Reviewer's report

Title: P.Re.Val.E.: Outcome Research Program for the Evaluation of Health Care Quality in Lazio, Italy

Version: 1 Date: 17 May 2011

Reviewer: ENRIQUE BERNAL-DELGADO

Reviewer's report:

P.Re.Val.e, Regional Outcomes Evaluation Program, is a valuable experience of performance measurement in Lazio region (Italy) that systematically analyzes, in an observational basis, multiple health performance indicators. In this article authors aimed to describe the program as well as the impact on releasing performance data to the public.

MAJOR COMPULSORY REVISIONS

From my point of view, two major questions are missed in the paper.

1) Cluster effect assessment. It is well known that patients within a provider, or hospitals within a particular area, tend to be or behave homogeneously, compared with those within another provider or area. This “cluster effect” might influence the estimation of expected events and standard errors —particularly with “small numbers”, and have to be modelled. The effect of modelling is on smoothing the estimates, being generally more demanding in finding differences among providers, if they exist.

Multilevel effects modelling will provide a sounder estimation of the “performance variation” attributable to individuals, and the fraction amenable to providers (rho statistic), allowing relative measures (median odds ratio, e.g.) useful to compare healthcare providers.

In my opinion, the cluster phenomenon has to be assessed before publication.

2) Measures Precision: On a different point, most of the P.Re.Val.e indicators might be considered infrequent events, what might jeopardize the precision of the estimates (critical when it comes to inform differences). Thus, p values although useful, are scarcely informative. Healthcare performance assessment requires insight on how precise the measures are. Robust confidence intervals should be reported.

3) Given the nature of the article –description of a project- I would expect a deeper discussion on internal and external validity threats: misclassification bias, risk adjustment limitations, difficulties to reach sounder and reliable measures, etc. and specially, how P.Re.Val.e is dealing with all of these challenges.
MINOR ESSENTIAL REVISIONS

1) It can be assumed that the work is based on all the discharges produced by the hospital providers in each geographic area. Some reference to this point would be helpful in order to determine to what extent the uneven records coverage across providers or regions may influence the estimates.

2) On a similar point, database linking is a major issue in P.Re.Val.e. Some detail on the effectiveness of the linking process would be useful (i.e., number of records properly merged, percentage of missing records, risk of double-counting, etc.). Some comment on the relevance of these figures on determining numerators and denominators would be also helpful.

3) It is not clear to me, whether population-based measures are counting just for patients treated in hospital-providers within their area of residence, or patients living in the area whichever the hospital they are treated. It would be worth mentioning to what extent the former option (ruling out patients treated in hospitals outside the residence area) may bias the estimates.

4) Risk adjustment: authors mention the use of comorbidities and/or severity measures for risk adjustment purposes. Some of them are detailed in the appendix. To my understanding - not clear though - b coefficients are used to standardize rates. It would be useful to know which regression models are chosen for each indicator, and its goodness of fit.

5) Also referred to risk adjustment models, it can be reckon the effort to reduce the risk of misclassifying acute conditions as comorbidities. However, an additional explanation on how was severity defined and measured would be valuable.

6) Although hospital to hospital comparisons are useful, models comparing observed to the expected cases have been proven efficient and more robust, when it comes to analyse “small numbers”. I do not understand the Aristotelian argument proposed by the authors rejecting the use of these indirect standardized methods, basically, because these analytical techniques do not seek hospital to hospital comparisons. Furthermore, these techniques are widely used in social epidemiology or healthcare services research.

7) Pag 6, last paragraph. Authors mention “most data were expressed as ratios...” Following the description, it would be said that they are referring to rates, or maybe cumulative risks. Is that correct?

8) It seems to me that OR conversion to RR is not needed given the small numbers expected in the non-exposed groups. In any case, to my knowledge the formulation used by the authors is that proposed by Zhang (see JAMA 1998, 280: 1690-1) If so, wouldn’t it be mistaken?

Level of interest: An article of limited interest
Quality of written English: Needs some language corrections before being published

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'