Author's response to reviews

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

Authors:

Emilio A L Gianicolo (emilio.gianicolo@ifc.cnr.it)
Antonella Bruni (antonella.bruni@ifc.cnr.it)
Enrico Rosati (enricorosat@tiscali.it)
Saverio Sabina (sabina@ifc.cnr.it)
Roberto Guarino (guarino@ifc.cnr.it)
Gabriella Padolecchia (g.padolecchia@asl.brindisi.it)
Carlo Leo (controllogestione@asl.brindisi.it)
Maria Angela Vigotti (vigomar@ifc.cnr.it)
Maria Grazia Andreassi (andreas@ifc.cnr.it)
Giuseppe Latini (gilatini@tin.it)

Version: 2 Date: 9 November 2012

Author's response to reviews: see over
Dear Dr. Gissler,

Many thanks for the opportunity reserved to our work.

Please find attached the revised version of our manuscript “Congenital anomalies in a polluted area. A ten-year retrospective study” - ID 8572986747116042. The changes we made are highlighted using bold and red coloured text.

On May 28, 2010 (reference number 38131), the ethics committee of the hospital of the city of Brindisi approved our research. Research was in compliance with the Helsinki Declaration. We revised the text accordingly, adding a statement in the Methods section of the manuscript, including the name of the body, which gave approval, with a reference number.

The revised manuscript includes all the reviewers’ comments.

Best regards,
Emilio Gianicolo on behalf of the authors

Responses to Reviewer 1

C1 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.

A1 We revised our manuscript, focusing on the validity analysis and the local and international comparison analysis. The rest of our supporting analysis (i.e. further hospital admissions) was actually not central to our article and we decided to focus on it in a separate paper, thus receiving the reviewer suggestion.

For that regards the comparison with EUROCAT, we used the same criteria to define cases with a congenital anomaly. As reported by Lechtat and Dolk (1993), registries in EUROCAT use the common system of the British Paediatric Association Classification of Diseases ICD-BPA. Thus, in order to have comparable datasets we used the corresponding ICD-9-CM codes.

The validity study, conducted together with expert neonatologists and a geneticist, supports the use of hospital discharge data (HDD) for estimating CA in the city of Brindisi. Nevertheless, the use of HDD in the rest of the country has to be further evaluated.

For what regards the local comparing analysis, the emission sources are mainly located in the city of Brindisi which is a government-designated “Environmental risk area”. However, the aim of logistic regression analysis was to verify if there was an association between the municipality of residence and the prevalence of birth defects. In this paper we didn’t have any personal or ecological exposure data. Therefore, it was impossible to infer about the possible role of pollution.

C2 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.

A2 The aim of our paper was to estimate prevalence of congenital anomalies in the city of Brindisi and give to the readers some indications about the validity of the data used and a comparison with local and international figures. Therefore, we decided to include in the paper both methods/results on validity and methods/results on national and international comparison.

C3 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.

A3 In the background section we added more information about the Province of Brindisi.

C4 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.

A4 The reviewer is right, we revised the text by including a table (now table1) with the sub-groups of congenital anomalies considered and corresponding ICD-9-CM diagnosis codes.

C5 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?

A5 Italian local health units are autonomous body of the National Health System. They organize and plan the health care systems for specific areas. Local health units hold all health records, which are mainly used for financial and economic purposes (i.e. billing). Hospital discharge data are part of health record and we used them in order to identify births with or without congenital anomalies. The Brindisi local health unit provide services to the population resident in the province of Brindisi. NICU is a division of the hospital sited in the city of Brindisi. In NICU a detailed and independent medical archive is held since 1986. We revised the text specifying these aspects.

C6 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)?

A6 HDD records were not matched to birth certificates. This could have brought to misclassification of cases in the different regional areas. However, this potential misclassification cannot be assumed as differential.

Reviewer is right, there could be births that did not occur in a hospital. Nevertheless, according to the Italian National Institute of Statistics – ISTAT (http://www.istat.it/it/archivio/7740), the percentage of birth not occurring at home is equal to 0.1%, mainly concentrated in the north part of Italy. Therefore, this phenomenon is rare in the investigated area and we can assume it had any effect on the estimated prevalence of CAs.
C7 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.

A7 Medical records were recovered in order to identify the gestational week and to correctly classify newborns affected by PDA. Validation of ICD-9-CM codes regarded cases admitted in the local NICU. Hence, in our paper active surveillance could be correctly referred only to cases with PDA and to cases admitted in NICU.

C8 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.

A8 This is a very important issue, which has been now addressed in the methods paragraph. Registries in EUROCAT use the code system of the British Paediatric Association Classification of Diseases, which is a five-digit extension of the ninth revision of the International Classification of Diseases (ICD-9). In order to have comparable data set we used the corresponding ICD-9-CM codes.

According to Lechtat and Dolk (1993), in EUROCAT “registration may take the form of a national system of specific notification of congenital anomalies (England and Wales, Finland, Sweden, Czechoslovakia, and Hungary), of centralized birth notifications (Norway, Belgium, and Sweden), or of neonatal discharge records (Scotland), or specialized (often regional) registries using multiple sources of information.” Therefore, the use of HDD as the only data source could represent a limitation of the study. Registries in EUROCAT register information about newborns and the mothers using different data sources (Greenlees R et al., Birth Defects Research (Part A): Clinical and Molecular Teratology, 2011). In work, in order to validate HDD, for a subgroup of records, we conducted a validity study with the support of expert neonatologists and a geneticist. Results confirm that HDD can be used in the city of Brindisi to estimate the prevalence of CAs.
C9 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.

A9 The table includes: cases with a CA ("All anomalies"); cases with a CHD and cases with other anomalies. Numbers do not add up because of newborns with more than one CA. For what regards CHD, we considered only those subcategories with at least 5 cases. The text has been revised accordingly.

C10 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"?

A10 Of 8,503 newborns from mother resident in Brindisi, 994 (11.7%) were admitted in the NICU division and registered in an independent and very detailed archive which was used for validation purposes. The text has been revised.

C11 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.)

A11 NICU admitted only a part of the newborns from mothers resident in Brindisi, we cannot assume that they are representative of all births in the city of Brindisi but we might suppose it represent a quite robust sample. It is true that in NICU we observed “the sickest children and their
conditions would likely be the most fully and rigorously diagnosed” but this may be not enough to assume NICU as the best-case scenario for correct recording of HDD. Further research is needed to investigate this aspect.

C12 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.

A12 The text has been revised accordingly.

C13 Missing period at the end of the sentence. "Sensitivity, specificity...were calculated."

A13 The text has been revised accordingly.

C14 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?

A14 Reviewer is right. Figures displayed in the O/E column are multiplied by 100. This inconsistency has been revised.

C15 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.

A15 The text has been revised accordingly.

C16 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.

A16 We agreed with the reviewer and revised the text accordingly.

C17 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?
The emission sources are mainly located in the city of Brindisi which is a government-designated “Environmental risk area”. However, the aim of logistic regression analysis was to verify if there was an association between the municipality of residence and the prevalence of birth defects. In this paper we didn’t have any personal or ecological exposure data. Therefore, it was impossible to infer about the possible role of pollution. Nevertheless, our working group is now focusing on exposure data and on the correlation between environmental data and health data. This will be the aim of our next paper.

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.?

We decided to focus on this aspect in a separate paper.

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.

We went through our paper and clarified the differences between city and province of Brindisi.

I am not sure that the 5 year follow up on readmissions is really central to your article and sufficiently discussed. May be 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.

We decided to focus on this aspect in a separate paper.

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.

The text has been revised accordingly.
C22 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.
A22 Reviewer is right. The column is not central to our paper. We excluded it from table 1.

C23 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?
A23 The ratio was multiplied by 100, the text has been revised accordingly.

C24 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.
A24 The title table has been revised accordingly. See also answer number 10.

C25 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?
A25 These case numbers do not match because in table 3 cases with PDA were excluded. Since it was not possible to identify the exact gestational age for each newborn to mother resident in municipality other than the city of Brindisi. For the city of Brindisi, we observed a case of PDA with 40 weeks gestational week and without any other defects. Hence, in order to have homogenous comparison, we decided to exclude this case from the analysis.

C26 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?
A26 We decided to focus on this aspect in a separate paper.

C27 Clarify "high probability of corrected recording." "Corrected recording" is unclear. Suggest "diagnoses were coded correctly" instead.
A27 The text was revised accordingly.

C28 The validity study of ICD-9 CMS from HDD was conducted at a single, specialty facility. CAs 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?

A28 This issue has been addressed in the revised discussion.

C29 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 from mother's residence? Would that be a logical next step?

Perhaps you can include the direction for your next research in the discussion.

A29 We have included in discussion the direction of our next research.

C30 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.

A30 Frohnert et al. was the reference number 15 (now number 21). It was cited in methods. In the revised paper it has been cited also in discussion.

Responses to Reviewer 3

Are the discussion and conclusions well balanced and adequately supported by the data? yes, but you must add these two article:


Both articles have been included among the references.