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: 25 July 2006

Reviewer: katherine berg

Reviewer's report:

General
This study raises very interesting questions concerning the measurement expectations of adverse event markers. The reader requires additional background information on the home care quality improvement initiatives by Center for Medicare and Medicaid Services (CMS) and generally how quality indicators are interpreted and validated. There is a need for describing how the OASIS quality measures are used, how they may be monitored and reported and how they may relate to adverse event markers. To be valid, quality indicators should be demonstrating variation amongst providers delivering different levels of quality of care and that variation should be attributed to quality of care as assessed by process measures, outcome information or other independent sources. It is not clear that adverse event measures should possess the same characteristics. In this present study, we are provided information for a single home health agency. Therefore, there is no indication that the adverse event rates vary by agency or that the variation is related to quality of care.

The objective was to compare 3 different operational definitions of substantial functional decline. More specifically, it was to examine the current CMS definition (decline of at least 2 points in 3 ADL tasks) with 2 less conservation definitions: a 2-point decline in 2 ADL tasks and a 2-point decline in one task. The stated objective was to examine the predictive validity of the 3 definitions and the test-retest reliability.

The methods are clearly presented but the nomenclature is somewhat confusing. Usually, the terms predictive validity and test-retest refer to characteristics of a measure or test. Predictive validity of a measure examines the ability of that measure to predict a future event. For example, whether SAT scores predict performance in university. Test retest of the measure would examine the consistency of scoring the measure at 2 points in time or for quality indicators the consistency of rates over different time periods.

The methods would be strengthened with a clearer hypothesis of what attributes the investigators believe should be associated with a better measure of adverse events. The reader is left to make her own assumptions of the hypotheses. The assumptions are that a more "valid" definition is reflected by a dependent variable that can be predicted by a multivariate logistic regression with a better "fit" as assessed by the c-statistic and the Hosmer-Lemeshow Goodness of Fit test. They also hypothesize that the logistic model predicting functional decline should be stable over time. They have referred to the latter attribute of the model as test-retest reliability. There is a need for clearer language to indicate that they are looking at the validity and reliability of statistical models not of the adverse event measure itself. The rationale and assumptions require explicit justification.

A fundamental question is whether an adverse event should be predictable or consistent over time periods. Or is it a rare event or a random occurrence. If it truly is random and rate, then the current CMS definition is the most appropriate as it was the least predictable. On the other hand, it raises the question as to the utility of adverse event markers.

The CMS home health website lists several quality measures, including 3 measures relating to ADLs: improvement in bathing, improvement in ambulation/locomotion, and improvement in transferring. It would appear that improvement in each area requires only a one point improvement in that task. There is no decline in function quality measure although not all patients can be expected to improve in function. If agencies wish to monitor their performance in preventing functional decline, it would be important to have a quality measure that is valid and stable over time. Adverse event measures that are rare are unlikely to be appropriate for this type of quality monitoring.

The authors have pointed out several drawbacks to the interpretation of the significant decline adverse event marker. The operational definition excludes individuals at the lower range of functioning because they cannot decline 2 points in 3 ADL areas. The degree of decline is substantial, particularly considering a 60 day time frame. Indeed, one would question whether such a substantial decline is within the control of the provider. It may represent a rather major health event such as stroke which may not be preventable in all
The rates of the current adverse event definition are low. Fewer patients are included than are patients excluded from the analyses due to death and hospitalization (Figure 1). Indeed, if the investigators choose to remain with their current analysis plan of validating the best fit of multivariate models and the stability of the models over time, they should consider a multinomial or competing risk model because they patients who demonstrate substantial functional decline are at competing risk of dying and of hospitalization.

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)
1. Present background information on the quality initiative used by CMS for home care and a framework for understanding how adverse event measures fit within the initiative
2. If there is an expectation that adverse events can be predicted, a multi-nomial model for competing risks of hospitalization and death and substantial functional decline is more appropriate.

Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)
3. There should be greater clarify in the use of predictive validity and test-retest validity. The terms are generally used for measures whereas the authors use them to denote the validity and stability of multivariate models.

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: No

Declaration of competing interests:
I declare that I have no competing interests