Reviewer's report

Title: Completeness and accuracy of crash outcome data in a cohort of cyclists: a validation study

Version: 1 Date: 27 March 2013

Reviewer: Gabrielle S Davie

Reviewer's report:

An interesting paper – on both work and personal levels given I am a commuting cyclist. Thanks for the opportunity to review.

I do not think any major compulsory revisions are required.

Minor Essential Revisions

1. Pg 5 para 3. Need to add “for the majority” to the sentence that starts “Recruitment was undertaken at the time of the 2006 event.”
2. Pg 6 para 2. Define ACC.
3. Pg 6 para 4. Contrary to what is stated, the NMDS does not contain information about all day patients discharged from all public hospitals. In 2008 only about half of all District Health Boards in NZ were reporting Emergency Department discharges where a patient received a minimum of three hours of treatment to the NMDS. This could well explain some of the disagreement in Table 5 as participants could rightly self-report a crash requiring hospital admission but if it was an ED discharge, the event may not have been submitted to the NMDS.
4. Pg 7 para 3. Contrary to what is implied, the Mortality Collection does code underlying cause of death. According to NZ’s Mortality Collection Data Dictionary, they classify “the underlying cause of death for all deaths registered in New Zealand… using the ICD-10-AM 2nd Edition and the WHO Rules and Guidelines for Mortality Coding.”
5. Pg 9 para 2. Crashes aren’t “attended by” ACC or hospital.
6. Pg 11 para 4. The word ‘agreement’ is used with reference to kappa which gets confusing given columns 2-3 of Table 5 present Agreement (N & %) which is different. I also noticed this on the last paragraph of pg 9.
7. Pg 18 Ref 7. Typo “inrastructure”.
8. Pg 25 Table 4. The AIC weights & Weighted estimates for Model 5 especially for All Crashes and On-road crashes should be checked as they are a magnitude of 10 higher than the other models.

Discretionary Revisions

1. Pg 4 para 2. “For each million hours……” The tense of this sentence is odd. Suggest adding after hours “that were……”
2. Pg 4 para 3. In addition to Police crash reports “disproportionately underestimating the actual number of bicycle crashes”, they are also biased since they focus on on-road crashes.

3. Pg 4 para 3. I would have also thought that validity of self-report would also be affected by individuals’ differences on how severe a crash has to be for it to warrant mentioning in a survey.

4. Pg 6 para 3. It is my understanding that ACC do not record “outcome of injury” or “length of time off work”. What specifically is the “outcome of injury” data you extracted? Was “length of time off work” estimated from number of days weekly compensation was paid given people can return to work before needing more time off for which they may or may not have additional weekly compensation paid?

5. Pg 6 para 4. I suggest adding references for the NMDS & Mortality Collection. In addition the estimate that 90% of private hospitals in NZ submit to the NMDS needs a reference.

6. Pg 7 para 1. It would be helpful to state who matched participant data to NHI. I assume it was the Ministry of Health. Given ACC collect NHI, was the NHI then used to check/improve the ACC-cohort linkage?

7. Pg 7 para 2. How were readmissions identified? Do you have a reference?

8. Pg 8 para 1. I would like to know what the actual question was re the ‘total number of bicycle crashes experienced over the preceding year’. Was any severity threshold provided? Was this question asked before the questions about number of crashes recorded by ACC/NMDS/police? This would also help clarify on Pg 11 para 4 where I wondered whether the 4.7% who reported at least one crash in the preceding year was any self-reported crash or just those who self-reported at least one crash coming to the attention ACC, NMDS or the Police.

9. Pg 10 para 3. I think better wording could be used than “majority”. The majority of participants actually do not live in ‘least deprived neighbourhoods’ (50.1%).

10. Pg 10 para 3. Number of hours cycled per week is obviously skewed. Could IQR be presented too?

11. Pg 10 para 4. For clarity I suggest rewording to say “There were a total of 784 SELF-REPORTED crashes”.

12. Pg 13 para 2. Has anyone assessed the impact of violation of the assumption that each individual has equal probability to be captured in each database? Given the databases have quite different administrative purposes, I would have thought the likelihood that this assumption is met would be low.

13. Pg 14 para 1. Suggest removing ‘number’ from “Although the match rate by NHI number was high....”

14. Pg 16 para 2. It was not clear to me from this paper that there were “inaccuracies of bicycle crash data”.

15. Pg 16 para 2. Are you really wanting a “complete picture of injuries”? Or
rather a complete picture of important injuries?

16. Pg 23 Table 1. I am surprised ethnicity was not missing for any participants. Were those missing ethnicity responses categorised as Non-Maori?

17. Pg 24 Table 1. Should be Table 3. To make it easier for the reader to marry up the text & this table, I think it would be helpful for a total row to be included for All crashes, On-road crashes and Collisions. A new column with % of total crashes would also be informative.

18. Pg 26 Table 5. Any suggestions on how to interpret a negative kappa value in the 95% CI?

19. Pg 26 Table 5. Thinking about the disagreement for “at least one crash reported to the police”, is it possible for a bicycle crash to be reported to the Police but not entered in CAS as the crash didn’t involve a motor vehicle? Cases like this would be SR yes, LD no.

20. Pg 27 Table 6. I think an overall % agreement (sum of both yes & both no) column would be helpful for readers to compare categories of variables. I understand there are already lots of columns in this table but I would prefer a % agreement column over the reported %s for both yes, both no.

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

**Quality of written English:** Acceptable

**Statistical review:** Yes, and I have assessed the statistics in my report.

**Declaration of competing interests:**

I declare that I have no competing interests.