Author's response to reviews

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

Authors:

Danilo Fusco (fusco@asplazio.it)
Anna P Barone (annapatrizia.barone@io-cittastudi.it)
Chiara Sorge (sorge@asplazio.it)
Mariangela D'Ovidio (dovidio@asplazio.it)
Massimo Stafoggia (stafoggia@asplazio.it)
Adele Lalio (lallo@asplazio.it)
Marina Davoli (davoli@asplazio.it)
Carlo A Perucci (perucci@agenas.it)

Version: 3 Date: 14 November 2011

Author's response to reviews: see over
Rome, November 14th 2011

To the Editor of

BMC Health Services Research

Object: point-by-point response to comments of the reviewers of the manuscript: “P.Re.Val.E.: Outcome Research Program for the Evaluation of Health Care Quality in Lazio, Italy”

Dear Flory Mae Calumpita
the point-by-point responses to comments of the reviewers are listed below.

No changes have been made in the manuscript.
We thank Rino Bellocco and Hester Floor Lingsma for their useful comments and suggestions.

Reviewer: ENRIQUE BERNAL-DELGADO
MAJOR COMPULSORY REVISIONS
Comment n.1
I do thank the authors for further detailed explanation, although they refer partially to what I was asking for. From my point of view, backed on recent research [*], the estimation of the cluster effect by using multilevel analysis is worthwhile to determine what the magnitude of variation in hospital (area) outcomes attributed to a hospital is, as opposed to that amenable to the patient. As compared to a “rho statistic” around 10% (just as a matter of example), a rho statistic including the null value (or close to it) would have quite a different relevance. In the latter, we cannot rule out that the differences among hospitals are just explained by differences in the case-mix of patients. Therefore, it is not just about estimating the standard errors accounting for non-independent observations (as suggested by the authors) but on estimating the magnitude of the variation amenable to the cluster – hospitals or areas in this case-study. This is a particularly useful measure when it comes to providing decision-makers with relevant information.

On a different point, authors are correct when highlighting the possibility of point-estimates “bias”. Although possible, in our experience the smoothing effect is more on the small numbers, having a negligible effect in the majority of hospitals, particularly as for studies with large sample size.

My concern on the cluster effect has not been sufficiently addressed along the reply provided by the authors. It will be worth adding an additional statistic estimating the magnitude of this effect (e.g. rho statistic, intra-class correlation index, median odds ratio) to each one of the indicators under study.

[*] Maybe of use in this specific piece of work:

Response n.1
We thank the referee for having raised a so interesting and debated methodological issue. We are aware of multilevel potentiality in variance decomposition, assessment of intraclass correlation and analysis of higher level variables. However, with regard to the hospital effect evaluation, our opinion is substantially different from that of the referee. As already mentioned in the previous answer, when hospital-specific random effects are used to estimate the hospital-specific risk ratios, the hierarchical cluster analysis introduces a bias, named shrinkage towards the mean, because of which “high performing” hospitals will be presented too negatively and “low performing” hospital will be presented too positively. The referee points out that this kind of bias could be negligible for studies with
large sample size. This is true “on average”, but we underline that sample size substantially varies between hospitals. Moreover, the magnitude of the bias is large not only for small hospitals but also for providers whose performance is far from the overall mean (see Bryk and Raudenbush, 1992; Raudenbush and Willms, 1991; Snijders and Bosker, 1999; Hox, 2002). Furthermore, we have strengthened the decision not to use the multilevel approach after having applied this method and tested it on our data. In fact, in the majority of hospital, the shrinkage towards the mean was clear and considerable. Therefore, after carefully weighing advantages and disadvantages of random-effects and fixed-effects modeling, we confirm the use of fixed-effects modeling to calculate hospital-specific effects and adjust for differences in patients’ case-mix. We finally underline that, according to the referee’s suggestion, in order to assess for clustering, sandwich variance-covariance estimators were already introduced in the revised analysis.

Comment n.2
a) I do thank the clarification, and I do agree on the importance of showing the size as a reference for discussion for decision-makers - a 2-fold difference has a different impact depending on the basal rate of events. But I do not see the argument provided to justify the use of p-values in providing a better insight for discussion, when guidelines are so clear in the other way around; and particularly, when your contribution is on the context of a scientific journal.

b) In any case, the adjusted risk-ratios are not expected to follow a normal distribution; conversely, they are expected to be skewed. If true: Would those areas/hospitals with p values below 10% change the statistical significance when estimating the appropriate (robust) confidence intervals? Confidence intervals provided by multilevel analysis are more conservative – i.e. reducing the possibility of flagging as bad performers hospitals affected (in fact) by small numbers. A table showing the two elements, p-values and confidence intervals, would be illustrative along this discussion.

Response n.2
a) As already mentioned in the previous response, P.Re.Val.E. is an outcome research program conceived mainly as a tool for promoting discussion among healthcare managers and professionals in the Lazio region, so that we choose to report P-values and sample sizes, because, for our purposes, they are clearer than confidence intervals: it can be judged whether a value is greater or less than the previously specified regional mean value. We decided not to report confidence intervals because they need a priori confidence level selection (90%, 95%, 99%) depending on how much conservative you would like to be. It means we are going to suggest a choice that is not the purpose of this program. As reported in the paper, we preferred to leave this choice to the policy-makers.

b) All statistical tests are based on log-transformed RRs which, as well known, follow a normal distribution.

MINOR ESSENTIAL REVISION
Comment n.3
The authors reply on risk adjustment methodology “all possible confounders have been identified” brings another issue up which is “over-adjustment”. Using all possible confounders may drive to over-adjustment, if no previous judgment upon the clinical rationale behind each indicator is made. I am not clear after reading the explanation, if authors have made or not this specific assessment. Over-adjustment, described by Patrick
Romano in several papers, may mislead conclusions by concealing real differences, particularly in those more complex providers – large hospitals treating complex patients.

Response n.3
Yes, we had already made this specific assessment, as described in the Method section:

"Coexisting medical conditions"

Chronic comorbidities and/or severity characteristics that were potentially associated with the outcomes under study were chosen using information in the literature7-9, 17, 18, 21-23, 25-30 and in the Mattoni-outcome project.13 The potential risk factors were identified on the basis of ICD-9-CM codes registered either during hospitalization for the condition under study (index hospitalization) or in previous hospital or emergency department admissions during the previous two years. Acute events occurring during the index hospitalization, which could be complications of care/treatments (i.e. on the causal pathway between exposure and outcome), were not included.

```
Preliminary disclosure of performance data to clinicians and providers
```

Before public release of data, we shared the P.Re.Val.E methods and results with different groups of clinicians and providers, to promote discussion and encourage contributions and critical assessments."