Reviewer's report

Title: Clinically Relevant Quality Measures for Risk Factor Control in Primary Care: a retrospective cohort study

Version: 3 Date: 24 March 2014

Reviewer: Asaf Bitton

Reviewer's report:

The paper by Weiler et al offers an interesting analysis of a retrospective cohort of 1002 patients form university-based primary care clinics in Switzerland. The authors measure proportions of patients with common chronic conditions in control, and then also add to this an assessment of whether appropriate clinical action was taken in response to the measures, and whether that resulted in potential overtreatment. They found that adding on whether clinicians responded "appropriately" to values outside the normal range significantly changed the view of whether or not clinical quality was achieved, while also not concomitantly increasing overtreatment.

Major Revisions:

My major feedback has to do with the overall conceptualization of the study, not its methods (which appear to be well-done). The question, to me, appears to center around the following: When evaluating quality of care for chronic conditions, should a clinician, or clinical team, get credit for trying? That may sound trite, but it is at the heart of the issue here. If, in our clinical quality measurement framework, we are looking to achieve certain thresholds of results because the epidemiological and trial data suggest better outcomes with higher levels of the population achieving those results, then then answer to this question, and the response to this study, should be negative. That is to say, it doesn't really matter, from the perspective of the end-outcome for the end-user, if the clinical team tried or took "appropriate clinical action" if the blood pressure is still 179/108mg/Hg or the LDL cholesterol is 211mg/dL because those, by and large, are dangerous values associated with harmful outcomes which the measurement system is intended to prevent, and for which the clinical team will be held accountable for not preventing.

However, if the goal of the measurement is to spur improvement and not accountability, then the allowability of the effort/other contextual explanatory factors would seem to be appropriate. This is a version of Campbell's law, which suggests that a you cannot use 1 instrument to serve 2 objectives (performance improvement and accountability) because by focusing on accountability, you corrupt the opportunity for improvement.

What is needed, then, is a framework for understanding how and when these measures are to be used, in order to evaluate both the face validity and performance utility. If they are to be used to establish the proportion of patients at
risk for a CVD event by virtue of not being in control (the true outcome here), and
the clinical teams will be held accountable to that end-outcome, then it does not
really matter if they took action unless it serves to reduce the proportion who are
out of control. If the measure framework is intended to spur clinical performance
through buy-in of the measure system and responsiveness to clinical action
taken that may or may not improve the end-outcome, then it makes sense. But
without a discussion of that tension, and those competing demands, it is hard to
take at face level the claims made by the authors that this is inherently a better
measure system and framework. The questions are: better at what, for whom,
and why? Of course if you include lots of points for trying to respond to out of
control values and take into view the large patient context, that will make
providers look like they are "doing better" at taking care of their patients. But will
that improve overall outcomes, reduce morbidity and suffering from CVD events,
and reduce mortality? This study is not designed to tell us that and thus is limited.
The conclusion on pg 14 says this is a "more accurate index" but I don't think that
they have proven this contention. The authors should address head-on this
philosophical and improvement challenge so that we can better understand their
results and judge for ourselves whether it is, or is not, "more accurate", and if it
is, what exactly it is more accurate at doing.

Another major limitation is the exclusion of diabetic patients on insulin. These
patients represent a major area of improvement (or lack thereof) and are
vulnerable to clinical inertia. The authors need to better address why they
excluded them.

Finally, the authors need to address on pg 9/10 the statements on how they new
patients with hypertension were less likely to receive appropriate clinical action
when they had other chronic co-morbidities, and how they knew diabetic patients
were not more likely to receive appropriate clinical action in response to elevated
A1c levels. Were these univariate correlations or multivariate models? I think
they were the latter based on the tables, so it should be more clearly specified.

Minor revisions:
pg 3: "among other differences" -- add "from US"
pg 8: typo on sentence with "52-55% of patients"

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.

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