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 Lallo (lallo@asplazio.it)
Marina Davoli (davoli@asplazio.it)
Carlo A Perucci (perucci@agenas.it)

Version: 2 Date: 24 August 2011

Author's response to reviews: see over
Rome, August 24th 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 Eadgar Austin Azul on behalf of Dr Judith de Jong,

the point-by-point responses to comments of the reviewers are listed below.

All changes made in the manuscript are in bold character.

General Response

P.Re.Val.E. is a program in progress as described in the discussion section: methodologies and results have been updated since the paper was submitted. Most of reviewer’s comments have been considered in the updated version available on line at http://www.epidemiologia.lazio.it/prevale10/index.php.

This updated version is in Italian but will be translated in English within a couple of months.

Please refer to the previous version only for Data sources, record linkage criteria and protocols.
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.

Response n.1
The referee raises an interesting and controversial methodological issue. In fact, the group-specific intercepts obtained from a multilevel regression analysis are shrunk back towards the overall mean coefficient. If intercepts are used as an indicator of the quality of hospitals, the shrinkage estimators could be biased: “high performing” hospitals will be presented too negatively and “low performing” hospitals too positively (Bryk and Raudenbush, 1992, chapter 5; Snijders and Bosker, 1999). In general, the bias should be largest analyzing hospitals with low number of admissions and far from the overall mean. In order to assess for cluster effects, we used the cluster sandwich (“robust”) variance-covariance estimators, relaxing the usual requirement that the observations are independent. According to this methodology, the observations are independent across groups but not necessarily within groups. We preferred this solution to the multilevel modeling, because it affects the standard errors of the estimators but not the estimated coefficients. Added in the statistical analysis section.

Comment n.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.

Response n.2
See Response n.1.
We agree with the reviewer, recent guidelines for presenting the results of clinical experiments or observational studies, suggest providing confidence intervals instead or together with the p-values, and giving the effect sizes of the investigated associations. P-values and confidence intervals, however, are not contradictory statistical concepts. If the size of the sample and the dispersion or a point estimate are known, confidence intervals can be calculated from p-values, and conversely. The two statistical concepts are complementary.
As reported in the discussion, 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.

Comment n.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.

Response n.3
We have discussed these issues in the discussion section.

MINOR ESSENTIAL REVISIONS
Comment n.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.

Response n.1
We have added these issues in the Study design and data sources section.

Comment n.2
On a similar point, database linking is a major issue in P.Re.Val.e. Some details 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.

Response n.2
We have added these issues in the Study design and data sources section.

Comment n.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.

Response n.3
We have added this issue in the Study population section.

Comment n.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.

Response n.4
For each indicator, all possible confounders have been identified, as reported in the operative protocols (see AMI example). The potential confounders were selected analyzing the multivariate relationship between the possible confounders and the outcomes by application of multivariate regression models. The investigation of the stability and predictability of the models is a significant part of the modelling process. To validate our findings we used a stepwise bootstrap procedure (added in the Statistical analysis section).

Comment n.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.

Response n.5
See Response n.4.
We added, as an example, the risk adjustment model used for 30-day mortality after IMA. (see Table 1 in the Results section)

Comment n.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.

Response n.6
That part has been omitted in the Discussion section

Comment n.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?

Response n.7
It depends on the indicators, this is the reason we use the more general term “ratios”

Comment n.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?

Response n.7
We overcome this formulation. See the Statistical analysis section.
Reviewer: Hester Floor Lingsma
MAJOR COMPULSORY REVISIONS
Comment n.1
The question and the conclusions do not adhere to the methods and results. If the aim is
to describe the program, I would like more details on how it was set up exactly, which data,
who were involved etc in the results. Now most emphasis in on the data of three specific
indicators. This is of course highly interesting, but fit a research question like: ‘describe
differences in quality of care between hospitals in Lazio’. To state it differently: the result
section currently almost seems to come from another paper.

Response n.1
We revised the Results section, reporting, as an example, only the 30 day-mortality after
AMI and Methods section.

Comment n.2
The current 'state of the art’ method in quality of care measurement is random effect
modelling instead of fixed effect modelling. Although I expect not many changes in the in
the results because of the high numbers of outcomes, the authors should make sure this is
the case and mention the consideration of random effect models in the paper.

Response n.2
The referee raises an interesting and controversial methodological issue. In facts, the
group-specific intercepts obtained from a multilevel regression analysis are shrunk back
towards the overall mean coefficient. If intercepts are used as an indicator of the quality of
hospitals, the shrinkage estimators could be biased: “high performing” hospitals will be
presented too negatively and “low performing” hospitals too positively (Bryk and
Raudenbush, 1992, chapter 5; Snijders and Bosker, 1999). In general, the bias should be
largest analyzing hospitals with low number of admissions and far from the overall mean.
In order to assess for cluster effects, we used the cluster sandwich ("robust") variance-
covariance estimators, relaxing the usual requirement that the observations are independent.
According to this methodology, the observations are independent across groups but not necessarily within groups. We preferred this solution to the multilevel
modeling, because it affects the standard errors of the estimators but not the estimated
coefficients. Added in the Statistical analysis section.

Comment n.3
The authors put a lot of emphasis on direct standardisation in the introduction and
discussion, while this is not a research question, neither is it related to the research
questions. The part in the introduction on standardization should be omitted, it adds
nothing there.

Response n.3
We revised the Introduction and discussion section.

Comment n.4
Although I know the discussion on indirect vs direct standardization, to my knowledge it is
practically impossible to apply direct standardization in combination with case-mix
adjustment, since the rates in the reference group should be observed for each specific
stratum (e.g. males between 50 and 55 with no previous MI but with diabetes and hypertension etc) I wonder how the authors have done this.

Response n.4
In order to estimate adjusted group-specific (hospital/area of residence) log odds of outcome, logistic regression models with no intercept and centered covariates were applied for each outcome. Adjusted risks were obtained for each group by back-transforming parameter estimates with the following formulas:

\[ \text{Adj risk} = \frac{\exp(\text{estimate})}{1 + \exp(\text{estimate})} \times k \]

where \( k \) is a correction coefficient introduced to take into account the nonlinear nature of the logistic model. \( k \) is calculated as follows:

\[ k = \frac{\text{actual number of events}}{\sum_{j=1}^{m} p_j \times n_j} \]

where \( p_j \) are the adjusted risks, \( n_j \) is the group size, and \( m \) is the number of groups. This approach allowed comparison of the outcome for a given facility or area of residence with that of the whole study population and with each of the other facilities/areas. We changed our reference group in the whole population in order to obtain more robust estimates. It has been added in the Statistical analysis section.

Comment n.5
In the discussion the authors state that indirect standardization should not be used for hospital comparisons. Besides the remark that I do not see the additional value of the Aristotelian deduction method here, it should be noted that indirect standardization leads to different results ONLY if the effect of risk factors varies between hospitals. This should be mentioned.

Response n.5
It has been added in the discussion section.

Comment n.6
There seems to be confusion on the distinction between outcome and process indicators. In the abstract Methods section the program is on '54 outcome/process indicators' while in the rest of the manuscript the authors refer to '54 outcome indicators'. But I would consider operation within 48 hours in hip fracture patients a process measure.

Response n.6
It has been changed in the text.

Comment n.7
I think the authors quote the literature on the merits of quality measurement a little selectively. Although many believe in the potential of quality measures and their public disclosure to improve quality of care, at least as many do not. Example (with most relevant references): Shahian et al., Variability in the measurement of hospital wide mortality rates, N Engl J Med 2010. 363;26. The discussion could be more balanced in that sense.

Response n.7
It has been added in the discussion and references section.
Reviewer: Rino Bellocco
MAJOR COMPULSORY REVISIONS

Comment n.1
Abstract: authors state they want to describe the P.Re.Val.E. program, but they only concentrate on three of the indicators of the program,

Response n.1
It has been changed in the abstract.

Comment n.2
Order properly the aims (number five is repeated twice). I also suggest that these aims are not bulleted, but described in a sentence. These are not the aims of the paper, which are actually clearly stated in the beginning of the results sections.

Response n.2
It has been changed in the Introduction section.

Comment n.3
Can the authors clear why direct standardization methods rather than indirect standardization methods were used for the comparative evaluation of outcomes. Is the assumption of lack of interaction being assessed?

Response n.3
Hospital league tables obtained by indirect standardization procedures should not be used for hospital-to-hospital comparisons. This technique can lead to biased conclusions unless the distribution of risk factors or their effects do not vary between the hospitals to be compared. It has been added in the Discussion section.

Comment n.4
The indicators were defined using information collected regional health information systems. Can the authors add on the completeness of such registries?

Response n.4
As reported in the Methods section, please see http://www.epidemiologia.lazio.it/vislazio_en/fonti.php.

Comment n.5
wait times on page 7 should be waiting times

Response n.5
It has been changed in the text.

Comment n.6
Acute events occurring during the index hospitalization, which could be complications of care/treatments, Can they clarify this statement?

Response n.6
It has been clarified in the Coexisting medical conditions section

Comment n.7
Conversion formula from OR to RR is highly questionable. Confidence intervals are too narrow, does produce bias if confounding is present. Another issue is the assumption of interaction of the OR scale and RR scale. The authors need to justify this method carefully, unless they are willing to use other methods (see MacNutt, AEJ, 2003).

Response n.7
In order to estimate adjusted group-specific (hospital/area of residence) log odds of outcome, logistic regression models with no intercept and centered covariates were applied for each outcome. Adjusted risks were obtained for each group by back-transforming parameter estimates with the following formulas:

\[
\text{Adj risk} = \left( \frac{\exp(\text{estimate})}{1 + \exp(\text{estimate})} \right) \cdot k
\]

where \( k \) is a correction coefficient introduced to take into account the nonlinear nature of the logistic model. \( K \) is calculated as follows:

\[
k = \frac{\text{actual number of events}}{\sum_{j=1}^{m} p_j \cdot n_j}
\]

where \( p_j \) are the adjusted risks, \( n_j \) is the group size, and \( m \) is the number of groups.

This approach allowed comparison of the outcome for a given facility or area of residence with that of the whole study population and with each of the other facilities/areas. We changed our reference group in the whole population in order to obtain more robust estimates. It has been described in the Statistical analysis section.

Comment n.8
Authors refer to http://www.epidemiologia.lazio.it/vislazio_en/metodi_statistici.php for the methodology used to select the reference group. Can they summarize in a few lines in the section?

Response n.8
See Response n.7

Comment n.9
Authors state that “For each indicator, trend analyses and comparisons of the 2006-2008 data versus the 2004-2006 data were developed for each hospital and area of residence. But they did not specify in the previous section that they also have data for 2004-2006

Response n.9
It has been changed in the text.

Comment n.10,11
Table 1b. Health care facilities: reference group Is not clear to me what this represents Same for Table 2b and Table 4b, Table 2 is after Fig 1 and Fig 2 legend

Response n.10,11
The tables and figures have been changed.