

Reviewer's report

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

Version: 1 **Date:** 20 February 2006

Reviewer: Ian Graham

Reviewer's report:

General

This paper provides an excellent example of making the most of a qualitative study by reanalyzing data initially collected on the meaning of practice guidelines, and barriers and facilitators to the use of guidelines to explore characteristics on the intervention of audit and feedback related to better guideline adherence. An important strength of the study, which is very much in keeping with the qualitative sampling strategy of maximum variation, is the purposefully selecting for comparison facilities that performed well on guideline adherence with those that did not to begin to identify characteristics differentiating to two types of facilities. A grounded theory approach is used although I would be more inclined to describe the design as a case study as 3 HPF are purposefully compared with 3 LPF and the unit of analysis is the facility.

The take-home message from the study is the greater need to carefully consider (and study) the characteristics or components of an effective audit and feedback approach and ensuring that these characteristics are actually delivered when this implementation intervention is being evaluated (ie the need to pay attention to the fidelity of the intervention in efficacy/effectiveness trials). The results of this paper offer suggestions for the sometimes differing findings of trials and systematic reviews that have evaluated the effectiveness of A&F.

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

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.

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)?

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- indeed 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.

Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

1. 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.
2. 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?
3. 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?
4. Was a codebook of the concept labels and definitions created? If not, how was consistency of coding assured?
5. 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).
6. 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?
7. define HPF and LPF in the abstract
8. p14, 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?
9. Are the initials of respondents actual initials or pseudonyms? If they are actual initials, could they be recognized given their position is also noted?
10. add CPRS to abbreviation definitions
11. 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)?

Discretionary Revisions (which the author can choose to ignore)

OTHER COMMENTS

A thought for another paper, the authors might consider trying to validate their model with the 9 facilities that had performance ratings between the top and bottom three facilities. If the model holds, there should be no consistent pattern related to the 4 characteristics of timely, individualized, non-punitive, customized with these middle 9 facilities. If there are patterns, then they may be able to modify the model to take these new findings into account.

What next?: Accept after minor essential revisions

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.