Author's response to reviews

Title: Screening for inter-hospital differences in cesarean section rates in low-risk deliveries using administrative data. An initiative to improve the quality of care

Authors:

Willem Aelvoet (willem.aelvoet@health.fgov.be)
Francis Windey (francis.windey@health.fgov.be)
Geert Molenberghs (Geert.Molenberghs@uhasselt.be)
Hans Verstraelen (hans.verstraelen@Ugent.be)
Patrick Van Reempts (Patrick.Van.Reempts@uza.be)
Jean-Michel Foidart (jmfoidart@ulg.ac.be)

Version: 4 Date: 29 June 2007

Author's response to reviews: see over
Dear Editor,

Please find herewith enclosed the revised manuscript, figure and supplementary tables. Please find as follows our point-by-point description of the changes made.

Answer review KS Joseph

1. In our view the 1-minute Apgar scores in our study population are rather a process indicator than an outcome: we consider that a newborn, having a 1-minute Apgar < 4 should have been delivered by a CS. (See page 7 in the resubmitted manuscript). Since Apgar scores may be confounded by socio-economic status of the mother it may be erroneous only to consider the associations between Apgar score and CSR group. Further since all the determinants are mutually adjusted one should also consider the associations between Apgar score and mode of delivery and between Apgar score and CSR group. It may then that the interaction term between CSR and mode of delivery (“departure from additivity”[1]) is also important. Such interactions may indicate that the association of the mode of delivery and Apgar score is not constant over CSR groups. For instance, when comparing the “average CSR” group and the “high CSR” group the OR of 1.22 means that caesarean sections in the “average CSR” group are associated with more Apgar 0-3 scores than vaginal deliveries in the “high CSR” group. In the resubmitted manuscript p 21 we merely state that this finding may indicate a problem of over-use. Regarding the “low CSR” we found that there were more 1-minute Apgar <4 scores than in the average group and no significant interaction as well as several other less favourable neonatal outcomes, and we mention on p. 20 that this group “seemed to perform worse”.

2. We agree that this type of data is insufficient to establish “overuse”. Therefore we conclude (p 21) that our results are a useful screening rather than firm statements.

3. Although the purpose of the study aims at identifying and comparing inter-hospital and inter-regional CSR variations, we presented in table 2 both the total number of cases of each neonatal endpoint irrespective of mode of delivery as well as the number and percentage of those cases, delivered by CS. We merely used neonatal endpoints, except 1-minute Apgar, to assess whether good processes led to good outcomes.

4. We defined outlying hospitals (p7) as hospitals with a departure of < +35 or < -25 and statistically significant. Regarding the trend the boundaries were set at +5 and –5. In the cited example the boundary criterion was met but not the criterion of statistical significance, see p 12 of the resubmitted manuscript.

5. We apologize for the misleading term of “transfer” to a specialized neonatal service. As a matter of fact the term “Admission to a specialized neonatal service” should have been used, by which we mean that a child born in a maternity ward of a hospital is admitted in this type of service in the same hospital. Our data do not allow the follow-up of transferred newborns. We adapted our manuscript accordingly (Tables 2 and 5, p7). We tried to address the concerns about the low rate of referral by showing that the value of administrative data with respect to an admission to a specialized neonatal service is questionable p13.
We also addressed the concern regarding the twofold variation in CSR by weekday not only in the Results (Section National CSR and its determinants p 12) but also in the main findings of the discussion (p16) and in the Conclusions (p 21). We hypothesized that obstetrical intervention and structural issues may intervene in the phenomenon (p 16)

6. According to the reviewers’ remarks elective caesarean has been defined in the Introduction p1.

7. The manuscript has been restructured and reedited, hoping it will be more readable by now.

8-9. We tried to avoid the inconsistent use of commas and periods, and to avoid typographical errors (see Table 2 for instance).

Minor essential revisions:

P14 and 15: the mentioned speculation has been removed from the Results Section

P6 The sentence has been reworded on page 5: “The sampling frame consisted of the 455,933 deliveries, involving live born singletons in vertex presentation that were registered from 2001 to 2004. Thereof 86,310 (18.96%) were classified as cesarean deliveries. By applying the aforementioned criteria the final data set, the study population, comprised 381,989 deliveries of which 49,578 (12.98%) by cesarean section. Out of the 73, 944 deliveries, not meeting these criteria, 36,732 (49.68%) were cesarean deliveries, the comparison population.”

The term IUGR has been replaced by full-term small-for-gestational age (defined as newborns born after 36 completed weeks of gestation with a birth weight of less than 2500g)

P8 According to the remark we regrouped the Apgar scores into three categories: “Apgar 0-3”, “Apgar 4-6”, and “Apgar > 6” (p 7)

The p-values have been corrected up to 2 decimals except for the Bonferroni correction where smaller p-values are required.

The definition of the p-value has been adapted to the remark of the reviewer (Table 3).

The tables have been changed according to the remark of the reviewer.
1. Elective has now been defined on p 1. Indeed our criteria of low-risk is a rather general one and omits to include placenta previa, previous uterine surgery, abruption, preeclampsia, and diabetes. We acknowledge these shortcomings on p.17 and explain them by the recognized, weak case identification from administrative data.

We also mentioned on the same page the medical necessity of systematically carrying out a CS in case of diabetes without macrosomia [2] and hypertensive disorders (except some cases of eclampsia with lasting fetal suffering) [3] has not yet univocally been established. Conversely, mothers suffering from pathologies such as placenta previa and congenital anomalies may be considered at risk of rightly undergoing a CS.

Since placenta previa is a rather uncommon disease (in about 0.5% of the pregnancies, of which almost half of the cases are delivered before 37 weeks)[4,5] and having no reason to suppose an important clustering of cases, we think that our analyses would not much have changed by excluding these cases. On the other hand, our analyses excluding congenital anomalies (data not shown) gave almost identical results with the analyses where these cases were included.

The semester has now been defined as a six-month period, the first semester comprising the first six months of the calendar year (p5)

The study population and comparison population have now been defined as follows on p 5: The sampling frame consisted of the 455,933 deliveries, involving live born singletons in vertex presentation that were registered from 2001 to 2004. Thereof 86,310 (18.96%) were classified as cesarean deliveries. By applying the aforementioned criteria the final data set, the study population, comprised 381,989 deliveries of which 49,578 (12.98%) by cesarean section. Out of the 73,944 deliveries, not meeting these criteria, 36,732 (49.68%) were cesarean deliveries, the comparison population.

On page 5 and 6 we have reformulated our definitions as follows: “It has been suggested that in analyses, founded on administrative databases, confounding cannot be ruled out as an explanation of rather small, yet statistically significant effect sizes, such as a relative risk (RR) of 0.75,[6] Therefore we defined a zone of non-interpretation, where the CS rate or trend of a hospital, compared with the national ones, should not be described as being “higher” or "lower". To determine the boundaries of this zone we firstly computed per hospital the relative risk (RR) of a hospital of having a higher/lower CS rate or trend than the national the national ones. We then calculated a departure D (expressed in %): with the formula $D = (RR - 1) \times 100$. Subsequently we defined the lower boundary as corresponding to a departure of minus 25 – which is equivalent to the afore-mentioned RR of 0.75 – and the upper boundary as a departure of plus 35, the lower boundary’s approximate, statistical counterpart.

In the absence of any references regarding the significance of departures from the CS trend and by assuming that data quality has remained constant over time, a similar zone of non-interpretation was defined to allow for a comparison in the evolution over time in the hospital-specific CS rates by which we arbitrarily allowed for a – 5 to + 5 departure from the national trend.”
2. In accordance with the reviewer’s concern we now stated on page 7 that we considered the 1-minute Apgar score a process rather than an outcome indicator.

4. We have replaced the expression of “anamnesis of a previous caesarean delivery” by “a history of ….” (p 10). According to table 1 an incompleteness of almost 40% of cases seemed to us an important disagreement.

5. This paragraph has been reformulated at page 10 as follows:” Comparing the study population (381,988 deliveries from which 12.98% CS) with the comparison population (73,944 deliveries of live born singletons (in vertex position) from which 49.68% CS), we found in the latter population a relative risk of being delivered by CS of 3.83, 95% CI (3.79; 3.87). Applying to our source population the basic triad of “mothers with singleton, full-term (37 weeks and more) births involving a vertex presentation”, recently used to describe maternal risk profiles[7], we would have had 395,021 low risk deliveries, giving rise to 52,611 CS and to a relative risk of 1.12, 95% CI (1.10; 1.13).”

We defined outlying hospitals (p7) as hospitals with a departure of < +35 or < -25 and statistically significant. Regarding the trend the boundaries were set at +5 and −5 and the significance criterion was kept. An example is provided on page 12.

6. Due to the privacy regulations in our country such data are not available in our MCD and without a long-lasting procedure they may not been obtained from other sources.

7. Since Apgar scores may be confounded by socio-economic status of the mother it may be erroneous only to consider the association between Apgar score and mode of delivery. It may then that the interaction term between CSR and mode of delivery is also important. Such interactions may indicate that the association of the mode of delivery and Apgar score is not constant over CSR groups. For instance, when comparing the “average CSR” group and the “high CSR” group the OR of 1.22 means that caesarean sections in the “average CSR” group are associated with more Apgar 0-3 scores than vaginal deliveries in the “high CSR” group. In the resubmitted manuscript p 21 we merely state that this finding may indicate a problem of over-use. Regarding the “low CSR” we found that there were more 1-minute Apgar <4 scores than in the average group and no significant interaction, and we mention on p. 20 that this group “seemed to perform worse”.

8. We have reorganised and shortened the Discussion thoroughly. Subheadings have been provided and, for instance, the limitations of the study have been regrouped. We recognize that we do not have medical and obstetrical diagnoses in detail as mentioned under 4. We further agree that this type of data is insufficient to establish “overuse”. Therefore we conclude (p 21) that our results are a useful screening rather than firm statements.

9. The tables have been reedited according to the reviewer’s remarks. Some of them have been omitted from the manuscript and are now made available as supplementary material. Period and trend have now been defined in the Methods section page 5 and 8. P-values have been reduced to 2-digits except for the Bonferroni correction where smaller p-values are required. They are similar to the classical bounds of a 95% CI, but this time the 0.05 alpha has been divided by the number of simultaneous comparisons made and hence rendering the interval much larger.
In this table the odds ratios are relative to the category preceded by the word “vs” and are mutually adjusted. P-values have now been added. In table 6 not only Apgar 0-3 but also Apgar 4-6 have been tested. Table 7 has now been removed from the manuscript.

According to the reviewer’s remark, the duplex graphs have been reduced to the one graph including the congenital anomalies.

Answer review ER Declercq

The manuscript has been reduced and reorganised. A more judicious comparison has been made regarding the characteristic of low risk pregnancy on page 10.

We have focussed on the quality aspect as explained in the main goals on page 6.

The last two pages have been skipped by now.

We explained why the rates on induction are so much lower in the sample on page 18: “Finally, part of the limitations of administrative data may be due to the basic tension which exists between using the same data for reimbursement and for measuring quality. “When the use is reimbursement, there is a tendency to perform coding quickly and to maximize the coding of complications and comorbidities. When the use is to assess quality, however, it is important for coders to have a complete record and to restrict diagnosis coding to conditions that affect patient care.” For instance, hypertension and diabetes may intervene in the algorithm used to determine the case mix of an admission and thus be rewarding in financial terms, whereas this may not be the case for labor induction, epidural anesthesia and history of a previous cesarean.”

In our dataset the history of a previous CS is largely under-registered. Hence we did not make an attempt at distinguishing primary or repeat CS. If we would have done this an overestimation of primary CS would undoubtedly have been the consequence and this is the reason why we did not make such an attempt.
Reference List


