Reviewer's report

Title: Congenital anomalies in a polluted area. A ten-year retrospective study

Version: 1 Date: 11 July 2012

Reviewer: Barbara Frohnert

Reviewer's report:

I have 30 comments, each one has the following components:

1. Comment number
2. Revision type (Major compulsory revision, Minor essential revisions, Discretionary revisions)
3. Section (header from the manuscript)
4. Paragraph/Table (number of the paragraph in each section or table title)
5. Comment

Comment number: 1 Revision type: (General comment) Section: Entire paper
Paragraph/Table: Entire paper Comment: Thank you, I enjoyed your manuscript and found your work very interesting and relevant to my work. I am not an experienced reviewer, so I hope you take my comments with that in mind. Nonetheless, I hope you find them helpful. Fundamentally, I think your premise is very interesting. I can see how the analyses build off one another. You have done many interesting analyses, but the quantity of analyses included in your paper distracted me from the main purpose. You have done quite a few related - but different - analyses. Would you consider focusing your paper on the main analyses and perhaps reference your subsequent, supporting analyses in the discussion where they have enhanced your interpretation of your result? Or, you could probably split the analyses into at least three separate papers so you could spend more time discussing the strengths/weaknesses of the assumptions made, the different data sources, etc. I think your EUROCAT comparison is interesting but are the two datasets (Brindisi ICD-9-CM estimates and EUROCAT) similar enough to make this comparison? The results are very compelling, but the limitations merit more discussion before I can understand if Brindisi is really experiencing greater rates of birth defects. The validity study is interesting, but I am not sure it is sufficient evidence to use the results from a single hospital to support using ICD-9-CM codes in general for estimating prevalence estimates. The third analysis comparing the three Brindisi regions also has a lot of promise, but the implications of the assumptions you have made regarding proximity to environmental pollutants should be described in greater detail. I think separate papers could provide more compelling arguments that would be more useful to the field of birth defects research.

Comment number: 2 Revision type: Minor Essential Revisions Section: Abstract
Paragraph/Table: Abstract Comment: The abstract describes only a small portion
of the various analyses included in the paper. I feel the paper does not do your work enough justice because it does not describe any of them in sufficient detail. I had to re-read sections of the paper several times to distinguish the different pieces. They seem like related, but separate projects, each worthy of their own abstract, their own methods, results and discussion.

Comment number: 3 Revision type: Major Compulsory Revisions Section: Background Paragraph/Table: Methods Comment: This background describes the city of Brindisi, but the different analyses extend beyond Brindisi city. The denominator changes in each analysis: Brindisi city 8503 births, NICU hospital: 900+ births, Brindisi province: 35000+ births. It is confusing that the background implies that we are talking about 8503 births throughout when the number/coverage actually changes.

Comment number: 4 Revision type: Discretionary Revisions Section: Methods Paragraph/Table: 1 Comment: Please define a Congenital Anomaly diagnosis more specifically. Briefly describe the number of anomalies, at a minimum general categories so that the reader understands the breadth and variety of the conditions included under the general "CA" abbreviation.

Comment number: 5 Revision type: Discretionary Revisions Section: Methods Paragraph/Table: 1 Comment: What is meant by a "local health unit?" Is this the governmental unit that holds birth records, i.e. that enumerates your denominator (all births in a particular region during a particular time frame)? Or is this the health department or a different entity that gathers hospital discharge data only? Please specify what you mean by "local" Brindisi city or the entire Brindisi region?

Comment number: 6 Revision type: Major Compulsory Revisions Section: Methods Paragraph/Table: 1 Comment: Please speak to the limits of your hospital discharge data. Is it complete? Is it matched with birth certificates to ensure you have included the entire population of births to Brindisi residents?(How do you know you are not missing cases?) What about births that do not occur in a hospital (i.e. no discharge data)?

Comment number: 7 Revision type: Major Compulsory Revisions Section: Methods Paragraph/Table: 1 Comment: Could you clarify that you did not validate the ICD-9-CM codes you obtained from the hospital discharge data, other than to exclude PDA in preterm infants ? It seems you are accepting the ICD-9-CM codes as a diagnosis of a CA without verification from the medical record (i.e. passive rather than active surveillance of birth defects). This is fine if that is what you did, but as a person who works in birth defects surveillance, this is important that you make this distinction.

Comment number: 8 Revision type: Major Compulsory Revisions Section: Methods Paragraph/Table: 1 Comment: I am not familiar with the EUROCAT data. If it is generally based on ICD-9-CMs from HDD, then this comparison is fairly safe, but if EUROCAT estimates are based on data that has been collected using more involved case-verification techniques and more specific case-definitions, it may exclude cases that the Brindisi study includes. Then it
wouldn't be a surprise that the Brindisi rates are higher than the EUROCAT estimates. I can't tell if the two methods of data collection can produce equivalent prevalence estimates, and therefore whether it means anything that the Brindisi rates appear higher. The assumptions you are making could be described more specifically.

Comment number: 9 Revision type: Major Compulsory Revisions Section: Methods Paragraph/Table: 2 Comment: This sentence/paragraph needs to be clarified. Please describe how counting each case with multiple anomalies only once affects the data presented in Table 1 and the total case counts. You had a total of 194 children diagnosed with at least one congenital anomaly in Table 1. If all cases with more than one diagnosed congenital anomaly were included in the "Total CAs " category only, then does the CHD section of Table 1 include only solitary CHDs? If so, your rates are for solitary CHDs only and should be displayed as such. If you are counting defects, not individual children, in that section, it needs to be explained. I don't understand why your numbers don't add up in Table 1 and this sentence does not explain enough to clarify why.

Comment number: 10 Revision type: Minor Essential Revisions Section: Methods Paragraph/Table: 3 Comment: From Table 2, it appears that 994 cases were available from the single NICU hospital, and you reviewed 11.7% which would be approximately 116-117 cases. You must have reviewed the 61 cases with CAs based on the ICD-9-CM codes from Hospital Discharge Data, but what was the selection criteria for the remaining cases that were reviewed? Why did you select the 11.7% you "observed"?

Comment number: 11 Revision type: Major Compulsory Revisions Section: Methods Paragraph/Table: 3 Comment: Why is this single NICU hospital representative of all the birth hospitals in Brindisi? You are assessing the validity of using ICD-9-CM codes from HDD based on what you found at a single hospital with a NICU. Won't the NICU see the sickest children and their conditions would likely be the most fully and rigorously diagnosed, therefore the sensitivity/specificity/PPV/NPV of their HDD would be a "best-case" scenario? I don't think that this is clear in your paper and it likely overstates how good this administrative dataset is for estimating the prevalence of CAs. Why do you think the other Brindisi hospitals would have similarly sensitive/specific ICD-9 CM codes in their HDD? Is this a reasonable assumption to make? (It might be the case in Italy, but I wouldn't make that assumption here in the USA.)

Comment number: 12 Revision type: Discretionary Revisions Section: Methods Paragraph/Table: 3 Comment: This paragraph seems to inform Table 2, yet paragraph 4 returns to Table 1. Perhaps this paragraph should follow the one describing the O/E ratios in Table 1.

Comment number: 13 Revision type: Minor Essential Revisions Section: Methods Paragraph/Table: 3 Comment: Missing period at the end of the sentence.
"Sensitivity, specificity...were calculated."

Comment number: 14 Revision type: Major Compulsory Revisions Section:
Methods Paragraph/Table: 4 Comment: The calculation for the observed/expected ratios with 95% Cis is clear, but the figures displayed in the O/E column in Table 1 appear to be multiplied by 100. The observed vs expected ratio for all anomalies = 194/165.5 = 1.172, not 117.2. Can you clarify this inconsistency?

Comment number: 15 Revision type: Minor Essential Revisions Section: Methods Paragraph/Table: 6 Comment: Please specify that this analysis is for the entire PROVINCE. The results of this analysis are in Table 3, which includes over 35,000 births.

Comment number: 16 Revision type: Minor Essential Revisions Section: Methods Paragraph/Table: 6 Comment: It is a bit confusing when you are discussing Brindisi province vs. Brindisi city. I would suggest that you use the same wording in both the text and the map displayed in Figure 1. Are these three "regional" categories described here the same as those displayed in Figure 1, Province of Brindisi map? Brindisi = Brindisi city, Group A = Neighboring municipalities, Group B = Rest of the municipalities in Brindisi province. Perhaps the balance of Brindisi province could be described as "outlying municipalities in Brindisi province". That might be clearer to the general, non-Italian readers.

Comment number: 17 Revision type: Minor Essential Revisions Section: Methods Paragraph/Table: 6 Comment: What is the connection between the three regional groupings? Is this the level at which you have the social deprivation index data, or are the emission sources from Figure 1 primarily located in the first category (Brindisi city) ? It seems you are implying that the municipality of residence is associated with environmental emissions and you have a map with their locations, can you clarify the relationship further?

Comment number: 18 Revision type: Major Compulsory Revisions Section: Methods Paragraph/Table: 7 Comment: How did you select the cases for the follow up? The results section says there were 109 cases observed from 2001-2005. I am confused to how much time the follow up included. Did you follow up on the 2001 cases for 4 years, 2002 cases for 3 years, 2003 cases for 2 years, etc.?

Comment number: 19 Revision type: Minor Essential Revisions Section: Methods Paragraph/Table: In general Comment: Consistency with definition of "Brindisi" - city, province? Can you please go through your article and clarify between Brindisi province vs. the municipality of Brindisi? This section includes 8,503 live births to women who were residents of the CITY/MUNICIPALITY of Brindisi. The validity study was at a SINGLE NICU HOSPITAL somewhere in Brindisi province, possibly in the municipality of Brindisi, and covers only 994 births in the same timeframe. The third analysis splits Brindisi PROVINCE into three sections, Brindisi municipality, neighboring municipalities, and outlying municipalities. It appears to cover over 35,000 births in this time period.

Comment number: 20 Revision type: Discretionary Revisions Section: Results Paragraph/Table: 6 Comment: I am not sure that the 5 year follow up on
readmissions is really central to your article and sufficiently discussed. Maybe write this up in a separate paper. What is the time-frame for follow up? Only 59.6% had further admissions, but some of these cases may only have been tracked for 1 year while others for close to 5. The selection of cases to be followed and the duration of follow up is not clear so I can't make sense of the 59.6% figure, nor that a minority (38.5%) were admitted for same diagnosis.

Comment number: 21 Revision type: Minor Essential Revisions Section: Results Paragraph/Table: all tables Comment: Table 1, 3, 4: In the titles, please clarify which specific region the data is from: HOSPITAL, CITY, or PROVINCE. It's confusing because they all say Brindisi, but they are all different parts/populations/denominators, and the differences matter.

Comment number: 22 Revision type: Discretionary Revisions Section: Results Paragraph/Table: Table 1 Comment: You include a column with the rates for live births in Italy but you do not describe it. Is Brindisi not included in this data? If the column isn't central to your discussion/argument, then maybe you should exclude this column from your table.

Comment number: 23 Revision type: Major Compulsory Revisions Section: Results Paragraph/Table: Table 1 Comment: As mentioned above, the O/E column and corresponding 95% CI don't correspond to the formula described in paragraph 4 of the methods section. Shouldn't the ratio be 1.17?

Comment number: 24 Revision type: Major Compulsory Revisions Section: Results Paragraph/Table: Table 2 Comment: Title: add "at a single NICU facility." From the methods section, these data are from a single NICU hospital with 994 births. Table 1 has results from the entire province and 8603 births. Please clarify.

Comment number: 25 Revision type: Minor Essential Revisions Section: Results Paragraph/Table: Table 3 Comment: Table 1 includes 194 CA cases, 84 CHDs in Brindisi (city?), Table 3 includes 193 CA cases, and 92 CHDs. Can you describe why these case numbers don't match?

Comment number: 26 Revision type: Discretionary Revisions Section: Discussion Paragraph/Table: 1 Comment: Why include the re-admission follow up if you do not include it in the discussion? What does it mean that only 38.5% of cases were readmitted with the same early-diagnosed birth defects? Were the diagnoses the kinds of birth defects that are less likely to be diagnosed by 28 days?

Comment number: 27 Revision type: Minor Essential Revisions Section: Discussion Paragraph/Table: 5 Comment: Clarify "high probability of corrected recording." "Corrected recording" is unclear. Suggest "diagnoses were coded correctly" instead.

Comment number: 28 Revision type: Major Compulsory Revisions Section: Discussion Paragraph/Table: 5 Comment: The validity study of ICD-9 CMS from HDD was conducted at a single, specialty facility. CA s that require less intensive
care may not go to NICU facilities and may not be coded as accurately. I feel the representativeness of this particular facility is not adequately addressed in the discussion. The study by Frohnert (that's me, by the way) was also at a single NICU facility. This was a major limitation to its applicability to all HDD data. If this is not also true for this study, why not?

Comment number: 29 Revision type: Major Compulsory Revisions Section: discussion Paragraph/Table: 7 Comment: This study uses mother's residence as a proxy for environmental exposures. Without any true exposure data, this is a large limitation to this attempt at an environmental association and it is not specifically mentioned as a limitation in the discussion. That would be an interesting discussion to read. If you do not have exposure data, could you improve upon this analysis by estimating the proximity to these polluting sites to from mother's residence? Would that be a logical next step? Perhaps you can include the direction for your next research in the discussion.

Comment number: 30 Revision type: Minor Essential Revisions Section: References Paragraph/Table: 15 Comment: You don't cite the article by Frohnert et al in the paper. It would guess it was originally used somewhere in the Methods/Paragraph 3, since they are very similar studies. It doesn't seem to be specifically referenced in the text or else the reference is unclear/missing. If you decide to remove it, you will need to renumber your references.

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Acceptable

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.