Author's response to reviews

Title: Development and Validation of an Algorithm to Recalibrate Mental Models and Reduce Diagnostic Errors Associated with Catheter-Associated Bacteriuria

Authors:

Barbara W Trautner (Trautner@bcm.edu)
Rupal D Bhimani (Rupal.Bhimani@va.gov)
Amber B Amspoker (Amspoker@bcm.edu)
Sylvia J Hysong (hysong@bcm.edu)
Armandina Garza (Armandina.Garza@va.gov)
Peter A Kelly (pakeelly@tulane.edu)
Velma L Payne (vpayne@bcm.edu)
Aanand D Naik (anaik@bcm.edu)

Version: 3 Date: 21 December 2012

Author's response to reviews: see over
December 21, 2012

Dear Editors:

We are pleased to submit a revised version of our manuscript, “Development and Validation of an Algorithm to Recalibrate Mental Models and Reduce Diagnostic Errors Associated with Catheter-Associated Bacteriuria.” We appreciate the helpful comments and suggestions of the reviewers and feel that the manuscript is much improved with these revisions. We have carefully reviewed and addressed all comments. All revisions made to our manuscript have been marked bold on the text, references, and tables.

Sincerely,

Barbara Trautner, MD, PhD
Associate Professor
Baylor College of Medicine
One Baylor Plaza
BCM 296, Room N-1319
Houston, Texas 77030-3498

Reviewer 1

Reviewer: Christopher Crnich

The manuscript submitted by Trautner and colleagues is a very well written description of an investigation of the factors that contribute to inappropriate antibiotic treatment of patients with asymptomatic bacteriuria.

I found the manuscript to be an enjoyable read on multiple levels. However, I did identify a number of issues that the authors should address before the final manuscript is published.

Author’s Response: We appreciate the reviewer’s generous comments and efforts to improve the manuscript through peer-review. We have attempted to address each of them below.

Major Compulsory Revisions
1. The manuscript makes no mention of how clinicians for phase 1 or 2 were selected for this study. Was it a convenience sample? Was it purposive? If the latter, what criteria were applied?

Author’s Response: Thank you as these are important edits for better understanding the study participants. We incorporated the participant’s selection criteria for phase 1 and 2 into the manuscript. Page 7 contains the criteria for phase 1 (physicians and allied health professionals). Page 9 contains the criteria for phase 2 (expert panelists and non-expert panelists).

2. Similarly, the authors provide no detail on how the cases in phase 1 and 2 were selected (a minor comment: the abstract alludes to the fact that phase 1 involved standardized cases but in the manuscript body it appears that these were real patients so I am not sure how they were standardized). Was the selection of these cases random? Was there a process for their selection? If the latter, please describe.

Author’s Response: We incorporated the case selection criteria for phase 1 and 2 into the manuscript and clarified the point that the cases in phase 1 were drawn from actual patient cases. Page 6 contains the criteria for phase 1. Page 10 contains the criteria for phase 2.

3. In the “Procedures” section, the authors provide a description of the cognitive interviewing exercise using examples of questions. It is not clear from this section if clinician participants were asked to respond to all 20 signs and symptoms (both guideline concordant and discordant). I assume this was the case but if not, please provide greater description on how the investigators chose which signs and symptoms to inquire about during the cognitive interview.

Author’s Response: We incorporated text into the manuscript to clarify the cognitive interview process (see text on page 7). Also we have included our script for the cognitive interviews to clarify this questions for reviewers (but not included in the manuscript).

4. In the “Analyses for Phase 1”: I am not a statistician but I was under the impression that Cohen’s kappa only applied to two independent raters. In this study, there are multiple (>2) raters and I did not think Cohen’s could be calculated in this type of study design. Is it possible that the authors calculated Fleiss’ rather than Cohen’s kappa?

Author’s Response: We appreciate the reviewer’s insight and comment and agree that the methodology for our reliability analyses need more clarification and in some cases key revisions. We originally calculated Cohen’s simple kappa for each pair of raters in both Phases. In Phase 1 we then took the average of Cohen’s kappas for each pair to assess overall inter-rater reliability and in Phase 2 (because we are specifically interested in each pair) we simply reported Cohen’s kappas for each pair (i.e., expert versus non-expert #1, expert versus non-expert #2, and non-expert #1 versus non-expert #2). However, we agree with the reviewer that Fleiss’ kappa (used in agreement for >2 raters) has advantages over averaging pairs of raters who rate the same cases. Therefore, we now report Fleiss’ kappas for Phase 1 inter-rater reliability. Because we are specifically interested in pairwise inter-rater agreement in Phase 2, we have retained Cohen’s simple kappas for these analyses as described in Phase 2 methods.
5. In the “Analyses for Phase 1”: is it possible to calculate a kappa with just 4 cases? It would seem that this estimate would be highly unstable with such a limited number of outcomes (in this instance, I am referring to the 1st sentence in this paragraph, not the subsequent sentences that refer to the signs and symptoms). I think it makes sense to include confidence intervals along with the point estimates of the kappa measures in every instance in this paper. Their inclusion would provide a great deal more insight on the uncertainty surrounding the values currently included in the manuscript.

Author’s Response: We agree about the amount of error that can be introduced with smaller numbers of cases. We have now reported the number of cases rated in all instances (to provide more context) and have included 95% confidence intervals following all kappa statistics (both Fleiss and Cohen). Although these point estimates (kappas) were calculated on a small number of cases, these cases are representative of a real-world sample of catheter-associated bacteriuria and is therefore still useful for a *preliminary* investigation of IR reliability.

Level of interest: An article of outstanding merit and interest in its field

Quality of written English: Acceptable

Statistical review: Yes, but I do not feel adequately qualified to assess the statistics.

Declaration of competing interests: I declare that I have no competing interests

Reviewer 2

Reviewer: Carol E Chenoweth

1. Reviewer’s report - This is a very interesting and timely paper on development of an innovative algorithm for improving treatment of catheter-associated bacteriuria. The description of the algorithm and how it was developed is fascinating and I look forward to trying to incorporate their ideas when this is published.

Author’s Response: We very much appreciate the reviewer’s comment as this is our purpose for conducting this study. We also appreciate the reviewer’s careful review and comments and have significantly revised the manuscript to address them as detailed below.

Major Compulsory Revisions

2. However, the paper overall becomes confusing as the reader tries to determine if the focus is on the mental models or development of the algorithm. Because the authors try to develop both ideas, the message becomes muddled. It almost seems that there are two different papers - first development of an algorithm and second a paper on how physicians make treatment decisions.

Author’s Response: We have implemented several manuscript changes that hopefully will enable the reader to clearly identify the connection of the two ideas. A single manuscript that more coherently connects these two ideas would be preferable (and more likely to be published) than two disparate manuscripts. The understanding we gained of clinicians’ mental models, especially diagnostic errors, enabled us to develop a novel algorithm. The
algorithm developed to address these errors is critical for developing interventions to
improve the accuracy and reliability of CA-UTI diagnoses.
Specific changes incorporated for this item include:
- Significantly rewrote the Introduction section
- Rewrote portions of the Abstract
- Rewrote the conclusion to connect the two points more clearly

3. **Specific Comments - Abstract:** If reading the abstract alone, without the paper, it would be
very difficult to discern the take away message. The background is about trying to improve
treatment of UTI, but the results focus mostly on reliability of the algorithm and mental
models for treatment. Conclusions don't seem to follow either background or results - the
first two sentences in conclusions don't appear to relate to anything above.

Author’s Response: We have made several changes in the Abstract that more explicitly
connects the objectives of the research to the methods, results and conclusions.

4. **Methods - Algorithm face validity and Inter-rater reliability:** Conclusions based on both of
these measures are very suspect. Both measures are based on extremely small numbers. I
am not certain that 4 residents, and NP and PA make a solid sample to test face validity of
the algorithm.

More worrisome is the use of 3 team members to test inter-rater reliability. These are
subjects who are designing the study. Using these people as subjects is scientifically
unsound. If authors want to include a section on inter-rater reliability, they must have a
larger numbers of unbiased subjects. Because of the flaws in design, there are no valid
conclusions that can be made about usefulness or effectiveness of the algorithm.

Author’s Response: We have clarified the text to better explain the purpose of the face
validity assessments. Face validity requires a lower threshold for confirmation and can be
established by a spectrum of non-expert clinicians (residents, nurse practitioners, acute and
extended care providers, etc.) who are expected to use the algorithm in routine settings with
actual cases. Regarding inter-rater reliability, we have modified the text in the methods
(pages 6, 10-11) and discussion (page 16) to clarify that our reliability assessments are
more to reflect measurements of improved diagnostic accuracy using the algorithm
compared to a criterion standard (clinical expert using the algorithm). We have also edited
the limitations (page 18) to better reflect the reviewer’s concern about small sample and
more modest comment about our findings regarding inter-rater reliability.

5. **Results:** Table 1: May be deleted.

Author’s Response: We have maintained Table 1 as we feel it is important to describe the
characteristics of the study participants and to clarify for the reader the different participants
in each phase

6. **Discussion - Limitations** are outlined; some are those I have included above, but I am not
certain, that they fully explain why limitations are acceptable.

Author’s Response: We have edited the limitations section to better describe the
significance of this work despite its limitations and to underscore the validity of the methods
used.
7. **Conclusions** - These conclusions don't follow methods or results. It would help to conclude key findings, and then make more general conclusions.

   Author’s Response: We have edited the conclusions section to better connect the introduction, results, and discussion.

8. **Minor Essential Revisions** - The *Introduction* is long and disjointed; again I believe because the authors are trying to merge two messages, first need for implementing evidence based UTI guidelines and then mental modeling. The long description of Representativeness appears out of place, since it never is referred to again in results or discussion.

   Author’s Response: We have rewritten the Introduction section of the paper, removing the discussion of Representativeness, as we agree with the reviewer that this description is not necessary given the problem we are addressing is clinicians’ improperly constructed (non-guideline compliant) mental models.

**Level of interest**: An article of importance in its field

**Quality of written English**: Acceptable

**Statistical review**: No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests**: None