Reviewer's report

Title: The discriminative power of patient experience surveys

Version: 2 Date: 7 February 2011

Reviewer: Jeannie Haggerty

Reviewer's report:

This would be a welcome report on an important area for many health services researchers, and I think it is important that this type of work be published. However, there are a variety of issues of reporting that need to be improved and two major issues that I think limit the scientific contribution of this paper. Overall, I found the paper quite difficult to apprehend, even though I have used multi-level analysis and thought I have a basic familiarity with the concepts.

- Major Compulsory Revisions

1. It is critical that the authors refer in a consistent way to the different types and levels of variance. The variance at the provider level is the between physician variance; the variance at level 1 or individual is a combination of between-individual variance PLUS random error. Multilevel modeling only allows partitioning between levels, but is does not allow further partitioning at individual level. The random error will contain variance from a variety of sources, including variance at the provider level (within-physician variance) and error arising from the instrument. The article does not communicate this nuance, leading the authors to conclude erroneously (I believe) that ‘patient experiences are fare more dependent on individual difference that on differences between providers’. I agree that the there appears to be limited capacity to detect between-provider differences, but I think that their interpretation is not entirely correct.

2. In the calculation of the number of participants to be approached the authors adjust the required sample size to achieve a given reliability by the expected response rate (presumably they divide). It seems that they are inappropriately mixing issues of statistical power (sample size) with that of validity (response rate). Low response rate raises concerns about systematic errors related to selection, and they cannot be corrected simply by approaching more people. Issues of statistical power assume representativeness, and this goes beyond simple numbers.

3. I think it is very important for the authors to operationally define what they mean by ‘discriminatory power’.

4. Methods, Participants paragraph 1: The description of the Dillman method is faulty and the reference ascribed to it does not match. As someone who conducts surveys and applies the Dillman method, I was consequently uncertain about how exactly the data was collected. Nonetheless, the response rates are
relatively good and within acceptable range.

5. Methods, Selection of patient experiences: It would be helpful to explain briefly why different communication items were asked to different patient groups. The idea of composite scores needs to be better explained. It is unclear how these relate to what is actually reported; do these refer to global rating and recommendation to others?

6. Methods, Data Analysis paragraph1:

Sentence3: ICC is the proportion of total variance that is attributed to between-provider differences.

7. The phrase ‘…range in which 95% of health care providers are expected to occur…’ is not understandable. I think here it is a problem of syntax, but the conceptual idea is not clear either. By ‘average’ I assume they mean the overall mean; by SD do they mean the square root of the between-physician variance (tau).

8. The notion of reliability absolutely needs to be explained better. It is very confusing.

9. The authors must reiterate in the text and table 3 that the estimated sample sizes they calculate are ‘per provider’.

10. The discussion must also talk about and discuss the various sources of variation that can contribute to total variance and especially at the individual level: within physician variance, measurement error, lack of internal consistency within the instrument. This seems especially important given the problems of precision that seem inherent with a four-point response scale or a three-point response scale for change in physical function.

11. Discussion, paragraph 4: I agree that we will really only be able to come up with robust ICC measures when we have an independent estimate of the true between-provider differences and consequently its variance.

- Minor Essential Revisions

1. I am presuming that this is a self administered survey. It would be helpful to state this.

2. I would appreciate having the titles of the tables presented with each table.

3. I am presuming that provider=doctor, but it would be good to clarify since the word can also be used to refer to an organizational delivery system.

- Discretionary Revisions

1. Table 1 would read better if it were transposed, so that columns were the different patient groups.

2. I would strongly suggest not calculating ICC when the between-provider
variance is not significantly different from 0.

Answers to questions:

1. Is the question posed by the authors well defined?
The objectives are slightly different as stated in the abstract and just before the Methods section. The term ‘discretionary power’ needs to be defined operationally.
The introduction is generally well written and clear.

2. Are the methods appropriate and well described?
I believe they are appropriate but especially the data analysis section needs to be better explicated.

3. Are the data sound?
As far as I can tell, yes.

4. Does the manuscript adhere to the relevant standards for reporting and data deposition?

5. Are the discussion and conclusions well balanced and adequately supported by the data?
I think that the interpretation does not necessarily reflect the data because the between-individual differences are overstated.
The discussion does not adequately address other sources of error and variance at the individual level.

6. Are limitations of the work clearly stated?
Not really addressed.

7. Do the authors clearly acknowledge any work upon which they are building, both published and unpublished?
Somewhat.

8. Do the title and abstract accurately convey what has been found?
It could be improved to indicate that the issue is the number of patients per provider as opposed to the precision and reliability of the instruments themselves.

9. Is the writing acceptable?
Generally, yes. I think it would just be helpful to be more pedagogical for the general readership.

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

Statistical review: Yes, but I do not feel adequately qualified to assess the statistics.

Declaration of competing interests:

'I declare that I have no competing interests'