Reviewer's report

Title: Validity of Functional Status Decline as a Measure of Adverse Events in Home Health Care: an observational study

Version: 1 Date: 31 July 2006

Reviewer: Gideon Caplan

Reviewer's report:

General
This article deals primarily with statistical analysis and therefore is quite technical in nature.

This article attempts to answer an interesting and important question regarding whether industry databases may be used to monitor quality of care. While being from outside North America means that I am not familiar with many of the nuances, it appears to be a reasonably new and well-defined question.

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

This is a highly technical article, and I am rather simple, so I may have missed the point. But unfortunately there is, in my perception, a major problem with the methods. Validity, in my book, is the degree to which a test measures what it is supposed to measure. That is, a comparison between a test and a gold standard. I may have completely misunderstood, but I cannot see where that is done. One way to describe this article is, as the authors have done, an evaluation of test-retest reliability, but that is generally taken to mean that the test is applied twice to the same sample. Again, I may have misunderstood, but I cannot see where it is clearly stated that the two sets of patient episodes are necessarily for the same patients. Furthermore, while this could be done with the ADL score (ie comparing an ADL score twice to assess test-retest reliability) I would need to be convinced that measuring change in a score during two different periods can be similarly assessed for test-retest reliability. This presents considerable conceptual difficulties.

So how I would best describe the article is, assessing the indices responsiveness to change. I agree entirely with the idea that ADL change is a vital measure for assessing frail older patients. But perhaps one should compare it to something else to determine whether it is measuring quality, as the authors state. For example, how does it relate to the other measures of adverse events.

From my statistical point of view, reducing the range of data available for analysis is wasteful. For example, when you have a continuous variable, and you compress it to quartiles, you lose a lot of data. Therefore, if I were doing the analysis, I would use the whole ADL index, rather than only one, two or three ADLs. But that is the authors choice.

Although the data appear sound, this is an article which would strongly benefit from some high level statistical input, for the reasons stated above.

But I do not believe, for the reasons stated above, that the authors have established validity, so the conclusions are inappropriate.

Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

The writing is clear enough. On p.6, last paragraph, I would like to see substantial defined. I assumed this meant two points, but from the discussion, it may possibly have been three points.

I believe there may be a significant typo at the top of page 8, in line two, two should be three.

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: Yes

Declaration of competing interests:

I declare that I have no competing interests.