Author's response to reviews

Title: The discriminative power of patient experience surveys

Authors:

Dolf de Boer (d.deboer@nivel.nl)
Diana Delnoij (d.delnoij@centrumklantenvragendzorg.nl)
Jany Rademakers (j.rademakers@nivel.nl)

Version: 3 Date: 17 May 2011

Author's response to reviews: see over
Dear Dr. Norton,

Editor-in-Chief
BMC Health Services Research

Utrecht, 17th of May, 2011

Dear doctor Norton,

Please find enclosed the revision of our manuscript entitled ‘The discriminative power of patient experience surveys’, which we hope will be suitable for publication in the journal BMC Health Services Research.

The paper reports on the extent to which patient experience surveys are able to detect differences between health care providers. This is currently an important issue since in a number of countries, patient experiences are used to measure provider performance. The paper presents data for various patient groups using corresponding measures of patient experiences and identical methods for data analyses. A number of previous papers have reported on this subject, but as far as the authors are aware, a comprehensive overview in different patient groups such as that provided by the present paper is lacking. We conclude that the differences between providers that are displayed by patient experiences vary between measures and between patient groups. Often, substantial sample sizes are required to detect these differences reliably. The latter particularly applies to measures of experienced change in physical ability following treatment.

Future projects that seek to develop patient experience surveys, may be informed by the present data on the discriminative power of such surveys in a variety of settings. These illustrations may guide expectations on the discriminative power of the survey under development, and, may help choose the unit of analyses at which providers are compared such that the expected number of respondents required per unit of analyses may be achieved.
Below we provide a point by point response to the reviewers’ comments. Changes in the text are marked by the usage of blue text. We thank you for the opportunity to resubmit our work and believe the manuscript has been strengthened by the reviewing process.

We look forward to hear from you,

Dolf de Boer
Diana Delnoij
Jany Rademakers

**Reviewer 1**

*General comments*

The reviewer states that ‘It is useful to have data on the unit level reliability of survey assessments published’ and that the manuscript is ‘…generally well written and interesting’. We thank the reviewer for these positive comments.

*Specific comments*

1. The reviewer states that managed competition has a very specific meaning and that perhaps we just meant competition. Indeed, the case we make in the introduction is not limited to managed competition and accordingly, we removed the word ‘managed’ (see page 3, 1st paragraph).

2. The reviewer notes that data on the performance of health care providers may still be used to facilitate consumer choice, even if that does not generate competition. We now mention the use of performance data to facilitate consumer choice in the 1st paragraph of the Background (see page 3, 1st paragraph).

3. The reviewer states that assessing individual differences within institutions is also important, even if there are no differences between institutions. The reviewer does not request we go into this issue but suggests we may be a little bit more cautious about the value of inter-unit discrimination. We agree there are useful applications for patient-experience data other than assessing inter-unit differences and we now specifically limit statements regarding assessing inter-unit differences to the context of
4. The reviewer suggests several papers that may be relevant for the introduction. We included nearly all of these suggestions in the Background section (see page 5, 2\textsuperscript{nd} paragraph) and we also added a reference to the discussion section (see page 15, 1\textsuperscript{st} paragraph, last reference).

5. The reviewer recommends we have the manuscript reviewed by an English speaking copy-editor. We have had the manuscript edited by a native English speaker which indeed resulted in a number of minor corrections. As these changes include many small adjustments of an arbitrary nature, these were not marked in the text. We do have a version of the manuscript where changes of the native English speaker are marked with track changes though, and we would be more than happy to send this version if the reviewer or editor would like to see this.

6. The reviewer notes that: ‘Cronbach’s alfa’s’ should be: ‘Cronbach’s coefficient alpha’. This is now changed (see page 7, 3\textsuperscript{rd} paragraph and page 8, 1\textsuperscript{st} paragraph).

7. The reviewer suggests we include footnote 1 in the main text. The footnote is now included in the main text (see page 4, 1\textsuperscript{st} paragraph).

8. The reviewer expresses a concern regarding the analyses of what we called ‘experienced change in physical functioning’. The reviewer notes that assessing differences regarding change in physical functioning is more complicated and requires another type of case mix adjustment than the other measures we report. We agree there are problems with the present strategy for measuring experienced change in physical functioning. However, despite the existing literature on measuring experienced change in physical functioning, as far as we are aware, the use of these measures to compare the performance of health care providers is particularly limited. Moreover, we presented the issue of using experienced change in physical functioning to compare hospitals to the Dutch tranch of the International Society for Quality of Life Research and the issue appeared remarkably novel for the audience. As such, we believe it is useful to present the available data, to stress the limitations of these data and to address the challenges for further research. This would stress the need for further development and may prevent that other researchers would pursue the present strategy, which is not unlikely given that this strategy is appealing from a practical point of view. In this context, we would also like to point out that we dedicated a paragraph of the discussion section to address the limitations of our measure of change.
in physical functioning and the need for further development (see page 13, 2\textsuperscript{nd} paragraph). In addition, we now also explicitly address the issue of case-mix adjustment for baseline health status in this paragraph (see page 13, 2\textsuperscript{nd} paragraph).

9. The reviewer requests that we state the number of providers and whether samples varied between providers. This is now addressed in the Methods section (see page 7, 1\textsuperscript{st} paragraph). Further, the reviewer requests that we make explicit what we mean by provider. This is now stated in the data analyses paragraph of the Methods section (see page 8, 2\textsuperscript{nd} paragraph).

10. The reviewer notes we report ICCs such as ‘0.1%’, while the ICC is a coefficient that ranges from 0 to 1. The ICC can be interpreted as the proportion of variance that is explained by the level of health care providers and we intended to present this proportion as a percentage. However, from other literature, it appears that the ICC is more commonly presented as a coefficient and accordingly, we now present the ICC as a coefficient (changes occur throughout the manuscript and in Table 3).

11. The reviewer requests we note that reliability depends partially on the heterogeneity of the sample of providers. This is now addressed in the Discussion section (see page 14, last paragraph).

\textbf{Reviewer 2}

\textit{General comments}

The reviewer provides a very positive evaluation of the manuscript, stating that the paper ‘….should be published, and it may be published without any major revisions’. We thank the reviewer for this assessment.

\textit{Specific comments}

1. The reviewer suggests we remove the word ‘slightly’ from the sentence ‘In this context, it should be acknowledged that measures of changes in physical functioning are successfully used to compare the effects of various health care interventions, albeit in a slightly different format using a pre- and post measurement’. The word slightly is now removed (see page 13, 2\textsuperscript{nd} paragraph).

2. The reviewer notes that it may be useful if we mention ‘the proportion of providers in each study which were statistically significant from others with the methods and
confidence intervals used’. We have considered this possibility extensively and concluded that this would be very difficult to do in an appropriate manner given the available data and the set-up of the manuscript. The reason is that the number of observations varies between studies, between providers within studies due to nonresponse and between measures within provides due to item-nonrespons. Since the number of observations is essential for the significance of differences, we believe that data presented along the lines suggested would be very difficult to interpret. For one patient group this could be solved by presenting differences, p-values and the number of observations for each provider, but with four patient groups and four different measures this would require more space than we think is appropriate.

**Reviewer 3**

**General comments**

The reviewer states that the paper ‘...would be a welcome report on an important area for many health services researchers....’ and the reviewer thinks that ‘.....it is important that this type of work be published’. The reviewer also provides a number of comments that need to be addressed in this revision.

**Major comments**

1. The reviewer notes that the variance at the individual level not only reflects individual differences but also includes variance from a number of other sources including error arising from the instrument or within physician variance. We agree that this may be the case and accordingly we rephrased two sentences in the article that lacked nuance in this respect (see page 4, last paragraph and page 12, 1st paragraph). Further we now specifically address this issue in the discussion section (see page 15, 1st paragraph).
2. The reviewer provides a comment on the calculation of the required number of patients to approach in order to achieve a given reliability. In particular, the reviewer is concerned that our approach may suggest that a low response rate may be compensated by approaching more people. In our opinion, a low response rate in the present context yields two methodological issues of which one may be addressed by approaching more people:
a. A low response rate poses a problem regarding the extent to which data from respondents can be generalized to the population from which they were drawn. This cannot be solved by approaching more people.

b. A low response rate also means that when approaching a given number of people, there may be insufficient data available for analyses. On the assumption that causes for nonresponse are broadly similar between providers and/or that possible response bias may be addressed through casemix adjustment, one may still be interested to compare the experiences from respondents between providers. In this context, we believe that it may be useful to adjust the number of people to approach to ensure there are sufficient observations per provider.

These issues are now addressed in the Discussion (see page 14, 3rd paragraph).

3. The reviewer recommends we operationally define what we mean by discriminatory power. This is now made explicit in the Methods (see page 8, last paragraph).

4. The reviewer states that our description of the Dillman method is incorrect. Since it is not entirely clear to us what exactly was incorrect, we now no longer refer to the Dillman method. In addition we provide more detail about the mailing procedure (see page 7, 1st paragraph).

5. The reviewer requests we explain why surveys differ in the items regarding communication. This is now explained in the method section (see page 7, 3rd paragraph).

6. The reviewer provides what we interpreted as a suggestion to rephrase a sentence in the method section. This suggestion has been integrated in the data analyses paragraph (see page 8, last paragraph).

7. The reviewer states that it is not clear what we mean by ‘…the range in which 95% of health care providers are expected to occur…’. In particular the reviewer would like us to clarify what we mean by the average across providers and how we calculated the SD. We rephrased the relevant sentences to that extent (see page 9, 1st paragraph). In addition, we noted that the way we referred to the 95% expected range in the table in a manner that was inconsistent with the text. We now also use the term ‘95% expected range of provider scores’ in the table (see Table 3).

8. The reviewer suggests we explain the reliability in more detail. We agree this may be clarified in more detail and we now provide a more detailed explanation in the Methods section (see page 9, 1st paragraph). In addition, some of the comments
regarding reliability have led us to believe that maybe we should rephrase some of the sentences regarding reliability slightly, particularly those in which we mention ‘reliable comparisons’. Essentially, the reliability reported concerns the ‘reliability of the provider scores’ or ‘reliability of estimates at the provider level’. It seems obvious that when estimates of provider scores are unreliable, comparisons between providers become unreliable. However, since most related literature does not really mention ‘reliable comparisons’, but rather speaks about reliability at the provider level, we have rephrased the relevant sentences (changes related to this issue occur throughout the manuscript).

9. The reviewer request we reiterate in the text and the table that the required sample sizes reported are per provider. This now reiterated (see Table 3 and page 11, 1st paragraph).

10. The reviewer states that the discussion should address various sources of variance at the individual level including measurement error, lack of internal consistency etc. This issue was already resolved in response to the first comment of the reviewer (see page 15, 1st paragraph).

11. The reviewer agrees that an independent estimate of the true between-provider variance is essential in the present context; this comment did not appear to require a response.

Minor comments

1. The reviewer suggests we state whether the surveys used are self-administered. This is now stated in the Methods section (see page 7, 1st paragraph).

2. The reviewer prefers tabel headings to be presented with the table. Table headings are now presented with the tables (see table 1, table 2 and table 3).

3. The reviewer requests we specify whether the health care providers in our models are individual doctors or an organizational delivery system. The latter is the case and this is now specified in the Methods section (see page 8, 2nd paragraph).

Discretionary revisions

1. The reviewer notes that table 1 may be easier to read if it were transposed. Although we would be willing to do this, we prefer the present layout of the table as this layout is consistent with Table 3: in both tables variables are columns and patient groups are rows. As such, we kept the layout as it was for the present revision.
2. The reviewer suggests we do not report ICCs in cases where the between-provider variance is not significant. We agree one should be careful with subsequent analyses based on a multi-level model if the variance on the level of providers is not significant. However, in the present context, the ICC is used to illustrate the discriminatory power of particular measures and we believe we should show that ICCs are particularly small when the variance on the level of providers is not significant. In addition the significance of variances on the level of providers is specifically mentioned in both the table and the text in the results section (see 10, 2nd paragraph).