Author's response to reviews

Title: Differential misclassification of confounders in comparative evaluation of hospital care quality: caesarean sections in Italy.

Authors:

Mirko Di Martino (m.dimartino@deplazio.it)
Danilo Fusco (d.fusco@deplazio.it)
Paola Colais (p.colais@deplazio.it)
Luigi Pinnarelli (l.pinnarelli@deplazio.it)
Marina Davoli (m.davoli@deplazio.it)
Carlo A Perucci (perucci@agenas.it)

Version: 4 Date: 3 September 2014

Author's response to reviews: see over
Title: Differential misclassification of confounders in comparative evaluation of hospital care quality: caesarean sections in Italy.

The corrections in the text are highlighted in yellow while the cancellations are in Strikethrough font. The sequence of the references has been modified. We uploaded a clean version of the revised manuscript as a supplemental file (additional file 2).

Reviewer: Timothy L Lash

Reviewer’s report:
The authors demonstrate the potential importance of differential misclassification of a confounder (high fetal head) in Italian surveillance data on the proportion of births by caesarean section within different hospitals and regions. Overall, the study is well conducted and well written.

Major compulsory revisions

1. The number of births at different hospitals is likely to be substantially different, with some hospitals having relatively few births. Typically, the lowest and highest proportions (of both HFH and CS) will be recorded at low volume hospitals. Semi-empirical Bayes shrinkage methods should be considered to avoid the potential for extreme proportions, measured with poor variance, to unduly influence the results.

The referee raises an interesting and controversial methodological issue.

In general, when performing comparative evaluation of health services outcomes, we prefer to use fixed-effects modeling to calculate hospital-specific quality indicators. In fact, the group-specific intercepts obtained from the Empirical Bayes estimators are shrunk back towards the mean. Therefore, if intercepts are used as a quality indicator, the shrinkage estimators introduce a bias, because “high performing” hospitals will be presented too negatively and “low performing” hospitals will be presented too positively (Bryk and Raudenbush, 1992, chapter 5).

However, in some circumstances, the problem raised by the referee may be very relevant, because the lowest and the highest proportions of CS, HFH (or any other outcome) may have been “recorded at low volume hospitals”.

After weighing advantages and disadvantages of the random and fixed effects methods in the specific setting of this paper, the Empirical Bayes estimator was now applied in a sensitivity analysis, for both CS and HFH proportions, in order to compare the findings deriving from the two different approaches [Statistical analysis section. Page 7; lines 18-24]. The Empirical Bayes estimator and the fixed-effects method produced very similar and consistent results [Results section. Page 8; lines 6-9 and 18-22], probably because, in our cohort, the hospital sample size is generally very high: estimates in large group are more reliable, and shrink less than estimates from small groups [Discussion section. Page 12; lines 11-14].

In the following figure (figure A), specifically performed according to the referee’s suggestion, the prevalence of HFH by maternity unit was estimated using both the fixed-effects approach (as reported in the Figure 1 of the paper) and the Empirical Bayes estimator. The shapes of the graphs as well as the distribution percentiles are very similar.
Figure A. Prevalence of HFH in 540 Italian hospitals, 2010, using fixed and random effects models.

**FIXED EFFECTS ESTIMATES**

* Min: 0.0%; Max: 69.6%; $\sigma^2_{u0} = 7.1$ (S.E. = 0.6; p-value < 0.0001).

**RANDOM EFFECTS ESTIMATES (EMPIRICAL BAYES SHRINKAGE METHOD)**

* Min: 0.0%; Max: 69.6%; $\sigma^2_{u0} = 7.1$ (S.E. = 0.6; p-value < 0.0001).
2. At least in the discussion section, if not implemented in the analysis, the authors should consider the potential for bias analysis to analytically address the problem of differential misclassification of HFH. The only apparently proposed solution is to avoid adjustment for HFH, but would it be possible to reduce the bias by using bias analysis to adjust for the misclassification?

The issue raised by the referee is very interesting. The authors believe that, in order to fully understand how the techniques based on Quantitative bias analysis can help to mitigate this particular type of confounding, further analyses are required, by making and comparing several and reliable assumptions about the bias parameter. However, we have mentioned and briefly described the "Quantitative analysis bias", and we emphasized that this methodology may provide a valuable alternative to the strategy we used in this paper [Discussion section. Page 14; lines 11-15]. Moreover, we added the following citation in the “References section”: Lash TL, Fox MP, Fink AK. Applying Quantitative Bias Analysis to Epidemiologic Data. Series: Statistics for Biology and Health. Springer, 2009.

3. It would also be useful to suggest, if not implement, a validation sub-study to examine the actual rates of misclassification of HFH. The widely variable proportion, in correlation with CS proportion and the historical information on demands to justify CS, make a compelling case. And the authors allude to the potential for HFH to be difficult to assess, so maybe validation is not possible. Nonetheless, it is generally good practice in a situation like this one to use validation sub-studies to document the classification errors, and that good practice should be mentioned even if not implemented in this study.

We fully agree with the referee. The anomalous cases highlighted by this study clearly denote a non-ethical conduct of some hospitals. After reporting our findings to the Italian Ministry of Health, there was a strong reaction, that led to the conduction of several validation sub-studies on a large sample of maternity units, aimed at identifying HFH classification errors and potentially opportunistic behaviors in reporting factors that are an indication for CS. Unfortunately, the results of validation sub-studies have not yet been disclosed by the Ministry. This issue was now addressed in the Discussion section, page 13; lines 1-5.

Minor essential revisions

4. In general and throughout, "misclassifications" (plural) should be "misclassification" (singular).

According to the report of the referee, the error has been corrected. In the current version of the paper, the word "misclassification" is always used as singular form.

5. In general and throughout, data are presented to too many digits beyond the decimal point. The sample size is not likely to be sufficient to support the reported level of accuracy.

According to the suggestion of the referee, in the current version of the paper, we reported only one digit beyond the decimal point, with the exception of the thresholds automatically identified according to the Jenks natural breaks optimization (used for the geographical maps) and with the exception of the Risk Ratios presentation. In fact, we think that two decimal places after the decimal point may be useful for the interpretation of Risk Ratios,
especially in the case of their confidence intervals (for instance, table 1), which are sometimes very narrow.

Level of interest: An article of importance in its field.

Quality of written English: Needs some language corrections before being published.

The paper was reviewed by a native English speaker.

Statistical review: No, the manuscript does not need to be seen by a statistician.

Declaration of competing interests: I am co-author of two textbooks that advocate for the use of quantitative bias analysis in analyses of epidemiologic data and for which I receive royalties.
Reviewer: Rossella Miglio

Reviewer's report:
The paper describes an interesting analysis that investigate misclassification of confounders in a comparative evaluation of hospital care with a focus on the appropriateness of Caesarian Sections in Italy. The paper is well organized, the method is applied correctly, and the results are presented appropriately.

Minor essential revision

The following are some comments that I hope may help improve presentation of the study:

Methods

1. The authors propose a two step procedure to select potential risk factors for CS deliveries. In the second step they adopt a modified Poisson regression model for prospective studies with binary data while for estimating the adjusted, hospital-specific, proportion of CS deliveries they use a multivariate logistic regression. I think that it could be useful to help the reader to understand this different model choice.

We thank the referee for this comment. In fact, we have introduced in the "Statistical analysis" section unnecessary confusion. The two models are equivalent. The modified Poisson regression model (Zou G. A Modified Poisson Regression Approach to Prospective Studies with Binary Data. Am J Epidemiol 2004, 159: 702-706) is a “convenient” solution for the predictive model because it directly estimates the risk ratios for the selected risk factors (the odds ratio would not be a good estimate of the risk ratio in this study, because the outcome of interest is not rare). However, when the dummy variables that “identify” the maternity units are included in the model, arise problems related to the high computational time of the modified Poisson model, because the standard errors are “corrected” using the generalized estimating equations. This does not happen when using the standard logistic regression model. In fact, we generally use the logistic model without intercepts, centering the covariates with respect to the overall mean, in order to obtain, as the first step, a series of “adjusted” hospital-specific log-odds (Renzi C, Sorge C, Fusco D, Agabiti N, Davoli M, Perucci CA. Reporting of quality indicators and improvement in hospital performance: the P.Re.Val.E. Regional Outcome Evaluation Program. Health Serv Res 2012, 47: 1880-901).

However, in order to avoid potential misunderstandings, we re-analyzed the data using the modified Poisson regression. The results we obtained were, as expected, identical.

Based on the referee’s consideration, the "Statistical analysis" section has been changed: in the current version of the paper, the paragraph only mentions the modified Poisson regression model for prospective studies with binary data.

2. In the statistical analysis section the authors should explain more clearly that the Jenks natural breaks optimization was used to choose the classes adopted for the geographical maps.
As suggested by the referee, the authors have explained better that the Jenks natural breaks optimization algorithm was used for the calculation of the classes used in the maps [Statistical analysis section. Page 6; lines 18-21].

Results

3. The authors report in Table 1 the predictive model obtained using the standard definition of malposition and malpresentation of the fetus. It could be interesting for the reader to see also the results obtained with the other model that adopt a modified definition of this confounding variable in order to evaluate differences.

As suggested by the referee, the adjusted risk ratio obtained with the model that adopt the “modified definition” of malposition and malpresentation of the fetus was now reported. The risk ratio decreased from 4.46 to 4.02 (95% CI: 3.98 - 4.07; p<0.001) [Results section. Page 10; lines 16-17]. However, the authors would prefer not to show the entire “new” predictive model to avoid burdening the paper with additional, large tables.

4. Figure 3 reports prevalence of HFH for individual Italian Local Health Units in 2010. Is it possible to report the same data with respect, for example, to another year like 2005? This is very helpful to the reader to see in more detail the change that have occurred in the coding of HFH globally. As an alternative it is of interest to report the change over time not only for one hospital but for all the 27 facilities with a HFH prevalence greater than the 95th percentile.

As suggested by the referee, we now reported the change over time related to the whole set of the 27 hospitals having HFH prevalences greater than the 95th percentile. HFH prevalence has increased progressively from 2005 to 2010, showing the following trend: 14.6%, 16.2%, 16.6%, 17.4%, 21.9% and 25.6% [Results section. Page 9; lines 19-21].

5. Moreover the two risk-adjustment models comparison is restricted to 27 maternity with HFH prevalence greater than the 95th percentile. It could be interesting to see the same results for a subgroup of facilities that behaved in virtuous manner. This could help to see the effects of a differential misclassified confounder also in another situation. I expect that no difference should be reported but it is interesting to see if unpredictable consequences with respect to the magnitude and direction of bias in the adjusted estimate occur.

The suggestion of the referee is very interesting. We selected a random sample of 35 maternity units that behaved in a “virtuous manner” (defined as having percentages of HFH in women that underwent CS delivery less than or equal to the analogous national percentage, in order to minimize the probability that the HFH diagnosis could have been over-coded, in an attempt to justify the surgical procedure). In the following table (table A), specifically performed according to the referee’s suggestion, we compared the “standard definition” to the “modified definition” within the sample of “virtuous” maternity units.

As expected, no substantial differences were observed between the resulting risk-adjusted CS proportions. As previously stated, the authors would prefer not to show all the results, to avoid burdening the paper with additional tables (the article already includes: 1 table, 6 figures plus the Appendix). However, in the current Results section, we have synthesized this important finding in a specific sentence. Page 10; lines 22-24 and page 11; lines 1-2.
Table A. The comparison of the standard and modified definition in a random sample of “virtuous” maternity units.

<table>
<thead>
<tr>
<th>Crude CS proportion (%)</th>
<th>Adjusted CS proportion (%)</th>
<th>Adjusted CS proportion (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Standard definition</td>
<td>Modified definition</td>
</tr>
<tr>
<td>22.46</td>
<td>26.59</td>
<td>25.53</td>
</tr>
<tr>
<td>35.29</td>
<td>27.27</td>
<td>27.33</td>
</tr>
<tr>
<td>24.68</td>
<td>18.28</td>
<td>19.79</td>
</tr>
<tr>
<td>17.89</td>
<td>12.82</td>
<td>13.60</td>
</tr>
<tr>
<td>12.62</td>
<td>13.89</td>
<td>13.43</td>
</tr>
<tr>
<td>14.32</td>
<td>13.08</td>
<td>12.84</td>
</tr>
<tr>
<td>79.41</td>
<td>74.43</td>
<td>74.73</td>
</tr>
<tr>
<td>30.85</td>
<td>36.23</td>
<td>34.93</td>
</tr>
<tr>
<td>18.61</td>
<td>17.30</td>
<td>17.16</td>
</tr>
<tr>
<td>58.50</td>
<td>63.55</td>
<td>62.04</td>
</tr>
<tr>
<td>22.38</td>
<td>23.65</td>
<td>23.32</td>
</tr>
<tr>
<td>43.07</td>
<td>42.51</td>
<td>43.47</td>
</tr>
<tr>
<td>43.80</td>
<td>48.73</td>
<td>46.10</td>
</tr>
<tr>
<td>44.62</td>
<td>52.98</td>
<td>51.11</td>
</tr>
<tr>
<td>36.37</td>
<td>42.25</td>
<td>41.09</td>
</tr>
<tr>
<td>27.43</td>
<td>32.45</td>
<td>31.09</td>
</tr>
<tr>
<td>32.75</td>
<td>36.18</td>
<td>35.51</td>
</tr>
<tr>
<td>11.16</td>
<td>11.97</td>
<td>12.02</td>
</tr>
<tr>
<td>58.31</td>
<td>66.58</td>
<td>65.08</td>
</tr>
<tr>
<td>44.57</td>
<td>58.86</td>
<td>56.77</td>
</tr>
<tr>
<td>23.33</td>
<td>17.34</td>
<td>18.15</td>
</tr>
<tr>
<td>10.99</td>
<td>5.37</td>
<td>5.51</td>
</tr>
<tr>
<td>22.97</td>
<td>22.48</td>
<td>21.98</td>
</tr>
<tr>
<td>58.15</td>
<td>61.14</td>
<td>60.23</td>
</tr>
<tr>
<td>17.27</td>
<td>18.18</td>
<td>17.66</td>
</tr>
<tr>
<td>26.36</td>
<td>20.38</td>
<td>21.48</td>
</tr>
<tr>
<td>31.89</td>
<td>35.41</td>
<td>34.63</td>
</tr>
<tr>
<td>31.86</td>
<td>35.25</td>
<td>35.60</td>
</tr>
<tr>
<td>26.01</td>
<td>34.41</td>
<td>32.42</td>
</tr>
<tr>
<td>33.75</td>
<td>34.11</td>
<td>33.84</td>
</tr>
<tr>
<td>14.96</td>
<td>12.46</td>
<td>11.73</td>
</tr>
<tr>
<td>29.66</td>
<td>27.49</td>
<td>29.81</td>
</tr>
<tr>
<td>40.96</td>
<td>46.55</td>
<td>45.58</td>
</tr>
<tr>
<td>15.68</td>
<td>16.60</td>
<td>16.50</td>
</tr>
<tr>
<td>17.46</td>
<td>13.86</td>
<td>13.78</td>
</tr>
</tbody>
</table>
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.

-------------------------------------------------------------------------------------------------------------------------

Additional editorial request:

Data availability - Please update your manuscript to clearly state whether the data used for your study are publicly available, and include the relevant link or reference to the data source. Alternatively, if the data are not publicly available, please revise your manuscript to clearly state whether you received permission to use these data, and who granted this permission.

As requested by the Editor, the modalities of access to data have been clarified [Data sources and study population section. Page 5; lines 5-6]. In addition, we added a bibliographical reference that indicates the link to the website. The results provided by the National Outcome Program are updated every year and are publicly available, including the data analyzed in this study. To access all the information, username and password may be directly requested to the authors of this paper.