

Author's response to reviews

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

Authors:

Mindy E Flanagan (meflanag@iupui.edu)
Rangaraj Ramanujam (ramanujr@purdue.edu)
Bradley N Doebbeling (bdoebbel@iupui.edu)

Version: 2 **Date:** 4 December 2008

Author's response to reviews: see over

Martin Eccles
Editor-in-Chief, **Implementation Science**
Newcastle University
United Kingdom

Brian Mittman, PhD
Editor-in-Chief, **Implementation Science**
VA Greater Los Angeles Healthcare System
Los Angeles, CA 91343

Dear Dr. Eccles and Dr. Mittman,
Please accept our revised manuscript, titled “The effect of provider- and workflow-focused strategies for guideline implementation on provider acceptance,” (MS: 2076539734181857) for consideration as a publication in Implementation Science.

We appreciate the reviewers’ thoughtful comments. In response, we have made substantial revisions to the manuscript throughout. Below, we provide a point by point response to the reviewers’ comments. For easier readability, we reproduce the reviewers’ comments and detail the corresponding changes that we have made in response (in bold).

The author responsible for all communication is Mindy Flanagan, Ph.D., HSR&D Center for Implementing Evidence-based Practice, Roudebush VAMC, 1481 W. 10th Street (11-H), Indianapolis, IN, 46202. My email address is meflanag@iupui.edu. My phone number is 816-746-5319; fax number is 317-988-3222. All authors have reviewed the response letter and revised manuscript and agree with its submission in this final form.

Thank you for your consideration.

Sincerely,

Mindy Flanagan, Ph.D.
Research Scientist
HSR&D Center for Implementing Evidence-based Practice
Roudebush VAMC

Reviewer: Nils Chaillet

This manuscript adheres to the relevant standards. However, it needs to be shortened. Maybe the figure 1 could be also removed.

Thanks for this suggestion. After careful consideration, we decided that retaining Figure 1 would be useful because we think it makes it easier for the reader to understand the form of the two-way interactions (compared to inferring it from the direction of the coefficients in the tables). Additionally, we think the figure illustrates the consistency in the observed interactions across three different sets of guidelines. That said, we remain open to suggestions for shortening the manuscript.

The background of the abstract needs to be revised. The objectives of this paper are not clear in this section (but are clear in the text). Research question need to be added. The abstract conclusion is unclear. Use the paper conclusion for the abstract could improve the understanding in regards to the goals of this paper.

We revised the abstract to clarify the objectives, research question, and the conclusion. The abstract now reads “Background: The effective implementation of clinical practice guidelines (CPGs) depends critically on the extent to which the strategies that are deployed for implementing the guidelines promote provider acceptance of CPGs. Such implementation strategies can be classified into two types based on whether they primarily target providers (e.g., academic detailing, grand rounds presentations) or the work context (e.g., computer reminders, modifications to forms). This study investigated the independent and joint effects of these two types of implementation strategies on provider acceptance of CPGs. Methods: Surveys were mailed to a national sample of providers (primary care physicians, physician assistants, nurses, and nurse practitioners) and quality managers selected from Veterans Affairs Medical Centers (VAMCs). A total of 2438 providers and 242 quality managers from 123 VAMCs participated. Survey items measured implementation strategies and provider acceptance (e.g., guideline-related knowledge, attitudes, and adherence) for three sets of CPGs—Chronic Obstructive Pulmonary Disease, Chronic Heart Failure, and Major Depressive Disorder. The relationships between implementation strategy types and provider acceptance were tested using multi-level analytic models. Results: For all three CPGs, provider acceptance increased with the number of implementation strategies of either type. Moreover, the number of workflow-focused strategies compensated (contributing more strongly to provider acceptance) when few provider-focused strategies were used. Conclusion: Provider acceptance of CPGs depends on the type of implementation strategies used. Implementation effectiveness can be improved by using both workflow-focused as well as provider-focused strategies.”

Reviewer: Martin Lee

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?

The original goal for this data collection (grant titled “Determinants of Clinical Guideline Implementation Effectiveness,” B. Doebbeling, PI) was to identify the implementation process factors that vary across VA facilities and the degree to which these factors are related to guideline compliance rates (with VA facilities divided into high and low guideline compliance groups). As such, the study was powered to detect such a difference between two groups of facilities.

The provider survey was instituted to validate, in a national sample, the structure and process factors identified as most critical to the level of provider compliance with CPGs. The analyses that we conducted should be considered as secondary data analysis with limited power as they do not represent the original goal of the study. However, we found several items to be of significance so we do not consider sample size a major limitation.

To explain the history of the project, we added on page 8: “At each facility, we sought to identify at least 8 physicians, 8 nurses, and 4 physician assistants (PA) and/or nurse practitioners (NP), if available. The primary goal of the parent project was to test organizational factors predicting guideline compliance. Rather than using individual provider data, these analyses were planned for facility level comparisons using aggregated data. Thus, the power calculations determined the number of responding facilities, and not the number of providers sampled at each facility. As a result, the number of providers sampled from each facility was intended to be adequate for understanding the implementation context. Additionally, the provider sample size per facility was based on pragmatic and budget concerns.”

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.)

Thanks for making this point. The average number of respondents per facility was about 15. The range for the number of respondents from each facility was from 5 to 33. In total, 18 facilities had fewer than 10 respondents. As indicated in the previous response, our analytic goals were secondary to the original study. We consider the number of providers included in this study to be quite large. Additionally, since we detected significant relationships in our multi-level models, we believe we have adequate power.

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?

For 42 facilities, one quality manager completed the survey. For 62 facilities, two quality managers completed the survey; and for 25 facilities, three quality managers returned the survey. To check that the facilities with more quality managers responding to the survey did not introduce systematic differences, we tested the effect of the number of quality manager respondents on the number of implementation strategies reported for the facility (which is an average value for those facilities with two or three quality manager respondents). This model was tested separately for COPD, CHF, and MDD. We did not find any significant effects of the number of quality managers responding on the total number of implementation strategies for a facility. The means for each guideline condition are listed in the table below by the number of quality managers completing the

survey. Additionally, the F values from the general linear models testing for differences among the three groups are included with the corresponding p-values.

On page 13, we added: “When multiple quality manager surveys from a facility were returned, the total number of implementation strategies reported was averaged to create a facility-level response. Using a general linear model for each condition, we tested the effect of the number of returned quality manager surveys (1, 2, or 3) on the total number of implementation strategies for that facility. No significant differences were detected.”

Guideline condition	Number of quality managers completing the survey			F (p-value)
	1 (n=42)	2 (n=62)	3 (n=25)	
COPD	4.24	3.80	4.72	2.05 (.13)
CHF	4.02	3.62	4.47	1.77 (.18)
MDD	4.24	3.45	4.04	2.07 (.13)

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).

Thanks for asking this clarifying point. The outcome variable, provider acceptance, was an average score of six items. Thus, it is a continuous variable. With a continuous outcome variable, the MLM approach was appropriate and does not violate the assumptions of the model.

We revised this section to read: “The response set for each item was a 5-point Likert-type scale that ranged from (1) “not at all” to (5) “very great”. The mean value of each provider’s responses to these six items with respect to each set of guidelines was used to measure the provider’s acceptance. A low score indicated low acceptance; and, a high score indicated high acceptance.” (page 10).

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).

Thanks for this recommendation. We created a variable to indicate the number of items (of the 6) for which responses were available to include in the outcome variable. The results are still the same with some slight changes in the parameter estimates (see Table 5). We added this variable in our Methods section. Specifically:

We added on page 11 that “since some of the providers did not respond to all six items about provider acceptance, the number of items that a provider responded to was included as a covariate.”

Also, on page 11, we added that “Lastly, we created a variable to indicate how many of the 6 items, which comprise the outcome variable, contained responses for each provider.”

In the Statistical Analyses section on page 12, we added “In the present MLM analyses, provider-level predictors included count of provider-focused and workflow-focused implementation strategies, general attitudes toward all CPGs, provider gender, provider tenure at the facility (in years), and number of items for the outcome variable that contained responses.”

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.

With the continuous outcome variable, the assumptions of the MLM approach are met. To test model fit, we calculated -2 Log Likelihood difference between the main effects only model and the full model (main effects and three interaction terms). For all three conditions, we found that the full model was an improvement over the main effects only model, demonstrated by a significant chi-square value. Additionally, in each case, the AIC decreased for the full model vs. the main effects only model.

We added this point on page 15: “To test whether the interaction terms improved each model, we compared the main effects only model to the full model (main effects and the three interaction terms). To conduct this comparison, we calculated the difference in -2 Log Likelihood values between the main effects only model and the full model for each condition. In all cases, the full model demonstrated a significant improvement.”

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".

Thanks for asking us to clarify this point. Our primary goal was to test hypotheses about specific interactions. Therefore, we included only the hypothesized interaction terms: 1) Number of facility-level implementation strategies X Number of provider-focused strategies; 2) Number of facility-level implementation strategies X Number of workflow-focused strategies; 3) Number of Provider focused strategies X Number of workflow focused strategies. Only the third of these interactions was significant.

We clarified this point on page 12 “Additionally, the three hypothesized 2-way interaction terms were included in the model (see Table 5).”

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 speciality 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.

This is an important point that you raise. We followed your suggestion to include all the

respondents in the analysis (Tables 3 and 4 were updated with descriptive statistics for entire sample). As shown in Table 5, although the parameter estimates changed slightly, our conclusions remain the same.

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.

We agree that the missing (or non-response) data issue needs to be addressed. To give more details on this, the table below includes the total number of respondents who did not respond to any of the 6 items used for computing each of the three outcome variables.

n=2438	COPD	CHF	MDD
Number of respondents not contributing data to outcome variable (% of total)	365 (15)	405 (16.6)	546 (22.4)

However, since the missing data is in the outcome variable and not the covariates, the missing data is less likely to create bias in our results. Additionally, as shown in the table below, the data is missing at random. That is, the data appears to be missing conditioned on observed data rather than missing data. Specifically, older respondents, female respondents, and those with a longer tenure at the facility were less likely to respond to the 6 items comprising the outcome variable. Additionally, respondents reporting fewer implementation strategies (of both types) were less likely to respond to the 6 items comprising the outcome variable. All of these covariates are included in the models. Hence, the multilevel models are considered valid.

	COPD	CHF	MDD
Provider age	*	*	ns
Provider gender	*	*	ns
Duration of employment at the facility	*	*	*
General attitudes toward CPGs	ns	ns	ns
Number of provider-focused implementation strategies	*	*	*
Number of workflow-focused implementation strategies	*	*	*

CPG=clinical practice guideline; ns=non-significant; * indicates significant ($p < .05$) difference between respondents not responding to any of the 6 items in the outcome variable and those who provided one or more responses.

On page 14, we added “Due to concerns with missing data in the outcome variables potentially creating bias in our results, we compared respondents and non-respondents. We identified a few differences between these two groups. Non-respondents were older, female, and had a longer tenure at the facility. Additionally, non-respondents reported fewer implementation strategies (of both types). The covariates representing these differences are included in all models. Hence, the results from the multilevel models are considered valid.”

Additionally, in the Discussion section (page 17), we added “Finally, the generalizability of the findings from this study to the target population may be limited due to the observed differences between respondents and non-respondents.”

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.

Thank you for pointing out this opportunity to clarify our analysis procedure. We only averaged responses from quality managers to create a facility-level covariate (total number of implementation strategies used). We did not average across provider responses from the same facility.

We clarified this point on pages 13: “A single quality manager survey was returned from 42 facilities. However, follow-up indicated that most represented a single institutional response, reached by consensus among those surveyed. When multiple quality manager surveys from a facility were returned, the total number of implementation strategies reported was averaged to create a facility-level response”.

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?

We removed this point due to its ambiguity. To indicate that the number of quality managers responding from a facility did not bias reports of implementation strategies used, we added on page 13: “When multiple quality manager surveys from a facility were returned, the total number of implementation strategies reported was averaged to create a facility-level response. Using a general linear model for each condition, we tested the effect of the number of returned quality manager surveys (1, 2, or 3) on the total number of implementation strategies for that facility. No significant differences were detected.”

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.

The low and high workflow and provider focused strategies were created based on 20% and 80% values from corresponding frequency distributions. This detail is added on page 14: “To plot the interactions, “low” and “high” workflow-focused and provider-focused was defined based on 20% and 80% values, respectively, of the corresponding frequency distribution.”

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.

We made this correction in the text (page 8), titles for Tables, and in Tables 3 and 4.