Reviewer’s report

Title: "It's hard to tell". The challenges of scoring patients on standardised outcome measures by multidisciplinary teams: a case study of neurorehabilitation

Version: 1 Date: 7 August 2008

Reviewer: Dan Bilsker

Reviewer’s report:

This well-written study examines a clearly-defined question, namely what are the processes by which a multidisciplinary healthcare team implements a routine outcome measure. This is a very important question because outcome rating of this kind has been proposed, indeed mandated in certain jurisdictions, to become a standard part of health care. It has been seen as contributing to key decisions regarding program planning, service continuation, allocation of cases to clinicians and even level of remuneration. By deepening our understanding of the processes which underlie this kind of outcome measurement, we are able to make better decisions regarding the appropriateness of using such data to make these decisions.

This study uses qualitative research methods that are appropriate to the question and well described. The manner in which data are presented is suitable for the approach and accurately communicates the findings. The authors do not overstate their findings, acknowledge limitations of generalizability and draw out the implications of their data in a useful manner. They correctly emphasize that attributes like reliability and validity are not inherent to a specific measure, but rather are attributes of a measurement context which includes the scale itself, the rater's training and degree of access to the patient, as well as any situational contingencies that may tend to bias ratings (for example, if the rater were receiving a reward for rating in a specific direction).

In a research situation, measurement context is carefully controlled. Raters are trained according to a standard protocol. They are given adequate access to the patient for the purpose of completing the rating scale -- but not given ongoing access to the patient, which would complicate the rating process. They receive no reward or penalty for rating in one direction and are thus "objective". Given this carefully defined and rigidly-controlled measurement context, one can then meaningfully speak of the reliability and validity of the rating scale itself. However, the clinical context examined in this study subverts some or all of these requirements. Raters vary considerably in professional training in a way that influences rating behaviour. Raters sometimes have limited access to the patient. Conversely, raters often have ongoing access to the patient that causes the rating process to become far more complex, e.g., as observed in this study, clinical raters may retrospectively adjust baseline ratings as they understand the patient better. Finally, raters in clinical situations may be subject to fluctuating rewards for differential rating -- for example, a clinician providing a form of
treatment whose efficacy is in question might be prone to inflated outcome ratings. The authors correctly stress that dimensions like reliability and validity should be considered as characteristics of the entire measurement context.

Discretionary Revision:

1. In the Conclusions section, the authors do not go quite far enough -- they point out that rating procedures remain clinically useful despite failing to meet standard criteria of reliability and validity, but should also point out that failing to satisfy these measurement criteria substantially limit how these data can be used. For example, one would not want to base program funding decisions on the data derived from the rating procedures observed in this study. (Especially if the program staff were aware that funding would be influenced by their ratings!)

No other revisions requested

**Level of interest:** An article of outstanding merit and interest in its field

**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.