Author's response to reviews

Title: Audit and Feedback and Clinical Practice Guideline Adherence: Making Feedback Actionable

Authors:

Sylvia J. Hysong (sylvia.hysong@med.va.gov)
Richard G. Best (rbest@satx.rr.com)
Jacqueline A. Pugh (pugh@uthscsa.edu)

Version: 2 Date: 17 April 2006

Author's response to reviews: see over
April 16, 2006

Brian Mittman, Ph.D.
Editor-in-Chief, Implementation Science
VA Greater Los Angeles Healthcare System
16111 Plummer St.
North Hills, CA 91343-2036

Dear Dr. Mittman:

Thank you very much for the many insightful comments you and the reviewers provided my colleagues and me with regards to manuscript # 9098232539155834, “Audit and Feedback and Clinical Practice Guideline Adherence: Making Feedback Actionable.”. We have carefully considered each of the comments and have incorporated most of them in the manuscript. Following is a detailed account of our responses to each of the reviewers’ comments which we hope will be found satisfactory. We look forward to seeing our manuscript in print.

If you have additional questions about the changes to the manuscript, or if I can be of any further assistance, please do not hesitate to contact me.

Sincerely,

Sylvia J. Hysong, Ph.D.
Houston Center for Quality of Care & Utilization Studies
Michael E. DeBakey VA Medical Center
2002 Holcombe Blvd.
Houston, TX 77030
713-794-8616
sylvia.hysong@med.va.gov
Responses to Reviewer 1: Ian Graham

Major Revisions

1. The section describing the study limitations is rather slim. The authors should consider discussing here or in the text the strategies they used to ensure study rigor - specifically what was done to minimize the threats to description and interpretation of the data? For example the audiotaping of the interviews and maintaining of field notes would have minimized threats to the description. Threats to interpretation may have been minimized for checking for representativeness of the data and coding categories, using two analysts with differing backgrounds, obtaining validation from participants, using debriefing methods or an audit trail- were any of these done? Another approach would be for the authors to explicitly discuss the issues of credibility, dependability, confirmability, and transferability of their findings.

Strategies to address study rigor are discussed throughout the paper. We have added new material throughout the text to more directly address questions of rigor. We have compiled both the old and new material here, along with page references to make them easier to evaluate. We used Miles & Huberman's (1994) standards for addressing credibility, dependability, confirmability, and transferability concerns as a guide for our discussions here.

**Dependability:**

- None of the interviewers in the research team were affiliated with any of the facilities in which interviews were conducted (new text added p.7); this guarded against any potential biases in interpreting the data differently due to having insider knowledge about a given facility. Although it can certainly be argued that insider knowledge is an important component of qualitative research, we wanted to avoid having inside information about one facility but not another; further we wanted to ensure consistency of protocol during data collection, hence the decision to have the research team conduct the interviews rather than recruiting local interviewers.

- A standardized interview and field note protocol was developed in advance, which included materials, instrumentation, and clear role and responsibility distributions between lead and assistant interviewers (readers are referred on p. 7 to the Appendix for protocol details).

- Interviewers were trained in interviewing and field note protocol in advance, specifically for this study (p. 7).

- Data were collected across a variety of sites that varied significantly in characteristics (see table below), so that they would be representative of the variety of facilities existing within VA. We did not include this table in the manuscript because, since these data are freely available to anyone within VA, it would not be difficult to deduce the identities of the facilities based on the data in the table.
Table 1. Facility Characteristics (Based on FY01 data)

<table>
<thead>
<tr>
<th>Fac no</th>
<th>Acad. Affil.</th>
<th>VISN (Geographic Region)</th>
<th>Unique Patients Qty.</th>
<th>%ile</th>
<th>Outpatient Visits Qty.</th>
<th>%ile</th>
<th>Facility Type</th>
<th># of PC clinics</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Y</td>
<td>16 - South Central</td>
<td>41,985 73</td>
<td>313,755 64</td>
<td>Tertiary Care</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 Y</td>
<td>15 - Heartland</td>
<td>29,964 52</td>
<td>216,425 42</td>
<td>GM&amp;S</td>
<td>19</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 Y</td>
<td>07 - Southeast</td>
<td>40,027 68</td>
<td>302,273 62</td>
<td>Tertiary Care</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 Y</td>
<td>15 - Heartland</td>
<td>20,124 26</td>
<td>135,317 15</td>
<td>GM&amp;S</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 Y</td>
<td>15 - Heartland</td>
<td>81,503 95</td>
<td>530,964 88</td>
<td>GM&amp;S</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 Y</td>
<td>07 - Southeast</td>
<td>13,837 6</td>
<td>128,845 11</td>
<td>Tertiary Care</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Automated word searches were conducted during open coding to efficiently, yet consistently capture construct of interest – these were then manually searched for relevance (p.8). Please see our response to your comment #5 for more detail on the automated word search strategy.

Confirability:

- Interviewers were from different backgrounds (p. 7), including medicine, nursing, organizational psychology, clinical psychology, and sociology; this provided different perspectives during interviews, field note compilation, and data analysis.

- Although a formal audit of the data was not conducted, justifications for coded passages were documented as they were being coded, as were assumptions, values, and biases throughout the process. (new text added on p. 9)

- Interviews were audio taped and transcribed for analysis (p.8).

- Interviewers were blind to the performance level of the facility in which they were conducting interviews (new text added, p. 7)

Credibility:

- The feedback characteristics emergent from the data are consistent with Feedback Intervention Theory (Kluger & DeNisi, 1996) and feedback research in both the organizational and health care literatures (p. 16)

- It was possible to discern a logical set of relationships among the emergent concepts in the model, which (see Figure 1 in the manuscript), which Miles & Huberman (1994) consider evidence of credibility of the concepts found.

- The emergent model was developed by this paper’s principal author, and independently reviewed by the study’s principal investigator for plausibility.

Transferability:

- The six sites varied significantly by size, geography, facility type (i.e., tertiary vs. general medicine & surgery), and primary care capabilities (see Table 1 above); this variation did not significantly differ between HPF and LPF. The presence of a pattern of feedback characteristics despite this variability in site characteristics supports the idea that this pattern may be transferable to other facilities (new text added p. 18).
The feedback characteristics emergent from the data are consistent with existing research and theory on feedback characteristics, which suggests that our model could be transferable not only to other VA clinics, but potentially to other outpatient settings as well (p. 16, new text added p. 18).

2. Given that qualitative studies are not designed to be generalizable (they are usually designed for internal validity not external validity to use quantitative terminology) why is this listed as a potential limitation (p17)?

Miles & Huberman (1994) conceptually equate “generalizability” with “transferability”, as did we for the purposes of this paper. Our intent was to note the limitations of transferability given the small number of sites used, given that VA sites are widely variable on many different dimensions. VA researchers commonly remark, “If you’ve seen one VA, you’ve seen one VA”. Thus, our intent was to caution the reader that these findings may or may not be as applicable in other VA settings. We have revised this portion of the limitations section to (a) discuss transferability rather than generalizability, (b) better reflect our intentions, and (c) more directly discuss the applicability of our findings to other settings.

3. In terms of the actual findings, I would be inclined to be slightly more tentative about the characteristic of customizability since there was no clear evidence this was present in the HPF- in deed in all three cases it was coded as insufficient evidence in Table 2. It might be better framed as ‘not’ customizing will reduce the effectiveness of AF but its presence may not be/is not essential for AF to be effective.

Good point. We have altered text in the results and discussion sections to soften the discussions about customizability, and emphasize the need for future research on this characteristic.

Minor Revisions

4. The purposeful sampling strategy is completely appropriate. The ranking of facilities based on CPG performance aggregate each facility’s performance on 20 indicators. Were there any major systematic differences within or between facilities on the 20 indicators that might raise concerns about aggregating them using the IRSUM? I presume not but stating this might be helpful.

You presume correctly. In looking at the rankings of the high and low performing facilities, high performers tended to rank consistently high across most disease conditions, and low performers tended to consistently rank low across most disease conditions, we thus felt comfortable aggregating the indicator rankings into an overall, IRSUM score. We have explicitly stated this in the text on pp. 5-6.

5. Identification of passages related to feedback was done automatically- why was not open coding done by the analysts as they read each transcript? What is the limitation (and advantages) of using an automatic search approach to identify relevant passages?

As noted in our response to your comment #8, this paper is a secondary analysis of existing transcripts. When we read the transcripts for their originally intended purpose, we did notice feedback as a recurring theme, and we also observed that EPRP reports were often associated with feedback; we also found that high performing facilities tended to place a greater value on feedback than low performing facilities (Hysong et al., 2005). These
observations formed the impetus of this study, and guided our search criteria for the automated text searches used during open coding.

The automatic word searches were performed as an initial step to cast a wide, but consistent net across all interview transcripts. As discussed by various sources (Ford et al., 2002; Weber, 1990; West, 2001), this strategy is commonly used in content analytic methods, and has the advantage of identifying passages extremely reliably and efficiently – computers don’t skip or overlook text. That said, this reliability is entirely dependent on providing valid search criteria. The search criteria we provided was very broad by design, so that we wouldn’t miss potentially relevant passages; however, this strategy has the disadvantage of yielding many false alarms (passages that meet the search criteria but are irrelevant to the concept in question). This is why we adopted a two-tiered approach. We identified potential passages with an automated search for reasons of reliability and efficiency, then we manually reviewed the results of the search (the search was designed to return the entire paragraph containing the search criteria in question) as is done more traditionally in open coding, to separate the hits from the false alarms. As discussed on p. 8, we selected only passages that directly related to the purpose of the study: passages had to specifically discuss feedback processes in relation to guideline adherence.

6. Re the coding process: Who did the coding? What were their backgrounds and disciplines and roles in the project? Were the analysts the interviewers? How many individuals were involved? Was there independent coding by two individuals? If so, how were disagreements resolved? How was dependability of the coding assessed or assured- ie. how were the codes verified?

All coding for this study was done by the lead author, who was also one of the interviewers. Several steps were taken to ensure the quality of the coding. First, code definitions were explicitly documented as soon as they emerged, thus creating a codebook that was continuously referred to throughout the coding process. Second, every passage assigned a particular code received written documentation justifying the code assignment. Third, coded passages were re-examined by the lead author to insure the code assignments were consistent with code definitions. Fourth, a sample of coded passages and the emergent model were independently reviewed by another investigator for concurrence. This is now explicitly discussed on page 10 of the manuscript.

7. Was a codebook of the concept labels and definitions created? If not, how was consistency of coding assured?

Yes. Please see our response to your comment #6 above.

8. As it does not appear that emergent concepts were followed-up in subsequent interviews (eg. the facilities were not asked about whether their approach was punitive or non-punitive), I am assuming that data gathering and analysis did not occur concurrently as is usually the case with a ground theory approach. Please clarify. If this did not occur then it could/should be listed as a limitation (this is quite justifiable if this was a secondary analysis of existing transcripts).

This is exactly the case. Data gathering and analysis did not occur concurrently, as an in-depth study of feedback was not the original intent of the data collection. This is now discussed in the limitations section of the manuscript.
In terms of format, I found the ‘findings’ section somewhat repetitive given the initial description of the 4 characteristics and then the reiteration by HPF and LPF. Perhaps slightly more detail and quotes could be used with the generic description of the 4 characteristics and then in the section comparing the two sets of facilities a table could be used to show the quotes related to each characteristic by HPF and LPF. Why is the order timeliness, individualization, customizability, punitiveness on p10 and then the rest of the paper discusses punitiveness before customizability?

Thank you for this excellent comment. Per your suggestion, we reorganized the results section so that each characteristic is discussed only once, and all that needs to be said about each characteristic appears together. Although we added no new quotes due to word limitations, we did reorganize the existing quotations so that they supported the appropriate findings.

The order of the characteristics on page 10 was an oversight – now that the first list is gone (due to the merging of the initial description and facility comparison sections), this is no longer an issue.

10. Define HPF and LPF in the abstract.
Done.

11. Page14, last line- as we move up the facility rankings from the lowest to the highest, fewer of the properties appear to be present’ – should this actually be that more properties appeared to be present as you move up the hierarchy?

That is correct. Thank you for bringing this to our attention. The text has been corrected to say that more properties are present as you move up the facility rankings.

12. Are the initials of respondents actual initials or pseudonyms? If they are actual initials, could they be recognized given their position is also noted?

All participant initials used in quotations are fictitious. This is now explicitly stated on p. 10 of the manuscript. (They are all initials of Food Network chefs – one of the authors, who shall remain nameless, has a weakness for this particular TV channel).

13. Add CPRS to abbreviation definitions
Done.

14. Figure 1. Each characteristic is presented in the positive light (timely, individualized, customized) except for the 3rd (punitive). Why is this not presented as non-punitive (as it is in table 2)?

Figure 1 has been revised to present punitiveness in a positive light, consistent with Table 2.

Responses to Reviewer 2: Brad Doebbeling

Major revisions
No major revisions to the manuscript were requested.

Minor revisions
1. Define HPF and LPF in the abstract.
Done.

2. Add word "employees at" or "providers and managers at" to description of sample of 6 VAMCs
   Done.

3. Clarify sentence in Methods section of abstract--is it "to practice guideline implementation" or to better understand practices involved in guideline implementation, or?
   The text has been modified to read: “they discussed strategies, facilitators, and barriers to implementing CPGs. We hope this is clearer.

4. Results section--high performing facilities or high performing individuals?
   Although we interviewed individuals, the subject of the interviews was what the facility was experiencing and doing to better implement CPGs. Additionally, the performance measures we used to classify facilities either as high or low performing were strictly facility level, not individual level data. Thus, unit of analysis in this study is the facility – discussion about high performing and low performing facilities is our intent.

5. Suggest splitting the conclusion into 2 sentences.
   Agreed. The conclusion is now two sentences.

6. Please add a conclusion for managers and policy makers about what they should do in response to these findings. Such as "Managers and policy makers need to ensure that their feedback of performance is actionable..."
   Both the implications and the conclusions sections have been edited to include suggestions for managers and policy makers.

7. Capitalize first word after RESULTS and CONCLUSION in the abstract.
   Done.

8. Replace & with 'and' on p. 4.
   Done.

9. Methods data analysis (p.7 ) refers to transcripts but data collection section refers to note taking.
   Interviews were audio recorded and then transcribed for analysis (see p. 8, just before data analysis section). In addition to these transcripts, interviewers took field notes which contained material such as salient themes in the interviews, informal observations about the facilities, and initial impressions of the facility’s “story”. Only the transcripts were coded, however, the field notes provided useful context that helped better interpret the material in the transcripts (see axial coding section, pp. 8-9). The following text has been added on p. 7, which we hope will clarify this issue: “Interviewers discussed their own observations after each interview, and compiled field notes for each facility based on these observations and discussions.”

10. Were tape recorders used and interviews transcribed?
    Yes (see our response to your comment #9 above).
11. Were more than 1 coder involved in coding the same sections to reach agreement on
the coding structure?

*Please see our response to reviewer 1’s comment #6 for information on this issue.*

12. The discussion could benefit from a discussion of how this paper’s findings add to or
build upon relevant literature. [Discretionary--For example, these findings are consistent
with a paper on barriers and facilitators to using computers in guideline implementation
we published recently (Lyons, S.S., Tripp-Reimer, T., Sorofman, B.A., Dewitt, J.E.,
BootsMiller, B.J., Vaughn, T.E., and Doebbeling, B.N. VA QUERI Informatics Paper:
Informatics technology for clinical guideline implementation: Perceptions of multiple

*The discussion section has been edited on p. 17 to include relevant literature, including the
Lyons et al. paper.*

13. Suggest adding a paragraph on policy and management implications of the study.

*Please see our response to your comment #6.*

**Discretionary Revisions (which the author can choose to ignore)**

14. Use ‘valence’ in place of ‘sign’ at bottom of p. 3

*We used the word “sign” because this is the term used in the organizational feedback
literature (feedback sign – did you get positive feedback or negative feedback). However, we
have changed it to “valence”, on the assumption that Implementation Science readers are
more likely to resonate with ‘valence’ than with ‘sign’.*

15. I was somewhat curious about the ‘difference’ in performance measures between the
HPF and LPF included in the study.

*Recall that our rankings reflect an aggregate of 20 different performance indicators, thus we
would advise caution in examining the absolute differences between the highest and lowest
performer on a single indicator. Nevertheless, differences between the highest and the
lowest facility on single indicators range from -20% to 56%, with a mean difference of 14%.
In only two out of the 20 indicators did a low performing facility outperform a high
performing facility. Looking at the facilities from a ranking perspective (which is what we
used to classify them as high or low), high performers tended to rank consistently high across
most disease conditions, and low performers tended to consistently rank low across most
disease conditions, which made classification into high and low groups straightforward.*

16. Would it be helpful to add they type of employees and more about the sample...in other
words were the CPGs ambulatory care prevention, chronic care, inpatient?

*The six guidelines used to measure performance (diabetes, depression, tobacco use
cessation, ischemic heart disease, cardiopulmonary disease, and hypertension, see p. 6)
reflected chronic conditions that would be treated in an outpatient setting; thus all the direct
care participants we interviewed (i.e., not the middle managers/support staff or facility
leadership) worked in outpatient settings. This is discussed on p. 6 of the manuscript;
further, text has been added on p. 5 to clarify that guidelines were for chronic conditions
customarily treated in outpatient settings,*
17. Consider adding a discussion of issues related to provider-level feedback (limited sample size, reliability problems, validity questions) versus work-group or clinic (may be more relevant to emphasis on improving the system, etc)

*Text has been added on page 18 to emphasize the importance of having good data (reliable and valid, based on appropriate sample sizes) to feedback to clinicians in order for the feedback properties to even become an issue.*

18. The introduction could also refer to the pay for performance movement as a major stimulus for audit and feedback There are research needs to improve methodologic standards, better understand how feedback is given influences organizational change, speed, sustainability.

*These data do not speak to pay for performance in any meaningful way. However, we agree that this is an important issue, and have mentioned it in the conclusion and future directions section of the text.*

19. Site selection (page 4) consider citation to original report describing methods.

*Done.*

20. Suggest adding a section on strengths of the study (for example Horton recommends Strengths and Limitations of the Study Study question Study design Data collection Analysis Interpretation).

*Although we agree that this would be a nice addition to the paper, word budget limitations precluded us from including such a section.*

21. Suggest adding more on "Interpretation and Implications in the Context of the Totality of Evidence Is there a systematic review to refer to? If not, could one be reasonably done here and now? What this study adds to the available evidence Effects on patient care and health policy Possible mechanisms" For example, what does this add to the recent Grimshaw review and what issues need further investigation

*Text has been added to the implications section (p. 17) suggesting that some of the mixed findings in the Grimshaw and Jamtvedt reviews could be due to the fact that feedback characteristics usually aren’t taken into consideration when researching audit & feedback.*

22. Consider potentially adding section on "Controversies Raised by This Study" using Horton’s recommendations on a structured discussion. Suggest expanding the next steps in terms of research.

*Additional next steps have now been added in both the future directions and other sections of the text, in response to earlier comments from both reviewers.*

References


