Reviewer’s report

Title: Dental practice satisfaction with preferred provider organizations

Version: 1 Date: 26 May 2007

Reviewer: Brent Shelton

Reviewer’s report:

General

-------------------------------------------------------------------

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

Review of “Dental Practice Satisfaction with Preferred Provider Organizations”

This article focuses on an examination of how satisfied dental practices are with their PPOs.

Major issues:

1.) The investigators ‘systematically’ select 6 PPOs for this analysis. But they fail to mention how these PPOs correspond to the larger population of PPOs they could have selected but possibly did not. I’m concerned about selection of the best or possibly the worst PPOs and as such these results would need to be tempered in regard to whom one is making applicability. The investigators seem to have chosen the 6 PPO plans that give them the largest sample size (N=4582). More comments/questions related to the selection of practices and PPOs for this paper are provided below.

2.) The way the investigators chose to handling the missing values for scales having only 1 item missing was to use single imputation by plugging in the mean based on values for other observations where the scale did not have to be imputed. This is referred to in the statistical literature as unconditional imputation using the mean. The investigators also indicate that this method was found to be reasonable in simulations and refer to a paper by Joe Schafer in an attempt to further justify this method. As indicated in Little and Rubin (their latest book edition on Missing Data), this is a poor strategy in analyzing datasets that are subject to missing data. The problem often is with bias but even if the authors can demonstrate (which I doubt they can) that they can get an unbiased estimate for the mean of that scale, what about it’s precision? This is the problem with single unconditional imputation. It always results in under-estimation of the standard errors of estimates. The investigators need to address the issue of under-estimation of the standard errors (which they did not address) and either a.) present a more compelling argument than a reference to Joe Schafer’s paper that analyses involving single unconditional imputation of scales will lead to unbiased estimation or b.) select a better method to address the missing items (such as multiple imputation for which there is readily available software, eg. PROC MI and MIANALYZE in SAS software).

The investigators also need to let the reader know how many scales had to undergo imputation and how many overall values were imputed amongst the 4508 subjects ultimately included in the analysis.

3.) The analysis was conducted using a linear model assuming, for sake of hypothesis testing (which was done to assess whether the beta coefficients were significant or not), that the errors were normally distributed. This is probably not the case given that there are only four possible levels for the outcome satisfaction variable. The authors present no regression diagnostics for these linear models. Were there any predicted values that were well below 1 or well above 4 (assuming the outcome is coded as 1 through 4)? Did residual analyses reveal any suggestion that an important covariate might have been left out, etc.?

A more appropriate model might be to treat the outcome as a categorical ordinal variable (4 outcome categories) and try to fit a proportional odds, equal adjacent odds or generalized logits model. How would the conclusions have differed had the investigators fit such a class of ‘categorical models’? The measure of association then becomes an odds ratio rather than a slope coefficient. There are assumptions with these models that also much be checked as well such as linearity in the logits for the measures of communication, compensation, and claims services.

4.) I am not as enthusiastic about these results as the investigators are because I believe that an adequate
adjustment for length of time in practice (for example) might have influenced some of the results the authors have observed. This may be true for some of the other possible covariates listed in table 1 as well. How this might have played out is a mystery since the authors did not attempt to adjust for such factors even though these factors are displayed in Table 1.

It appears that in the 2002 survey data, none of the practice characteristics displayed in Table 1 were collected. The authors indicate this would result in removal of nearly 1/3 of the observations from the subsequent analysis. My question is, if adjustments are made for these practice covariates (realizing that nearly 1/3 of the observations will be eliminated), do the results change? The authors don’t address this (other than to say that the covariates were not significantly related to the outcome in ‘preliminary’ analyses) but they need to.

If the results are disparate, then again, use of multiple imputation might be a reasonable approach to attempt to incorporate as much information as possible. If the outcome or one of the three covariates that are completely observed are related to the covariates not collected in 2002, the imputation process might very well lead to reasonable adjustment for those incomplete covariates.

At a minimum, the authors should re-run the analyses adjusting for as many potentially important covariates as possible.

Quite honestly, it would surprise me to know that length of time in practice was not associated with overall satisfaction especially since the practice has had time to potentially ‘weed out’ the PPOs they are most dis-satisfied with. I would also posit that the fewer the # of affiliated PPOs, the more satisfied the practice; especially amongst practices that have been in operation for some time.

5.) More details are needed to describe the sampling of practices as well as how the authors ended up with the 6 PPOs mentioned. The following two sentences are particularly confusing:

Quota sampling was used to allocate practices to dental PPOs with which they were affiliated. The first 100 dental practices participating in the survey were asked to rate their experience with MetLife; the next 50 practices were ‘randomly’ assigned to rate one of the PPOs to which they had submitted claims in the past 12 months.

This description of how the practices were selected for inclusion amongst the 4582 is very confusing the way it is currently described. It needs to be made clearer. For instance, how is the practice ‘randomly’ assigned to rate one of the PPOs? It seems to me the six PPOs had to be known in advance of the calls to the practices. Was a PPO randomly chosen for the practice being surveyed? If so, I would think that this would lead to far more than the 6 chosen in the final analysis. Again, there is a distinct dis-connect in the description of how we go from surveying practices to getting only practices from these 6 PPOs. If I don’t understand it, then other readers might have the same problem.

Were “N” total practices telephoned, and of these “N”, did you just choose the top 6 PPOs that happened to be connected to 4582 interviews? If so, how big was “N”? (ie. how many were in PPOs characterized by a smaller # of practices interviewed, i.e. < 75?)

It seems that a list of practices were provided to the authors by Met-Life and practices were then called and asked to name other PPO providers they were affiliated with. If, as the paper states, the practice was ‘randomly assigned’ to one of the PPOs (ie. I think what the authors mean is that one of the listed/mentioned PPOs was randomly chosen to be evaluated by that practice, no?), then I would expect that far more than just 6 PPOs were represented amongst ALL OF THE PRACTICES that were called during this 2 year period.

As far as sampling goes, it would be helpful to know:

a.) how many practices in total from 2002-2004 were called for an interview.
b.) How many practices refused?
c.) How many PPOs are represented (even if it only represents 1 practice) in all of the practices called during 2002-2004?

6.) How were the 3 scales for compensation, communication, and claims services constructed? I think the authors are just summing the individual items within each domain but this should be clearly stated.

7.) Are the correlations presented based on Pearson, Spearman, or polyserial correlations? Polyserial correlations may be more appropriate when an ordinal outcome is involved such as is the case with Overall satisfaction.
Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

Discretionary Revisions (which the author can choose to ignore)

What next?: Unable to decide on acceptance or rejection until the authors have responded to the major compulsory revisions

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Acceptable

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

Declaration of competing interests:

I declare that I have no competing interests.

The editorial office should also know that I tried twice to send the above review via email and evidently you did not get it either time. First time sent was on May 22. I know that you have received one email from me in the past (the email that said I would review the article). I don't know why you didn't get the latest 2 emails. Could you please check this and get back with me via my email address? It's frustrating to get a review in on time only to have the editorial office send an email saying it was not on time...