# The Use of Propensity Score Stratification and Synthetic Data to Address Allocation Bias when Assessing Bicycle Helmet Effectiveness.

Jake Olivier, Frances Terlich

***Abstract*** Case-control studies have found bicycle helmet use significantly mitigates the risk and severity of head injury in a motor vehicle collision. However, critics argue the decision to wear a helmet is confounded with other factors related to cycling safety such as cycling speed. If such an allocation bias exists, results from case-control studies may be invalid if confounding factors are ignored. Although allocation bias and bicycle helmet effectiveness is frequently mentioned in the literature, there is a paucity of research that has explored this relationship. This study aims to examine bicycle helmet effectiveness in a motor vehicle collision using the propensity score stratification method, which removes allocation bias from case (head injury) and control (no head injury) groups to allow for direct comparison of helmet effectiveness in reducing head injury. Due to privacy and data accessibility issues, synthetic data was created from a recently published Australian study of linked hospital and police data over a nine-year period. In a motor vehicle collision, helmet use was associated with factors that have been argued to influence estimates of helmet effectiveness; however, using propensity score stratification, there is no evidence these confounding factors influence estimates of helmet effectiveness.

***Keywords*** allocation bias, bicycle helmets, case-control study, propensity score, synthetic data.

## I. INTRODUCTION

Cycling is considered beneficial to one's health and the environment, whether pursued for sport, leisure or transportation. Although there are numerous benefits, cycling can lead to bodily harm in a crash or fall. Many jurisdictions around the world have invested in dedicated cycling infrastructure to separate cyclists from motorised traffic as a measure of crash avoidance. Despite these efforts, many cycling injuries are due to collisions with motor vehicles. Head injuries are particularly common among cyclists in a crash or fall.

Bicycle helmets are designed to attenuate the kinetic energy directed to the head in a crash or fall. Three systematic reviews and meta-analyses of case-control studies have found bicycle helmet use in a crash or fall is associated with odds reductions in head, serious head and facial injury [1-3]. The most recent meta-analysis [3] estimates odds reductions of 51% for head injuries, 69% for serious head injuries and 33% for facial injuries associated with helmet use.

Randomised controlled trials are not ethically possible to study bicycle helmet effectiveness and case-control studies are the highest quality of evidence possible for a single study. The lack of randomisation in a case-control study increases the possibility of allocation bias where the probability of selecting cases or controls differ by exposure group. As it relates to bicycle helmets and cyclists in a crash or fall, an allocation bias can occur if there is a systematic difference between cyclists who choose to wear or not wear a helmet [4]. A selection bias may also occur in case-control studies because not all cyclists in a crash or fall are recorded in hospital, emergency department, police or other databases that are available for analysis. It is unknown if this potential selection bias is associated with helmet use.

The risk compensation hypothesis has been proposed by some authors who posit helmeted cyclists may exhibit riskier riding behaviour thereby offsetting any protection offered by the helmet [5]. There is no clear evidence supportive of risk compensation and bicycle helmets and much of the current literature has been limited to commentaries from authors who are either supportive or opposed to the hypothesis [5-6]. The very few real-world studies on risk compensation and bicycle helmets have used speed as a proxy for risky behaviour

J. Olivier is Associate Professor (+61293856656, j.olivier@unsw.edu.au) and F. Terlich is a MSc graduate in Biostatistics in the School of Mathematics and Statistics at the University of New South Wales in Sydney, Australia.

[7-8]. Neither study found clear evidence supportive of risk compensation [9]. Another study found motor vehicles overtook at closer distances on average when the researcher wore a helmet [10]; however, a re-analysis of this data found that helmet wearing was a trivial effect size, the statistical significance from the original study was likely the result of an overly large sample size and the association disappeared for distances less than one metre when adjusting for other factors [11-12]. Additionally, a follow-up study by the original author could not reproduce the helmet effect on overtaking distance [13]. Although there is no strong evidence supportive of risk compensation and bicycle helmets, behavioural differences may exist between helmeted and unhelmeted cyclists which could then influence estimates of helmet effectiveness.

Several studies of bicycle helmet effectiveness have addressed potential confounders through model stratification. Using data from a French road trauma registry, an analysis adjusted for age, gender, crash opponent, injury severity score (ISS) for those injured below the neck, road type and crash setting [14]. An analysis of linked New South Wales (NSW) police and hospital data adjusted for posted speed limit, type of vehicle in collision, age of cyclist, serious injuries other than the head, whether the cyclists disobeyed a traffic control, whether the cyclist's blood alcohol content (BAC) was greater than 0.05 and whether the cyclist was riding on the footpath [15]. Model stratification, however, may not adequately remove confounding if predictors are associated with the treatment allocation mechanism, i.e., the decision to wear a helmet. This was true in the NSW study where helmet use was positively associated with higher posted speed limits and increasing age. Helmet use was negatively associated with disobeying a traffic control, having blood alcohol content (BAC) greater than 0.05 and riding on a footpath.

Propensity score stratification is a statistical method often used in observational studies to remove or lessen the effect of treatment allocation bias to strengthen causal inference [16-17]. In this setting, the propensity score is the probability of wearing a helmet given a vector of observed covariates. Propensity scores are usually estimated by logistic regression with treatment allocation (i.e., helmet use) as the outcome and the predictors taken from the remaining variables except the primary outcome (i.e., head injury severity). The goal of propensity score estimation is to accurately estimate the probability of treatment allocation and not to determine associated factors; therefore, it is common to choose which predictors to include in the model through stepwise regression or similar model selection methods and to include second order interaction terms or higher. Once propensity scores are estimated, the association between treatment allocation and the primary outcome is assessed while adjusting for propensity score quintiles. Previous research has demonstrated five subclasses of propensity scores can remove over 90% of allocation bias [18].

Most case-control studies do not collect information regarding the crash or fall, which limits the applicability of propensity score stratification or other methods of accounting for allocation bias. The NSW data is unique in this area of research since it includes details regarding the crash (police reports), medically diagnosed injuries (hospital records) and is one of the largest case-control studies ever published [3]. Due to data accessibility issues, the linked NSW data is not available for analysis; however, synthetic data can be created from published summary data. Therefore, the aims of this study are to (1) assess whether helmeted and unhelmeted cyclists differ on factors related to safety using published NSW data and (2) to assess the effect of bicycle helmets to mitigate head injury after removing or lessening potential allocation bias through propensity score stratification using synthetic data.

## II. METHODS

The reported NSW data includes information on NSW cyclists involved in a motor vehicle collision from 2001-2009 from linked police records and hospitalisation data. In total, there were n=6745 cases with 75.4% (n=5087) wearing a helmet. The combined data sources contain information on the cyclist (head injury severity, age category: 0-12, 13-19, 20-29, 30-39, 40-49, 50+ years, other serious injuries, whether the cyclist disobeyed a traffic control, BAC>0.05, gender, and helmet use) and information on the collision (posted speed limit, collision vehicle, whether the cyclist was riding on a footpath, and indicators for whether the collision occurred at an intersection, a metropolitan area, a curve, a highway/freeway, a sealed roadway, a dry roadway or in daytime). Injuries were identified using International Classification of Diseases version 10, Australian Modification (ICD-10-AM) and injury severity was determined using survival rate ratios [15].

The NSW study used multinomial logistic regression on four categories of head injury severity as the outcome and the predictors helmet use, posted speed limit, collision vehicle, age group, other serious injuries, whether the cyclist disobeyed a traffic control, BAC>0.05 and riding on a footpath. The crude model with only helmet use has the form

$$\log\left(\frac{\pi_{ij}}{\pi_{i0}}\right) = \beta_{0j} + \beta_{1j} helmet_i \qquad (1)$$

where $j = 0,1,2,3$ represent possible minor, moderate, serious and severe head injury respectively and $helmet_i$ is an indicator for helmet use (1=helmet, 0=no helmet). The multinomial odds ratio $\pi_{ij} / \pi_{i0}$ is a comparison of the probability of moderate, serious or severe head injury to possible minor injury for cyclists with the same covariates. The adjusted model has the same form plus covariates for other included variables.

***Synthetic Data of Linked NSW Police and Hospital Data***

Synthetic data was created using published summary data and model estimates as a guide. At each step, the data was simulated to incorporate known data structures including differences in demographic and crash information for helmeted and unhelmeted cyclists. The steps taken to generate synthetic data are given below.

1. Helmet status (yes or no) was randomly generated for n=6745 cyclists from a Bernoulli random variable with $\hat{p} = 0.754$.
2. Conditional on helmet status, demographic and crash information were randomly generated from Bernoulli (binary variables) or multinomial distributions.
3. Using the model estimates from [15] and the randomly generated covariates from (1) and (2), multinomial odds ratios were computed for moderate, serious and severe head injury, i.e., $\pi_{i1} / \pi_{i0}$, $\pi_{i2} / \pi_{i0}$ and $\pi_{i3} / \pi_{i0}$.
4. The probabilities of possible minor, moderate, serious and severe head injury were computed for each cyclist from the multinomial odds ratios and the constraint $\pi_{i0} + \pi_{i1} + \pi_{i2} + \pi_{i3} = 1$.
5. The probabilities of head injury severity categories were then used to generate a head injury severity from a multinomial distribution for each cyclist.
6. Steps (1)-(5) were repeated 200 times to account for additional variability associated with synthetic data.

At each step in the process, the synthetic data were checked against the published data for significant deviations.

***Propensity Score Stratification***

The propensity score $e(x_i)$ for a cyclist with covariate vector $x_i$ is defined as

$$e(x_i) = P(helmet_i = 1 | x_i). \qquad (2)$$

For each synthetic data set, a multivariable logistic regression model of the form

$$\log\frac{e(x_i)}{1 - e(x_i)} = x_i^T \beta \qquad (3)$$

was fit and propensity score estimates $\hat{e}(x_i)$ were obtained from model estimates $\hat{\beta}$ as

$$\hat{e}(x_i) = \frac{e^{x_i^T \hat{\beta}}}{1 + e^{x_i^T \hat{\beta}}}. \qquad (4)$$

Models were chosen for each synthetic data set by stepwise regression for main effects and two-way interactions from the variables posted speed limit, collision vehicle, cyclist's age, other serious injuries, whether the cyclist disobeyed a traffic control, BAC>0.05, whether the cyclist was riding on a footpath, cyclist's gender, and whether the collision occurred at an intersection, a metropolitan area, a curve, a highway/freeway, a sealed roadway, a dry roadway or in daytime. The estimated propensity scores were then categorised into quintiles. Forest plots of crude and propensity score quintile adjusted odds ratios were constructed to visually compare the associations between helmet use and the other covariates. Multinomial logistic regression was used to assess the association between head injury severity and helmet use while adjusting for propensity score quintile.

The synthetic data were created and analysed using SAS. Forest plots were created in R Studio using the metafor package [19].

## III. RESULTS

Two hundred synthetic data sets were created for cyclists in a motor vehicle collision. To compare with the source data, one synthetic dataset was randomly chosen and the percentages of responses for each variable were similar to the original data (see Table A1). The propensity scores estimated from this data set using stepwise logistic regression included main effects for posted speed limit, collision vehicle, cyclist's age, other serious injuries, whether the cyclist disobeyed a traffic control, BAC>0.05, whether the cyclist was riding on a footpath, cyclist's gender and whether the collision occurred on a highway/freeway or in daytime. Also in this model were two-way interactions terms for cyclist's age/whether the cyclist disobeyed a traffic control, cyclist's age/collision in daytime, and BAC>0.05/collision on highway/freeway. The p-values for all interaction terms and main effects not included in a two-way interaction were less than 5%. The estimated propensity scores were then categorised into quintiles.

For the same randomly chosen data set, the percentage of responses for each variable by propensity score quintile is given in Table A2. In comparison to the other quintiles, cyclists in the first quintile were more likely to not wear a helmet (57.7%), have a known head injury of any severity (15%), travel in 60 km/h areas or slower (94.4%), be less than 30 years of age (95.8%), have other serious injuries (12.5%), disobey a traffic control (14.7%), BAC>0.05 (11.2%), ride on a footpath (47.6%), male gender (93.9%) and not cycle on a highway (4.4%). Cyclists in the last quintile, on the other hand were less likely to not wear a helmet (4.8%), have a known head injury of any severity (6.5%), travel in 60 km/h areas or slower (78.3%), be less than 30 years of age (4.7%), have other serious injuries (3.1%), disobey a traffic control (0.7%), BAC>0.05 (0.6%), ride on a footpath (1.5%), male gender (66.7%) and not cycle on a highway (25.0%).

Forest plots of crude and propensity score adjusted odds ratios for helmet use are given in Figure 1. Helmet use was associated with posted speed limit, cyclist's age, whether the cyclist disobeyed a traffic control, BAC>0.05 and riding on the footpath. These associations were largely removed following propensity score stratification.

The crude, adjusted and propensity score adjusted multinomial odds ratios using the source and synthetic data sets are given in Table 1. There was very little deviation among the multinomial odds ratios irrespective of adjustment method or whether the data were real or synthetic.

TABLE I

CRUDE, ADJUSTED AND PROPENSITY SCORE ADJUSTED MULTINOMIAL ODDS RATIOS USING SOURCE AND SYNTHETIC DATA

| | Source Data[1] | | Synthetic Data[2] | | |
|---|---|---|---|---|---|
| | Crude | Adjusted[3] | Crude | Adjusted[3] | PScore[4] |
| *Moderate* | 0.513 | 0.506 | 0.504 | 0.501 | 0.531 |
| | 0.411-0.640 | 0.388-0.659 | 0.411-0.644 | 0.395-0.670 | 0.424-0.704 |
| *Serious* | 0.330 | 0.378 | 0.321 | 0.378 | 0.416 |
| | 0.248-0.440 | 0.267-0.539 | 0.252-0.427 | 0.294-0.539 | 0.321-0.580 |
| *Severe* | 0.259 | 0.257 | 0.233 | 0.264 | 0.285 |
| | 0.165-0.407 | 0.148-0.448 | 0.143-0.381 | 0.151-0.470 | 0.167-0.509 |

[1]OR and 95% CI, [2]Median OR, 2.5th and 97.5th Quantile OR, [3]Adjusted for speed, collision vehicle, age group, other serious injuries, disobeying a traffic control, BAC>0.05 and riding on a footpath, [4]Adjusted for propensity score quintile

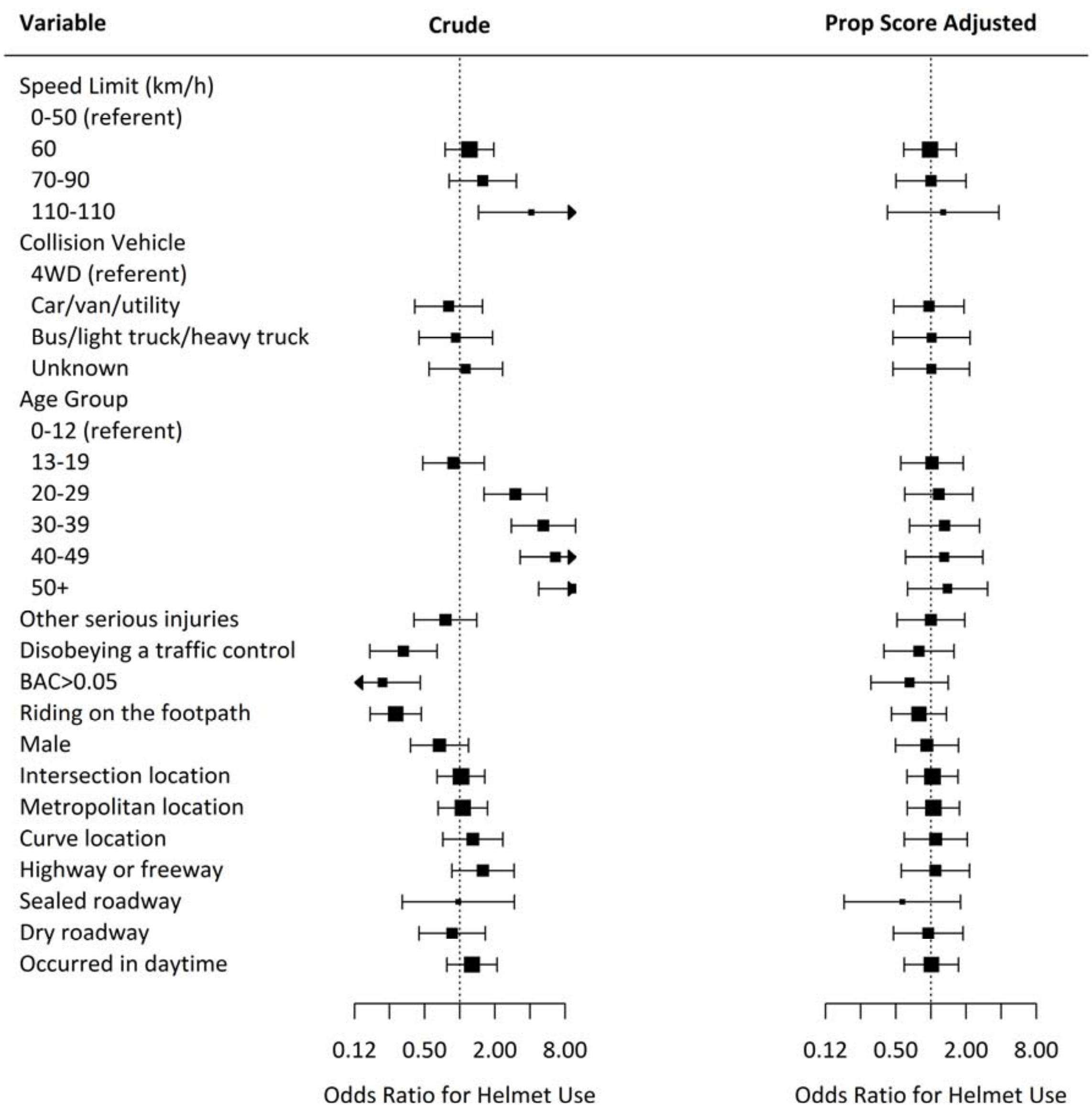


Fig. 1. Forest plot of crude and propensity score adjusted odds ratios for helmet use from multivariable logistic regression analysis using the source (crude) or randomly selected synthetic data set (propensity score adjusted).

## IV. Discussion

In this study, we found evidence of treatment allocation bias for a case-control study of helmet effectiveness and then adjusted for this bias using propensity score stratification. Synthetic data was created in lieu of the actual data due to lack of data access. Helmet use was associated with posted speed limit, age, whether the cyclist disobeyed a traffic control, BAC>0.05, and riding on the footpath. The direction of these estimates indicate helmet users are both risk takers (i.e., more likely to cycle with fast moving motorised traffic and less likely to cycle on footpath) and risk averse (i.e., less likely to disobey a traffic control or drink alcohol and cycle) which provides no evidence supportive of the risk compensation hypothesis. These associations were largely removed after adjusting for propensity score quintiles.

The propensity score quintiles seem to represent different types of cyclists. Cyclists in the first quintile tend to be younger, male, cycle with slower motorised traffic, perform illegal activities at a higher rate and are less likely to wear a helmet. The upper quintile is largely the antithesis of the lower quintile and exhibits many characteristics of middle-aged men in Lycra (MAMIL), i.e., more likely to cycle with fast moving traffic and older age. Importantly, estimates of helmet effectiveness are comparisons within these categories. In other words, the effectiveness of helmets was estimated by comparing similar cyclists or at least cyclists in similar conditions in a motor vehicle collision. Although helmet use was associated with several factors and these associations were largely removed after propensity score stratification, the estimates of helmet effectiveness changed very little. When these results are combined, they indicate (1) cyclists in a motor vehicle collision differ on several factors according to helmet wearing and (2) these differences have minimal influence on estimates of helmet effectiveness.

Our results estimate odds reductions of 47% for moderate, 58% for serious and 71% for severe head injury compared to possible minor head injury after propensity score adjustment. To compare these results with the most recent meta-analysis, the data were re-analysed for the binary variables head (moderate or greater severity) and serious head (serious or greater severity) injury with possible minor head injuries as the controls for both analyses. The results were similar for head injury (OR=0.49 vs OR=0.46) and serious head injury (OR=0.31 vs OR=0.38).

Bicycle helmet legislation is a controversial topic and behavioural factors such as risk compensation are often cited as a reason against enacting such laws [9]. Previous research has assessed the short- and long-term impact of helmet legislation in NSW [20-21] and this study is not a direct or indirect assessment of that legislation. However, our results do suggest helmet use mitigates the risk of a head injury in a crash or fall and estimates of helmet effectiveness are not influenced by behavioural factors. Still, policy makers should be cautious enacting such legislation since helmets are not a panacea and crash avoidance strategies such as segregated cycling infrastructure have been shown to reduce the incidence of all cycling injuries.

This study has several limitations. The primary analysis was performed on randomly generated data. Although known data structures were maintained in the creation of the synthetic data, it is unknown if an analysis of the real data would produce similar results. Verification is not possible due to data accessibility issues. Propensity scores were estimated by fitting models with up to two-way interaction terms, although higher order terms could estimate models with better predictive properties. Models with three- and four-way interactions were also estimated with similar results (results not shown). A model allowing five-way interaction terms was also attempted, but computations did not complete after seven days. The analysis was performed on 200 synthetic data sets and it is possible a greater number of data sets would improve estimates. In a sensitivity analysis, the analysis was repeated on 1000 synthetic data sets and the propensity score adjusted odds ratios did not change appreciably (possible minor head injury vs moderate, OR=0.533; serious, OR=0.414; severe, OR=0.279). The use of propensity scores in case-control studies can lead to artefactual effect modification [22] and it is presently unknown if the problem is exacerbated for synthetic data sets. However, data simulation suggests the magnitude of the effect is at worst modest for large sample sizes (n=3000). The quintiles created from propensity scores group similar cyclists together; however, these categories do not necessarily correspond to any specific type of cyclist. For example, the fifth quintile exhibits many characteristics of MAMILs; however, this quintile had the largest proportion of females. Propensity score methods have been shown to improve causal inference by removing or lessening the effect of allocation bias; however, these methods do not assure the inferences are causal in nature. A selection bias whereby helmet use is associated with being captured in police or medical records is still possible, although this has been somewhat addressed through model adjustment. Hence, our results do not prove helmet use *causes* a decrease in the risk of head injury in a crash or fall, although it does rule out one type of bias that prevents that statement being true. Finally, this is the only study of which we are aware that has accounted for treatment allocation biases in an analysis of bicycle helmet effectiveness, so it is unclear if our results are generalisable to other populations. Still, we have given a detailed analytic framework for other researchers to follow.

## V. CONCLUSIONS

In a motor vehicle collision, helmet use is associated with factors that have been argued to influence estimates of helmet effectiveness. After propensity score stratification, there is no evidence confounding factors influence estimates of helmet effectiveness. However, some caution should be taken in interpreting results from synthetic data that has not been verified on an analysis of real data.

## VI. Acknowledgement

The authors would like to thank Tim Churches for discussions around propensity score stratification and bicycle helmet use.

## VII. References

[1] Thompson DC, Rivara FP, Thompson R. Helmets for preventing head and facial injuries in bicyclists. *Cochrane Review*, 2000, CD001855.

[2] Attewell RG, Glase K, McFadden M. Bicycle helmet efficacy: A meta-analysis. *Accident Analysis and Prevention*, 2001, **33**(3):345-352.

[3] Olivier J, Creighton P. Bicycle injuries and helmet use: A systematic review and meta-analysis, *International Journal of Epidemiology*, in press.

[4] Sedgwick P. Selection bias versus allocation bias. *British Medical Journal*, 2013, **346**:f3345.

[5] Adams J, Hillman M. The risk compensation theory and bicycle helmets. *Injury Prevention*, 2001, **7(2)**:89-91.

[6] Thompson DC, Thompson RS, Rivara FP. Risk compensation theory should be subject to systematic reviews of the scientific evidence. *Injury Prevention*, 2001, **7(2)**:86-8.

[7] Messiah A, Constant A, Contrand B, Felonneau ML, Lagarde E. Risk compensation: a male phenomenon? Results from a controlled intervention trial promoting helmet use among cyclists. *American Journal of Public Health*, 2012, **102 Suppl 2**:S204-6.

[8] Fyhri A, Phillips RO. Emotional reactions to cycle helmet use. *Accident Analysis and Prevention*, 2013, **50**:59-63.

[9] Olivier J, Wang JJ, Walter S, Grzebieta R. Anti-helmet arguments: lies, damned lies and flawed statistics. *Journal of the Australasian College of Road Safety*, 2014, **25(4)**:10-23.

[10] Walker, I. Drivers overtaking bicyclists: Objective data on the effects of riding position, helmet use, vehicle type and apparent gender. *Accident Analysis and Prevention*, 2007, **39**:417-425.

[11] Olivier J, Walter SR. Bicycle helmet wearing is not associated with close motor vehicle passing: A re-analysis of Walker, 2007. *PLOS ONE*, 2013, **8(9)**:e75424.

[12] Olivier J, Walter SR. Too much statistical power can lead to false conclusions. *Injury Prevention*, 2015, **21**:289.

[13] Walker I, Garrard I, Jowitt F. The influence of a bicycle commuter's appearance on drivers' overtaking proximities: an on-road test of bicyclist stereotypes, highvisibility clothing and safety aids in the United Kingdom. *Accident Analysis and Prevention*, 2014, **64**:69-77.

[14] Amoros E, Chiron M, Martin J-L, Thélot B, Laumon B. Bicycle helmet wearing and the risk of head, face, and neck injury: a French case-control study based on a road trauma registry. *Injury Prevention*, 2012, **18**:27-32.

[15] Bambach M, Mitchell R, Grzebieta R, Olivier J. The effectiveness of helmets in bicycle collisions with motor vehicles: A case–control study. *Accident Analysis and Prevention*, 2013, **53**:78-88.

[16] Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 1983, **70(1)**:41-45.

[17] D'Agostino Jr RB. Tutorial in biostatistics: Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Statistics in Medicine*, 1998, **17**:2265-2281.

[18] Rosenbaum PR, Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association*, 1984, **79**:516-524.

[19] Viechtbauer W. Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 2010, **36(3)**:1-48.

[20] Walter SR, Olivier J, Churches T, Grzebeita R. The impact of compulsory cycle helmet legislation on cyclist head injuries in New South Wales, Australia. *Accident Analysis and Prevention*, 2011, **43**:2064–2071.

[21] Olivier J, Walter SR, Grzebieta RH. Long-term bicycle related head injury trends for New South Wales, Australia following mandatory helmet legislation. *Accident Analysis and Prevention*, 2013, **50**:1128–1134.

[22] Månsson R, Joffe MM, Sun W, Hennessy S. On the estimation and use of propensity scores in case-control and case-cohort studies. *American Journal of Epidemiology*, 2007, **166**:332-339

## VIII. APPENDIX

TABLE AI
PERCENTAGE OF RESPONSES BY HELMET USE FOR ORIGINAL AND SYNTHETIC DATA

| | Helmet | | No Helmet | |
|---|---|---|---|---|
| | Original (%) | Synthetic (%) | Original (%) | Synthetic (%) |
| *Head Injury Severity* | | | | |
| *Possibly Minor Injury* | 92.7 | 93.5 | 83.9 | 82.4 |
| *Moderate* | 4.6 | 3.9 | 8.1 | 9.5 |
| *Serious* | 2.0 | 2.0 | 5.6 | 6.3 |
| *Severe* | 0.7 | 0.7 | 2.5 | 1.9 |
| *Speed Limit (km/h)* | | | | |
| *0-50* | 50.0 | 49.3 | 56.9 | 55.8 |
| *60* | 38.2 | 38.7 | 35.9 | 36.2 |
| *70-90* | 8.8 | 9.0 | 6.3 | 7.2 |
| *100-110* | 3.0 | 3.1 | 0.8 | 0.8 |
| *Collision Vehicle* | | | | |
| *4WD* | 7.1 | 7.0 | 6.2 | 7.4 |
| *Car/van/utility* | 64.4 | 64.4 | 69.6 | 68.3 |
| *Bus/light truck/heavy truck* | 13.0 | 13.4 | 12.2 | 11.7 |
| *Unknown* | 15.5 | 15.2 | 12.0 | 12.6 |
| *Age Group* | | | | |
| *0-12* | 7.1 | 7.5 | 19.5 | 19.9 |
| *13-19* | 11.4 | 11.9 | 35.5 | 36.8 |
| *20-29* | 21.4 | 20.3 | 19.5 | 18.4 |
| *30-39* | 26.5 | 26.4 | 13.9 | 13.4 |
| *40-49* | 16.9 | 17.5 | 7.0 | 6.4 |
| *50+* | 16.7 | 16.5 | 4.6 | 5.1 |
| *Other serious injuries* | 7.3 | 6.5 | 9.5 | 9.3 |
| *Disobeying a traffic control* | 3.3 | 3.6 | 9.4 | 10.0 |
| *BAC>0.05* | 1.7 | 1.8 | 7.2 | 7.6 |
| *Riding on the footpath* | 12.9 | 13.0 | 34.4 | 35.4 |
| *Male* | 83.7 | 83.1 | 88.4 | 89.3 |
| *Intersection location* | 60.1 | 61.2 | 60.0 | 60.4 |
| *Metropolitan location* | 71.2 | 70.6 | 69.9 | 68.1 |
| *Curve location* | 12.8 | 12.1 | 10.2 | 11.0 |
| *Highway or freeway* | 12.6 | 12.3 | 8.3 | 7.2 |
| *Sealed roadway* | 99.2 | 99.0 | 99.2 | 99.2 |
| *Dry roadway* | 92.2 | 92.3 | 93.2 | 92.6 |
| *Occurred in daytime* | 76.7 | 76.5 | 72.1 | 72.1 |

TABLE A2
PERCENTAGE OF RESPONSES BY PROPENSITY SCORE QUINTILE OF SYNTHETIC DATA

| | Propensity Score Quintile | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| *Sample Size* | 1303 | 1394 | 1411 | 1287 | 1350 |
| *Helmet* | 42.3 | 65.4 | 85.8 | 90.8 | 95.2 |
| *Head Injury Severity* | | | | | |
| *Possibly Minor Injury* | 85.0 | 88.7 | 93.1 | 93.7 | 93.5 |
| *Moderate* | 7.9 | 6.0 | 4.1 | 3.9 | 4.2 |
| *Serious* | 5.5 | 4.0 | 2.1 | 1.8 | 1.7 |
| *Severe* | 1.6 | 1.3 | 0.7 | 0.6 | 0.7 |
| *Speed Limit (km/h)* | | | | | |
| *0-50* | 63.5 | 46.1 | 63.5 | 47.8 | 33.2 |
| *60* | 30.9 | 44.3 | 27.3 | 43.1 | 45.1 |
| *70-90* | 5.4 | 8.9 | 7.9 | 7.4 | 13.0 |
| *100-110* | 0.3 | 0.7 | 1.3 | 1.8 | 8.7 |
| *Collision Vehicle* | | | | | |
| *4WD* | 8.3 | 6.0 | 8.4 | 7.9 | 5.0 |
| *Car/van/utility* | 71.5 | 62.6 | 70.1 | 66.7 | 55.9 |
| *Bus/light truck/heavy truck* | 10.1 | 14.6 | 9.9 | 14.2 | 16.1 |
| *Unknown* | 10.1 | 16.7 | 11.6 | 11.1 | 23.0 |
| *Age Group* | | | | | |
| *0-12* | 24.3 | 25.5 | 1.5 | 0.7 | 0.2 |
| *13-19* | 56.3 | 32.1 | 0.9 | 0.7 | 0.1 |
| *20-29* | 15.2 | 12.9 | 49.9 | 15.5 | 4.4 |
| *30-39* | 2.8 | 17.3 | 32.3 | 37.4 | 26.2 |
| *40-49* | 1.2 | 6.8 | 9.7 | 20.3 | 36.5 |
| *50+* | 0.2 | 5.5 | 5.7 | 25.5 | 32.7 |
| *Other serious injuries* | 12.5 | 6.4 | 9.5 | 4.2 | 3.1 |
| *Disobeying a traffic control* | 14.7 | 5.4 | 4.1 | 0.8 | 0.7 |
| *BAC>0.05* | 11.2 | 3.8 | 0.5 | 0.2 | 0.6 |
| *Riding on the footpath* | 47.6 | 31.9 | 8.7 | 2.4 | 1.5 |
| *Male* | 93.9 | 82.6 | 93.1 | 86.9 | 66.7 |
| *Intersection location* | 61.1 | 60.5 | 60.2 | 61.4 | 61.9 |
| *Metropolitan location* | 67.4 | 68.8 | 70.9 | 71.2 | 71.9 |
| *Curve location* | 11.1 | 12.1 | 12.3 | 11.1 | 12.7 |
| *Highway or freeway* | 4.4 | 10.6 | 6.5 | 8.8 | 25.0 |
| *Sealed roadway* | 98.5 | 99.1 | 99.3 | 99.1 | 99.0 |
| *Dry roadway* | 91.9 | 93.4 | 91.6 | 92.4 | 92.4 |
| *Occurred in daytime* | 62.5 | 84.5 | 72.4 | 79.6 | 77.8 |