Reviewer's report

Title: The effect of provider- and workflow-focused strategies for guideline implementation on provider acceptance

Version: 1 Date: 18 March 2008

Reviewer: Martin Lee

Reviewer's report:

Major Compulsory Revisions

The article is an interesting one and the conclusions have an important effect. However, before accepting the article, there are a few analytical issues that I feel need to be addressed or expanded upon:

1) In the Methods section, "Sample": The authors refer to "power calculations", but do not indicate what these are other than to state that "8 physicians, 8 nurses and 4 PAs/NPs" were needed at each facility. The basis for this determination is important. Since a hierarchical model was used here, it is presumed that ICC(s) were assumed for this calculation. Where did these come? What was the primary calculation on which these sample size requirements based, i.e. what was the statistical goal of the study? Moreover, it is important to know whether the sample size goal per facility was achieved in the actual sample. Were there any consequences to the study if this was not the case? (It appears from the statistics in Table 3, that the goals for provider distribution were not quite achieved.)

2) In the Methods section, "Surveys": It is noted that if more than one quality manager survey was completed for a facility, then the average number of implementation strategies was computed at that facility for each CPG. How often did this happen? Was there a large disparity at times when this occurred, and, if so, was there any concern about averaging the results under these circumstances?

3) In the Methods section, "Surveys": The primary outcome of provider acceptance consisted of an average of six items assessing provider acceptance of the guidelines. It is noted at the end of the paragraph that the score ranges from 1 to 5. It is not clear to me whether this is a score for each item, which are then averaged or this is the overall score for acceptance. If the latter, how is the average made to concur with an ordinal scale of 1 to 5? Is it a rounding process? Why not just use the median of the values, which is analytically more appropriate? It should be of some concern that the outcome variable in the MLM models is a simple ordinality given the assumptions required of that model for inferential purposes (more on this below). I also am interested to know whether all respondents answered all 6 items for each of the three guidelines? If there were incomplete responses for some providers, then it might be appropriate to include the number of items responded to as a covariate (similar to a time
covariate in a Poisson regression model).

4) In the Methods section, "Statistical Analyses": I agree that the MLM is a reasonable approach here. Can the authors expand on the particular model fitting strategy used in SAS? I am concerned about the potential violation of assumptions because of the ordinal nature of the outcome variable. An ordinal logistic regression model with adjustment for facility-level clustering might have been an alternative. Were the model fit statistics from the output considered and evaluated? If there were no issues here, it would be important to say so. Also, in including the three two-way interactions noted, were these specifically pre-selected or were all of the interactions between the covariates considered. Right now, the text just indicates that "2-way interaction terms were included in the model".

5) In the Results section: There is considerable missing data (as there always is in survey sampling). Moreover, the providers who did not report the provision of primary or specialty care in an ambulatory clinic on a weekly basis were excluded. Does this mean that they did not provide this information, or that they did and their work was not in this category? Please clarify and provide a sound basis for excluding another 15-20% of your sample. However, regardless of the circumstances, there is a significant missing value problem that needs to be addressed either through weighting of the responders based on available characteristics of the responders and non-responders (via a logistic regression model, for example) or by weighting to a target population. If the authors do not feel any weighting is necessary under these circumstances, then this needs to be justified and comments to this effect need to be included in the paper.

6) In the Results section: I do not understand the use of an average of surveys from a facility to create a facility-level response. What was this used for? I thought you were using each survey from a provider as your sample. Also, in looking the respondents from a given facility, it is noted that a "reasonable concordance" was achieved. What does "reasonable" mean in this context?

7) In the Results section: The graphs of the interactions in Figure 1 refer to "low" and "high" workflow and provider focused strategies. How were these two categories defined given that the variables in question were quantitative? If these were dichotomized in the model fitting, then this should be stated.

All in all, I think the paper explores a very interesting and useful subject, but I believe the analytical tools used require a little more elaboration.

Discretionary Revision

The authors use "N" to refer to sample size instead of "n". I like to see the former used for population size and the latter used for sample size.

What next?: Unable to decide on acceptance or rejection until the authors have responded to the major compulsory revisions
Level of interest: An article of importance in its field

Quality of written English: Acceptable

Statistical review: Yes, and I have assessed the statistics in my report.