Author's response to reviews

Title: Redesign of a Computerized Clinical Reminder for Colorectal Cancer Screening: A Human-Computer Interaction Evaluation

Authors:

Jason J Saleem (Jason.Saleem@va.gov)
David A Haggstrom (dahaggst@iupui.edu)
Laura G Militello (Lmilitello@applieddecisionscience.com)
Mindy Flanagan (meflanag@iupui.edu)
Chris L Kiess (ckiess@iupui.edu)
Nicole Arbuckle (Nicole.Arbuckle@udri.udayton.edu)
Bradley N Doebbeling (bdoebbel@iupui.edu)

Version: 2 Date: 26 October 2011

Author's response to reviews: see over
October 26, 2011

Editorial staff
BMC Medical Informatics and Decision Making

Dear Editors:

We just electronically submitted a revised manuscript for, “Redesign of a Computerized Clinical Reminder for Colorectal Cancer Screening: A Human-Computer Interaction Evaluation” (MS: 2047317195840051) by Saleem, et al. We were encouraged by the thoughtful, helpful reviews and the opportunity to submit a revision.

Below, we provide a detailed, point-by-point, list of how the manuscript was revised to address each specific item of each reviewer's critique, including mention of the corresponding changes made in the revised text.

As requested, we have also added a statement at the beginning of the Methods section about the ethical approval obtained for this study.

Thank you for your time and consideration. I am happy to answer any questions.

Sincerely,

Jason J. Saleem, PhD
Reviewer: Ayse Gurses

Reviewer's report:
This is a great example of applications of HCI in health care. I like the fact that the authors used a “prototyped” version of the current system, rather than the current system to ensure a better experimental design. Here are some comments.

• Major Compulsory Revisions

The authors state that the redesigned prototype includes a patient education resource (one-page synopsis of CRC screening). Is this customized/tailored based on the patient (demographics, patient’s previous adherence history with the recommended screening guidelines etc.). If not, I definitely understand the one participant’s comment about having these forms pre-printed as the redesigned system does not have much advantage over the old system with regards to patient education.

Response: The one-page synopsis of CRC screening was not customized based on the patient. We agree the benefit of having this education resource available through the computerized CRC reminder was limited. We the edited the following sentence to the manuscript to clarify:

Methods, Prototypes, p. 9, top:
“…a general resource (i.e., non-patient specific) was added to assist the clinician in providing patient education.”

The authors state that they are using the Workflow Integration Survey as a measure, which is in press. It would be a good idea to include this accepted manuscript along with this submission, or at least the instrument itself with its psychometric testing results for the reviewer to make a more informed judgement.

Response: As requested, we have included the accepted manuscript for the Workflow Integration Survey with this submission (Flanagan et al AMIA 2011 Accepted.pdf), which includes the survey validation information. We have also included the survey questions as an additional file, referenced in the manuscript:

Methods, Procedures and Scenarios, p. 12:
“...see additional file 1: Workflow Integration Survey for a list of the 12 items that comprise the instrument.”

Also, what are the three questions they appended to CSUQ? Are these the questions in Table 2? Not clear.

Response: Yes, the three questions appended to the CSUQ are the questions in Table 2. We have revised the following sentence in the manuscript to be more clear:

Results, Dependent Measures, p. 15:
“However, PCPs rated the redesigned CRC screening reminder significantly higher (better) for the three items that were appended to the CSUQ (Table 2) by the Wilcoxon Signed Ranks Test.”
In their statistical analysis, the authors state that “they grouped similar usability questions together”? How? Did they do any factor analysis? If not, after they grouped the items, did they treat them as scales (if so, need psychometric properties) or as index? Please clarify.

Response: We did not use factor analysis to group similar usability questions together. Essentially, we have one large group for all but two of the questions in the standard CSUQ. The two questions that were treated as separate ‘groups’ were the two overall/summary questions. For the other 17 questions in the standard CSUQ, we summed responses to the 17 Likert items to create a single latent variable for all 17 questions (see below for the specific groups). This was done to avoid 22 individual statistical tests, which would have been inappropriate for multiple comparisons to avoid committing a Type 1 error (showing statistical significance when it is really not significant). An alternative approach could have been to apply a Bonferroni correction for multiple corrections. However, we feel consolidating the number of statistical comparisons for the usability survey was appropriate for this study, since the usability survey was only one of three instruments used (including the NASA TLX and Workflow Integration Survey).

**Group 1**
1. Overall, I am satisfied with how easy it is to use this system.

**Group 2**
19. Overall, I am satisfied with this system.

**Group 3**
2. It was simple to use this system.
3. I can effectively complete my work using this system.
4. I am able to complete my work quickly using this system.
5. I am able to efficiently complete my work using this system.
6. I feel comfortable using this system.
7. It was easy to learn to use this system.
8. I believe I became productive quickly using this system.
9. The system gives error messages that clearly tell me how to fix problems.
10. Whenever I make a mistake using the system, I recover easily and quickly.
11. The information provided with this system is clear.
12. It is easy to find the information I needed.
13. The information provided for the system is easy to understand.
14. The information is effective in helping me complete the tasks and scenarios.
15. The organization of the information on the systems screens is clear.
16. The interface of the system is pleasant.
17. I like using the interface of this system.
18. This system has all the functions and capabilities I expect it to have.

*Appended items related specifically to colorectal cancer screening:*

**Group 4**
20. It is easy to find information about the patient's colorectal cancer screening history in this system.

**Group 5**
21. It is easy to find the patient's current status with regard to colorectal cancer screening in this system.

**Group 6**
22. The system provides helpful patient education materials for CRC screening.

We have included the above questions and groupings as an additional file to the manuscript and added the following text to the manuscript to clarify:

Methods, Statistical methods, p. 13, top:
“See additional file 2: Usability Survey Questions and Groupings for a list of the usability survey questions and how they were grouped.”

Under the “Qualitative Analysis” section, the authors talk about “the open-ended debrief interview notes.” Can the authors explain this in more detail please? How was the interview conducted, was there an interview guide etc.? I don’t remember seeing this. If it is already described in the manuscript in detail and I missed it, then please ignore this comment.

Response: The open-ended debrief interview was truly unstructured (i.e., there was no interview guide). After completing the final scenario and survey instruments, the debrief interview was conducted to probe participants about specific issues that occurred during the scenarios to better understand comments they made while interacting with the new design features or to clarify their decision making process. Participants were also invited to comment on anything they wished to provide further feedback on. In the manuscript, we added additional detail on the debrief interview:

Methods, Procedure and scenarios, p. 12, bottom:
“After the experimental conditions, we conducted an unstructured, open-ended debrief interview to gather additional feedback on the redesigned interface. The debrief interview was conducted to probe participants about specific issues that occurred during the scenarios to better understand comments they made while interacting with the new design features or to clarify their decision making process.”

The authors state that they clarified CSUQ instructions after a participant’s confusion. Can the authors please comment on whether they have pilot tested these instructions specifically for this experiment (before they conducted the experiment)?

Response: We did have one pilot participant before we conducted the experiment to dry run the experimental instructions and procedure. Unfortunately, this issue did not surface in our pilot test.

Table 1: I am curious about how the “workload” measure in this table differs from NASA-TLX. The authors comment on the issue that these two measures give different results and include this in their discussion. But it would help greatly to the reviewer if they can explain the “workload” measure included in their Workflow Integration Survey in more detail so that one can judge the similarities/differences with NASA-TLX better.

Response: The items from the Workflow Integration Survey that comprise the “workload” subscale are:

4. Using CPRS Design 1 during face-to-face patient encounters adds effort (e.g., typing, clicks).
8. Using CPRS Design 1 during face-to-face patient encounters increases workload.
12. CPRS Design 1 helps you complete face-to-face patient encounters efficiently.

In contrast, The NASA TLX has several items that measure specific dimensions of workload: mental demand, physical demand, temporal demand, performance, effort, and frustration.
There are several differences between the “workload” construct in the Workflow Integration Survey and workload as measured by the NASA TLX. The NASA TLX includes specific constructs not covered by the Workload Integration Survey subscale for workload (e.g., frustration). Also, the Workload Integration Survey does not distinguish between mental effort and physical effort. These differences suggest the instruments measure different but related constructs and may explain the difference in results between the two.

We have added the following to the manuscript:

Discussion, p. 21:
“The items from the Workflow Integration Survey that comprise the “workload” subscale are items 4, 8, and 12 (see additional file 1: Workflow Integration Survey). In contrast, The NASA TLX has several items that measure specific dimensions of workload: mental demand, physical demand, temporal demand, performance, effort, and frustration. There are several differences between the “workload” construct in the Workflow Integration Survey and workload as measured by the NASA TLX. The NASA TLX includes specific constructs not covered by the Workload Integration Survey subscale for workload (e.g., frustration). Also, the Workload Integration Survey does not distinguish between mental effort and physical effort. These differences suggest the instruments measure different but related constructs and may explain the difference in results between the two.”

The authors state that they had a hard time recruiting participants. I was wondering whether the authors can describe what type of incentive(s), if any, they used for recruitment. And also how did you overcome these difficulties (what strategies you used)?

Response: VA providers are prohibited from receiving monetary compensation as an incentive for participation in VA research. The only incentive we able to provide was lunch prior to or after their participation. The most helpful strategy we used for recruitment was to enlist the assistance of the Chief of Primary Care for our facility. He encouraged the other primary care providers to consider participating. However, these efforts resulted in only 12 participants. We have added the additional detail to the manuscript regarding recruitment challenges:

Discussion, Limitations and challenges, p. 23:
“However, recruiting providers for the project was quite challenging, due to multiple competing demands. Also, VA providers are prohibited from receiving monetary compensation as an incentive for participation in VA research.”

How did you come up with the redesigned prototype? If this is already described in another paper, maybe you can include just couple sentences describing this in this manuscript to out things into context for the reader.

Response: We have added much more detail in the revised manuscript about the previous field study and how these insights from the field study led to the redesigned prototype:

Introduction, p. 4-6:
“Based on an extensive field study conducted across multiple sites to understand common barriers to the use of CDS for colorectal cancer (CRC) screening [8,9], we tested a redesigned CRC screening clinical reminder in a controlled laboratory simulation experiment. Our previous field study resulted in nine themes that relate to integrating CDS into workflow.
Themes included the following: 1) coordination of outside results; 2) coordination between primary and specialty care; 3) data organization and presentation; 4) just-in-time provider and patient education; 5) interface flexibility; 6) technological enhancements; 7) role assignments; 8) organizational issues; 9) and connecting decision support to quality reporting. Each of the nine themes that emerged corresponds to barriers to integrating CDS into workflow. Based on these barriers, three design features were selected as the most promising. They were judged to be likely to demonstrate high pay-off in terms of improvements to both design of CDS prototypes and their likely integration into workflow, and to be feasible in the short-term. These needed design features include the following: (1) integrating outside results; (2) improving data organization and presentation and (3) providing just-in-time education for patients and cognitive support to providers when and where needed. We prototyped a redesign of the Veterans Health Administration’s (VHA’s) computerized clinical reminder for CRC screening that included a timeline visual to address (1) and (2), as well as a patient education resource to address (3).

The computerized clinical reminders are the main form of CDS in the VHA’s EHR, known as the Computerized Patient Record System (CPRS). We chose to focus this work on CRC screening because there is a robust evidence-base to support the efficacy of CRC screening [10]; yet, rates of CRC screening are sub-optimal [11]. In our previous field study [8,9], we observed providers searching through numerous screens in the EHR to obtain the information they needed, which in many cases was characterized as frustrating and time-consuming by providers. Providers sometimes missed important information (e.g., previous CRC screening results) or relied on patient memory, which could be inaccurate. Our goal was to reduce cognitive load, represent all the high-level information in one location with a visual timeline, providing intuitive pathways to more detailed information and upcoming needed evaluation and follow-up. We took a similar approach for incorporating a patient education resource, which we designed to be readily available for the provider so that they could use it to guide real-time informed decision-making with the patient. In our field observations, we found providers using such educational materials; although they were not incorporated electronically as part of the CRC screening CDS. We designed a laboratory simulation experiment to understand whether our design changes to the CRC screening clinical reminder would result in improved usability, workload, and integration into workflow. We hypothesized that the redesigned CRC screening clinical reminder would be (1) perceived as easier to use, (2) perceived to have lower workload during its use, and (3) given higher ratings for workflow integration, compared to the current CRC screening clinical reminder.”

• Minor Essential Revisions

It would be great if either the authors or the journal submission system inserts page numbers. That would have made the reviewer’s job much easier. Under the Results section, second paragraph, the authors refer to Table2. Have they referred to Table 1 earlier? Again, I may have missed it but the authors may want to make sure the ordering is correct.

Response: We have inserted page numbers in the revised manuscript. Also, we have added a reference to Table 1 (first sentence in Results section) to ensure the ordering of the Tables is correct.
• Discretionary Revisions

The authors highlight the importance of iterative design. This manuscript would have been much stronger if the authors “redesigned” the redesigned prototype based on their findings and test it further. But I understand that this take time and resources.

Response: This is an excellent point. Unfortunately, we were only budgeted for a single test. However, for future grant proposals, we will consider budgeting for more than one lab simulation, since iterative testing is indeed stronger.

Reviewer: Monique W Jaspers

Reviewer's report:

Major Compulsory Revisions

This paper describes a human factors and usability evaluation of a redesigned computerized clinical reminder for use by primary care providers to support them in colorectal screening of their patients. As confirmed by the authors, uptake of human factors research and input into development and design of interactive healthcare technology is slow whereas it is needed to increase user acceptance of these technologies in daily work settings. The study however lacks lessons on usability issues and design of healthcare IT that system designers could profit from. This is due to the fact that:

1. the reminder tool under study is not adequately described and the fact that it takes it data from the CPR. I would highly recommend the authors to refer to the STAR-HI standard for reporting on evaluation studies in healthcare to improve their system description. How would a typical user interact with the tool.

Response: Consistent with the STARE-HI standard for health informatics evaluations reports, we have now added several new paragraphs under a distinct section in the manuscript for “Study context” (starting on p. 6), which includes a description of the organizational setting and system, as well as a descriptions of system details for the prototypes.

2. the insights following from the pre field study are not described. Apparently, these insights led to the decision to redesign the screening reminder tool but are not discussed in such a way that the reader can understand the rationale for choices made in the redesign of the tool. Why would the redesigned tool be easier to use, perceived as having a lower workload, better integrate in work flow compared to the current tool?

Response: We have added much more detail in the revised manuscript about the previous field study and how these insights from the field study led to the redesign decisions:

Introduction, p. 4-6:

"Based on an extensive field study conducted across multiple sites to understand common barriers to the use of CDS for colorectal cancer (CRC) screening [8,9], we tested a redesigned CRC screening clinical reminder in a controlled laboratory simulation experiment. Our previous field study resulted in nine themes that relate to integrating CDS into workflow. Themes included the following: 1) coordination of outside results; 2) coordination between primary and specialty care; 3) data organization and presentation; 4) just-in-time provider and patient education; 5) interface flexibility; 6) technological enhancements; 7) role assignments; 8)
organizational issues; 9) and connecting decision support to quality reporting. Each of the nine themes that emerged corresponds to barriers to integrating CDS into workflow. Based on these barriers, three design features were selected as the most promising. They were judged to be likely to demonstrate high pay-off in terms of improvements to both design of CDS prototypes and their likely integration into workflow, and to be feasible in the short-term. These needed design features include the following: (1) integrating outside results; (2) improving data organization and presentation and (3) providing just-in time education for patients and cognitive support to providers when and where needed. We prototyped a redesign of the Veterans Health Administration’s (VHA’s) computerized clinical reminder for CRC screening that included a timeline visual to address (1) and (2), as well as a patient education resource to address (3).

The computerized clinical reminders are the main form of CDS in the VHA’s EHR, known as the Computerized Patient Record System (CPRS). We chose to focus this work on CRC screening because there is a robust evidence-base to support the efficacy of CRC screening [10]; yet, rates of CRC screening are sub-optimal [11]. In our previous field study [8,9], we observed providers searching through numerous screens in the EHR to obtain the information they needed, which in many cases was characterized as frustrating and time-consuming by providers. Providers sometimes missed important information (e.g., previous CRC screening results) or relied on patient memory, which could be inaccurate. Our goal was to reduce cognitive load, represent all the high-level information in one location with a visual timeline, providing intuitive pathways to more detailed information and upcoming needed evaluation and follow-up. We took a similar approach for incorporating a patient education resource, which we designed to be readily available for the provider so that they could use it to guide real-time informed decision-making with the patient. In our field observations, we found providers using such educational materials; although they were not incorporated electronically as part of the CRC screening CDS. We designed a laboratory simulation experiment to understand whether our design changes to the CRC screening clinical reminder would result in improved usability, workload, and integration into workflow. We hypothesized that the redesigned CRC screening clinical reminder would be (1) perceived as easier to use, (2) perceived to have lower workload during its use, and (3) given higher ratings for workflow integration, compared to the current CRC screening clinical reminder.”

3. the results presented are for the most quantitative that do not provide a detailed insight into the usability problems users encountered in interaction with the two alternative designs. This is a laboratory simulation study with use of Morae software that allows the logging of all mouse, keyboard input and users verbalizations. I would therefore recommend to provide more details on the usability data produced and underlying causes for problems encountered by users. Only then, lessons learned, aside from that iterative design of these technologies is essential, could be given. The authors state in the discussion that: ‘the more general usability statements about simplicity, efficiency, learnability, error recovery, overall satisfaction and other dimensions of usability did not produce significant improved ratings for the redesigned prototype’. Important lessons could only be drawn from this statement when the reader knows in detail the underlying causes for this!

**Response:** This is an excellent point. We have gone back to analyze the mouse clicks and mouse movement data from Morae. However, we do not collect key stroke data (i.e., we turn that option off for all new Morae study configurations in our lab so that we don’t inadvertently capture actual clinicians’ passwords for the electronic health record).

Adding the additional design features in our redesigned prototype resulted in greater average mouse clicks and greater average mouse movement compared to the original design. As noted in the original manuscript, 2 participants did not complete scenarios for both the current design
(design A) and the redesigned prototype (design B). Thus the comparison of the mouse clicks and mouse movement data is based on data from 10 participants. The average number of mouse clicks for all scenarios with design A was 96 compared to 106 for design B. The average mouse movement for all scenarios with design A was 58,189 pixels compared to 70,960 pixels for design B. We have added this important, new information to the manuscript:

Discussion, p. 20, middle:
Adding the new design features in the redesigned prototype corresponded to additional mouse clicks and mouse movement, as recorded by the Morae software, and likely contributed to the lack of significant improvement in general usability. In terms of usability, potential advantages of the new timeline visual and patient education resource may have been off-set by the additional steps needed to access these features.

The users’ verbalizations recorded by Morae were a critical and primary source of qualitative data for us to understand their interaction with the new design features. We have created a new Table 3 (Results, Qualitative results, p. 18) to show representative user verbalizations about the timeline visual for patient history of colorectal cancer screening. This additional detail provides important context for interpreting some of the quantitative findings related to the CRC specific questions we appended to the CSUQ usability survey.

In addition, we have added the following to the manuscript to further interpret the general usability survey quantitative findings:

Discussion, p. 20, bottom:
“Specific participant verbalizations provide some insight why general usability ratings for simplicity, efficiency, and other usability constructs, were not significantly improved with the redesigned prototype. For example, two participants directly commented that the timeline visual for CRC history should be integrated at the same level as the CRC reminder (i.e., without having to click on an additional button). One participant stated: “It would be better if you had it [the timeline] over here on the window itself [next to the reminder dialog] because otherwise you have to click back and forth and it’s hard to remember. Especially which somebody with a complicated history and you got to go from one thing to another, that doesn’t work. So why not put it over here, you know, same window.””

The introduction lacks reference to literature providing evidence that acceptance and successful integration of CDS tools in work flow requires high usability, low cognitive work load and mapping on work flow. Please refer to the work of Bates, Ash, Patel, Kushniruk to mention a few here…

Response: We have now included several references to the appropriate literature for successful implementation of CDS:

Introduction, p. 6 top:
“Successful implementation and end-user acceptance of CDS requires high usability, low cognitive workload and integration of the CDS into workflow [12-16]. To this end, we designed a laboratory simulation experiment to understand whether our design changes to the CRC screening clinical reminder would result in improved usability, workload, and integration into workflow.”
I would recommend going into the reasons why the validated CSUQ survey provided results that differed from the questions added. CSUQ assesses usability as a multi-construct and this was the main focus of the study in comparing the two designs. Provide potential reasons for finding no significance differences here between the two designs. Is power an issue here (the results of merely 9 participants were included in the analyses).

Response: While it is possible that power was an issue, we feel it is more likely that the specificity of the added questions regarding colorectal cancer screening was the main reason why the results differed from the validated CSUQ survey. The three questions added were specifically targeted to get at the issues regarding colorectal cancer screening that our design changes were meant to address. In contrast, the wording of the validated CSUQ questions is non-specific, meant to be applicable to any system. Adding the new design features in the redesigned prototype corresponded to additional mouse clicks and mouse movement (see earlier response regarding mouse clicks and mouse movements) and likely contributed to the lack of significant improvement in general usability (as measured by the validated CSUQ survey). In terms of usability, potential advantages of the new timeline visual and patient education resource may have been off-set by the additional steps needed to access these features. We added this interpretation to the manuscript in our previous response about the Morae data. In addition we have added the following to the manuscript:

Discussion, p. 19, bottom and p. 20 top:
“The specificity of these appended items to the CSUQ regarding CRC screening was likely main reason why the results differed from the validated CSUQ survey, since the three questions added were specifically targeted to assess the issues regarding CRC screening that our design changes were meant to address.”

I would recommend to provide the questions of the Work flow integration questionnaire and discuss differences found between the NASA TLX and this survey in more detail in the discussion.

Response: We have added the Workflow Integration Survey questions as an additional file, referenced in the manuscript:

Methods, Procedures and Scenarios, p. 12:
“...see additional file 1: Workflow Integration Survey for a list of the 12 items that comprise the instrument.”

Discussion, p. 21:
“The items from the Workflow Integration Survey that comprise the “workload” subscale are items 4, 8, and 12 (see additional file 1: Workflow Integration Survey). In contrast, The NASA TLX has several items that measure specific dimensions of workload: mental demand, physical demand, temporal demand, performance, effort, and frustration. There are several differences between the “workload” construct in the Workflow Integration Survey and workload as measured by the NASA TLX. The NASA TLX includes specific constructs not covered by the Workload Integration Survey subscale for workload (e.g., frustration). Also, the Workload Integration Survey does not distinguish between mental effort and physical effort. These differences suggest the instruments measure different but related constructs and may explain the difference in results between the two.”
Minor Essential Revisions

I would recommend to change the structure of the paper and provide the system description and pre field study insights as rationale for the redesign of the tool in a distinct section.

**Response:** This was a helpful recommendation and we have now restructured the manuscript this way (see our first two responses under “Major Compulsory Revisions”).

Provide a rational for redesign choices such as the visual timeline.

**Response:** We have included addition rationale for our redesign decisions, including the visual timeline, as informed by our previous field study to the Introduction of the manuscript (see second response under “Major Compulsory Revisions”).

Please be consistent in terminology: EHR vs CPR

**Response:** We have revised the manuscript throughout to be consistent. We mention the VA’s EHR, called the Computerized Patient Record System (CPRS), initially. We then use the term EHR throughout the manuscript.

The participants experience in the VHA shows a wide varation from 1-25 years. Though the authors claim that average participant experience spanned the existence of VA’s EHR (or CPR), I would be interested to learn whether participants with longer EHR experience performed differently in using the two designs than less experienced users..

**Response:** Six of the participants has between 1-10 years of experience with the VA’s EHR. The other six participants had between 11-25 years of experience with the EHR. Although there is not enough participants to perform statistical comparisons based on level of experience, we looked at the data for descriptive trends. The only trend based on experience that we can discern is that the participants with 11-25 years of experience consistency rated the usability of the current EHR higher than the participants with 1-10 years of experience. This same trend was not observed for the redesigned prototype. As we mentioned in the original manuscript (Methods, Participants section), we considered recruiting non-VA PCPs to remove the potential bias of users having experience with the current design of the CRC clinical reminder being more likely to perform well with the current system. However, information obtained from first-time users of a system is limiting, since it would not be possible to assess whether the new design enhancements affect the performance measures after the participants move past an initial learning phase.

We have added the following to the manuscript:

**Results, Study population, p. 14:**

“Six participants had 1-10 years of experience with the VA’s EHR and the other six had 11-15 years of experience. Participants with 11-25 years of experience consistency rated the usability of the current EHR higher than the participants with 1-10 years of experience on all items from the CSUQ survey. This same trend was not observed for the redesigned prototype. We suspect those with greater experience using the VA’s system corresponded to higher usability due to their familiarity with the current system.”
The authors discuss that some providers questioned the validity of data provided by the CDS tool and give alternative solutions for that. I yet up to now do not know whether these data come from the HER-CPR or not. How do the authors envision to improve the quality of the underlying data or transparently provide information regarding the data source (and how would that solve the problem of distrust?)

Response: We added a section for system details for the prototypes (p. 8), as guided by the STARE-HI standard (see first response above). Since the prototypes used for this evaluation were mock-ups, made to mimic actual integration with the EHR, no real patient data was used to populate the CRC reminder prototypes (fictitious patient data was developed for our scenarios – see Methods section for scenario development), although the actual clinical reminder system is populated with data from the EHR. Regarding the qualitative findings of lack of trust in the quality or authenticity of the underlying data, we have added the following to the manuscript:

Discussion, p. 22, top:
“To increase trust in the data, we envision providing additional options in the visual timeline to increase the transparency of the data. In this way, providers can investigate, for example, the specific details of a patient’s colonoscopy that was performed outside of the VA. These additional details may include the contact information for the facility and physician who performed the test. The additional details may also improve the overall quality of the data.”

It would be worthwhile to readers if the authors could go into these matters in a new version of their paper.

Response: We have carefully considered each critique and revised the manuscript accordingly. We feel the review critiques have substantially strengthened our manuscript and we appreciate the time it took for the reviewer to provide this quality review.