



OPEN ACCESS

# Estimating bias from loss to follow-up in a prospective cohort study of bicycle crash injuries

Sandar Tin Tin, Alistair Woodward, Shanthi Ameratunga

Section of Epidemiology and Biostatistics, School of Population Health, University of Auckland, Auckland, New Zealand

## Correspondence to

Dr Sandar Tin Tin, Section of Epidemiology and Biostatistics, School of Population Health, University of Auckland, Private Bag 92019, Auckland 1142, New Zealand; s.tintin@auckland.ac.nz

Received 19 August 2013  
Revised 22 October 2013  
Accepted 24 October 2013  
Published Online First  
11 December 2013

## ABSTRACT

**Background** Loss to follow-up, if related to exposures, confounders and outcomes of interest, may bias association estimates. We estimated the magnitude and direction of such bias in a prospective cohort study of crash injury among cyclists.

**Methods** The Taupo Bicycle Study involved 2590 adult cyclists recruited from New Zealand's largest cycling event in 2006 and followed over a median period of 4.6 years through linkage to four administrative databases. We resurveyed the participants in 2009 and excluded three participants who died prior to the resurvey. We compared baseline characteristics and crash outcomes of the baseline (2006) and follow-up (those who responded in 2009) cohorts by ratios of relative frequencies and estimated potential bias from loss to follow-up on seven exposure-outcome associations of interest by ratios of HRs.

**Results** Of the 2587 cyclists in the baseline cohort, 1526 (60%) responded to the follow-up survey. The responders were older, more educated and more socioeconomically advantaged. They were more experienced cyclists who often rode in a bunch, off-road or in the dark, but were less likely to engage in other risky cycling behaviours. Additionally, they experienced bicycle crashes more frequently during follow-up. The selection bias ranged between -10% and +9% for selected associations.

**Conclusions** Loss to follow-up was differential by demographic, cycling and behavioural risk characteristics as well as crash outcomes, but did not substantially bias association estimates of primary research interest.

estimates differ slightly in the retained sample compared to the full cohort.

While the prospective cohort design has been employed frequently in injury research,<sup>20</sup> little is known about the magnitude and direction of bias from loss to follow-up. The Taupo Bicycle Study is a prospective cohort study of cyclists designed to examine factors associated with regular cycling and injury risk. Research questions of primary interest include injury risks associated with cycle commuting, bunch riding, inconspicuity, distraction and previous crash experience. Outcome data were collected for all participants through linkage to four administrative databases. The participants were also resurveyed 3 years after the study commenced, but only three-fifths responded. Loss to follow-up will not be an issue in analyses using exposures measured at baseline in this study but may impact analyses using those measured in the resurvey. We therefore investigated the impact of (simulated) loss to follow-up on seven relevant exposure-outcome associations of primary research interest.

## MATERIALS AND METHODS

### Design, setting and participants

The sampling frame comprised cyclists, aged 16 years and over, who enrolled online in the Lake Taupo Cycle Challenge, New Zealand's largest mass cycling event held each November and attracting about 10 000 cyclists. Participants have varying degrees of cycling experience ranging from competitive sports cyclists and experienced social riders to relative novices of all ages.

We recruited the majority of participants at the time of the 2006 event for the majority of participants, as described in detail elsewhere.<sup>21</sup> Briefly, we sent email invitations containing a hyperlink to an information page describing the study, to 5653 contestants who provided their email addresses at registration for the event. Those who agreed to take part in the study were taken to the next page containing a web questionnaire and asked about demographic characteristics, general cycling activity and previous crash experience in the preceding year and habitual risk behaviours. A total of 2438 cyclists completed and submitted the questionnaire (43.1% response rate). We recruited another 190 cyclists from the 2008 event by including a short description about the study in the event newsletter. We resurveyed all participants in December 2009 using a web questionnaire containing similar questions as in the baseline survey. A total of 1537 participants completed the questionnaire. We obtained

## INTRODUCTION

By explicitly incorporating the passage of time, prospective cohort studies overcome methodological limitations of many other observational designs, and have arguably provided more credible evidence for decision making.<sup>1</sup> Nevertheless, non-response to baseline and follow-up surveys may occur and if related to exposures, confounders and outcomes of interest, may bias association estimates.<sup>2</sup>

Survey non-response rarely occurs at random and has been associated with sociodemographic factors as well as health outcomes.<sup>3-5</sup> Reassuringly, several prospective studies have found bias on association estimates due to initial non-response to be minimal.<sup>6-10</sup> This may be because exposure information is collected in advance of outcomes of interest. Drop out during follow-up appears more of a concern, but several studies suggest that its impact on associations is modest.<sup>11-19</sup> That is, association



Open Access  
Scan to access more  
free content



CrossMark

**To cite:** Tin Tin S, Woodward A, Ameratunga S. *Inj Prev* 2014;**20**:322-329.

ethical approval from the University of Auckland Human Participants' Ethics Committee.

For this analysis, we restricted the study sample to 2590 participants who were resident in New Zealand at recruitment as the crash outcome data for the overseas participants were not available. We also excluded three deaths that occurred prior to the resurvey. As a result, there were 2587 participants in the baseline cohort, of whom 1526 responded to the second questionnaire administered in 2009. We termed the latter group the 'follow-up cohort' and its members the 'responders', assuming that outcome data were not available from those who did not complete the second questionnaire. Figure 1 presents the flow of study participation and losses.

### Crash outcome data

We collected crash outcome data through record linkage to insurance claims, hospital discharge and mortality data and police reports, covering the period from the date of recruitment to 30 June 2011. All participants consented to link their data to these databases.

### Insurance claims

In New Zealand, the Accident Compensation Corporation (ACC) provides personal injury cover for all residents and

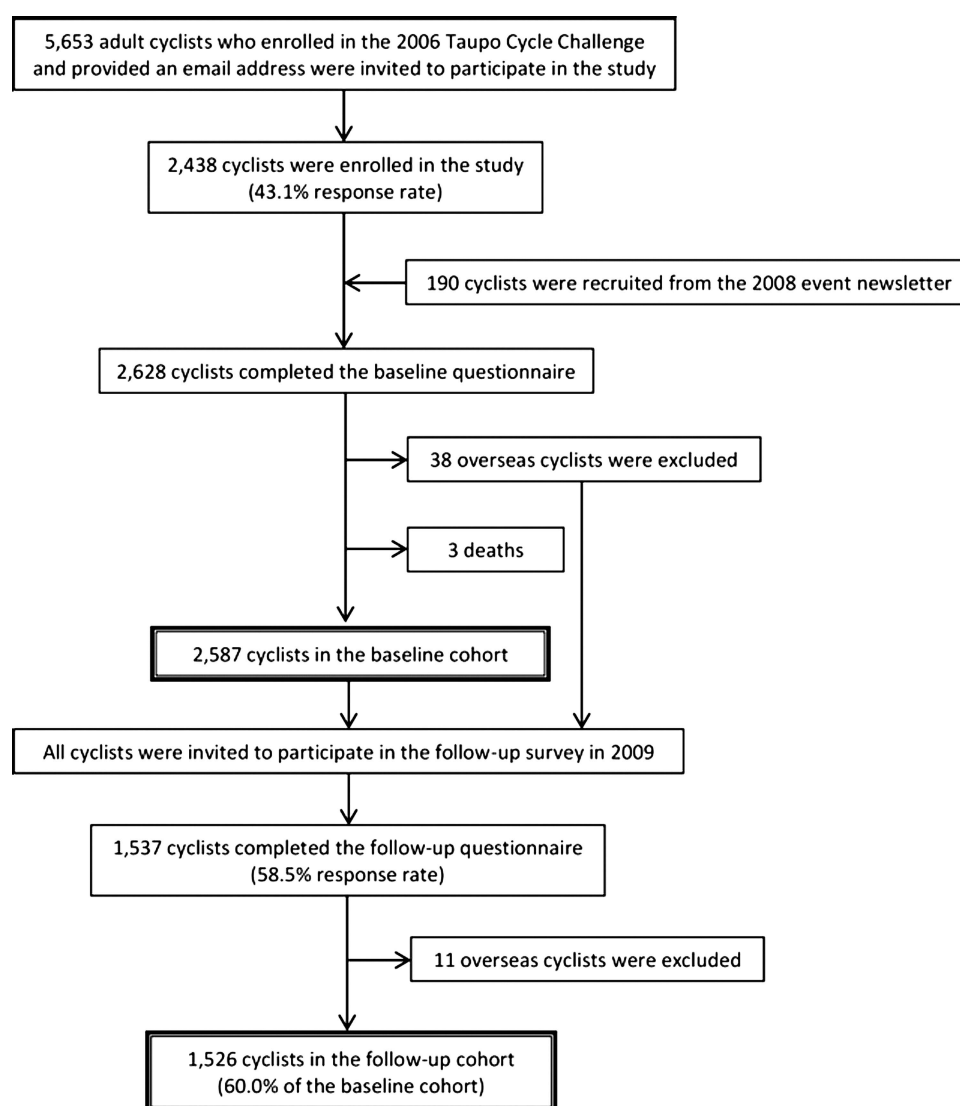
temporary visitors to New Zealand no matter who is at fault. The claims database is a major source of information on relatively minor injuries with over 80% of the claims relating to primary care (eg, general practitioners, emergency room treatment) only.<sup>22</sup> We obtained approval for record linkage from the ACC research ethics committee.

### Hospital discharge and mortality data

The hospital discharge data contains information about inpatients and day patients discharged after a minimum stay of 3 h from all public hospitals and over 90% of private hospitals in New Zealand.<sup>23</sup> The mortality data includes information about all deaths registered in the country.<sup>24</sup> Diagnoses in each hospital visit and underlying causes of death are coded under ICD-10-AM. We identified bicycle crashes using the E codes V10-V19; those that occurred on public roads using the E codes V10-V18.3-9, V19.4-6, V19.9; and those that involved a collision with a motor vehicle using the E codes V12-V14, V19.0-2 and V19.4-6. We then identified readmissions as described previously<sup>25</sup> and excluded them.

### Police reports

In New Zealand, it is mandatory that any fatal or injury crash involving a collision with a motor vehicle on a public road be



**Figure 1** Flowchart of recruitment and loss to follow-up in the Taupo Bicycle Study.

reported to the police. This database, therefore, contains information on all police-reported bicycle collisions.

For each participant, we matched bicycle crashes identified across different databases based on the date of crash allowing for a two-day difference, so as to avoid double-counting of the same crash.

## Analyses

We presented baseline characteristics and crash outcomes of the baseline and follow-up cohorts by relative frequencies (RFs) and made comparisons by ratios of relative frequencies (RRFs) which were calculated as:  $RF_{\text{follow-up}}/RF_{\text{baseline}}$ .

We investigated potential bias from loss to follow-up on seven exposure-outcome associations. These include: associations between previous crash experience, bunch riding, listening to music and using lights in the dark and the risk of all bicycle crashes, and associations between cycle commuting, using fluorescent colours and using reflective materials in the dark and the risk of on-road crashes. For each exposure-outcome association, we calculated individual cell follow-up response rates and

cross-products, using the dichotomous outcome (one or more crashes vs no crash). The cross-products are equivalent to ratios of unadjusted ORs (RORs) which were calculated as  $OR_{\text{follow-up}}/OR_{\text{baseline}}$ .

As more than a single crash may be experienced during follow-up, we performed Cox proportional hazards regression modelling for repeated events using a counting process approach to examine the associations in the baseline and follow-up cohorts. We adjusted HRs for all demographic and cycling characteristics at baseline. All baseline data were complete for 2435 participants (94.0%). Assuming that the data was missing at random, we computed missing values using multiple imputation (PROC MI) with 25 complete datasets created by the Markov chain Monte Carlo method<sup>26</sup> incorporating all baseline covariates (ie, all demographic and cycling characteristics presented in tables 1 and 2, respectively) and the number of crash outcomes. We presented crude and adjusted HRs and estimated the magnitude and direction of bias by ratios of HRs (RHRs) which were calculated as  $HR_{\text{follow-up}}/HR_{\text{baseline}}$ .

**Table 1** Demographic characteristics of the baseline and follow-up cohorts

Participant characteristics	Baseline cohort		Follow-up cohort		RRF (95% CI)*
	n	Per cent	n	Per cent	
Total	2587		1526		
Age (yrs)					
16–35	579	22.4	268	17.6	0.78 (0.73 to 0.84)
36–50	1351	52.2	815	53.4	1.02 (0.99 to 1.05)
51+	657	25.4	443	29.0	1.14 (1.09 to 1.20)
Gender					
Male	1871	72.3	1109	72.7	1.00 (0.98 to 1.02)
Female	716	27.7	417	27.3	0.99 (0.94 to 1.04)
Ethnicity					
Māori	104	4.0	47	3.1	0.77 (0.61 to 0.92)
Non-Māori	2483	96.0	1479	96.9	1.01 (1.00 to 1.02)
Education					
High school (secondary) or less	534	20.6	278	18.2	0.88 (0.82 to 0.95)
Polytechnic	654	25.3	368	24.1	0.95 (0.90 to 1.01)
University	1393	53.8	876	57.4	1.07 (1.04 to 1.10)
Missing	6	0.2	4	0.3	
Body Mass Index					
<25	1359	52.5	824	54.0	1.03 (1.00 to 1.06)
25–30	999	38.6	582	38.1	0.99 (0.95 to 1.03)
30+	214	8.3	111	7.3	0.88 (0.77 to 0.99)
Missing	15	0.6	9	0.6	
Urbanicity of residence					
Main urban area	2010	77.7	1218	79.8	1.03 (1.01 to 1.04)
Others	541	20.9	293	19.2	0.92 (0.86 to 0.98)
Missing	36	1.4	15	1.0	
NZDep2006 score†					
1–3	1290	49.9	786	51.5	1.03 (1.00 to 1.07)
4–7	918	35.5	542	35.5	1.00 (0.96 to 1.04)
8–10	343	13.3	183	12.0	0.90 (0.82 to 0.99)
Missing	36	1.4	15	1.0	
Region of residence					
Auckland	918	35.5	557	36.5	1.03 (0.98 to 1.07)
Wellington	533	20.6	326	21.4	1.04 (0.97 to 1.10)
Others	1100	42.5	628	41.2	0.97 (0.93 to 1.01)
Missing	36	1.4	15	1.0	

\*95% bootstrap CI.

†2006 New Zealand Deprivation Index with decile 10 the most deprived neighbourhood and decile 1 the least.

RRF, ratios of relative frequencies.

**Table 2** Cycling characteristics of the baseline and follow-up cohorts

Participant characteristics	Baseline cohort		Follow-up cohort		RRF (95% CI)*
	n	Per cent	n	Per cent	
Years of cycling					
<1 year	204	7.9	102	6.7	0.85 (0.74 to 0.96)
1–4 years	1307	50.5	715	46.9	0.93 (0.90 to 0.96)
5+ years	1065	41.2	702	46.0	1.12 (1.08 to 1.16)
Missing	11	0.4	7	0.5	
Time spent cycling per week					
0–3	708	27.4	410	26.9	0.98 (0.93 to 1.03)
4–6	1048	40.5	626	41.0	1.01 (0.97 to 1.05)
7+	822	31.8	485	31.8	1.00 (0.95 to 1.05)
Missing	9	0.3	5	0.3	
Ever ride off-road					
Yes	990	38.3	602	39.4	1.03 (0.99 to 1.07)
No	1583	61.2	915	60.0	0.98 (0.95 to 1.01)
Missing	14	0.5	9	0.6	
Ever ride in the dark					
Yes	1730	66.9	1042	68.3	1.02 (1.00 to 1.04)
No	851	32.9	482	31.6	0.96 (0.91 to 1.01)
Missing	6	0.2	2	0.1	
Ever ride in a bunch					
Yes	1836	71.0	1113	72.9	1.03 (1.01 to 1.05)
No	731	28.3	400	26.2	0.93 (0.88 to 0.98)
Missing	20	0.8	13	0.9	
Cycle to work at least once a week					
Yes	770	29.8	461	30.2	1.01 (0.97 to 1.06)
No	1766	68.3	1028	67.4	0.99 (0.97 to 1.01)
Missing	51	2.0	37	2.4	
Type of bike most commonly used					
Road	2242	86.7	1323	86.7	1.00 (0.99 to 1.01)
Mountain	195	7.5	107	7.0	0.93 (0.82 to 1.04)
Others	138	5.3	90	5.9	1.11 (0.97 to 1.24)
Missing	12	0.5	6	0.4	
Crash in the past 12 months					
Yes	801	31.0	482	31.6	1.02 (0.97 to 1.07)
No	1781	68.8	1042	68.3	0.99 (0.97 to 1.01)
Missing	5	0.2	2	0.1	
Always wear helmet					
Yes	2552	98.6	1506	98.7	1.00 (1.00 to 1.00)
No	25	1.0	16	1.0	1.08 (0.76 to 1.41)
Missing	10	0.4	4	0.3	
Wear fluorescent colours					
Always	758	29.3	433	28.4	0.97 (0.92 to 1.02)
Sometimes	1309	50.6	808	52.9	1.05 (1.01 to 1.08)
Never	499	19.3	275	18.0	0.93 (0.87 to 1.00)
Missing	21	0.8	10	0.7	
Always use lights in the dark					
Yes	1431	55.3	877	57.5	1.04 (1.01 to 1.07)
No	298	11.5	164	10.7	0.93 (0.85 to 1.02)
Never ride in the dark	851	32.9	482	31.6	0.96 (0.91 to 1.01)
Missing	7	0.3	3	0.2	
Use reflective materials in the dark					
Always	849	32.8	530	34.7	1.06 (1.01 to 1.10)
Sometimes	489	18.9	304	19.9	1.05 (0.99 to 1.12)
Never	386	14.9	203	13.3	0.89 (0.82 to 0.97)
Never ride in the dark	851	32.9	482	31.6	0.96 (0.91 to 1.01)
Missing	12	0.5	7	0.5	
Ever listen to music while riding					

Continued

Table 2 Continued

Participant characteristics	Baseline cohort		Follow-up cohort		RRF (95% CI)*
	n	Per cent	n	Per cent	
Yes	422	16.3	218	14.3	0.88 (0.80 to 0.95)
No	2151	83.1	1301	85.3	1.03 (1.01 to 1.04)
Missing	14	0.5	7	0.5	

\*95% bootstrap CI.

RRF, ratios of relative frequencies.

Accounting for the interdependency between the two cohorts, we computed CIs using a non-parametric bootstrapping method with 2000 resamplings (with replacement) of the baseline cohort. We calculated the  $\ln(\text{RHR})$  in each replicate as  $\beta_{\text{follow-up}} - \beta_{\text{baseline}}$  and calculated bias-corrected  $\ln(\text{RHR})$  estimates as  $2 \times \ln(\text{RHR})_{\text{observed}} - \text{mean}(\ln(\text{RHR})_{\text{replicates}})$  as described previously.<sup>17</sup> Around each bias-corrected estimate, we constructed 95% confidence limits by using the SD of the  $\ln(\text{RHR})_{\text{replicates}}$ .<sup>27</sup> We used SAS V9.2 (SAS Institute, Cary, North Carolina, USA) for all analyses.

## RESULTS

The follow-up cohort (n=1526) constituted 60.0% of the baseline cohort (n=2587).

### Baseline characteristics

Groups that were over-represented in the follow-up cohort include cyclists aged over 50 years, non-Māori and university graduates (table 1). The responders were more likely to have a normal Body Mass Index, and more likely to reside in urban areas and in least deprived neighbourhoods.

Additionally, there were differences in some cycling characteristics (table 2). The responders were more often experienced cyclists and bunch riders. Protective behaviours, such as use of conspicuity aids were more prevalent in the follow-up cohort, whereas, distracting behaviours, such as listening to music, were less prevalent. Although not significant, the responders were more likely to ride off-road and in the dark.

### Bicycle crash outcomes

The follow-up cohort consistently experienced more bicycle crashes throughout the follow-up period (figure 2 and table 3).

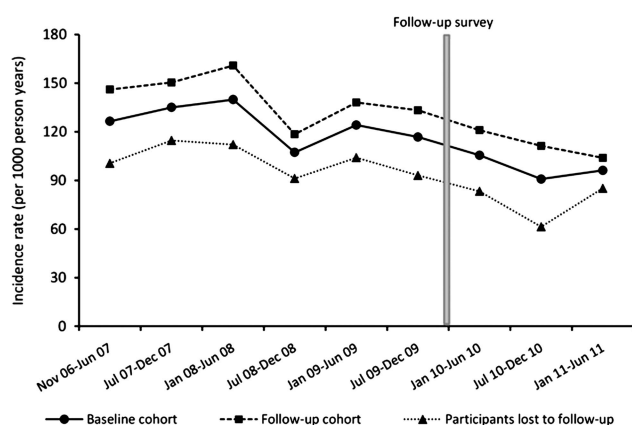


Figure 2 Incidence of bicycle crashes experienced during follow-up.

### Exposure-outcome associations

Individual cell follow-up response rates show how selection bias could occur in selected associations (table 4). Response was not differential in general, but the less balanced distribution of exposures in the non-crash group resulted in slightly biased estimates. This was evident when individually comparing bunch riding, listening to music and using lights with the risk of all crashes. The cross-products, that is, crude RORs, were similar to crude RHRs (where repeated crash events were taken into account) presented in table 5.

Crude RHR estimates ranged from 0.88 (95% CI 0.78 to 0.98) in the bunch riding-all crashes association to 1.14 (95% CI 1.00 to 1.29) in the listening to music-all crashes association (table 5). When we adjusted for all baseline covariates, the HRs changed markedly but in a similar direction in the two cohorts, resulting in modest changes in their relative sizes. Adjusted RHRs ranged between 0.90 and 1.09. We observed the largest positive bias in the listening to music-all crashes and using fluorescent colours-on-road crashes associations, and it was away from the null. We found the largest negative bias in the using lights-all crashes, and using reflective materials-on-road crashes associations, and it was away from the null. Similar results were observed in complete case analysis (ie, restricted to 2435 participants with complete data).

## DISCUSSION

### Main findings

In this prospective cohort study involving 2587 cyclists, 60% responded to a questionnaire administered 3 years after establishment of the study. Failure to respond was associated with

Table 3 Crash outcomes of the baseline and follow-up cohorts

Crash outcome	Baseline cohort		Follow-up cohort		RRF (95% CI)*
	n	Per cent	n	Per cent	
Overall crashes					
One or more	855	33.0	553	36.2	1.10 (1.05 to 1.14)
None	1732	67.0	973	63.8	0.95 (0.93 to 0.97)
On-road crashes					
One or more	553	21.4	360	23.6	1.10 (1.04 to 1.16)
None	2034	78.6	1166	76.4	0.97 (0.96 to 0.99)
Other crashes					
One or more	427	16.5	276	18.1	1.10 (1.02 to 1.17)
None	2160	83.5	1250	81.9	0.98 (0.97 to 1.00)
Collisions with a motor vehicle					
One or more	104	4.0	66	4.3	1.08 (0.92 to 1.23)
None	2483	96.0	1460	95.7	1.00 (0.99 to 1.00)

\*95% bootstrap CI.

RRF, ratios of relative frequencies.



**Table 4** Individual cell follow-up response rates for selected exposures and outcomes

	One or more crashes	None	Cross-product
<i>All crashes</i>			
Crash history at baseline			
Yes	66.7	55.8	1.06
No	63.6	56.3	
Ever ride in a bunch			
Yes	64.8	58.2	0.90
No	64.1	52.0	
Ever listen to music while riding			
Yes	61.0	47.0	1.15
No	65.5	58.0	
Always use lights in the dark			
Yes	66.1	58.6	0.87
No	64.2	49.7	
<i>On-road crashes</i>			
Cycle to work at least once a week			
Yes	65.9	58.2	0.99
No	64.4	56.5	
Wear fluorescent colours			
Always	63.6	55.6	1.01
Sometimes	67.5	60.0	1.00
Never	60.6	53.7	
Use reflective materials in the dark			
Always	66.5	61.3	0.91
Sometimes	70.4	59.6	0.99
Never	59.8	50.2	

demographic, cycling and behavioural risk characteristics as well as crash outcomes. However, the selection bias relating to the seven associations of interest appeared to be small with the adjusted RHRs ranging between 0.90 and 1.09.

### Strengths and limitations

This is one of the very few prospective cohort studies involving cyclists. Baseline data were collected in advance of the crash outcomes and were complete for almost all participants, as mandatory fields and validation checks were incorporated in the web questionnaire. Data on crash outcomes were collected from four administrative databases (including the government-funded universal no-fault injury compensation claims database) and, therefore, were available for all participants in the baseline cohort and covered injuries across the spectrum of severity. This provided us a unique opportunity to estimate bias from (simulated) loss to follow-up in the injury field.

This analysis, however, excludes very minor crashes not requiring either medical or police attention, which represents approximately 70% of self-reported crashes in this study.<sup>28</sup> Ascertainment of crash outcome data may also be affected by personal, social and health service factors<sup>29</sup> as well as the quality of individual data sources and record linkage.<sup>28</sup> Self-reported exposure data may not be accurate and may change over time. Nevertheless, potential misclassifications of crash outcomes and exposures are likely to be non-differential<sup>2</sup> and resulted in underestimation of association estimates in our previous analysis.<sup>30</sup> Our participants are not representative of all New Zealand cyclists; however, this may have minimal impact on the association estimates.<sup>31</sup> Moreover, our

**Table 5** Crude and adjusted HRs in the baseline and follow-up cohorts

	Crude HR (95% CI)			Adjusted HR (95% CI)		
	Baseline cohort	Follow-up cohort	Crude RHR (95% CI)*	Baseline cohort	Follow-up cohort	Adjusted RHR (95% CI)
All crashes						
Crash history at baseline						
Yes	1.43 (1.28 to 1.60)	1.46 (1.27 to 1.67)	1.02 (0.93 to 1.12)	1.27 (1.13 to 1.42)	1.31 (1.14 to 1.51)	1.05 (0.95 to 1.16)
No	1.00	1.00		1.00	1.00	
Ever ride in a bunch						
Yes	1.89 (1.64 to 2.18)	1.68 (1.41 to 2.00)	0.88 (0.78 to 0.98)	1.48 (1.27 to 1.72)	1.41 (1.17 to 1.70)	0.95 (0.84 to 1.08)
No	1.00	1.00		1.00	1.00	
Ever listen to music while riding						
Yes	1.11 (1.00 to 1.28)	1.28 (1.07 to 1.52)	1.14 (1.00 to 1.29)	1.13 (0.98 to 1.31)	1.25 (1.04 to 1.50)	1.09 (0.96 to 1.25)
No	1.00	1.00		1.00	1.00	
Always use lights in the dark						
Yes	0.91 (0.77 to 1.06)	0.82 (0.67 to 1.00)	0.90 (0.77 to 1.05)	0.81 (0.69 to 0.96)	0.74 (0.61 to 0.92)	0.91 (0.78 to 1.07)
No	1.00	1.00		1.00	1.00	
On-road crashes						
Cycle to work at least once a week						
Yes	1.19 (1.02 to 1.38)	1.24 (1.03 to 1.49)	1.04 (0.94 to 1.16)	1.05 (0.89 to 1.24)	1.13 (0.92 to 1.38)	1.07 (0.95 to 1.21)
No	1.00	1.00		1.00	1.00	
Wear fluorescent colours						
Always	0.89 (0.72 to 1.10)	0.87 (0.67 to 1.14)	0.98 (0.83 to 1.16)	0.93 (0.74 to 1.16)	0.96 (0.72 to 1.28)	1.05 (0.87 to 1.26)
Sometimes	1.04 (0.86 to 1.25)	1.05 (0.83 to 1.33)	1.02 (0.89 to 1.18)	1.00 (0.82 to 1.22)	1.06 (0.83 to 1.36)	1.09 (0.94 to 1.27)
Never	1.00	1.00		1.00	1.00	
Use reflective materials in the dark						
Always	0.90 (0.73 to 1.11)	0.84 (0.65 to 1.08)	0.88 (0.74 to 1.05)	0.93 (0.74 to 1.18)	0.89 (0.66 to 1.18)	0.90 (0.75 to 1.09)
Sometimes	0.92 (0.73 to 1.15)	0.87 (0.65 to 1.15)	0.99 (0.81 to 1.20)	0.85 (0.66 to 1.08)	0.84 (0.62 to 1.13)	1.03 (0.84 to 1.26)
Never	1.00	1.00		1.00	1.00	

\*95% bootstrap CI.  
RHR, ratios of HRs.

participants represented a wide variation with regard to demographics, cycling exposure and experience. Finally, our analysis was limited to seven exposure-outcome associations of primary interest to this study.

### Interpretation

In this study, the response rate to a follow-up survey was 60%. If attrition does not depend on exposures, confounders and outcomes (ie, missing completely at random) or depends on exposures and confounders but not on outcomes (ie, missing at random), there is no evidence of serious bias with up to 60% of attrition, according to one study.<sup>32</sup> But if otherwise, even low levels of attrition (20% or less) can bias association estimates.

As with many other studies,<sup>3 5</sup> loss to follow-up did not occur at random in this cohort. The follow-up responders were older, more likely to be university educated, and more likely to reside in urban areas and in least deprived neighbourhoods. This socioeconomic disparity may be attributed to the web-based data collection used in the study; however, similar findings have been reported from other cohorts, regardless of the method used to collect data.<sup>12–14 17–19 33–35</sup> Other factors, such as participants' IQ scores, cognitive functioning and personality characteristics, may also influence follow-up response as reported previously,<sup>35–37</sup> but we were not able to assess these variables in our cohort.

Cycling characteristics also predicted follow-up response. The responders were more experienced cyclists and often rode in a bunch, off-road or in the dark. It is plausible that cycling enthusiasts are more willing to continue participation in research on cycling and its safety. Our findings showed that the responders were less likely to engage in risk behaviours such as not using conspicuity aids and listening to music while riding. This is consistent with previous research, although unrelated to cycling, showing that lifestyle risk factors, such as smoking,<sup>9 16 17 38 39</sup> alcohol abuse,<sup>12 40</sup> physical inactivity<sup>16 40 41</sup> and poor diet,<sup>40</sup> are more common in non-responders or late responders.

Additionally, bicycle crash outcomes differed by response status. The responders were more likely to experience crashes during follow-up. By contrast, other studies not concerned with injury have reported poorer health outcomes and higher mortality among non-responders<sup>11–15 17 33 42</sup> although this was not always the case.<sup>39</sup>

Loss to follow-up, however, caused only modest bias in the selected associations. The magnitude of the bias may depend on the strength of associations of exposures and outcomes to attrition, and also on whether such associations are direct or via other common causes.<sup>43</sup> In this study, the largest bias in crude estimates appeared to be due to unbalanced distribution of exposures among the follow-up participants who did not experience a crash, but the bias was attenuated after adjustment for baseline covariates. This may be because most of the covariates predicted follow-up response and were also related to exposure and/or outcome, and adjustment for common causes (or their proxies) will attenuate the bias.<sup>2 44 45</sup> For example, in the case of M-bias where follow-up response is influenced by factors that also determine exposure and/or outcome, adjustment for those factors will block the back-door pathway opened up by conditioning on follow-up response. In the adjusted estimates, the bias was modest and ranged between –10% and +9%, suggesting that effect measures to be estimated in our future analyses based on exposures measured in the resurvey may not be substantially biased. This is also in accordance with the findings from various cohorts in fields other than injury.<sup>11–19</sup>

### CONCLUSIONS

Loss to follow-up was systematic and differential by demographic, cycling and behavioural risk characteristics as well as crash outcomes. This overestimated the incidence of bicycle crashes but did not substantially bias association estimates. The findings are reassuring, but strictly speaking, apply only to selected associations in the Taupo Bicycle Study. Attempts should be made to minimise attrition and to estimate associated biases in any prospective study.

#### What is already known on the subject

- ▶ Loss to follow-up, if related to exposures, confounders and outcomes of interest, may bias association estimates.
- ▶ Several studies have attempted to estimate the magnitude and direction of such bias but rarely involved cohorts followed-up for injury outcomes.

#### What this study adds

- ▶ We investigated bias from loss to follow-up in a prospective cohort study involving 2590 New Zealand cyclists.
- ▶ Loss to follow-up was differential by demographic, cycling and behavioural risk characteristics as well as crash outcomes.
- ▶ Yet, loss to follow-up did not substantially bias seven association estimates of primary research interest.

**Acknowledgements** We thank the participating cyclists and organisers of the Lake Taupo Cycle Challenge for their support, and Professor John Langley, Professor Anthony Rodgers and Dr Simon Thornley for their initial contribution to the study. Our thanks also go to the Accident Compensation Corporation, Ministry of Health and New Zealand Transport Agency for the provision of bicycle crash data, and to Mr Roy Miodini Nilsen from the University of Bergen, Norway and Ms Silvia Stringhini from the Institute of Social and Preventive Medicine, Lausanne University Hospital, Switzerland, for their kind assistance with bootstrap analyses.

**Contributors** STT contributed to the conception and design of the study, acquisition, analysis and interpretation of data and drafting of the manuscript. AW and SA contributed to the conception and design of the study, interpretation of data and revision of the manuscript. All authors read and approved the final manuscript.

**Funding** This work was supported by the Health Research Council of New Zealand (grant number 09/142).

**Competing interests** None.

**Patient consent** Obtained.

**Ethics approval** University of Auckland Human Participants' Ethics Committee.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Open Access** This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/3.0/>

### REFERENCES

- 1 Samet JM, Muñoz A. Evolution of the cohort study. *Epidemiol Rev* 1998;20:1–14.
- 2 Rothman K, Greenland S, Lash TL. *Modern epidemiology*. 3rd edn. Philadelphia: Lippincott Williams & Wilkins, 2008.
- 3 Chatfield MD, Brayne CE, Matthews FE. A systematic literature review of attrition between waves in longitudinal studies in the elderly shows a consistent pattern of dropout between differing studies. *J Clin Epidemiol* 2005;58:13–19.

- 4 Fan W, Yan Z. Factors affecting response rates of the web survey: a systematic review. *Comput Human Behav* 2010;26:132–9.
- 5 Brinkhof MWG, Pujades-Rodriguez M, Egger M. Mortality of patients lost to follow-up in antiretroviral treatment programmes in resource-limited settings: systematic review and meta-analysis. *PLoS ONE* 2009;4:5790.
- 6 de Winter A, Oldehinkel A, Veenstra R, et al. Evaluation of non-response bias in mental health determinants and outcomes in a large sample of pre-adolescents. *Eur J Epidemiol* 2005;20:173–81.
- 7 Nohr EA, Frydenberg M, Henriksen TB, et al. Does low participation in cohort studies induce bias? *Epidemiology* 2006;17:413–18.
- 8 Harald K, Salomaa V, Jousilahti P, et al. Non-participation and mortality in different socioeconomic groups: the FINRISK population surveys in 1972–92. *J Epidemiol Community Health* 2007;61:449–54.
- 9 Nilsen RM, Vollset SE, Gjessing HK, et al. Self-selection and bias in a large prospective pregnancy cohort in Norway. *Paediatr Perinat Epidemiol* 2009;23:597–608.
- 10 Pizzi C, De Stavola B, Merletti F, et al. Sample selection and validity of exposure–disease association estimates in cohort studies. *J Epidemiol Community Health* 2011;65:407–11.
- 11 Kempen GJM, van Sonderen E. Psychological attributes and changes in disability among low-functioning older persons: Does attrition affect the outcomes? *J Clin Epidemiol* 2002;55:224–9.
- 12 Osler M, Kriebbaum M, Christensen U, et al. Does loss to follow-up in a cohort study bias associations between early life factors and lifestyle-related health outcomes? *Ann Epidemiol* 2008;18:422–4.
- 13 Osler M, Kriebbaum M, Christensen U, et al. Loss to follow up did not bias associations between early life factors and adult depression. *J Clin Epidemiol* 2008; 61:958–63.
- 14 Wolke D, Waylen A, Samara M, et al. Selective drop-out in longitudinal studies and non-biased prediction of behaviour disorders. *Br J Psychiatry* 2009;195:249–56.
- 15 Batty GD, Gale CR. Impact of resurvey non-response on the associations between baseline risk factors and cardiovascular disease mortality: prospective cohort study. *J Epidemiol Community Health* 2009;63:952–5.
- 16 Bjertness E, Sagatun A, Green K, et al. Response rates and selection problems, with emphasis on mental health variables and DNA sampling, in large population-based, cross-sectional and longitudinal studies of adolescents in Norway. *BMC Public Health* 2010;10:602.
- 17 Greene N, Greenland S, Olsen J, et al. Estimating bias from loss to follow-up in the Danish national birth cohort. *Epidemiology* 2011;22:815–22.
- 18 Carter KN, Imlach-Gunasekara F, McKenzie SK, et al. Differential loss of participants does not necessarily cause selection bias. *Aust N Z J Public Health* 2012;36:218–22.
- 19 Howe LD, Tilling K, Galobardes B, et al. Loss to follow-up in cohort studies: bias in estimates of socioeconomic inequalities. *Epidemiology* 2013;24:1–9.
- 20 Rivara FP, Cummings P, Koepsell TD, et al. *Injury control: a guide to research and program evaluation*. Cambridge, New York: Cambridge University Press, 2001, 304p.
- 21 Thornley SJ, Woodward A, Langley JD, et al. Conspicuity and bicycle crashes: preliminary findings of the Taupo Bicycle Study. *Inj Prev* 2008;14:11–18.
- 22 Accident Compensation Corporation (ACC). *Annual report 2012*. Wellington: ACC, 2012.
- 23 Ministry of Health. *National minimum dataset (Hospital inpatient events): data mart —data dictionary V7.5*. Wellington: Ministry of Health, 2012.
- 24 Ministry of Health. *Mortality collection data dictionary version 1.3*. Wellington: Ministry of Health, 2009.
- 25 Davie G, Samaranyaka A, Langley JD, et al. Estimating person-based injury incidence: accuracy of an algorithm to identify readmissions from hospital discharge data. *Inj Prev* 2011;17:338–42.
- 26 Schafer J. *Analysis of incomplete multivariate data*. London: Chapman & Hall, 1997.
- 27 Greenland S. Interval estimation by simulation as an alternative to and extension of confidence intervals. *Int J Epidemiol* 2004;33:1389–97.
- 28 Tin Tin S, Woodward A, Ameratunga S. Completeness and accuracy of crash outcome data in a cohort of cyclists: a validation study. *BMC Public Health* 2013;13:420.
- 29 Cryer C, Langley J. *Developing indicators of injury incidence that can be used to monitor global, regional and local trends*. Dunedin: Injury Prevention Research Unit, University of Otago, 2008.
- 30 Tin Tin S, Woodward A, Ameratunga S. Incidence, risk, and protective factors of bicycle crashes: findings from a prospective cohort study in New Zealand. *Prev Med* 2013;57:152–61.
- 31 Lash T, Fox M, Fink A. *Applying quantitative bias analysis to epidemiological data*. New York: Springer, 2009.
- 32 Kristman V, Manno M, Cote P. Loss to follow-up in cohort studies: how much is too much? *Eur J Epidemiol* 2004;19:751–60.
- 33 Goldberg M, Chastang JF, Leclerc A, et al. Socioeconomic, demographic, occupational, and health factors associated with participation in a long-term epidemiologic survey: a prospective study of the French GAZEL cohort and its target population. *Am J Epidemiol* 2001;154:373–84.
- 34 Strandhagen E, Berg C, Lissner L, et al. Selection bias in a population survey with registry linkage: potential effect on socioeconomic gradient in cardiovascular risk. *Eur J Epidemiol* 2010;25:163–72.
- 35 Nishiwaki Y, Clark H, Morton S, et al. Early life factors, childhood cognition and postal questionnaire response rate in middle age: the Aberdeen Children of the 1950s study. *BMC Med Res Methodol* 2005;5:16.
- 36 Salthouse TA. Selectivity of attrition in longitudinal studies of cognitive functioning. *J Gerontol B Psychol Sci Soc Sci*. Published Online First: 2 June 2013. doi: 10.1093/geronb/gbt046
- 37 Beaver KM. Intelligence and selective attrition in a nationally representative and longitudinal sample of Americans. *Pers Individ Dif* 2013;55:157–61.
- 38 Barchielli A, Balzi D. Nine-year follow-up of a survey on smoking habits in Florence (Italy): higher mortality among non-responders. *Int J Epidemiol* 2002;31:1038–42.
- 39 Verlato G, Melotti R, Olivieri M, et al. Asthmatics and ex-smokers respond early, heavy smokers respond late to mailed surveys in Italy. *Respir Med* 2010;104:172–9.
- 40 Kypri K, Samaranyaka A, Connor J, et al. Non-response bias in a web-based health behaviour survey of New Zealand tertiary students. *Prev Med* 2011;53:274–7.
- 41 Van Loon AJM, Tijhuis M, Picavet HJ, et al. Survey non-response in the Netherlands: effects on prevalence estimates and associations. *Ann Epidemiol* 2003;13:105–10.
- 42 Ferrie JE, Kivimäki M, Singh-Manoux A, et al. Non-response to baseline, non-response to follow-up and mortality in the Whitehall II cohort. *Int J Epidemiol* 2009;38:831–7.
- 43 Greenland S. Quantifying biases in causal models: classical confounding vs collider-stratification Bias. *Epidemiology* 2003;14:300–6.
- 44 Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. *Epidemiology* 2004;15:615–25.
- 45 Daniel RM, Kenward MG, Cousens SN, et al. Using causal diagrams to guide analysis in missing data problems. *Stat Methods Med Res* 2012;21:243–56.