attending physician interviews were conducted to obtain end-user input on the measures to include and the design and functionality of the dashboard as well as to identify barriers to its implementation.” (line 136-138)

We agree with the reviewer that one of the goals of the dashboard that we ended up developing, is to reduce test volume. We recognize the limitation of looking at volume and not at appropriateness of tests and added this limitation in the discussion to make this more explicit: “This is especially important when appropriateness measures are not readily available for metrics
such as test utilization and LOS in the ED; reductions in test ordering may not always be appropriate and reductions in LOS might not always improve quality of care even if they may increase patient satisfaction.” (line 486-489; also see comment above and later comments in the discussion section)

BACKGROUND

As per previous comments regarding some suggested additional information to consider including

Line 71: should clarify that you are refereeing to LOS 'in the ED'

Response: we added this. (line 73)

It might be useful to clarify that for the purposes of this paper the term 'diagnostic tests' refers to laboratory and imaging tests

Response: we added this specification in the introduction: “[…] and the number of diagnostic tests (e.g. lab and imaging tests) ordered.” (line 75-76)

Do you have any general data regarding the over ordering of diagnostic tests in the US and if there are particular conditions/settings where this is more likely to occur and for which tests. Or if there are biases present e.g. low rates of ordering for low socio-economic groups? Also is there any information for the information where the study took place?

Response:

While a lot of systematic work has been done on this very important subject in other fields, this subject has not been as extensively evaluated in this setting. This is an important consideration that needs to be addressed in future studies. While we acknowledge the importance of the subject, we feel that this subject is much more nuanced than what can be covered in the introduction of this paper.
Regretfully, no data on appropriateness of test ordering at the study site was available for this study. However, we did identify substantial differences between providers’ practice patterns. We added this to the methods section: “Additionally, there was substantial variation between providers in LOS and utilization metrics; for example, the interquartile range (IQR) of the median LOS per provider was 6-9 hours and the IQR of the % patients with a CT scan was 10-30%.” (line 125-127)

Lines 102 - 3: the 'how' needs to be added to this sentence i.e. by reducing diagnostic test ordering.

Response: Please also see the response to the first comment on page 2 of this document. The initial goal was to reduce LOS. Test utilization was identified later in the interviews as one of the measures to incorporate on the dashboard to achieve this goal (time to disposition decision was also identified as an important metric in this process). We clarified our aims and methodology in the introduction and methods sections (line 106-109 and 136-138; see above).

METHODS
See earlier comments re use of descriptors for medical staff that are more global in nature

Response: see above; we agree with the reviewer that the language we used was not specific enough and changed the word ‘provider/s’ to ‘physician/s’ throughout the paper.

Line 113: can abbreviate 'Los Angeles County' at the end of this sentence to 'LAC'

Response: we changed this. (line 119)

Would the hospital where the study took place be considered a 'public' hospital in other settings e.g. UK, Australia, NZ, Canada. It might be useful to signal that for the non-US reader
Response: We agree with the reviewer and clarified this for the international reader: “LAC+USC is a large public hospital owned and operated by the LAC Department of Health Services and is a major teaching hospital.” (line 118-119); and also made the terminology consistent in: “which is well above the LOS observed in other EDs serving un- and underinsured populations” (line 124-125)

Line 124: the changing of practice would appear to be in relation to the ordering of diagnostic tests - this context should be added.

Response: Please see our responses above related to this subject.

Line 125: change 'sixty-minute' to 60 minute

Response: we changed this. (line 133)

Line 126: goals in relation to what all aspects of the unit of just diagnostic tests?

Response: we discussed the strategic goals of the ED as well as more specific goals related to specific performance measures. We clarified this in the methods: “… to understand the strategic goals of the department and how a dashboard with specific performance measures could help attain those goals” (line 134-136)

Line 130: were these leadership people also clinicians or were they unit management type people - useful to clarify

Response: the leadership people were ED physicians who occasionally practice in the ED. We clarified this in the methods: “The ED leaders were ED physicians whose primarily responsibility was leadership but who also practiced occasionally as attending physicians in this ED.” (line 140-141)
Line 200: did the list of 'ED provider-names' include medical staff and the leadership team?

Response: The list included names of physician-leadership but not of medical staff. We clarified this in the methods: “Up to seven names could be selected from a pre-specified list of ED physician-names including those in leadership roles.” (line 213-214)

Ethics: currently no mention of if ethics approval was obtained for the study, if so from where? Was informed consent to take part obtained from participants? If ethics approval was not obtained then this needs to be explained

Response: we apologize for the misunderstanding. Per the journal’s policy, details about the ethics approval procedure are listed in a separate ‘DECLARATIONS’ section on page 25 (line 515-517).

RESULTS

This section was very well presented, and clearly presented.

Some very interesting findings in relation to workplace culture and the adoption in innovations

Response: We thank the reviewer for this comment.

Useful to identify respondent quotes in a de-identified manner e.g. 01, 02. Otherwise it is difficult for the reader to know if it was the same respondent providing all the quotes.

Response: We added these.

% in tables should be avoided unless the n is close to 100. Suggest removing %'s from the tables and text. Instead refer to the proportions a fractions (e.g. more than a half) or by overall size (e.g. most (n = x/x)
Response: we agree with the reviewer and made the requested changes in Table 2.

There are a lot of tables. I wonder if table 3 could be simply described in the text.

Response: we removed Table 3 and described the numbers in the text instead: “… the percentage of physicians who would order imaging after seeing the prototype dashboard (12/19 pre and 14/19 post; p=0.69). There were also no significant changes in the self-determination theory-based motivation measures; median pre- and post-values for value/usefulness were 4.4 (IQR 4.2 - 5.2) and 5.0 (IQR 4.0 - 5.8), respectively (p=0.21); the perceived competence values were 5.5 (IQR 4.3 - 6.0) and 5.0 (IQR 4.0 – 6.3), respectively (p=0.88); and the autonomy support values were 5.0 (IQR 2.8 – 5.7) and 4.0 (IQR 3.0 – 6.0), respectively (p=0.83)” (line 390-395)

DISCUSSION

Again, well-structured and clearly articulated. A good discussion of the strengths and limitations of this work and the implications

Response: Thank you

In the first paragraph, mention of the desire to reduce the number of diagnostic tests should be made explicit

Response: Please see our comments above on page 2 and 3 related to this subject. We hope this clarifies what we did and why.

There is currently insufficient mention of the shortfalls of dashboards
Response: We added a paragraph to discuss shortfalls in the Discussion: “Lastly, we would like to acknowledge that caution should be exercised when attempting to implement a performance dashboard such as the one developed in this study. Both intended and unintended consequences should be monitored closely, for example through the inclusion of balancing measures. This is especially important when appropriateness measures are not readily available for metrics such as test utilization and LOS in the ED; reductions in test ordering may not always be appropriate and reductions in LOS might not always improve quality of care even if they may increase patient satisfaction. While we included 72-hour return rates as a balancing measure on the dashboard, several other potential unintended consequences were identified in the interview process, including physician stress levels, dehumanization of the patient experience, effects on teaching, and effects on quality and safety of care.” (line 483-492)

What are the processes in place to audit the appropriateness of diagnostic test ordering in the organisation? E.g. a "quick disposition decision" may have a negative outcome - patient unnecessarily admitted/discharged; key diagnostic information not obtained that results in delays in treatment.

Response: We incorporated this in the paragraph that we added about the shortfalls (See comment above).

ABSTRACT:
Revise in light of feedback given for specific sections above.

Response: In the abstract, we changed the word ‘provider/s’ for ‘physician/s’ per the reviewer’s suggestion. We also added that inappropriate diagnostic tests not only impact LOS but also have a negative effect on quality of care (line 34).

Methods: Suggest indicating where the study took place i.e. USA, public hospital ED

Response: we added this (line 44).
Line 42: when I first read this I envisaged you meant drs, nurses, allied health etc. It would be more accurate to refer to them as senior and trainee drs (or something similar)

Response: We modified this: “We performed 13 semi-structured interviews with attending physicians in different roles within a single public ED in the U.S.”

Noah Michael Ivers (Reviewer 2): Thank you for the opportunity to review this interesting paper

Response: We thank Dr. Ivers for sharing his expertise in these insightful comments and concerns. We hope we addressed these concerns sufficiently in the responses below.

This was a small, local study that aimed to gather end-user input to develop a dashboard to encourage efficiency/throughput in the emergency department and then sought to understand end-user perceptions of the dashboard that was designed.

I believe that the findings related to the social network analysis are - to my knowledge - relatively novel in this space and therefore represent a potentially important contribution, even if the results must be considered quite tentative due to the limitations of small sample size, incomplete response rates, and single site design. These limitations are appropriately noted in the paper. It may be the case that involvement in the design process of the Dashboard was important moreso for galvanizing support for the importance of the metrics included in the dashboard, rather than for increasing the effectiveness of said Dashboard.

Response: We agree with the reviewer that there might be several different explanations for the observed effects, which we did not formally test in the present study. As shown in table 2, the physicians involved in the design process were more likely to have higher in-degree scores and were therefore more central to the social network, which might make them more likely to understand and share core beliefs and values common to individuals within the network. Since the metrics on the dashboard were chosen based on the interviews, it is definitely possible that this could have allowed for better accepted features and a greater likelihood of adoption by other
members of the network. The exact mechanisms will need to be clarified in further research as we cannot draw any firm conclusions based on the data presented in the current study.

We clarified the critical fact that metric were chosen based on end-user input more clearly in the introduction: “We employ a human-centered design approach to identify the most suitable measures to include on the dashboard and to ensure essential design requirements are met.” (line 107-109); And in the methods section: “Attending physician interviews were conducted to obtain end-user input on the measures to include, the design and functionality of the dashboard and to identify barriers to its implementation” (line 136-138)

Additionally, after reading the reviewer’s comment and re-reading our results, we realized that we described the results of our social network effects in a potentially confusing way, as our analysis measured exposure to other physicians involved in the development process, which only indirectly translates into exposure to the development process itself (which is how it was phrased in certain part of the results section). We modified this in several sentences in the results section to ensure clarity: “Physicians with less exposure to those involved in the development process tended to have less experience […]” (line 382-383); “We then assessed changes in sub-groups based on the level of exposure to people involved in the development process as determined by the social network analysis.” (line 395-397); “[…] we found that physicians with little exposure to those involved in the development process […]” (line 412-413)

The other qualitative findings are not novel. The concept that end-users need to trust the data and the specific requests they make certainly echo in the literature. What would be more interesting is laying out exactly how the designers responded to these requests. This is best addressed in Supplementary Table 2 which I think should be part of the paper, not supplementary.

Response: We thank the reviewer for this thoughtful insight. We followed his recommendations and moved Supplemental Table 2 into the main text.

I also worry that while the authors claim a UXD approach, the number of iterations is not clear. I'd encourage the authors to clarify how the design changes over the course of iterations and how they decided to stop iterating. The above-mentioned table may be used for this purpose as well, I suppose, possibly with an extra column.
Response: We thank the reviewer for this comment and made sure to provide additional information about the iterations in the results section: “[The dashboard] was then iteratively improved in two additional design cycles based on additional end-user feedback from five physicians in November and December 2017 (Supplemental Table 1). The first major iteration was developed after three additional interviews; the second and last major iteration was developed after another two interviews that revealed only minor required changes (Table 1), after which the design process was finalized.” (line 324-328); We also specified which changes were made in each of the round in Supplemental Table 2, which was moved into the main manuscript and is now Table 2.

The sample size could be better justified - was saturation achieved?

Response: thematic saturation was achieved after interview 7; no new themes emerged in interview 8. We clarified this in the results section: “[Thematic saturation was achieved after seven interviews; no new themes emerged in the eighth interview.”] (line 249-250)

I'm not sure that the self-determination theory is quite right. Typically, the literature on audit and feedback initiatives such as dashboards start with the assumption that motivation is present but that awareness of a discrepancy between desired and actual processes or outcomes is limited. A&F is also only likely to be effective in scenarios when recipients are capable of addressing the undesired discrepancy once they become aware of it. Other theories that may have been more directly relevant include Feedback Intervention Theory (kluger and denisi), or Control Theory (carver and sheier).

Response: We thank the reviewer for this comment and will consider altering our framework for future evaluations. We are aware that A&F is only thought to be effective when recipients are capable of addressing the underlying discrepancy, and ensured the selection of actionable measures for that reason. We clarified this in the section about the human-centered design approach: “We specifically ensured the selection of measures that are actionable to physicians as this is a known requirement for effective audit and feedback interventions.” (line 343-344)

Minor suggestions:

Be sure to define acronyms at bottom of each table.
Response: we added the acronyms at the bottom of each table.

Use consistent language (physician/provider/interviewees)

Response: we made our language more consistent: we no longer use the word provider and refer to physicians instead as this is more in line with what we actually did (see comments reviewer 1). In the context of the interview process we use the word interviewee consistently to identify physicians who participated in the interview process.

The N for figure 2 should be added

Response: We added the n’s to the figure. Each dot in the figure represents a single observation; we clarified that in the legend as well.

Table 1 could be just explained in text to make space for another table.

Response: we moved this table to the supplemental materials.

Line 419 - I would have assumed they would have had high scores on one and low on the other

Response: thank you for pointing this out; in the calculation of the SAQ the scoring for the question “In this clinical area, it is difficult to discuss errors” is reversed so that higher scores still represent a higher teamwork climate. For clarity we exchanged this sample question with another for which the scoring is not reversed in the calculation: “I would feel safe being treated here as a patient.” (line 454-455)

Line 423 and elsewhere - I don't think you can say anything much definitive about behaviour change.
Response: we modified the language throughout the paper to more clearly describe that we measured motivation and not behavior change.

I would have liked to see the whole survey all together in a supplementary table. Hard otherwise to make sense of all the data that were collected and whether there was some or was not some cherry picking in reporting of results.

Response: We uploaded the surveys as an appendix for your review. As you will see almost all of the variables have been presented in the paper. The most important question that was not presented is a the second social network question: “Who do you go to for advice about difficult clinical decisions?”.

The question we describe in the paper (“Who do you discuss problems with at work?”) identifies people who have the trust of the community and are good communicators, the question we did not describe in the paper (“Who do you go to for advice about difficult clinical decisions?”) identifies people who are seen as experts in the field and have the technical (i.e. medical) knowledge to persuade others (Saint-Charles et al. Social Networks; 2009). The first network is important if barriers to adoption are cultural, the latter is important if barriers are technical (Valente et al. PLoS One; 2015).

Because we expected that barrier would be primarily cultural (which is also the primary focus of the paper), we opted to include this analysis in the paper. For brevity we opted to exclude the second analysis. We ran the analysis for the second network and included the results hereunder for your reference. As expected, the results are similar though slightly less pronounced, which is in line with our expectations as we expected barriers to be primarily cultural.

(See figure in supplemental file with reviewer responses)

Legend: results of the analysis based on the network analysis using the question “Who do you go to for advice about difficult clinical decisions?”.

Compared to the data presented in the paper:

(See figure in supplemental file with reviewer responses)
Legend: results of the analysis based on the network analysis using the question “Who do you go to for advice about difficult clinical decisions?”.