Major Compulsory Revisions

1. Throughout, the manuscript treats the particular global rating question used in the study as generalizable to all global rating questions. But this generalisation is not justified in the manuscript and it is not obvious that the same results would be obtained with a different choice of global rating scale. Whilst the ‘strengths and limitations’ section of the discussion briefly acknowledges that the findings may not be generalizable to “other survey data”, it is important to be specific about the exact question used and the implications of this. Some of the comments below relate to specific instances where this issue arises, but it should be addressed throughout and the authors should comment on the degree to which findings can or cannot be generalised. This is important for the interpretation of the findings and their implications.

Minor Essential Revisions

2. Introduction, para 4: "In many surveys, patients are asked to rate the overall quality of the healthcare provider e.g. on a scale from 0 to 10." This para also references findings from research with a specific global rating question. However, there are other global rating questions with different wordings used in other settings - and it is not clear that they have the same characteristics (eg that they "largely represent patients' experiences with the process of care"). It would be clearer to state the exact question wording being referenced in this para.

3. Introduction, para 6 - composite patient experience scores have been used extensively in the UK since around 2005. But it would be reasonable to say that there is limited peer reviewed evidence on the statistical properties of these.

4. Methods, final para before 'Data analyses' - “The Global Rating of quality consisted of a single question”. Again, the specific question should be presented.

5. Discussion, para 4 – “any ranking of providers should preferably be done using composite scores rather than global ratings”. The implication is that the poor performance of the global rating is responsible for the differences in ranks – but I wonder whether an alternative explanation might be that providers close to the median have very small absolute differences on either global rating or composite scores. In other words, could the differences in ranking be explained by rankings making somewhat arbitrary distinctions between very similarly performing
providers?

6. Discussion – There is limited coverage of the impact of nonresponse on different composite strategies, but it would be useful to reflect on this. Particularly if different questions have substantially different means, then systematic differences in the proportions of missing data by provider can potentially be a source of bias. For example, if a composite consists of three questions to be averaged, two with a very high and one a very low mean score, then a provider with a high proportion of missing data for the one item with a very low score may have positively biased composite scores. To what extent is this likely to be an issue with the approaches described, and how can it be addressed?

Discretionary Revisions (minor issues not for publication)

7. Abstract, Background, third sentence - this would be much clearer if the phrase "and compared them to the global rating" was deleted.

8. Abstract, Methods - this would be much clearer if the last sentence ("Data from...") was moved to the start of this section.

9. Methods, Data selection, first para - typo - "98% of respondents was 60 years of age or older" - should be "were"

10. Results, second para – typo – “results were give a weighting of 0.80” – should be “given”

11. Results – missing close parenthesis “)” after ref [27]

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

Quality of written English: Needs some language corrections before being published

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

Declaration of competing interests:

I employed by the Picker Institute, a not-for-profit organisation that is partly funded via conducting surveys of peoples' experiences of health and social care. Neither I nor my organisation have a direct involvement in any of the survey collections or instruments described in the manuscript.