Author's response to reviews

Title: Adjusting for under-identification of Aboriginal and Torres Strait Islander births in time series produced from birth records: using record linkage of survey data and administrative data sources

Authors:

David Lawrence (dlawrence@ichr.uwa.edu.au)
Daniel Christensen (dchristensen@ichr.uwa.edu.au)
Francis Mitrou (francism@ichr.uwa.edu.au)
Glenn Draper (glenn.draper@abs.gov.au)
Geoff Davis (Geoff.Davis@health.wa.gov.au)
Sybille McKeown (sybille.mckeown@abs.gov.au)
Daniel McAullay (d.mcaullay@ecu.edu.au)
Glenn Pearson (glennp@ichr.uwa.edu.au)
Stephen R Zubrick (steve@ichr.uwa.edu.au)

Version: 3 Date: 18 May 2012

Author's response to reviews: see over
We would like to thank all four reviewers for their helpful and thoughtful comments on our paper. We have revised the manuscript in line with the reviewers’ suggestions. We feel that the paper has been significantly improved thanks to the valuable feedback received from each of the reviewers. We have detailed below the specific changes that have been made to the manuscript in response to each of the individual points raised by the reviewers.

**Reviewer 1**

1. **Use of the terms “Indigenous” and “Aboriginal” or “Aboriginal and/or Torres Strait Islander” should be made consistent**

Based on consultations with the Aboriginal Steering Committee that oversaw the WAACHS we have been advised to use the term “Aboriginal and/or Torres Strait Islander” unless referring to broad concepts such as Indigenous status. We have corrected the language in the paper to reflect this standard, unless a direct quotation from other work.

2. **A definition of Aboriginal and/or Torres Strait Islander infant needs to be stated.**

We have added a definition of Aboriginal and/or Torres Strait Islander infant, and note that the two administrative data sources used in this paper do not conform to this standard, particularly the Midwives data which only records status of the mother.

3. a) **Results paragraph 1. Much of the first part of this paragraph refers to methods. I would suggest the text be moved to the methods section. The rest of the paragraph should be moved to the discussion.**

   The first part of the paragraph has been moved to the methods section. The remaining material has been moved to the discussion.

   b) **Paragraph 2. The first sentence repeats information in the methods and should be removed.**

   The first sentence has been removed.

   c) **Paragraph 3. This paragraph does not refer to any results and should be removed.**

   The paragraph has been removed.

4. **Results. Characteristics of children who would be identified differently between administrative and survey data sources.**

   a) **Paragraph 1. Sentence 1 and 2 refer to “the survey”—which of the two survey sources is this?**

   This has been clarified to refer to the WAACHS.

   **Sentence 2 also refers to “both” sources. However, there are three data sources mentioned in this paragraph—it is a bit confusing where “administrative data sources” are referred to as a single “data source”. The wording here should be clarified.**

   This has been clarified to note that all three data sources were consistent.

   b) **Paragraph 2. References to “the survey” and “both” as per comment 5. above.**

   This has been fixed.
5. Results. Inconsistent identification of non-Aboriginal children in the 1993 WA Child Health Survey. The first part of this paragraph (up to “as shown in Table 5”) repeats information that was previously presented in the Methods. This should be replaced by a short introductory sentence or phrase.

This redundant material has been removed.

6. Results. Time series of low birth weight and low gestational age babies. Information on methods that has previously been described in the Methods section should be removed.

This material has been deleted.

7. Discussion. As per 3.a) above, it is a bit confusing where “administrative data sources” are referred to as a single “data source”. The wording here should be clarified.

The wording has been clarified throughout this section to make clear that there are two administrative data sources.

8. Discussion. Paragraph 10. Last sentence It is not clear what is meant by “neonatal status” as an outcome variable. Does this refer to the low birth weight and prematurity outcomes previously discussed? If so, then either a definition of neonatal status should be included in the Methods or the outcomes should be specifically stated here.

This has been changed to specifically refer to low birth weight and gestational age. The term “neonatal status” has been removed.

Minor essential revisions

1. On occasion, the term “demographic data” is used to refer to “reporting of Aboriginal and Torres Strait Islander peoples” (e.g. Background, paragraph 4). The term “demographic” should be replaced by “Aboriginal and Torres Strait Islander peoples” or a similar descriptive term where relevant.

“demographic” has been replaced by “Aboriginal and/or Torres Strait Islander status”

2. Background, paragraph 4. In order to clarify the study aims, the term “administrative data collection/s” should be replaced or supplemented by the name/s of the data collection/s. The names of the two administrative data collections have been added for clarity.

3. Background, paragraph 8. “As there is…”, should be replaced by “As there are…”.

This has been fixed.

4. Methods, Analysis methods, paragraph 1. The last sentence refers to “children and young people”. However, this study concerns neonates. The wording here needs clarification.

This has been fixed.

5. Methods, Analysis Methods, paragraph 3. Information on the measurement of “level of isolation of residential address” and “family structure” should be included in the Methods.

Details of the derivation of both level of relative isolation and family structure have been included.
6. Methods. Adjusting administrative-based time series, last paragraph. “…summing the probabilities within categories of neonatal outcomes” should be changed to “…summing the relevant probabilities” as the outcomes have been previously referred to in the same sentence.

This has been amended as suggested.

7. Results. Characteristics of children who would be identified differently between administrative and survey data sources, paragraph 1. The first sentence refers to “outcomes”. However, the variables listed are not all outcomes. The term “outcomes” should be changed to “indicators” or “measures” or a similar term.

This has been amended as suggested.

8. Discussion. The heading “Limitations” should be removed as there are no other headings in the Discussion. An introductory phrase could be included at the beginning of the next paragraph.

This heading has been removed as suggested.

9. Tables and Figures. The location and relevant years should be included in the title. For example, Table 1 should be modified to read “…weighted survey estimates, Western Australia [YEAR] to [YEAR].

We have added “Western Australia” to the titles. As the WAACHS data collection was spread over a two year period, and the children ranged in age at recruitment into the survey, the survey estimates do not pertain to a specific range of years.

10. Tables and Figures. Footnotes to each table indicating the scope of data reported in the table would be very helpful e.g. Linked records of [X] and [Y] datasets, or All records of [X] dataset and linked records of [Y] dataset. For tables where weighted survey data are reported, a footnote should refer to the number of records in the survey that are represented in the table.

Footnotes have been added to each table to indicate the number of sample records contributing to the estimates, and the scope of the data set. We have also added an additional figure to describe the linkage process and sample count in more detail.

11. Tables 1, 2 and 5. The inclusion of row or column percentages as relevant will facilitate interpretation of the data.

Because of the size of the tables, we have been unable to fit the additional percentages in, although we have included the most pertinent comparisons in the text.

12. Tables 3 and 4. The numbers in these tables are presumably all percentages. This needs to be clarified in the column headings or elsewhere. It needs to be clear whether the percentages are row or column percentages or if there is some other denominator.

The data in table 3 are percentages. The title and column headings have been changed to reflect this. We have added footnotes to clarify the denominators of the percentages. The units for the data in table 4 have been indicated in the row descriptions.
13. Table 4. The table would be more readable if the confidence intervals were located in columns alongside the point estimates, rather than below, with a relevant column heading.

This presentation could fit nicely in the final formatted version of the paper, but won’t fit in the specification of the table format for submitting papers to the journal.

_Discretionary Revisions_

1. Methods, Data Sources, paragraph 4. This paragraph may be deleted and the references moved to the previous paragraph, as this single-sentence paragraph adds little to the paper.

The paragraph has been removed, and the references added to the previous paragraph as suggested.

2. Discussion, paragraph 3, last sentence. The word “less” could be replaced with “fewer”.

This has been amended as suggested.
Reviewer 2

We thank the reviewer for her positive comments on our paper.

Reviewer 3

We thank the reviewer for his generous comments on our work.

(1) The totals in Tables 1 and 2 are wrong in several cases

Consistent with all published work from the WA Aboriginal Child Health Survey, we have rounded all survey estimates to 3 significant digits. This reflects the level of accuracy of the survey estimates. Because of this procedure, some totals differ slightly from the sum of the rounded components.

(2) There is a formatting error in the references. The first eleven have the title in italics, but the remainder have the title bold

We have endeavoured to follow the BMC Medical Research Methodology referencing style which variously applies italics or bold formatting to the title depending on the nature of the piece being cited.

Discretionary revisions

(1) This study not only shows that under-ascertainment is quite significant, but also that it is biased (for instance, resulting in an improvement in the proportions of low birth weight and low gestational age births). I think you should seriously consider making this point in your conclusions.

We have added an additional paragraph to the conclusion to re-iterate the finding that Aboriginal births that are not identifiable as such via the Midwives data are systematically different, so the figures based on the administrative data only are biased.

(2) Did you consider looking to see how many births there were in the WAACHS collection districts that were recorded as Aboriginal or Torres Strait in the administrative collections, but not included in the survey?

We agree that this would be very interesting to investigate. Unfortunately with the data we have available for this project we aren’t able to address this question. The data we have is the linkage of those children whose parents consented to linkage, and we were unable to link the remaining births, and therefore unable to link births identified from the register to the survey data as part of this project. We are hoping to be able to do more work on this question in the future.

(3) the reference [8] could be better positioned

We have changed this as suggested.

(4) Should the work “Indigenous” really start with a capital “I” all the time?

We have been advised that this is preferred, at least by the Aboriginal Steering Committee overseeing the survey, although we acknowledge that views on this differ.
Reviewer 4

I would have liked to see more information regarding:

a) Phrasing of the questions about Indigenous status in the Register of Births, Midwives’ Notification System and WAACHS survey.

The exact wording of the questions and form fields have been included in the paper.

b) A reference and 1-2 sentence description about the level of relative isolation, socioeconomic disadvantage. Has the latter derived from the ABS SEIFA?

We have added in descriptions of the level of relative isolation and socio-economic disadvantage variables and references to their detailed descriptions.

c) I also suggest reporting sensitivity, specificity and agreement measures

We have not included measures of sensitivity or specificity as these would imply, at least for some readers, an inappropriate interpretation of the results of this study. While the data would suggest that identifying Aboriginal and/or Torres Strait Islander births from Midwives or Births data leads to under-enumeration because of errors in these data, we acknowledge that Indigenous identification in Australia is based on self-identification and people can validly choose to identify differently in different situations or over time. The terms sensitivity and specificity and their definition in terms of false positives and false negatives and number of cases correctly identified would suggest that one measure is wrong and another is right, while we acknowledge that there isn’t necessarily a “right” identification.

I feel that the discussion section could benefit from a slight rewrite, as well as comparing the study results to the Australian literature – please see a list of relevant papers in the revision section below.

The discussion has been re-written and shortened.

Perhaps a mention about non-response to WAACHS survey could be beneficial. Are the non-responders more likely to be families with poorer child outcomes, living in more disadvantaged areas, and living in more remote communities?

The WAACHS response rate (84%) has been described in the text, along with a citation to a detailed published description of non-response patterns and weighting of the survey data.

The paper would benefit from the inclusion of recent relevant articles (please see Major compulsory revision for details). It might also be worthwhile indicated how the Midwives’ data is used currently in WA, in terms of the standard reporting (e.g. Mothers and babies reports in New South Wales). Does WA Health publish any regular reports?

We have included additional references as noted below, and referred to the Department of Health’s regular reporting.

As a means of context, what % of populations/children in WA are Indigenous? And where do the majority of Indigenous people live? Have there been any other studies in WA that looked at Indigenous reporting?
We have cited all the other studies that we are aware of that look at Indigenous status in WA collections. A detailed description of the population of Aboriginal children in WA from these data has previously been published and has been cited in the text.

a) Please report sensitivities, specificities, positive predictive values and negative predictive values. The tables with weighted crude numbers are informative but it might be worthwhile incorporating other measures.

As noted above, we have chosen not to include these measures as they have the potential to be misinterpreted and possibly cause offence to key stakeholders. With the results that we have included in the paper it is clear that there is a substantial discrepancy between survey and administrative data sources with the difference being almost entirely in one direction. The survey identifies significantly more children as being of Aboriginal and/or Torres Strait Islander descent than the register data, with very few cases of children being identified as of Aboriginal and/or Torres Strait Islander descent on the registers that are not identified as such in the survey. Specific figures for sensitivity and specificity would add little to this clear conclusion. Additionally the approach we have taken recognises that with available register data it is not possible to identify at the individual level which children would be likely to be of Aboriginal and/or Torres Strait Islander descent in a survey context that aren’t identified as such on the register. Our approach is to make aggregate adjustments to the derived statistics.

b) The paper needs some editing, to remove excessive (redundant) detail and improve continuity. Although the authors have given a comprehensive outline of the methods and findings, perhaps shortening certain sections might improve readability and clarity of the paper. This could be achieved by removing superfluous and repetitive sentences, usually found in both the methods and results sections. Some examples are outlined in the Minor Essential Revisions section.

The paper has been edited to remove repetition and improve clarity.

c) It could be useful to add information about other validation studies in Australia or overseas in the Background/Introduction section, especially relating to under-identification of Aboriginal status in administrative datasets. At present, a literature review is missing in this section. The authors might like to consult some of the references below:

Xu F, Sullivan EA, Madden RC, Black D, Jackson Pulver LR. Improvement of maternal Aboriginality in NSW birth data. BMC Medical Research Methodology 2012; 12:8
ABS 2006. Births, Australia, 2006. ABS cat. no. 3301.0. Canberra: ABS. (95% of Indigenous mothers were identified as Indigenous in the WA Registry of Births from 2002 to 2006)


We have included most of these publications, where we could make a link with our study.
d) It is not clear how information from siblings is treated in the analysis. Are the models adjusted for obvious clustering effects (children from the same mother)?

The models adjust for clustering by area and within family. We have added an extra paragraph to the methods section to clarify the estimation techniques used with the WAACHS data.

a) What number and proportion of 5,289 children had mothers as carers?

This is noted in the methods section and the additional figure 1.

b) How is level of relative isolation measured?

We have added a description and citation for this measure.

c) What does family structure comprise of?

We have added a description and citation for this measure.

d) How is relative socioeconomic disadvantage measured? Which geographical level was used to attach SEIFA scores (CD, SLA, LGA, postcode)? Part of this information is included on page 20 in the Results section; I suggest moving this information to the Methods section.

We have added a description and citation for this measure.

e) It is unclear which data source(s) was used to obtain birth weight, gestational age, and rented accommodation and test scores. Presumably the first two are obtained from the Midwives notification system, and the latter two from WAACHS? Perhaps stating this in the methods could be useful for the reader.

We have added this clarification as suggested.

f) Adjusting administrative-based time series section (page 13) could potentially be broken into 2 sections, as it mentions different models/results using different data sources:
   a. Factors contributing to correct recording
   b. Adjustments of time series data

We identified factors contributing to correct recording in the previous analysis, but were limited in the adjustment model to variables available from the register data. We have amended the text to resolve this confusion.

h) Which statistical package was used to analyse the data?

Analysis was undertaken using SAS.

i) Perhaps a diagrammatic representation of data sources and overlaps can be incorporated to distinguish which data sources/years/available information and respondent numbers were available for analysis at different stages.

We have included an additional figure to describe the process for combining the survey and administrative data.

*From the Results section:*
a) Page 15 – first sentence is repetitive, please remove. Potentially remove sentences 2, 3 and 4 and combine that information with sentence 5.

The sentences have been removed.

b) Page 16, first paragraph – this information is better suited in the Discussion section.

This has been moved to the discussion.

First sentence of the second paragraph is repetitive, please remove.

The sentence has been removed.

Last sentence – father’s status is missing in 33% of records, not 31%.

This has been corrected.

c) What % of children had either Indigenous mother or father using Register of Births?

77%. We have added this figure to the results.

d) Page 18, inconsistent identification section – first 2-3 sentences are repetitive, please remove.

The sentences have been removed.

e) Page 20 – first paragraph should be moved to the Methods section. Second sentence can be deleted, as it is repetitive.

This has been changed as suggested.

f) Please fix figures – they are hard to read at the present. Perhaps use black/white/grey instead of colour, and solid and dashed lines? What do the lines represent (e.g. Lowess smooth)?

The figures have been re-drawn using both solid and dashed lines. The lines represent the smoothed trend series using LOESS smoothing. This has been noted in the methods.

From the Discussion section:
a) The authors might consider structuring the discussion section as follows:
1. Re-iterate the main findings of the study
2. Discuss the findings in light to previous research
3. Outline strengths and weaknesses of the study
4. Potential future work

The discussion has been restructured as suggested.

Discretionary Revisions

a) Ethics approval information generally appears in the last section of Methods

This has been moved as suggested.

b) Perhaps Aboriginal and/or Torres Strait Islander can be mentioned first up, and for the remainder of the paper, this could be replaced by Aboriginal for the ease of the readership?
Perhaps the same could be applied to terms ‘WA Register of Births’ and ‘WA Midwives’ Notification System’, if possible?

We appreciate that the language may appear cumbersome, but it is significant to at least some Aboriginal and/or Torres Strait Islander people and we have employed the standard agreed language throughout to respect the wishes of the owners of the data. The custodians of the administrative data sets have requested that they be referred to by their proper titles, and we are honoring their request.