

## PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

### ARTICLE DETAILS

<b>TITLE (PROVISIONAL)</b>	Impact of participant attrition on child injury outcome estimates: a longitudinal birth cohort study in Australia
<b>AUTHORS</b>	Cameron, Cate; Osborne, Jodie; Spinks, Annelies; Davey, TM; Sipe, Neil; McClure, Rod

### VERSION 1 - REVIEW

<b>REVIEWER</b>	Vicki Anderson Murdoch Childrens Research Institute, Australia
<b>REVIEW RETURNED</b>	28-Jan-2017

<b>GENERAL COMMENTS</b>	I enjoyed reading this paper. It was well written, research questions and analyses are clear, and analyses appeared appropriate. Further, the issue addressed is an important one - longitudinal studies are frequently criticised due to attrition, with reviewers making unsupported conclusions about how such attrition will impact findings. This manuscript has addressed the various factors critical to such arguments well. The use of both active and passive data collection methods is a strength, and the findings are well stated and significant
-------------------------	---

<b>REVIEWER</b>	Mariana Brussoni and Amy Schneeberg University of British Columbia, Canada
<b>REVIEW RETURNED</b>	06-Feb-2017

<b>GENERAL COMMENTS</b>	<p>This is an interesting paper using a rare mix of both active and passive data collection allowing for the investigation of a pertinent question in childhood injury research. The manuscript is appropriately concise and well written. Below is one major methodological question/concern and a few minor editorial comments.</p> <p>1. The number/proportion of children from each of the 4 birth cohorts was not provided. Nor does it appear that there is an attempt to control for birth cohort. One might anticipate that more person years were contributed by younger children (later birth cohort) in the complete follow up group relative to those lost to follow up. Children who were in the earlier birth cohorts would have been followed longer thus being more likely to have partial or incomplete data. In addition, younger children would be at different risk of injury relative to older, more independent and mobile children. Did you consider controlling for year of birth?</p> <p>Editorial comments</p> <p>1. Page 5 – in the description of the EFHL it would be helpful to the reader to indicate what the follow up protocol was/how often follow-up surveys were distributed to participants. Although this information</p>
-------------------------	--

	<p>is available in a cited manuscript it is relevant enough to the current methods to be included here.</p> <p>2. Pg 7 – line 55 – Injury related TO a hospital or emergency department....” Missing the word “to” in manuscript</p> <p>3. Pg 9 line 47 – include n with percentage for clarity around missing data</p> <p>4. In the discussion you indicate that generalizability might be limited based on variables measured and injury outcome focus of analysis however also pertinent here is the follow up protocol. A more demanding follow up protocol might result in more impactful loss to follow up/attrition. It is prudent to put your results into context of what the follow up in your study looked like and how different types of follow up might impact these results in order to avoid readers over generalizing.</p> <p>5. Note that “data” should be plural throughout manuscript.</p> <p>This review was completed as a mentored peer review with Amy Schneeberg.</p>
--	--

<b>REVIEWER</b>	Carl Bonander Karlstad University, Sweden
<b>REVIEW RETURNED</b>	06-Mar-2017

<b>GENERAL COMMENTS</b>	<p>This is a well-written and interesting paper that adds to the current knowledge regarding differential attrition and its impact on exposure-outcome associations in cohort studies of injury outcomes. I find no reasons to question the validity of the results, and the procedures are clearly defined and easy to follow. I only have two minor suggestions that the authors might want to consider:</p> <p>1. On Page 6, line 5-6, you write that cases of maternal and child death were excluded. The rationale behind this choice could be stated more explicitly. I assume it is because they are lost due to death rather than attrition?</p> <p>2. On Page 7, line 45-50, you note that deterministic and probabilistic methods was used for record linkage. Perhaps this could be elaborated with e.g. details regarding the variables used in the matching procedure, or other ways in which the quality of the matching could be assessed.</p>
-------------------------	---

### VERSION 1 – AUTHOR RESPONSE

#### REVIEWER 1

I ENJOYED READING THIS PAPER. IT WAS WELL WRITTEN, RESEARCH QUESTIONS AND ANALYSES WERE CLEAR, AND ANALYSES APPEARED APPROPRIATE. FURTHER, THE ISSUE ADDRESSED IS AN IMPORTANT ONE - LONGITUDINAL STUDIES ARE FREQUENTLY CRITICISED DUE TO ATTRITION, WITH REVIEWERS MAKING UNSUPPORTED CONCLUSIONS ABOUT HOW SUCH ATTRITION WILL IMPACT FINDINGS. THIS MANUSCRIPT HAS ADDRESSED THE VARIOUS FACTORS CRITICAL TO SUCH ARGUMENTS WELL. THE USE OF BOTH ACTIVE AND PASSIVE DATA COLLECTION METHODS IS A STRENGTH, AND THE FINDINGS ARE WELL STATED AND SIGNIFICANT.

We thank Reviewer 1 for the positive review and comments. No changes

#### REVIEWER 2

THIS IS AN INTERESTING PAPER USING A RARE MIX OF BOTH ACTIVE AND PASSIVE DATA COLLECTION ALLOWING FOR THE INVESTIGATION OF A PERTINENT QUESTION IN CHILDHOOD INJURY RESEARCH. THE MANUSCRIPT IS APPROPRIATELY CONCISE AND WELL WRITTEN. BELOW IS ONE MAJOR METHODOLOGICAL QUESTION/CONCERN AND A FEW MINOR EDITORIAL COMMENTS.

We thank Reviewer 2 for the positive review and comments.

The number/proportion of children from each of the 4 birth cohorts was not provided. Nor does it appear that there is an attempt to control for birth cohort. One might anticipate that more person years were contributed by younger children (later birth cohort) in the complete follow up group relative to those lost to follow up. Children who were in the earlier birth cohorts would have been followed longer thus being more likely to have partial or incomplete data. In addition, younger children would be at different risk of injury relative to older, more independent and mobile children. Did you consider controlling for year of birth?

We have now included in Figure 1, the number of households from each of the birth cohort years included and the proportion of each with complete follow-up. There was no statistically significant difference in the number of complete and incomplete follow-up households across the cohort waves. The following has been added to the results.

“There was no significant difference in the proportion of households with complete or incomplete follow-up across the 5 recruitment waves ( $p=0.20$ ).”

We completely agree with the issues raised by Reviewer 2 with regards to the importance of individual length of exposure time for attrition and injury risk for this study. In order to control for the issue of different lengths of exposure (age of child) and subsequent injury risk, the rates and regressions analyses conducted in this study have included the individual person-years of exposure time for each participant. To better clarify this, we have amended the text in the methods to the following.

“The ages of the children and length of follow up for injury related hospital treatment varied considerably across participants due to the 5 recruitment waves. As such, individual person-years (PYs) of exposure time were calculated for each child, based on the time between birth and 31 December 2013 in which he or she was residing in the state of Queensland, alive and eligible for health care.” Figure 1 and Page 10

#### EDITORIAL COMMENTS

1. PAGE 5 – IN THE DESCRIPTION OF THE EFHL IT WOULD BE HELPFUL TO THE READER TO INDICATE WHAT THE FOLLOW UP PROTOCOL WAS/HOW OFTEN FOLLOW-UP SURVEYS WERE DISTRIBUTED TO PARTICIPANTS. ALTHOUGH THIS INFORMATION IS AVAILABLE IN A CITED MANUSCRIPT IT IS RELEVANT ENOUGH TO THE CURRENT METHODS TO BE INCLUDED HERE.

The following text has been added to the description of EFHL on Page 5 to indicate the follow-up and contact periods with participants.

“Participants completed a questionnaire at baseline and then follow-up surveys were sent 1, 3 and 5 years after the birth of the child. In between these scheduled contacts, regular newsletters were sent and additional sub-study project contacts occurred. Returned mail or contact difficulties triggered the use of alternative contact mechanisms supplied at enrolment and updated at follow-ups, including a relative or friends contact details, email and Facebook.” Page 5

2. PG 7 – LINE 55 – INJURY RELATED TO A HOSPITAL OR EMERGENCY DEPARTMENT....”  
MISSING THE WORD “TO” IN MANUSCRIPT

Changed Page 8

3. PG 9 LINE 47 – INCLUDE N WITH PERCENTAGE FOR CLARITY AROUND MISSING DATA

The text has been amended to read as follows:

“Automated linkage and manual searching by the state health department data linkage unit found records in QHAPDC for 97.1% of the child participants (n=2245), including their birth record.” Page 10

4. IN THE DISCUSSION YOU INDICATE THAT GENERALIZABILITY MIGHT BE LIMITED BASED ON VARIABLES MEASURED AND INJURY OUTCOME FOCUS OF ANALYSIS HOWEVER ALSO PERTINENT HERE IS THE FOLLOW UP PROTOCOL. A MORE DEMANDING FOLLOW UP PROTOCOL MIGHT RESULT IN MORE IMPACTFUL LOSS TO FOLLOW UP/ATTRITION. IT IS PRUDENT TO PUT YOUR RESULTS INTO CONTEXT OF WHAT THE FOLLOW UP IN YOUR STUDY LOOKED LIKE AND HOW DIFFERENT TYPES OF FOLLOW UP MIGHT IMPACT THESE RESULTS IN ORDER TO AVOID READERS OVER GENERALIZING.

We agree with the issues raised by Reviewer 2 regarding the impact that follow up protocols may have on the generalizability of the study findings. In line with the reviewers comments, we have modified the limitations section of the discussion as follows:

“The second limitation relates to the nature of the study sample, the determinant variables measured, the injury outcome focus of the analysis and the follow up protocols used. While the research results could be expected to apply to similar circumstances, it is unknown how validly the study findings should be applied beyond the study group we examined. It is possible that in studies with a demanding follow up protocol, participant attrition may be higher in people whose determinant variables are already affecting outcomes, and hence in these studies attrition may be more likely to be associated with a biased effect estimate.” Page 13-14

5. NOTE THAT “DATA” SHOULD BE PLURAL THROUGHOUT MANUSCRIPT.

Use of the term ‘data’ has been changed to plural throughout the manuscript.

REVIEWER 3

THIS IS A WELL-WRITTEN AND INTERESTING PAPER THAT ADDS TO THE CURRENT KNOWLEDGE REGARDING DIFFERENTIAL ATTRITION AND ITS IMPACT ON EXPOSURE-OUTCOME ASSOCIATIONS IN COHORT STUDIES OF INJURY OUTCOMES. I FIND NO REASONS TO QUESTION THE VALIDITY OF THE RESULTS, AND THE PROCEDURES ARE CLEARLY DEFINED AND EASY TO FOLLOW. I ONLY HAVE TWO MINOR SUGGESTIONS THAT THE AUTHORS MIGHT WANT TO CONSIDER.

We thank Reviewer 3 for the positive review and comments.

1. ON PAGE 6, LINE 5-6, YOU WRITE THAT CASES OF MATERNAL AND CHILD DEATH WERE EXCLUDED. THE RATIONALE BEHIND THIS CHOICE COULD BE STATED MORE EXPLICITLY. I ASSUME IT IS BECAUSE THEY ARE LOST DUE TO DEATH RATHER THAN ATTRITION?

As suggested by the reviewer, cases of maternal and child death were excluded from the analysis as these children had died (which is a known outcome), and were not lost to follow up. The following text has been added to the Methods section of the paper:

“Cases of child and maternal death were excluded from the analysis as these participants had died (which is a known outcome) and were not lost to follow up.” Page 6

2. ON PAGE 7, LINE 45-50, YOU NOTE THAT DETERMINISTIC AND PROBABLISTIC METHODS WAS USED FOR RECORD LINKAGE. PERHAPS THIS COULD BE ELABORATED WITH E.G. DETAILS REGARDING THE VARIABLES USED IN THE MATCHING PROCEDURE, OR OTHER WAYS IN WHICH THE QUALITY OF THE MATCHING COULD BE ASSESSED.

The following text has been added to the methods section to include more details on the matching procedures used. The data linkage is performed by the State Health Department as the custodians of the data. We have referenced the Queensland Data Linkage Framework procedures they use.

“The matching procedure was undertaken by the Queensland Department of Health (the custodians of the data) using linkage software, based on deterministic and probabilistic methods, to link demographic, child and maternal information to hospital records. Deterministic linkage involves the linking of data sets through comparing fields such as name, year of birth and street name with the requirement that the records agree on all characters. Probabilistic linkage involves the use of statistical models and algorithms to estimate the probability of data from different data sets having commonality (e.g. the same person/event). Clerical reviews of the data were undertaken to manually inspect uncertain matches in probabilistic linkage.” Page 8.

#### VERSION 2 – REVIEW

<b>REVIEWER</b>	Carl Bonander Karlstad University, Sweden
<b>REVIEW RETURNED</b>	24-Mar-2017

<b>GENERAL COMMENTS</b>	I think that all comments raised by me (reviewer 3), as well as the other reviewers, have been appropriately addressed by the authors. I recommend publication.
-------------------------	---

<b>REVIEWER</b>	Mariana Brussoni and Amy Schneeberg University of British Columbia, Canada
<b>REVIEW RETURNED</b>	24-Mar-2017

<b>GENERAL COMMENTS</b>	The authors have adequately addressed the reviewer concerns and this manuscript has the potential to make a valuable contribution to the literature.
-------------------------	--