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

TITLE (PROVISIONAL)	Bicycling injury hospitalization rates in Canadian jurisdictions: Analyses examining associations with helmet legislation and mode
	share
AUTHORS	Teschke, Kay; Koehoorn, Mieke; Shen, Hui; Dennis, Jessica

VERSION 1 - REVIEW

REVIEWER	James Woodcock
	University of Cambridge, UK
REVIEW RETURNED	18-Mar-2015

	The second
GENERAL COMMENTS	I o me the study is almost ready to publish. One comment I is on the
	difference between men & women. Generally the authors are right
	difference between mental when the constrainty the database of significant
	nowever, they may want to indicate that some studies have found
	that for fatalities associated with HGVs risks are higher for women
	e.g. Woodcock J, Tainio M, Cheshire J, O'Brien O, Goodman A.
	Health effects of the London bicycle sharing system: health impact
	modelling study. BMJ 2014;348:g425 doi: 10.1136/bmj.g425 and
	previous work for London. This difference in risk would not be
	expected to come out in the present analysis. Perhaps more
	generally the authors should explain how fatalities are dealt with in
	the data? One concern could be that the absolute number of
	hospitalisations is not different but injury severity changes with
	helmet use. The authors should consider if there data would capture
	this.

REVIEWER	lan Walker
	University of Bath
	United Kingdom
REVIEW RETURNED	31-Mar-2015

GENERAL COMMENTS	This is an interesting and generally well-written manuscript that makes a useful contribution to the literature. The idea to use Canada as a semi-controlled naturalistic experiment makes sense, and this cross-sectional approach complements earlier longitudinal studies of helmet laws being introduced. The analyses suggest that there is no clear association between the presence of a helmet law (which is in part a proxy for helmet-use rates) and the levels of serious injuries to body regions that a helmet might be expected to protect. As such, the paper agrees with other studies, such as Robinson's work in Australia, that have suggested that, at the population level, it is difficult to see effects of helmet use on safety.

 This brings me to my first suggestion, which is that it might be useful to add a paragraph to the Discussion to explore the mechanisms of this apparent paradox. We know from laboratory tests that helmets do dissipate energy in the event of an impact, and yet we also know that there seem to be few signs of this making a difference to population-level injury rates - a message reinforced by this study. It would be nice to see the authors explicitly state their throughts about how this disconnect might occur. The following observations are me playing Devil's Advocate. Addressing these points now might help strengthen your paper against any later attempts to dismiss it, thanks to its implication that mandatory helmets are not a panacea for bicycle safety. First, could the decision to use hospital admissions data have introduced a bias in your analysis because helmets shift a proportion of injuries from fatal injury to serious injury? Imagine the most extreme case, where all non-helmeted riders die on impact and all helmeted riders survive, albeit requiring a night in hospital. In this case, hospitalization data would make waring a helmet look really dangerous, as the only people to be admitted would be those who war them! If a commending adding some text to head off this line of argument - perhaps with reference to statistics on death v injury outcomes in orthly regions - perhaps they represent extreme risk-takers, for whom flouting this law is part of a wider constellation of risk yebaiviours that makes them non-comparable to a non-helmeted ridiviours that makes them non-comparable to a non-helmeted ridiviours that makes them con-comparable to a non-helmeted ridiviours that makes them con-comparable to a non-helmeted ridiviours that makes them on-comparable to a non-helmeted ridiviours that makes them on-comparable	
 The following observations are me playing Devil's Advocate. Addressing these points now might help strengthen your paper against any later attempts to dismiss it, thanks to its implication that mandatory helmets are not a panacea for bicycle safety. First, could the decision to use hospital admissions data have introduced a bias in your analysis because helmets shift a proportion of injuries from fatal injury to serious injury? Imagine the most extreme case, where all non-helmeted riders circle on impact and all helmeted riders survive, albeit requiring a night in hospital. In this case, hospitalization data would make wearing a helmet look really dangerous, as the only people to be admitted would be those who wear them! I'd recommending adding some text to head off this line of argument - perhaps with reference to statistics on death v injury outcomes in collisions. It is striking from Table 3 that even in places where it is mandated, helmet-wearing is -70% and not closer to 100%. I suspect that some will suggest non-wearers in these regions are qualitatively different from non-wearers in these regions are qualitatively to the anothelmet-user in, say, Quebec. Consider addressing this potential criticism. Third, is there anything you can add to help us focus more on helmet use rather than legislation - that is, make more use of the data in Table 37 The lack of a legislation, and 1'd suggest you give some thought to expanding this. Either draw attention more strange that legislation and non-legislation provinces. There's an analysis briefly alluded to in p. 10 para 1, to predict injury from wearing rates rather than legislation, and 1'd suggest you give some thought to expanding this. Either draw attention more strangly to the between-province rates in Table 3 early on, to make it abundantly clear that legislation status is a proxy for helmet-use rates, or perhaps even better, consider including wearing rates in your analyses (like this example on	This brings me to my first suggestion, which is that it might be useful to add a paragraph to the Discussion to explore the mechanisms of this apparent paradox. We know from laboratory tests that helmets do dissipate energy in the event of an impact, and yet we also know that there seem to be few signs of this making a difference to population-level injury rates - a message reinforced by this study. It would be nice to see the authors explicitly state their thoughts about how this disconnect might occur.
 First, could the decision to use hospital admissions data have introduced a bias in your analysis because helmets shift a proportion of injuries from fatal linjury to serious injury? Imagine the most extreme case, where all non-helmeted riders die on impact and all helmeted riders survive, albeit requiring a night in hospital. In this case, hospitalization data would make wearing a helmet look really dangerous, as the only people to be admitted would be those who wear them! I drecommending adding some text to head off this line of argument - perhaps with reference to statistics on death v injury outcomes in collisions. It is striking from Table 3 that even in places where it is mandated, helmet-wearing is ~70% and not closer to 100%. I suspect that some will suggest non-wearers in these regions are qualitatively different from non-wearers in other regions - perhaps they represent extrem erisk-takers, for whom flouting this law is part of a wider constellation of risky behaviours that makes them non-comparable to a non-helmet-user in, say, Quebec. Consider addressing this potential criticism. Third, is there anything you can add to help us focus more on helmet use rather than legislation effect in the analyses would become a lot less interesting if helmet-wearing rates were comparable in legislation and non-legislation provinces. There's an analysis briefly alluded to in p. 10 para 1, to predict injury from wearing rates rather than legislation, and I d suggest you give some thought to expanding this. Either draw attention more strongly to the between-province rates in Table 3 early on, to make it abundantly clear that legislation status is a proxy for helmet-user rates, or, perhaps even better, consider including wearing rates in your analyses (like this example on p 10). MINOR SUGGESTIONS: The use of percentage cycling mode share as a predictor in a regression is slightly unusual. Moreover, the range of this predictor is very limi	The following observations are me playing Devil's Advocate. Addressing these points now might help strengthen your paper against any later attempts to dismiss it, thanks to its implication that mandatory helmets are not a panacea for bicycle safety.
It is striking from Table 3 that even in places where it is mandated, helmet-wearing is ~70% and not closer to 100%. I suspect that some will suggest non-wearers in these regions are qualitatively different from non-wearers in other regions - perhaps they represent extreme risk-takers, for whom flouting this law is part of a wider constellation of risky behaviours that makes them non-comparable to a non-helmet-user in, say, Quebec. Consider addressing this potential criticism. Third, is there anything you can add to help us focus more on helmet use rather than legislation - that is, make more use of the data in Table 3? The lack of a legislation effect in the analyses would become a lot less interesting if helmet-wearing rates were comparable in legislation and non-legislation provinces. There's an analysis briefly alluded to in p. 10 para 1, to predict injury from wearing rates rather than legislation, attention more strongly to the between-province rates in Table 3 early on, to make it abundantly clear that legislation status is a proxy for helmet-use rates, or, perhaps even better, consider including wearing rates in your analyses (like this example on p 10). MINOR SUGGESTIONS: - The use of percentage cycling mode share as a predictor in a regression coefficients tell us how the outcomes changed in these data as the mode share changed from, say 0.5 to 1.5, it's not clear to what extent you believe the coefficients would still be valid estimates if mode share exorting would still be valid estimates if mode share considered in these data. Please say something to address this.	First, could the decision to use hospital admissions data have introduced a bias in your analysis because helmets shift a proportion of injuries from fatal injury to serious injury? Imagine the most extreme case, where all non-helmeted riders die on impact and all helmeted riders survive, albeit requiring a night in hospital. In this case, hospitalization data would make wearing a helmet look really dangerous, as the only people to be admitted would be those who wear them! I'd recommending adding some text to head off this line of argument - perhaps with reference to statistics on death v injury outcomes in collisions.
 Third, is there anything you can add to help us focus more on helmet use rather than legislation - that is, make more use of the data in Table 3? The lack of a legislation effect in the analyses would become a lot less interesting if helmet-wearing rates were comparable in legislation and non-legislation provinces. There's an analysis briefly alluded to in p. 10 para 1, to predict injury from wearing rates rather than legislation, and I'd suggest you give some thought to expanding this. Either draw attention more strongly to the between-province rates in Table 3 early on, to make it abundantly clear that legislation status is a proxy for helmet-use rates, or, perhaps even better, consider including wearing rates in your analyses (like this example on p 10). MINOR SUGGESTIONS: The use of percentage cycling mode share as a predictor in a regression is slightly unusual. Moreover, the range of this predictor is very limited (it pretty much ranges from 0 to 2). Whilst the regression coefficients tell us how the outcomes changed in these data as the mode share changed from, say 0.5 to 1.5, it's not clear to what extent you believe the coefficients would still be valid estimates if mode share went from 2% to 4%, let alone up to Dutch levels. It's likely that non-linearities become an issue beyond the limited range of mode share considered in these data. Please say something to address this. At the bottom of p. 10 you write "with a 1% increase in mode share, 	It is striking from Table 3 that even in places where it is mandated, helmet-wearing is ~70% and not closer to 100%. I suspect that some will suggest non-wearers in these regions are qualitatively different from non-wearers in other regions - perhaps they represent extreme risk-takers, for whom flouting this law is part of a wider constellation of risky behaviours that makes them non-comparable to a non-helmet-user in, say, Quebec. Consider addressing this potential criticism.
 MINOR SUGGESTIONS: The use of percentage cycling mode share as a predictor in a regression is slightly unusual. Moreover, the range of this predictor is very limited (it pretty much ranges from 0 to 2). Whilst the regression coefficients tell us how the outcomes changed in these data as the mode share changed from, say 0.5 to 1.5, it's not clear to what extent you believe the coefficients would still be valid estimates if mode share went from 2% to 4%, let alone up to Dutch levels. It's likely that non-linearities become an issue beyond the limited range of mode share considered in these data. Please say something to address this. At the bottom of p. 10 you write "with a 1% increase in mode share, 	Third, is there anything you can add to help us focus more on helmet use rather than legislation - that is, make more use of the data in Table 3? The lack of a legislation effect in the analyses would become a lot less interesting if helmet-wearing rates were comparable in legislation and non-legislation provinces. There's an analysis briefly alluded to in p. 10 para 1, to predict injury from wearing rates rather than legislation, and I'd suggest you give some thought to expanding this. Either draw attention more strongly to the between-province rates in Table 3 early on, to make it abundantly clear that legislation status is a proxy for helmet-use rates, or, perhaps even better, consider including wearing rates in your analyses (like this example on p 10).
- At the bottom of p. 10 you write "with a 1% increase in mode share,	MINOR SUGGESTIONS: - The use of percentage cycling mode share as a predictor in a regression is slightly unusual. Moreover, the range of this predictor is very limited (it pretty much ranges from 0 to 2). Whilst the regression coefficients tell us how the outcomes changed in these data as the mode share changed from, say 0.5 to 1.5, it's not clear to what extent you believe the coefficients would still be valid estimates if mode share went from 2% to 4%, let alone up to Dutch levels. It's likely that non-linearities become an issue beyond the limited range of mode share considered in these data. Please say something to address this.
	- At the bottom of p. 10 you write "with a 1% increase in mode share,

· · · · · · · · · · · · · · · · · · ·	
	the rate was lower by about one-quarter". I'd be careful here - talking about quarters gets you dangerously close to the situation where some would say you should have used logistic regression or a generalized linear model for proportional data.
	- In the Results section, Figure 3 might be expanded. First, the text notes that Figure 3 would not look substantially different if only traffic-related injuries are considered, but I think it would be useful to let us see this for ourselves. Second, consider adding to Figure 3 some body regions where you WOULDN'T expect the helmet to make a difference, to show the "protected" and non-protected regions look much the same in these data.
	- Was the clear outlier in Figure 2d included in the regression?
	- Table 4. The mode share result is really interesting. Please add something to the table caption or table notes to remind the reader what the coefficient represents (presumably 801 fewer hospitalizations per percentage mode share increase?)
	 p 13 para 2 "and factors that encourage cycling" this wording implies it is the factors that encourage cycling which affect safety, and not the safety in numbers effect that you've previously discussed
	 Conclusion para 1 - the very end of the last sentence would be a good place to re-emphasize that there was also no association with wearing rate, not just with legislation.
	- The second paragraph of the conclusion is ambiguous. It could be read as saying that female risk choices are something that need addressing by policy makers, when of course if policy makers are going to address anything then it ought to be male risk choices, given male riders are hurt much more often. Rephrase this for clarity?

REVIEWER	Jake Olivier University of New South Wales
	Sydney, Australia
REVIEW RETURNED	21-Apr-2015

GENERAL COMMENTS	This was an interesting article about about factors related to cycling injury at a province level in Canada. The authors found female sex
	and cycling mode share were associated with less cycling injuries per trip. Although I'm generally positive about the aims of this study, I think there are problems with the study's methodology. The authors argue cycling safety is much more than helmets (which I strongly agree with, although I think they have a place); however, I'd be more convinced if their results are maintained after addressing the methodological issues below.
	The study spans the years 2006 to 2011; however, it is unclear the data actually spans those years. Hospitalizations cover Canadian financial years (April to March) while the surveys ran over the calendar years (January to December). It also seems that data collection for the surveys was every two years. This isn't clear to me and I wonder if 2005 estimates were used for 2006, 2007 estimates for 2008, etc. This issue becomes more problematic in that

work/school cycling is not in the 2005/06 survey. Note years and the data don't align in the 2013 Dennis BMJ either and I am not sure why the authors haven't tried correcting that problem (or adapting an appropriate analytic plan).
Was the population data for calendar or fiscal years? The temporal overlap for data and how multiple surveys were synthesized for analysis would be helpful.
Most analyses in the manuscript make a normal distribution of some kind. How were modeling assumptions checked? Count data don't usually follow a normal distribution (especially when counts are small) and a Poisson-type analysis (Poisson, over-dispersed Poisson or negative binomial) is a much better choice. Poisson-type models also allow for the inclusion of an offset like number of cycling trips. This modelling choice would also eliminate the problem with zero cell counts as the log Poisson mean is modeled instead of the observations themselves. PROC GLIMMIX in SAS could be used to account for dependence in the data for a Poisson response.
Is a random intercept for jurisdiction enough? This is essentially fitting an intercept for each jurisdiction and the other factors are assumed "parallel" across jurisdictions. Were other random effects considered?
The use of an indicator for helmet legislation is equivalent to a model that assumes everybody in jurisdictions with legislation wear helmets and no one wears them in non-legislation jurisdictions. This is clearly not true and perhaps the inclusion of helmet wearing estimates as a covariate is a better approach. But, if you're going to use helmet legislation as an indicator, shouldn't there just be two levels? The correlation between adults and kids in a jurisdiction have already been accounted for in the analysis. Both models could be fit and the best chosen based on AIC or some other information criterion.
Since females have such lower injury rates, I don't like the idea of collapsing sex for traffic-related injuries. Sex could very well explain a lot of the observed variability and a better modeling strategy would allow the inclusion of the sex variable.
Assuming the analytic results are valid, I also don't agree with the study's conclusions. Certainly, this study and others have found a gender association with cycling injury. However, is there any evidence males can be persuaded to make "female cycling choices"? Certainly men have a lot to learn from women about assessing personal risk, but is there any evidence men can or will change their behavior?
There have also been several studies that have found evidence for the so-called "safety in numbers" effect. However, there is no evidence this can be achieved by simply increasing cycling numbers. In fact, we published a paper last year using NSW data on cycling hospitalizations and cycling participation estimates, and found that cycling injuries increased at roughly the same rate as increases in cycling (i.e., no evidence to support the safety in numbers effect).
http://acrs.org.au/wp- content/uploads/ACRSjournalVol25NoNov14WEB-1.pdf

Another recent Australian study suggested cycling density was more important than sheer numbers (note: data was simulated and not 'real'). In places with low cycling mode share like Canada and Australia, benefits may not be noticeable until mode share gets about some threshold like say 10% where there is enough cycling "density" to make a difference.
http://www.ncbi.nlm.nih.gov/pubmed/24761795
Discussion: The apparent disconnect between case-control studies and population studies assessing helmet legislation is discussed. We discussed that in our 2013 paper (p. 206) where we estimate the expected decrease in head injury assuming Elvik's estimates are correct in his 2011 paper and the observed change in helmet wearing in NSW. The estimated range is a drop of 27-31.4% while our estimated impact of the NSW helmet law was 27.5% or 31.0% depending on the comparison time series. At least in our isolated instance, the disconnect vanished when accounting for changes in helmet wearing.
http://www.ncbi.nlm.nih.gov/pubmed/23339779
Minor Issues
If you are going to cite the Elvik paper, it should be the 2013 corrigendum or cite the 2001 Attewell paper as it has no known errors. To my knowledge, there were 3 published versions of the Elvik paper and the results from the version cited is known to be wrong.
http://www.ncbi.nlm.nih.gov/pubmed/24344450
Did hospitalizations include deaths in hospital?
Page 6. Was bootstrapping necessary for confidence intervals? Standard error estimates should be readily available using the sampling weights (could use SAS PROC SURVEYFREQ).
Page 6. I tend to agree that since helmet wearing is legislation for "on-road" use, some effort should be made to compare hospitalizations from similar circumstances. However, is there evidence helmet use differs much from on- to off-road cycling? My guess is more mountain bikers wear helmets than those on-road.
Page. 6. "Data on population size??"
Page 8. I highly recommend not using Excel to compute rates to minimize data errors.
Data availability: I'm sure it would be illegal for the authors to share the raw data; however, could the aggregated data used in the analysis be made available along with SAS code (or similar) to perform the analyses?

VERSION 1 – AUTHOR RESPONSE

Reviewer Name James Woodcock

Institution and Country University of Cambridge, UK

1. To me the study is almost ready to publish. One comment is on the difference between men & women. Generally the authors are right, however, they may want to indicate that some studies have found that for fatalities associated with HGVs risks are higher for women e.g. Woodcock J, Tainio M, Cheshire J, O'Brien O, Goodman A. Health effects of the London bicycle sharing system: health impact modelling study. BMJ 2014;348:g425 doi: 10.1136/bmj.g425 and previous work for London. This difference in risk would not be expected to come out in the present analysis.

Thanks for pointing out these exposure-based risk estimates. We have now cited this article as offering evidence of male vs. female risk differences, in both directions. We have not provided details of the specific HGV risk difference, since this is not an element investigated in our study.

2. Perhaps more generally the authors should explain how fatalities are dealt with in the data? One concern could be that the absolute number of hospitalisations is not different but injury severity changes with helmet use. The authors should consider if their data would capture this.

Our data includes anyone who died subsequent to hospital admission. We have now indicated this in the Methods. We did not ask for deaths to be separately tabulated for us by the Canadian Institute for Health Information (CIHI), but in a previously reported CIHI dataset, 0.4% of all hospitalizations ended in death [Dennis et al, 2013]. This suggests about 15 deaths/year in our dataset. Data from Transport Canada for the years of our study show an average of 57 bicycling deaths/year (representing about 1.5% of our hospitalizations), indicating that the majority die prior to hospital admission, so would not be included in our data. Note that the Transport Canada data differs in the following ways from our data: it includes deaths of children < 12 years of age (~5% of deaths in a recent coroner's report on cycling deaths in our most populous province, Ontario), but excludes deaths that did not involve motor vehicles (~20% of deaths in the same coroner's report). We have now included some information about cycling deaths in Canada in the Discussion.

Reviewer Name Ian Walker

Institution and Country University of Bath, United Kingdom

This is an interesting and generally well-written manuscript that makes a useful contribution to the literature. The idea to use Canada as a semi-controlled naturalistic experiment makes sense, and this cross-sectional approach complements earlier longitudinal studies of helmet laws being introduced. The analyses suggest that there is no clear association between the presence of a helmet law (which is in part a proxy for helmet-use rates) and the levels of serious injuries to body regions that a helmet might be expected to protect. As such, the paper agrees with other studies, such as Robinson's work in Australia, that have suggested that, at the population level, it is difficult to see effects of helmet use on safety.

1. This brings me to my first suggestion, which is that it might be useful to add a paragraph to the Discussion to explore the mechanisms of this apparent paradox. We know from laboratory tests that helmets do dissipate energy in the event of an impact, and yet we also know that there seem to be few signs of this making a difference to population-level injury rates - a message reinforced by this study. It would be nice to see the authors explicitly state their thoughts about how this disconnect might occur.

We have added a little text in the Discussion, though it is very incomplete. There are so many

possibilities. The issue deserves a full review, but we are concerned that too much more discussion of the absence of a helmet law effect will detract from the main results – that other factors do make a difference at the population level and warrant a diversion of attention away from helmet laws.

The following observations are me playing Devil's Advocate. Addressing these points now might help strengthen your paper against any later attempts to dismiss it, thanks to its implication that mandatory helmets are not a panacea for bicycle safety.

2. First, could the decision to use hospital admissions data have introduced a bias in your analysis because helmets shift a proportion of injuries from fatal injury to serious injury? Imagine the most extreme case, where all non-helmeted riders die on impact and all helmeted riders survive, albeit requiring a night in hospital. In this case, hospitalization data would make wearing a helmet look really dangerous, as the only people to be admitted would be those who wear them! I'd recommending adding some text to head off this line of argument - perhaps with reference to statistics on death v injury outcomes in collisions.

We have tried to address this in the Discussion (see response to James Woodcock above).

3. It is striking from Table 3 that even in places where it is mandated, helmet-wearing is ~70% and not closer to 100%. I suspect that some will suggest non-wearers in these regions are qualitatively different from non-wearers in other regions - perhaps they represent extreme risk-takers, for whom flouting this law is part of a wider constellation of risky behaviours that makes them non-comparable to a non-helmet-user in, say, Quebec. Consider addressing this potential criticism.

We have added a little to the discussion, but far from complete.

4. Third, is there anything you can add to help us focus more on helmet use rather than legislation - that is, make more use of the data in Table 3? The lack of a legislation effect in the analyses would become a lot less interesting if helmet-wearing rates were comparable in legislation and non-legislation provinces. There's an analysis briefly alluded to in p. 10 para 1, to predict injury from wearing rates rather than legislation, and I'd suggest you give some thought to expanding this. Either draw attention more strongly to the between-province rates in Table 3 early on, to make it abundantly clear that legislation status is a proxy for helmet-use rates, or, perhaps even better, consider including wearing rates in your analyses (like this example on p 10).

We have added a graphic (Figure 2) illustrating the differences in reported helmet use by helmet legislation to make clear that helmet laws are associated with increased helmet use.

We have not emphasized the results related to helmet use, because this style of study (data at the jurisdiction level not the individual level) is not an appropriate design for testing helmet use effectiveness. Our examination of associations with helmet use was simply a surrogate for checking the potential impact of municipal helmet bylaws within provinces or territories with no helmet laws. We have tried to make this clearer in the Methods, Results, and Discussion.

MINOR SUGGESTIONS:

5. The use of percentage cycling mode share as a predictor in a regression is slightly unusual. Moreover, the range of this predictor is very limited (it pretty much ranges from 0 to 2). Whilst the regression coefficients tell us how the outcomes changed in these data as the mode share changed from, say 0.5 to 1.5, it's not clear to what extent you believe the coefficients would still be valid estimates if mode share went from 2% to 4%, let alone up to Dutch levels. It's likely that non-linearities become an issue beyond the limited range of mode share considered in these data.

Please say something to address this.

Good point, we have added a footnote to Table 4 to make clear that this result can only apply within the range of the data.

6. At the bottom of p. 10 you write "with a 1% increase in mode share, the rate was lower by about one-quarter". I'd be careful here - talking about quarters gets you dangerously close to the situation where some would say you should have used logistic regression or a generalized linear model for proportional data.

We do now have logistic regression models, but in any case, given your previous point, we've tried to remove over-interpretation by removing references to the magnitude of differences (for both sex and mode share) in the text.

7. In the Results section, Figure 3 might be expanded. First, the text notes that Figure 3 would not look substantially different if only traffic-related injuries are considered, but I think it would be useful to let us see this for ourselves. Second, consider adding to Figure 3 some body regions where you WOULDN'T expect the helmet to make a difference, to show the "protected" and non-protected regions look much the same in these data.

Good idea - have added both these items to a revised figure - now numbered 4.

8. Was the clear outlier in Figure 2d included in the regression?

Yes, it was included because we could find no sound reason to exclude it. We reran the original model without it and though coefficients changed somewhat in magnitude, none changed in direction or significance, so our interpretation of the results would not have been altered.

The new modeling method (in response to Reviewer 3, see below), using the logit transformation, resulted in normally distributed data and this point is no longer remarkable.

9. Table 4. The mode share result is really interesting. Please add something to the table caption or table notes to remind the reader what the coefficient represents (presumably 801 fewer hospitalizations per percentage mode share increase?)

Explanation now added to a table footnote – now with a different interpretation because of the logit modeling.

10. p 13 para 2 "...and factors that encourage cycling..." this wording implies it is the factors that encourage cycling which affect safety, and not the safety in numbers effect that you've previously discussed.

We believe that the mechanism for the safety in numbers relationship is not clear and likely includes circularity and many complexities. We have tried to express this. We have also cited some of the factors that have been shown to be associated with increases in cycling, particularly safer infrastructure. We have altered some of the wording related to this issue to try to make ourselves better understood.

11. Conclusion para 1 - the very end of the last sentence would be a good place to re-emphasize that there was also no association with wearing rate, not just with legislation.

We hesitate to emphasize this, since the reason we examined relationships with helmet wearing

rates was simply to check that the results related to provincial legislation were not influenced by municipal bylaws having a separate effect on helmet use.

12. The second paragraph of the conclusion is ambiguous. It could be read as saying that female risk choices are something that need addressing by policy makers, when of course if policy makers are

going to address anything then it ought to be male risk choices, given male riders are hurt much more often. Rephrase this for clarity?

Clearly needs rewording. Have made an attempt.

Reviewer Name Jake Olivier

Institution and Country University of New South Wales, Sydney, Australia

This was an interesting article about factors related to cycling injury at a province level in Canada. The authors found female sex and cycling mode share were associated with less cycling injuries per trip. Although I'm generally positive about the aims of this study, I think there are problems with the study's methodology. The authors argue cycling safety is much more than helmets (which I strongly agree with, although I think they have a place); however, I'd be more convinced if their results are maintained after addressing the methodological issues below.

1. The study spans the years 2006 to 2011; however, it is unclear the data actually spans those years. Hospitalizations cover Canadian financial years (April to March) while the surveys ran over the calendar years (January to December). It also seems that data collection for the surveys was every two years. This isn't clear to me and I wonder if 2005 estimates were used for 2006, 2007 estimates for 2008, etc. This issue becomes more problematic in that work/school cycling is not in the 2005/06 survey. Note years and the data don't align in the 2013 Dennis BMJ either and I am not sure why the authors haven't tried correcting that problem (or adapting an appropriate analytic plan).

You are right, there is a mismatch. We did not have control of either administrative dataset to create more of a match. We have tried to be more explicit about the lack of a perfect match in the Methods. We also agree that the mismatch is a limitation of the study, one that we should have addressed in the Discussion. We have now added such text.

We do not believe this is a major limitation. We pooled 6 years of numerator data and 6 (partially mismatched) years of denominator data to calculate the hospitalization rates in each stratum. The hospitalization data is complete for the period. The denominator data is based on very large samples representative of the Canadian population. Differences in the number of trips by cycle of the Canadian Community Health Survey were small (averaging 2% for the country as a whole and from 1 to 16% for individual jurisdictions), especially in comparison to between-strata differences, and do not suggest a temporal trend. Smaller jurisdictions had larger period-to-period differences, suggesting that denominator errors are more likely from sampling than from the imperfect temporal match.

2. Was the population data for calendar or fiscal years? The temporal overlap for data and how multiple surveys were synthesized for analysis would be helpful.

Good point. The population size data is from the census, conducted on a single day mid-May each census. This is now indicated in the Methods.

3. Most analyses in the manuscript make a normal distribution of some kind. How were modeling

assumptions checked? Count data don't usually follow a normal distribution (especially when counts are small) and a Poisson-type analysis (Poisson, over-dispersed Poisson or negative binomial) is a much better choice. Poisson-type models also allow for the inclusion of an offset like number of cycling trips. This modelling choice would also eliminate the problem with zero cell counts as the log Poisson mean is modeled instead of the observations themselves. PROC GLIMMIX in SAS could be used to account for dependence in the data for a Poisson response.

We did check the normality and log-normality of the rate variables during the initial analyses. All were slightly positively skewed, with geometric standard deviations in the range of 1.5 to 1.9, medians

slightly lower than arithmetic means, and most did not reject a goodness of fit test for log-normality. A few did not reject a test of normality. The residuals were generally symmetrically distributed. We thought that the manuscript would be more understandable to the many policy makers and ordinary citizens interested in bicycling issues if we used linear regression, so followed the frequent guidance that regression is robust to the assumption of a normal dependent variable.

But you are right (as we inferred in the original manuscript), this was not ideal. So we tried your recommendation of Poisson modeling. Unfortunately it was not an improvement. All of the count variables rejected goodness of fit tests for Poisson distribution. We tried Poisson modelling of the count variables, with adjustment for overdispersion and using an offset (In-transformed) for trips, but these models had skewed distributions of their residuals (also non-constant over the range of predicteds) and estimates for the female-male rate ratios that were outside the range of the data.

We consulted one of our statisticians, Dr. Hui Shen, and she recommended we transform the 0,1 bounded hospitalization rates using the logit, and this worked very well. None of the resulting dependent variables rejected the Shapiro-Wilks goodness of fit test for normality, nor did the model residuals. We redid all the analyses using this method, and this has now been reported.

4. Is a random intercept for jurisdiction enough? This is essentially fitting an intercept for each jurisdiction and the other factors are assumed "parallel" across jurisdictions. Were other random effects considered?

We didn't consider random slopes because there were few repetitions per province (4 for all causes, 2 for traffic-related causes) and a fixed effect that varied only with province (mode share). We were concerned about overfitting. In the redo of the modeling described above, we decided to check whether we might be overfitting with even the random intercept modeling. The differences in AICc and BIC between the mixed models and the fixed effects models suggested that we were. We removed the intercept random effect when its component of variance was not substantial (>20%) or statistically significant.

5. The use of an indicator for helmet legislation is equivalent to a model that assumes everybody in jurisdictions with legislation wear helmets and no one wears them in non-legislation jurisdictions. This is clearly not true and perhaps the inclusion of helmet wearing estimates as a covariate is a better approach. But, if you're going to use helmet legislation as an indicator, shouldn't there just be two levels? The correlation between adults and kids in a jurisdiction have already been accounted for in the analysis. Both models could be fit and the best chosen based on AIC or some other information criterion.

We did not mean the helmet legislation variable to imply 100% vs. 0% helmet wearing. We have tried to make this clearer with the addition of Figure 2. It shows that helmet laws in Canadian jurisdictions are associated with higher helmet use proportions, but the difference is certainly far short of complete adoption vs. no use.

We used the helmet legislation variable because we wanted to examine the impact of the policy. We agree that our choice of indicator variable obscured our intent and we have adopted your recommended approach of 2 levels. The purpose of the secondary analysis using helmet wearing estimates was to check whether the null results for the law might be related to within-province municipal bylaws. We have tried to make this more clear in the text. We also took your suggestion of trying the helmet-wearing variable as a co-variate in the models. In all cases, the associations with hospitalization rates were positive – opposite to expectation. Some of these relationships were significant, but we have not reported this, since the hypothesis is in the opposite direction and these were secondary analyses meant to check the main results.

6. Since females have such lower injury rates, I don't like the idea of collapsing sex for traffic-related injuries. Sex could very well explain a lot of the observed variability and a better modeling strategy would allow the inclusion of the sex variable.

Unfortunately, we do not have the raw data, but rather received our data in tabulated format from CIHI. We contacted CIHI to find out whether we would need to re-enter their months-long queue to get this additional data and the answer was yes. In addition, they informed us that, this year, they do not have data sharing agreements in place for all provinces, so they could not assure us that we could receive complete data, even after the wait. We have decided to proceed without the additional data.

In addition, as we described in the paper, we collapsed male and female data for traffic-related injuries to minimize the number of strata with zero hospitalizations. Even if we had sex-specific data for traffic-related injuries, we worry that sex-specific effect estimates would be unstable.

7. Assuming the analytic results are valid, I also don't agree with the study's conclusions. Certainly, this study and others have found a gender association with cycling injury. However, is there any evidence males can be persuaded to make "female cycling choices"? Certainly men have a lot to learn from women about assessing personal risk, but is there any evidence men can or will change their behavior?

Thank you for raising this. The wording used in the conclusions was clearly unsatisfactory, implying meaning we did not intend. We have attempted improved wording.

As an aside, education in Canada about cycling safety has not been linked to evidence, so it is possible that many people do not know what factors are safer and less safe, and might make different choices with more information.

8. There have also been several studies that have found evidence for the so-called "safety in numbers" effect. However, there is no evidence this can be achieved by simply increasing cycling numbers. In fact, we published a paper last year using NSW data on cycling hospitalizations and cycling participation estimates, and found that cycling injuries increased at roughly the same rate as increases in cycling (i.e., no evidence to support the safety in numbers effect).

http://acrs.org.au/wp-content/uploads/ACRSjournalVol25NoNov14WEB-1.pdf

Another recent Australian study suggested cycling density was more important than sheer numbers (note: data was simulated and not 'real'). In places with low cycling mode share like Canada and Australia, benefits may not be noticeable until mode share gets about some threshold like say 10% where there is enough cycling "density" to make a difference.

http://www.ncbi.nlm.nih.gov/pubmed/24761795

We agree that the positive association between safety and numbers of cyclists may not arise directly from increases in the numbers of cyclists of unknown origin. We have tried to express some of the controversy about causal mechanisms and about the direction of the relationship. We have also cited some of the factors that have been shown to be associated with increases in cycling, particularly safer infrastructure. We have altered some of the wording related to this issue to try to make ourselves better understood.

9. Discussion: The apparent disconnect between case-control studies and population studies assessing helmet legislation is discussed. We discussed that in our 2013 paper (p. 206) where we estimate the expected decrease in head injury assuming Elvik's estimates are correct in his 2011 paper and the observed change in helmet wearing in NSW. The estimated range is a drop of 27-31.4% while our estimated impact of the NSW helmet law was 27.5% or 31.0% depending on the

comparison time series. At least in our isolated instance, the disconnect vanished when accounting for changes in helmet wearing.

http://www.ncbi.nlm.nih.gov/pubmed/23339779

We have added wording along these lines to the Discussion, but because of the positive relationship between helmet wearing and hospitalizations in our data, this cannot explain our result.

Minor Issues

10. If you are going to cite the Elvik paper, it should be the 2013 corrigendum or cite the 2001 Attewell paper as it has no known errors. To my knowledge, there were 3 published versions of the Elvik paper and the results from the version cited is known to be wrong.

http://www.ncbi.nlm.nih.gov/pubmed/24344450

Good point. Change made.

11. Did hospitalizations include deaths in hospital?

Yes. Have now included this information in the Methods.

12. Page 6. Was bootstrapping necessary for confidence intervals? Standard error estimates should be readily available using the sampling weights (could use SAS PROC SURVEYFREQ).

Statistics Canada specifies a protocol for estimating confidence intervals of their survey data; we followed their protocol. This is now indicated in the Methods.

13. Page 6. I tend to agree that since helmet wearing is legislation for "on-road" use, some effort should be made to compare hospitalizations from similar circumstances. However, is there evidence helmet use differs much from on- to off-road cycling? My guess is more mountain bikers wear helmets than those on-road.

We agree, but could not find data confirming this. The high all cause hospitalization rates for British Columbia (and its big relative drop in the traffic-related rates) may indicate that the risks of mountain biking there trump the high use of sophisticated protective gear (helmets with neck rolls and chin bars, neck braces, etc.) ... but without more data, "may" is the operative word.

14. Page. 6. "Data on population size??"

Good point - change made.

15. Page 8. I highly recommend not using Excel to compute rates to minimize data errors.

Agree that Excel can be dangerous. All rates have now been recalculated using programming in the statistical software. (Pleased to say the triple checking we did of the original work in Excel meant there were no changes.)

16. Data availability: I'm sure it would be illegal for the authors to share the raw data; however, could the aggregated data used in the analysis be made available along with SAS code (or similar) to perform the analyses?

We received the data in tabulated format from CIHI, not raw data. The tabulated data, perhaps because of small numbers in some cells, is restricted by CIHI.

VERSION 2 – REVIEW

REVIEWER	Jake Olivier School of Mathematics and Statistics University of New South Wales
	Sydney, Australia
REVIEW RETURNED	10-Jun-2015

GENERAL COMMENTS	The authors have taken great care and effort in addressing all the reviewers' comments. In the revision, the authors state
	"In our view, the most important implication of our results is that factors other than helmet legislation influenced bicycling hospitalization rates, whereas helmet legislation did not. Females had
	slowly, and to choose routes on quiet streets and with bike-specific infrastructure.[16, 39-41] We also found lower traffic-related
	hospitalization rates with higher cycling mode shares. Here too there is a reasonable link to safer bicycling infrastructure, since it has been shown to draw more people to bicycling.[43,44]"
	This appears to be the key finding in this paper. I would agree with this statement if I didn't still believe there were data and analytic shortcomings. I've addressed each separately below and a list of other issues.
	DATA ISSUES
	I raised data issues in my initial review regarding the alignment of data sources, although the authors "do not believe this is a major limitation."
	From my reading of the manuscript, hospitalization data covers the period April 2006 - March 2011 and is aggregated in 1 year periods (i.e., April to March); and bike trip survey data is from January to December of years 2005, 2007/8, 2009/10, and 2011/12. There is data on school/work days, but it is unclear if it spans calendar of fiscal years (it seems to be calendar years from the last review).
	First, it is clear the hospital and bike trip data do not align. As an example, consider the years 2007-2008. Hospitalizations would be observed from April 2007 - March 2009, while bike trips cover the period January 2007 - December 2008. Out of 24 possible months of data, 6 months of data do not overlap (Jan-Mar 2007, Jan-Mar 2009). The authors believe this is ignorable because their estimated comparison of provinces/territories with or without legislation was in the opposite, unexpected direction. I mentioned in my initial review this can bias the analytic results and this was also a problem in the 2013 Dennis paper (Dennis is a co-author on this paper). Let me elaborate more on that issue using the Ontario helmet law for children as a demonstration (injury count data provided by Dennis).
	The Ontario helmet law for kids was October 1995 and Dennis used yearly aggregated hospital data from 1994-2008 (or more correctly Apr 1994 – Mar 2009), one of their data points (Apr 1995 – Mar 1996) is neither pre- or post-helmet law, i.e., the pre- and post-

helmet law periods do no align with their hospital data. The authors defined this observation as "pre-law" although only 5/12 months of data occurred before the law. If I fit an overdispersed Poisson interrupted time series model with t=0 corresponding to year 1996 and treat 1995 data as "pre-law", I get a non-significant helmet law effect (IRR=1.13, 95% CI: 0.85, 1.51). However, if 1995 is treated as "post-law" with t=0 corresponding to year 1995, I get a significant helmet law effect (IRR=0.75, 95% CI: 0.65, 0.86).
I have attached a scatterplot of the rates of head injury for children in Ontario by year along with the estimated pre-law trajectories for the two scenarios. The significance (or insignificance) of the helmet law depends on whether the 1995 data point is considered pre- or post- law. Note that neither of these analyses adequately addresses the research question and a better analysis would use data that clearly separated pre- and post-law periods.
Another issue, albeit separate from this paper, is pre-law trends cannot be estimated from one or two pre-law observations (even three would be questionable). Any estimated model would go through the observed values as there aren't enough model degrees of freedom to estimate a variance. This issue would also be true for the New Brunswick, BC and Nova Scotia analyses.
The 2006 bike trip data is missing and it is unclear how the authors dealt with that missing information. In the methods, 2006 data is imputed using the 2005 observations as a last observation carried forward (page 26, lines 21-24), while the discussion indicates the 2006 data was interpolated from the 2005 and 2007/8 surveys (page 35, lines 13-14). Note that last observation carried forward or mean replacement are considered poor imputation strategies as they artificially minimise variance estimates.
I am concerned about Bike Score since data from one city is used over an entire province. Does Bike Score of one city accurately describe the rideability of a whole province? That seems unlikely to me.
ANALYTIC ISSUES
I mentioned in the previous review the analytic choice (logit transformed normal-type models) is the reason strata with zero hospitalizations are not analysed or the reason analytic choices were made to minimize the occurrence of zero cell counts.
In my original review, I stated
"Count data don't usually follow a normal distribution (especially when counts are small) and a Poisson-type analysis (Poisson, over- dispersed Poisson or negative binomial) is a much better choice. Poisson-type models also allow for the inclusion of an offset like number of cycling trips. This modelling choice would also eliminate the problem with zero cell counts as the log Poisson mean is modeled instead of the observations themselves. PROC GLIMMIX in SAS could be used to account for dependence in the data for a Poisson response."
The authors respond they tried Poisson and overdispresed Poisson models without much luck. Have they tried the negative binomial or any other GLM-type model that doesn't allow for zero cell counts? It

still seems strange to me data is eliminated because of a modelling choice when there are other analytic options.
Page 11, lines 14-21. The inclusion of helmet wearing rates will not change the sign of the effect unless, perhaps, helmet wearing was less in helmet law provinces which is highly unlikely. Instead, it serves to adjust the effect size relative to the difference in helmet wearing across those provinces. Also keep in mind the assumed relationship between helmet wearing and injury rates is linear on a logit scale (based on the authors model choice). Is there any evidence that functional form is true? There are boundary issues with helmet wearing data as it must be in the interval [0%, 100%].
Page 50, line 46. The authors claim to have performed a logistic regression, although this doesn't appear to be the case elsewhere (page 27, lines 33-38). Using the logit transformation on the injury rate is not a logistic regression. However, I think a logistic regression approach would be an improvement here as it would deal with zero cell counts. Again, in a generalised linear model, it is a function of the mean that is modelled and not the transformed data. This alleviates many of the issues.
OTHER ISSUES
Page 24, line 21. Most helmet efficacy studies are case-control designs. Therefore, the parameter of interest is the odds ratio and not the relative risk. (also page 33, line 27)
Page 6, lines 36-41. Why weren't injury rates calculated per year (or I presume two year periods when the bike trip data spanned two year periods)? The authors indicate no trend in bike trips; however, that doesn't mean there wasn't a temporal effect for hospitalisations and/or the rates of hospitalisations.
Table 1. 64.7% of hospitalisations were "V18: non-collision transport accident". Although I'm very supportive of the authors' conclusion cycling infrastructure is the primary factor related to safety here, how will infrastructure help if most hospitalisations do not involve other vehicles? Does this imply infrastructure should not just be about separation of cyclists from motorised traffic, but also improvements with how the cyclist interacts with the environment like placement of street furniture?
Page 29, line 15. Is cycling mode share and Bike Score correlated by design? I don't know how Bike Score is computed (is it proprietary?), but presumably cycling mode share could be one of the inputs to the Bike Score equation. If so, that raises issues of including them both in the same model.
Figure 3. Why are separate models fit to injury types? A piece-meal approach seems inadequate to me. It is possible to get individual estimates while modelling them jointly.
Page 33, lines 27-37. It is unclear how case-control studies and population studies should align (assuming helmets mitigate head injury). There is less control of confounders in population studies than case-control studies, so I don't understand why some have taken the apparent disconnect between case-control and population studies about helmets as an indication case-control studies are wrong. It certainly wouldn't be improved by "the selection of a control

group unexposed to neimet legislation" as in the 2013 Dennis study. This may only serve to increase problems with uncontrolled confounding as jurisdictions will differ in more ways than just legislation. Studies that use matching, on the other hand, are better suited for dealing with uncontrolled confounding as pairs would presumably have similar confounders (measures and unmeasured). This also includes selecting controls (via other injuries) from the same population as the cases. This is what we did in our 2011 and 2013 papers comparing trends in head and limb injuries. The uncontrolled confounding (like changes in cycling numbers or injury risk) would affect them similarly. Other jurisdictions are important to study as they give more information about the (in)effectiveness of an intervention, but they are not necessarily better controls.
Page 34, line 15. The treatment of risk compensation is very one sided. The Messiah study found a 3kph difference in cycling speed for males at one site (14.5kph vs 17.5kph), but this difference disappeared as the average speed of the site increased (20.9kph vs 20.7kph). Is a 3kph increase in speed in a lower speed area an important increase in injury risk? I don't think I would be able perceive that little of difference while cycling unless I had a speedometer. Also, the overtaking study by Walker (who is also a named reviewer to this submission) has been challenged by me and Scott Walter
http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0075 424
and Walker's follow up study found no helmet effect (albeit a different study design).
http://www.ncbi.nlm.nih.gov/pubmed/24333770
Page 49-50. The hypothesis helmets shift the injury severity in the real world is an open ended question. There is evidence of this in biomechanical studies (e.g., Cripton 2014 and the simulation studies by Donal McNally), but there has never been a study designed to answer this question in humans. It would require more information than just injury severity or fatality. Randomised controlled trials would better answer this question, but such a study would never pass research ethics.
Page 54, line 25. There is very little on mountain biking and anything to do with helmets. Here are a couple I've found.
http://www.wemjournal.org/article/S1080-6032(12)00015-4/fulltext http://bmb.oxfordjournals.org/content/85/1/101.full
1995: Pre-Law 1995: Post-Law
0 0 Observed Rate 0 Pre-law trajectory 9 - 9 - 000001 - 9 - 0 0 0 0 0 0
0
Year Year

VERSION 2 – AUTHOR RESPONSE

Responses to reviewer Reviewer(s)' Comments to Author: Reviewer Name Jake Olivier Institution and Country School of Mathematics and Statistics University of New South Wales Sydney, Australia

The authors have taken great care and effort in addressing all the reviewers' comments. In the revision, the

authors state

"In our view, the most important implication of our results is that factors other than helmet legislation influenced bicycling hospitalization rates, whereas helmet legislation did not. Females had lower rates in our study and they have been shown to cycle more slowly, and to choose routes on quiet streets and with bike-specific infrastructure.[16, 39-41] We also found lower traffic-related hospitalization rates with higher cycling mode shares. Here too there is a reasonable link to safer bicycling infrastructure, since it has been shown to draw more people to bicycling.[43,44]" This appears to be the key finding in this paper. I would agree with this statement if I didn't still believe

This appears to be the key finding in this paper. I would agree with this statement if I didn't still believe there

were data and analytic shortcomings. I've addressed each separately below and a list of other issues. DATA ISSUES

A. I raised data issues in my initial review regarding the alignment of data sources, although the authors

"do not believe this is a major limitation."

From my reading of the manuscript, hospitalization data covers the period April 2006 - March 2011 and is aggregated in 1 year periods (i.e., April to March); and bike trip survey data is from January to December of years 2005, 2007/8, 2009/10, and 2011/12. There is data on school/work days, but it is unclear if it spans calendar of fiscal years (it seems to be calendar years from the last review). First, it is clear the hospital and bike trip data do not align. As an example, consider the years 2007-2008. Hospitalizations would be observed from April 2007 - March 2009, while bike trips cover the period January 2007 - December 2008. Out of 24 possible months of data, 6 months of data do not overlap (Jan-Mar 2007, Jan-Mar 2009). The authors believe this is ignorable because their estimated comparison of provinces/territories with or without legislation was in the opposite, unexpected direction. I mentioned in my initial review this can bias the analytic results and this was also a problem in the 2013 Dennis paper (Dennis is a co-author on this paper). Let me elaborate more on that issue using the Ontario helmet law for children as a demonstration (injury count data provided by Dennis).

We have tried to make clearer in the Methods that we received tabulated hospitalization data for all 6 study years combined from the Canadian Institute for Health Information. We have added a supplementary file that shows the tabular format of the hospitalization data request (i.e., also the format of the data received). Each table included data for all six years combined, not stratified by year. This means that the mismatch in years for the rate calculations is three months at the beginning of the study and three months at the end (6 months of 72 months). We could not find where we suggested that this mismatch is ignorable or that the helmet legislation direction of effect was a reason to ignore it. A section of the limitations highlights the mismatch. The reason we believe the mismatch is not a major limitation is that bicycling trip estimates (data which we did have at a finer temporal scale than 6 years) changed relatively little within-strata from one survey time period to the next.

B. The Ontario helmet law for kids was October 1995 and Dennis used yearly aggregated hospital data

from 1994-2008 (or more correctly Apr 1994 – Mar 2009), one of their data points (Apr 1995 – Mar 1996) is neither pre- or post-helmet law, i.e., the pre- and post-helmet law periods do no align with their hospital data. The authors defined this observation as "pre-law" although only 5/12 months of data occurred before the law. If I fit an overdispersed Poisson interrupted time series model with t=0 corresponding to year 1996 and treat 1995 data as "pre-law", I get a non-significant helmet law effect (IRR=1.13, 95% CI: 0.85, 1.51). However, if 1995 is treated as "post-law" with t=0 corresponding to year 1995, I get a significant helmet law effect (IRR=0.75, 95% CI: 0.65, 0.86).

I have attached a scatterplot of the rates of head injury for children in Ontario by year along with the

estimated pre-law trajectories for the two scenarios. The significance (or insignificance) of the helmet law depends on whether the 1995 data point is considered pre- or post-law. Note that neither of these analyses adequately addresses the research question and a better analysis would use data that clearly separated pre- and post-law periods.

Another issue, albeit separate from this paper, is pre-law trends cannot be estimated from one or two pre-law observations (even three would be questionable). Any estimated model would go through the observed values as there aren't enough model degrees of freedom to estimate a variance. This issue would also be true for the New Brunswick, BC and Nova Scotia analyses.

The above comments refer to a different paper, published in BMJ in 2013, so we are unable to respond. It used a different Canadian hospitalization dataset (1994 to 2008, all ages) and a different denominator (census population). It had a different study purpose and design (temporal change in hospitalization rates with helmet legislation, accounting for baseline trends). The only elements these two papers have in common are one author (JD) and the source of hospitalization data.

C. The 2006 bike trip data is missing and it is unclear how the authors dealt with that missing information. In the methods, 2006 data is imputed using the 2005 observations as a last observation carried forward (page 26, lines 21-24), while the discussion indicates the 2006 data was interpolated from the 2005 and 2007/8 surveys (page 35, lines 13-14). Note that last observation carried forward or mean replacement are considered poor imputation strategies as they artificially minimise variance estimates.

We used 2005 data to estimate 2006 leisure cycling trips. In the years prior to 2007 when Statistics Canada administered the Canadian Community Health Survey in a single year biennially, it was meant to apply to a two-year period. We followed standard practice by using the 2005 data for 2006 estimates.

The question about bicycling to work or school was not asked in any of the Canadian Community Health Surveys until 2007, so our only option was to use 2007 data for 2006. D. I am concerned about Bike Score since data from one city is used over an entire province. Does Bike

Score of one city accurately describe the rideability of a whole province? That seems unlikely to me. Bike Score data was available only for one city in each province, and it was not available for every province. These are the reasons it was not used in the analyses of hospitalization rates. We still felt it was useful data to report descriptively, since readers may wonder about differences in cycling conditions between provinces. The cities for which Bike Score was available often comprised about half of the provincial population. We have revised the wording about Bike Score in the Methods.

ANALYTIC ISSUES

E. I mentioned in the previous review the analytic choice (logit transformed normal-type models) is the reason strata with zero hospitalizations are not analysed or the reason analytic choices were made to minimize the occurrence of zero cell counts.

In my original review, I stated

"Count data don't usually follow a normal distribution (especially when counts are small) and a Poisson-type analysis (Poisson, over-dispersed Poisson or negative binomial) is a much better choice. Poisson-type models also allow for the inclusion of an offset like number of cycling trips. This modelling choice would also eliminate the problem with zero cell counts as the log Poisson mean is modeled instead of the observations themselves. PROC GLIMMIX in SAS could be used to account for dependence in the data for a Poisson response."

The authors respond they tried Poisson and overdispresed Poisson models without much luck. Have they tried the negative binomial or any other GLM-type model that doesn't (should this be "does"?) allow for zero cell counts? It still seems strange to me data is eliminated because of a modelling choice when there are other analytic options.

In the reviewer's comments about the analytical methods in the first review, two issues were raised. One was that the distributions of the dependent variables (hospitalization rates) may not conform to the distributional assumptions of the method. We checked the goodness of fit of the rate data untransformed (normal distribution), count data (Poisson distribution), and the rate data logit-transformed (normal distribution). The logit-transformed rate data fit its distributional assumptions best and the count data least well. Even accounting for overdispersion, Poisson models did not perform well (e.g., distributions of residuals), whereas the models using the logit-transformed rates did.

The other issue raised was zero hospitalization counts. Zero counts are handled via the modeling method in Poisson or negative binomial regression (by estimating the mean). In the

models using the logit-transformed rates, we handled the zero counts manually instead. Please note that we did not eliminate any data.

We prioritized the better distributional fit over concerns about how to handle the zero counts, for two reasons. First, for the main rates reported (Figure 3, 132 rates plotted) and analysed (Table 4), there was only one zero count (1 of 44 for injuries to the head, scalp, skull or face, all causes). Second, we tested two substitution methods for affected rates, a rate calculated with a count of 0.1 instead of zero and the mean rate for the body region, and these methods had nearly identical results. This was also true for the analyses reported in Figure 4, even where more substitutions were needed (i.e., face and neck injuries). We have now also checked our choice of substitute value using GLM (Poisson), by comparing analyses with zero counts to analyses substituting 0.1 for the zero counts. Again, the results were nearly identical. F. Page 11, lines 14-21. The inclusion of helmet wearing rates will not change the sign of the effect unless, perhaps, helmet wearing was less in helmet law provinces which is highly unlikely. Instead, it serves to adjust the effect size relative to the difference in helmet wearing across those provinces. Also keep in mind the assumed relationship between helmet wearing and injury rates is linear on a logit scale (based on the authors model choice). Is there any evidence that functional form is true? There are boundary issues with helmet wearing data as it must be in the interval [0%, 100%]. Helmet legislation resulted in consistently higher helmet wearing rates, as shown in Table 3 and Figure 2, so these variables were strongly correlated (r = 0.86). None of the helmet wearing rates was close to the boundaries of 0 or 1. The distribution was bimodal, with peaks around the means for helmet legislation and no helmet legislation. The directions of relationships between helmet wearing and injury rates were the same in the original analyses (linear regression, injury rates untransformed as the dependent variable) and in the revised analyses (linear regression, injury rates logit-transformed as the dependent variable), suggesting they were robust to functional form. Only the direction of the relationships was reported in the paper.

G. Page 50, line 46. The authors claim to have performed a logistic regression, although this doesn't appear to be the case elsewhere (page 27, lines 33-38). Using the logit transformation on the injury rate is not a logistic regression. However, I think a logistic regression approach would be an improvement here as it would deal with zero cell counts. Again, in a generalised linear model, it is a function of the mean that is modelled and not the transformed data. This alleviates many of the issues.

We mentioned "logistic" in our response to reviewer Walker, not in the paper itself. In the paper, we describe what we did as linear regression with a logit-transformed dependent variable. The use of the word "logistic" in our response to Walker was meant in the sense that he discussed (a model resulting in results on a geometric rather than additive scale). OTHER ISSUES

H. Page 24, line 21. Most helmet efficacy studies are case-control designs. Therefore, the parameter of

interest is the odds ratio and not the relative risk. (also page 33, line 27)

We used the term relative risk because we felt it would be more well known and understood by diverse readers of transportation literature and because odds ratios are good estimates of relative risk when rates are very low (below 0.001 is the usual criterion) and this is the case for bicycling hospitalization rates.

But you are right, it is not as precise, so we have searched the paper for the use of the term relative risk and changed it to indicate relative odds where case-control studies were the source of estimates.

I. Page 6, lines 36-41. Why weren't injury rates calculated per year (or I presume two year periods when the bike trip data spanned two year periods)? The authors indicate no trend in bike trips; however, that doesn't mean there wasn't a temporal effect for hospitalisations and/or the rates of hospitalisations.

We did not receive annual data for hospitalizations so could not calculate rates for each year (see A above). We did not request this data because trends were not the subject of the paper. J. Table 1. 64.7% of hospitalisations were "V18: non-collision transport accident". Although I'm very supportive of the authors' conclusion cycling infrastructure is the primary factor related to safety here, how will infrastructure help if most hospitalisations do not involve other vehicles? Does this imply infrastructure should not just be about separation of cyclists from motorised traffic, but also improvements with how the cyclist interacts with the environment like placement of street furniture? Good point. It is difficult to know how well hospital ICD coding captures the nuances of cycling crash circumstances. A bicycling injury study done by an author of this paper (KT)

collected more detailed data and found that about a third of transport cycling crashes were direct collisions with motor vehicles and another 15% indirectly involved motor vehicles when cyclists crashed as they were trying to avoid a vehicle collision. In addition, as you suggest, another third of crashes involved bicyclists only but were collisions with infrastructure (e.g., street car tracks, bollards, curbs, street furniture). This suggests that a large majority of crashes could be avoided or mitigated with improved infrastructure.

We searched for a spot to add this detail to the paper, but we worry that it strays from the results, which are related to female route choices and infrastructure that increases cycling. To our knowledge, there is no data relating other infrastructure components to either of these. K. Page 29, line 15. Is cycling mode share and Bike Score correlated by design? I don't know how Bike

Score is computed (is it proprietary?), but presumably cycling mode share could be one of the inputs to the Bike Score equation. If so, that raises issues of including them both in the same model. Bike Score for U.S. cities includes cycling mode share as one component, but for Canadian cities, it does not. The components of Bike Score for Canadian cities are listed in the Methods under "Other data sources". Bike Score is proprietary, but there is a group (including KT) preparing an article describing it for an academic audience. Unfortunately it is not yet ready for citation to provide more information to readers.

L. Figure 3. Why are separate models fit to injury types? A piece-meal approach seems inadequate to me. It is possible to get individual estimates while modelling them jointly.

We are not certain what is meant by injury "type". Perhaps "injury cause"? If so, traffic causes are a subset of all injury causes. Figure 3 was meant to show rates descriptively, not to report models.

Perhaps this comment refers to Figure 4 and injury "type" refers to "body region". If this is the case, is this suggesting multivariate regression with rates for every body region as multiple dependent variables? It is not clear that the complexity of a multivariate model would add value. The intent of Figure 4 was to show simple associations before providing the modeling results of Table 4.

M. Page 33, lines 27-37. It is unclear how case-control studies and population studies should align (assuming helmets mitigate head injury). There is less control of confounders in population studies than case-control studies, so I don't understand why some have taken the apparent disconnect between case-control and population studies about helmets as an indication case-control studies are wrong. It certainly wouldn't be improved by "the selection of a control group unexposed to helmet legislation" as in the 2013 Dennis study. This may only serve to increase problems with uncontrolled confounding as jurisdictions will differ in more ways than just legislation. Studies that use matching, on the other hand, are better suited for dealing with uncontrolled confounding as pairs would presumably have similar confounders (measures and unmeasured). This also includes selecting controls (via other injuries) from the same population as the cases. This is what we did in our 2011 and 2013 papers comparing trends in head and limb injuries. The uncontrolled confounding (like changes in cycling numbers or injury risk) would affect them similarly. Other jurisdictions are important to study as they give more information about the (in)effectiveness of an intervention, but they are not necessarily better controls.

We believe this is not a request for changes, but rather musings on the issue that we tried to address in the Discussion – why there are differences in results related to helmet legislation among the various studies.

We certainly hope that readers don't interpret our remarks as implying that "case-control studies are wrong". We believe that in areas of research where observational studies may be the only option, examining issues using many different observational designs is an aid to understanding (though understanding doesn't necessarily lead to clarity).

N. Page 34, line 15. The treatment of risk compensation is very one sided. The Messiah study found a 3kph difference in cycling speed for males at one site (14.5kph vs 17.5kph), but this difference disappeared as the average speed of the site increased (20.9kph vs 20.7kph). Is a 3kph increase in speed in a lower speed area an important increase in injury risk? I don't think I would be able perceive that little of difference while cycling unless I had a speedometer. Also, the overtaking study by Walker (who is also a named reviewer to this submission) has been challenged by me and Scott Walter

http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0075424

and Walker's follow up study found no helmet effect (albeit a different study design). http://www.ncbi.nlm.nih.gov/pubmed/24333770

We have made clearer in the Discussion that results in this area aren't always consistent (but

since we were not trying to review the risk compensation literature, haven't cited other papers). We have also made clearer that the two papers cited were each single papers on their topic. We rechecked the Messiah paper. We reported as the authors did, but reported the direction of effect rather than specifics. We have made a revision to make clear it wasn't a universal effect (pointing out the lack of effect in women rather than the difference in male effect in the top speed category). We were aware of the reviewer's critique of Dr. Walker's paper, so worded the Walker results in a way that was supported by the reanalyses reported in Table 3 of the PLoS One critique.

O. Page 49-50. The hypothesis helmets shift the injury severity in the real world is an open ended question. There is evidence of this in biomechanical studies (e.g., Cripton 2014 and the simulation studies by Donal McNally), but there has never been a study designed to answer this question in humans. It would require more information than just injury severity or fatality. Randomised controlled trials would better answer this question, but such a study would never pass research ethics. Agree, we would love to see a randomized trial of helmet use. (A trial of helmet laws does not seem possible or at least would involve a whole other level of difficulty – randomizing jurisdictions.)

There have been some lovely randomized trials recently in Denmark – of day time running lights and high visibility clothing. I wonder if the same group might be trying one of helmets (easier to pass ethics review where there is less emphasis on helmets). The main difficulty may be that the sample size would have to be so much greater, since the outcome would be considerably rarer (i.e., not just crashes, but crashes causing injuries; and not all injuries, just head injuries).

P. Page 54, line 25. There is very little on mountain biking and anything to do with helmets. Here are a couple I've found. http://www.wemjournal.org/article/S1080-6032(12)00015-4/fulltext http://bmb.oxfordjournals.org/content/85/1/101.full

Thank you for reminding us about these articles. We reread them and thought that the main item that might be of interest to readers is that many of the Canadian ski resorts that sell summer lift access to mountain bikers require them to wear helmets. There is no helmet requirement for riders who access the mountain back country independently. We have not added these points because they don't seem to fit in the flow of the paper, and they don't aid understanding of the differences in hospitalization rate distributions for all causes vs. trafficrelated causes.

Q.

These graphics refer to adifferent paper, published inBMJ in 2013, so we are unableto respond (see B above).

REVIEWER	Jake Olivier UNSW
	Sydney, Australia
REVIEW RETURNED	13-Jul-2015

VERSION 3 - REVIEW

GENERAL COMMENTS	This is now my third review of this manuscript as I reviewed the original submission and the first revision. I find the current version to be the most disappointing of the three and I do not recommend publication.
	In each of my previous reviews, I have commented about the lack of data alignment for the hospital and bike trip data. In the current version, the authors claim there is a "3-month discrepancy at either end of the 6-year study period (6 of 72 months)." This is not even remotely true. I repeat below what I wrote in my last review.
	From my reading of the manuscript, hospitalization data covers the period April 2006 - March 2011 and is aggregated in 1 year periods (i.e., April to March); and bike trip survey

data is from January to Dec 2009/10, and 2011/12.	ember of years 2005, 2007/8,
For years 2007/08, there is hospital March2009 and bike trip data spann 2008. There are three months where hospital data (Jan 2007-Mar2007) a there is hospital data but no bike trip same is true for years 2009/10 and 2 discrepancy.	data spanning April 2007- ing January 2007-December e there is bike trip data but no nd another three months where o data (Jan 2009-Mar2009). The 2011/12 for a total 18 month
The 2006 data is unique as there is be imputed but, given the source da January-December 2006. There is the data discrepancy which is 33.3% of since there is no 2006 data, it is real	no bike trip data. This data can ta used for imputation, will span herefore at least 24 months of all months (24/72). However, Ily 41.7% (30/72).
This is a major limitation of this stud reviews. I am perplexed as to why th correct this problem. There is nothin trip data; however, hospital data can including following a calendar year. virtually eliminates this limitation (the be missing) and therefore my criticis	y as I've stated in the past two ne authors have not sought to ig that can be done to the bike in be aggregated different ways Using more appropriate data the 2006 bike trip data would still sm.
In my last review, I provided a detail demonstrating the problems that car not align. The data misalignment this legislation in Ontario. The date of the making it impossible to separate pre- aggregated data on an April-March y observation is neither pre- nor post- apparent effectiveness of the helme observation is included as a pre-law graph below). I find that result thorou unfortunate the authors of the curren address this issue.	ed example using Canadian data n occur when data sources do s time relates to bicycle helmet e legislation was October 1995 e- and post-law periods for yearly year. The April 1995-March 1996 law and, importantly, the t law depends on whether this or post-law observation (see ughly disturbing and I find it nt submission have refused to
1995: Pre-Law	1995: Post-Law
Pre-law trajectory Pre-law trajectory Heimet Law Date 0 0 0 0 0 0 0 0 0 0 0 0 0	No Observed Rate Pre-law trajectory Helmet Law Date Pre-law trajectory O </td
1994 1998 2002 2006 Year	1994 1998 2002 2006 Year

I also repeat my concerns about choice of modelling. Poisson or logistic models are standard for data such as this. It is highly unusual to logit transform computed rates and then perform a linear regression. The statistical inferences from these approaches are not the same. One issue with logit transformed rates in a linear regression model is cells with zero counts cannot be transformed. The authors have attempted to deal with this issue by ignoring the influence of gender. On page 6, the authors state

> Because traffic-related injuries were only about half of all injuries, these data were not stratified by sex, to minimize the number of strata with zero hospitalizations. For each body region, rates were calculated for 22 strata: 11 jurisdictions * 2 age groups.

I find the aggregation across genders confusing considering the paper's conclusions are supportive of "female cycling choices". However, we know nothing about the influence of gender on trafficrelated injuries because the authors have insisted on an archaic modelling choice instead of a Poisson-type model (or any other categorical analysis for larger contingency tables like Cochran-Mantel-Haenszel, for example).

Here is a simple demonstration of the problem for cycling injuries by gender and bikelane.

	Bikelane	No
Female	100	0
Male	50	50

Although there were no injuries to females outside of bikelanes, the chi-square test for independence can be performed since the expected count for that cell is 25 (=100*50/200). This test is statistically significant (p<0.0001) indicating gender is associated with location injury occurred. In particular, there is a difference in injuries between bikelane/no bikelane for females but not males. If this table is aggregated across gender, the result is also highly significant (p<0.0001) for the difference in injuries in a bikelane or not. However, the interpretation would be incorrect as there is no such difference for males.

I am concerned this is also what has happened in this submission. This is a limitation that is completely avoidable if the authors would do a proper analysis.

There is certainly a bicycle helmet controversy. It existed long before I started doing research into cycling injury and will continue until I die if we continue at the current rate of progress. This paper has the potential to shed some real light on this subject and, as I've mentioned in each review, I'd believe the authors' results if the

correctable limitations are corrected. Since the authors have refused to do that, I cannot recommend publication.
Other Issues
Page 33, line 17. How can you demonstrate inconstancies in research regarding risk-related behaviours if you only discuss one side? I've mentioned this in my other two reviews as well.
Page 40. Standard practice for SAS is to code binary variables as - 1/1. This is called GLM coding and some statistical packages use reference coding instead (0/1). A logistic model in SAS would have presented the correct odds ratios irrespective of coding.
Response A. The data discrepancy is not ameliorated by aggregating over all 6 years. You should be analysing year-based rates and incorporating that into an analysis. If this does not change model estimates, then you are correct. However, this is all unknown until you do that analysis.
Response B. How are baseline trends estimated with one or two pre- law data point(s)?
Response C. I find it unusual that last observation carried forward would be considered standard practice. The missing data literature vehemently argues against such practice.
Response H. Why not make the argument the probability of an event is small so that the odds ratio is similar to the relative risk in the paper? Calling an odds ratio a relative risk is just wrong and contributes to the lack of understanding of what an odds ratio is.
Response I. Why not ask the Canadian government for aggregated data that matches the bike trip data? This is the core problem with this paper, in my opinion.

REVIEWER	Sara Farchi
	Department of Epidemiology, Lazio regional Health Service, Rome,
	Italy
REVIEW RETURNED	12-Oct-2015

GENERAL COMMENTS	This is an interesting paper on bicycle injuries. It's well written, and the problem is well addressed. I have no
	comments

VERSION 3 – AUTHOR RESPONSE

Dr. Olivier repeatedly overestimated the temporal mismatch in our data (a calendar year to fiscal year difference). Initially we thought that we were unclear about our methods. We received all six years of numerator data as pooled tabulated data and summed all six years of denominator data, then

calculated hospitalization rates. But in the third review, Dr. Olivier repeated his second review statement. In that statement, he understands the effect of pooling 2 years of data (2007 and 2008) such that in 24 months of data, there would be three months at either end with a mismatch. But he does not allow that by pooling all six years of data before the rate calculation, there is still only a three-month offset, affecting the three months at either end of the 6-year period. Additional mismatch was created because of Statistics Canada's survey methodology prior to 2007, but there was data in both the numerator and denominator that completely corresponded from January 1 2007 though December 31 2011 (60 of the 72 months; 83% of the data).

In his second review, Dr. Olivier compared our study to one that is not similar: a study published in 2013 in BMJ by one of our co-authors (Jessica Dennis) and others.1 The 2013 paper examined temporal trends in hospitalization rates (per unit population) before, during and after helmet laws were implemented (or not) within each Canadian province. It was fundamentally different from the current study. Our study compares hospitalization rates (per unit bicycling exposure) between provinces with and without helmet laws, during a time period of helmet law stability (i.e., when no new helmet laws were implemented). Our study did not examine temporal trends, so the argument he raised could not apply. Our explanation of the differences between the studies was not acknowledged. In his third review, Dr. Olivier indicated that he wanted us to respond to his criticism of the 2013 paper: "The April 1995- March 1996 observation is neither pre- nor post-law and, importantly, the apparent effectiveness of the helmet law depends on whether this observation is included as a pre-law or postlaw observation (see graph below). I find that result thoroughly disturbing and I find it unfortunate the authors of the current submission have refused to address this issue." He also separately listed criticism of the 2013 paper within a list of comments on our manuscript: "Response B. How are baseline trends estimated with one or two pre-law data point(s)?" Dr. Olivier's concerns with the 2013 Dennis et al. paper are completely separate from our current manuscript.

Dr. Olivier did not acknowledge extensive analyses we conducted to address concerns he raised in his first review:

 To address his concern about model fit, we consulted a statistician (now a co-author) and completely reanalyzed the data before our second submission. We tested the fit of the count and rate data with four different distributional assumptions, we reanalyzed the data using three modelling methods, including two GLM methods (Poisson and Poisson correcting for overdispersion) as he recommended, and we checked the fit characteristics of each of these models, as he recommended. In the third review, he commented that "the authors have insisted on an archaic modelling choice instead of a Poisson-type model." Yet the Poisson model did not perform as well as the one we used. He recommended other GLM methods in his second review, but we hesitated to redo all analyses and model checking again because we were not 1 Perhaps Dr. Olivier thought that we constituted the same team that authored the 2013 Dennis paper. This was not the case (we explained this in our response to his second review). I (Kay Teschke) was a reviewer of the 2013 Dennis paper. I had not met any of the authors of that paper at the time. After that study was published, I approached Jessica Dennis to discuss Canadian hospitalization data. Those discussions led to the current collaboration. 3 confident it would make a difference to the reviewer. We have checked negative binomial models for the four main analyses presented in Table 4. The results were comparable to the logit-transformed models.

• To check his concern about using 2005 bicycling trip data for 2006 (beyond explaining that this is standard practice for Canadian Community Health Survey data prior to 2007), we compared bicycling trip data estimates between surveys and found that differences in the number of trips by bicycle were small, especially in comparison to between-strata differences, and did not indicate a temporal trend. This was not acknowledged.

In the second review, Dr. Olivier raised two statistical issues that seemed odd:

• He stated that "There are boundary issues with helmet wearing data as it must be in the interval [0%, 100%]." While it is true that helmet wearing is proportion data with potential for boundary issues, the manuscript included a figure (2) that showed the data for % helmet wearing and none of the data came close to the 0% or 100% boundaries.

• He stated "Figure 3. Why are separate models fit to injury types? A piece-meal approach seems inadequate to me. It is possible to get individual estimates while modelling them jointly." This was difficult for us to interpret because Dr. Olivier used terminology (injury type) that was not used in our manuscript and might refer to either injured body region or injury cause. Figure 3 included both, but the data included subsets that were not independent, so not appropriate to analyze together. It is possible that he meant Figure 4, but the method that would be required to model the body region hospitalization rate variables jointly (true multivariate, i.e., multiple dependent variables) seems unnecessarily complex. In Figure 4, we simply presented results of unadjusted analyses.

Dr. Olivier made a number of comments that did not reflect what we did or said:

• In his second review, he stated that "The authors believe this [the temporal mismatch] is ignorable because their estimated comparison of provinces/territories with or without legislation was in the opposite, unexpected direction." We did not say this. Our comment about opposite, unexpected direction results was in reference to another issue in a different section of the manuscript.

• In his second review, he stated that "the analytic choice (logit transformed normal-type models) is the reason strata with zero hospitalizations are not analysed or the reason analytic choices were made to minimize the occurrence of zero cell counts." and "It still seems strange to me data is eliminated because of a modelling choice when there are other analytic options." This appeared to be the reason for dismissing our reanalyses (described above). We did not eliminate any data or make decisions about data based on modeling choice. We described how we treated zero counts in every version of the paper and checked its effects in multiple ways (including Poisson models, described in our second response). This was not acknowledged.

• In his second review, he stated that he "mentioned in the previous review the analytic choice (logit transformed normal-type models)", but in his first review, he did not mention logit-transformed models, nor indicate they were a concern. We did not use logit-transformed data in the initial submission. 4

• In his second review, he stated "I don't understand why some have taken the apparent disconnect between case-control and population studies about helmets as an indication case-control studies are wrong. It certainly wouldn't be improved by "the selection of a control group unexposed to helmet legislation" as in the 2013 Dennis study." We are unsure whether he thought we implied that case-control studies are wrong. We did not describe any study designs as being wrong, we did not specifically name casecontrol designs, and we did not raise the 2013 Dennis study as being correct in comparison to others. We outlined some design differences between studies with differing results related to helmet laws.

• In the third review, he stated "How can you demonstrate inconstancies in research regarding riskrelated behaviours if you only discuss one side? I've mentioned this in my other two reviews as well." We could not find a comment of this nature in his first review. We did not write about risk-related behaviours in our initial submission. This was added in the second submission in response to reviewer Walker. After Dr. Olivier raised one-sidedness in his second review, we added a sentence to make clear that studies of risk-related behavior do not have consistent results. This was not acknowledged.