Reviewer's report

**Title:** Personal and trip characteristics associated with safety equipment use by injured adult bicyclists: A cross-sectional study

**Version:** 1  **Date:** 10 July 2012

**Reviewer:** Brent Hagel

**Reviewer's report:**

The authors have conducted a solid cross-sectional study examining the personal and trip characteristics associated with visibility aids and helmet use among injured adult bicyclists seen in emergency departments. A total of 690 bicyclists consented to participate at 5 hospitals in both Vancouver and Toronto, Canada (93.1% confirmed eligible; 66.5% estimated to be eligible). The authors used a structured questionnaire to ascertain information on demographic and trip characteristics, safety equipment use and injury circumstances. Logistic regression was used to relate cyclist and trip characteristics to 3 different outcome variables: light use, highly visible clothing and helmet use. A number of characteristics were related to each of these outcomes and the results were comprehensively discussed in the context of previous literature and relevant limitations were identified. I do, however, have the following comments for the authors' consideration:

**MAJOR COMPULSORY REVISIONS:**

1. Page 4: Methods: Though the authors indicate that the “Methods of study conduct and reliability testing have been described...elsewhere.” It would be useful to provide some indication of the reliability of these data in the present manuscript in order to assess the potential for misclassification bias in the effect estimates presented.

2. Page 5: Methods: “Don’t know’ or ‘refused’ responses for all questions were grouped with the ‘no’ category.” I think the investigators need to be careful in assuming that all of these ambiguous responses represent absence of the characteristic. Did the authors examine the characteristics of the individuals with missing information to be sure that the covariate patterns were similar to the “no” category? Have the authors attempted any imputation approaches on these data? In any case, it would be useful to show in the tables the percentage of missing data for each predictor variable.

3. Page 5: Methods: The authors indicate that they used “backwards selection to construct multiple logistic regression models, starting by offering all variables of interest. Based on the Wald test for each variable, the variable with the highest p-value was removed and the model refit...until all variables in the model were statistically significant...” I have some comments related to this analytic approach:

   a) My first comment relates to the number of variables entered into the different
regression models. A rule of thumb would be no more variables than the minimum of 10% of the least frequent outcome category (Harrell Jr. FE, Lee KL, Mark DB. Tutorial in biostatistics. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Statistics in Medicine. 1996;15:361-87). Based on the variables listed in tables 3-5, there were at least 37 indicator variables included in the full model for each outcome. However, there were only 135 individuals with at least one bike light turned on and so based on the rule, a maximum of 13 variables could be included in the full model. Can the authors address this issue?

b) Backward selection based on p-values likely results in overly narrow confidence limits and potentially biased estimates for the effects of interest. Can the authors present the unadjusted estimates, the fully adjusted estimates, and then the estimates based on the backward selection process? This would be helpful to the reader in determining how robust the associations are to the modeling strategy; though, I would argue the fully adjusted estimates would represent the gold standard as they are the least confounded (but again, the authors need to consider my comment on the frequency of the least frequent outcome category in determining the number of predictor variables considered).

4. Page 6: Results: The authors state that “Participants represented 93.1% of those confirmed to be eligible and 66.5% of those estimated to be eligible.” I am confused as to where the 66.5% figure is derived from? Can the authors comment on this and provide any information on the characteristics of those who were excluded either in the results section or speculate on their characteristics and how excluding them may have affected the results in the limitations section?

MINOR ESSENTIAL REVISIONS:

5. Page 7: Results: In referring to highly visible clothing, the authors state that “Older adults (#50 years of age) and those who were more frequent cyclists were more likely to wear such clothing.” However, in table 5, the 95% confidence interval for the 60+ age group includes the null value of 1.0 (95% CI: 0.68-2.38). Can the authors clarify?

6. Page 8: Discussion: There is a grammar problem with the sentence beginning “A new development in this areas is bike share systems…”.

DISCRETIONARY REVISIONS:

7. Page 11: Discussion: The authors could include references for the sentence “Our results compared favourably to studies using observations of cyclists in the field.”, unless they are only referring to the alcohol results presented later in the paragraph.

Level of interest: An article of importance in its field

Quality of written English: Acceptable

Statistical review: No, the manuscript does not need to be seen by a statistician.
Declaration of competing interests:

I was asked to be a collaborator on a Canadian Institutes of Health Research Strategic Teams in Applied Injury Research grant that Dr. Cusimano successfully obtained in 2009 to look at the issue of traumatic brain injury. However, we have not, as yet, had the opportunity to work together on any related research projects and I have not published with Dr. Cusimano.