

Reviewer's report

Title: Using a summary measure for multiple quality indicators in primary care: the Summary Quality Index (SQUID)

Version: 1 **Date:** 20 December 2006

Reviewer: Steven Asch

Reviewer's report:

The problem of aggregating quality indicators is a very important one, especially as indicators proliferate. Individual providers, consumers and even managers have difficulty interpreting multiple simultaneous data on various aspects of their performance. As the authors point out, there is a need for summative measures to screen for problems and guide efforts toward quality improvement. They have developed an aggregate measure of the 30 or so common primary care indicators of quality by simply adding them up and averaging either for an individual patient or a practice. The description of the simple methods for calculating these scores is unnecessarily long. This approach has been used before both by CMS and RAND, as the authors note, as well as by the VA, as they do not.

This approach has problems, some of which the authors acknowledge. The first is that not all indicators are created equal with regard to clinical outcomes. It is more important to measure LDL in a diabetic than measure total cholesterol in all other adults. Worse yet, because there are many more adults without diabetes than with, the latter measure will have a much higher weight in the overall practice score. Various approaches have been tried to address this, and the authors could discuss them more. Some research teams have developed weights based on either expert opinion or truncated decision trees predicting health related quality of life. Others have partitioned overall scores into conditions e.g. diabetes, coronary artery disease, and weighted care for each disease equally. The authors might rightly say this was beyond their scope, but they should at least point the reader in that direction, or provide some basic simulation studies. Another problem is that of interaction between indicators, which the authors fail to address. A third problem is that even if one were to assume equal weights, some indicators are inherently more difficult to pass. There are statistical methods for adjusting for difficulty of passing derived from item response theory. The authors could have addressed both of these last two problems more thoroughly.

The component indicators are simple ones, purposively, so that they might be abstracted from an electronic medical record without a lot of clinical data on contraindications. For implementation purposes, this was a good idea. Still it would be good to know to what extent the unmeasured contraindications are common or nonrandomly distributed, which would make them more worrisome, an analysis that could have been accomplished with a limited chart review.

The analyses of the statistical properties have some strengths and weaknesses. The idea of reliability testing by evaluating sequential measures during a quality improvement initiative seems misguided. Ideally, two different measures would have taken place at the same time, though if these are just abstracted from a database, how they could change from measure to measure is unclear. I would suggest removing the reliability section of the manuscript. Measuring the responsiveness of SQUID to quality improvement is more clearheaded, yet the data here are unconvincing. Only a very small improvement was noted. The internal consistency of SQUID is high, but this reviewer is unclear as to what that means. Do we really expect that simple quality measures should be highly intercorrelated? If they were not, would that mean that we should not aggregate them? There has been a lot of debate about whether there is an underlying construct called quality or whether there are subdomains (see Katherine Kahn's work on this), and the authors don't touch upon it. More encouraging is the variation in SQUID scores across practices and patients.

The strength of this manuscript is underdeveloped. Especially for this journal, the readers would benefit greatly from an expanded version of the subsection of Methods entitled "Use of SQUID in quality improvement" and the results section beginning at the bottom of page 11. Knowing more about how providers reacted to the idea of summative measures, the problems the authors had in implementing the measures would help others who plan similar efforts. This qualitative information is at the core of the real message of this paper and I would encourage the authors to develop it further.

What next?: Unable to decide on acceptance or rejection until the authors have responded to the major

compulsory revisions

Level of interest: An article of importance in its field

Quality of written English: Acceptable

Statistical review: Yes, and I have assessed the statistics in my report.