# What influences crash risk and crash prevalence for e-scootering? Insights from a naturalistic riding study

Transportation Research Part F: Psychology and Behaviour 114 (2025) 160–170

Contents lists available at ScienceDirect

Transportation Research Part F: Psychology and Behaviour

journal homepage: www.elsevier.com/locate/trf

# What influences crash risk and crash prevalence for e-scootering? Insights from a naturalistic riding study☆

Rahul Rajendra Pai *, Marco Dozza

*Department of Mechanics and Maritime Sciences, Division of Vehicle Safety, Chalmers University of Technology, Hörsalsvägen 7, 41258 Göteborg, Sweden*

A B S T R A C T

Naturalistic data, i.e. data collected in real traffic by road users attending their daily routines, are the gold standard for crash causation analyses. In fact, these data can show the pre-crash road-user behaviour that is hard to observe from other crash data. Naturalistic data from 6868 trips by 4694 distinct participants, collected over a period of 1.5 years from 17 e-scooters, were used to estimate crash risk by means of odds ratios (OR) and crash prevalence by population attributable risk percentage (PARP). We computed OR and PARP, comparing crashes and near-crashes to baseline events from normal riding. The baselines were selected through both matching and random sampling strategies in order to expand and increase the statistical significance of previous results—while also providing new methodological insights for future research on crash causation. This study also investigated the impact of different baseline–to–safety–critical event ratios for the assessment of crash risk.

From a safety perspective, our findings suggest that safety interventions that reduce leisure trips, exposure to intersections, trips on Fridays and Saturdays, pack riding, and inexperienced riding should be prioritised. From a methodological perspective, we showed how combining random and matched baselines can help quantify the crash risk and crash prevalence for micromobility vehicles. The results from this study may encourage policymakers to make data-driven decisions regarding e-scooter regulations. Future research should combine data from naturalistic studies and crash databases with data from the perspective of other road users to provide a more holistic view of e-scooter safety.

## 1. Introduction

The rise in e-scooter usage, particularly in urban areas, has brought both advantages and new safety challenges. Given the relatively recent introduction of e-scooters, most research has concentrated on data from crash databases, hospital records, and police reports (Cicchino et al., 2021a; Sanders & Nelson, 2023; Stigson et al., 2021; Trivedi et al., 2019). These studies have provided initial insights into e-scooter crash patterns, often highlighting the prevalence of head injuries and the involvement of inexperienced riders (Austin Public Health, 2019; Cicchino et al., 2021a). However, analysis based on the crash reports does not capture all incidents, focussing on severe injury crashes. Further, the behavioural mechanisms leadings up to the safety–critical event (SCE; i.e., crashes and near-crashes) are not captured. Research on temporal patterns has shown increased crash rates on weekends (Stigson et al., 2021; Uluk et al., 2022), and during certain times of the day. Some studies report a higher proportion of crashes during daylight hours (Shah et al., 2021; Uluk et al., 2022) and others finding a greater risk at night (Stigson et al., 2021). In traffic safety research, naturalistic data analysis provides a unique understanding of rider behaviour and the leading causes of SCEs. Despite their potential, only a few naturalistic studies have been conducted on e-scooters so far (Pai & Dozza, 2025; White et al., 2023).

In the context of naturalistic data analysis, baselines are segments of trips that do not involve SCEs and can thus serve as reference

☆ This article is part of a special issue entitled: 'Electric personal mobility devices' published in Transportation Research Part F: Psychology and Behaviour.

* Corresponding author.
*E-mail addresses:* rahul.pai@chalmers.se (R.R. Pai), marco.dozza@chalmers.se (M. Dozza).

https://doi.org/10.1016/j.trf.2025.05.030
Received 14 January 2025; Received in revised form 7 March 2025; Accepted 27 May 2025
Available online 1 June 2025

points for comparison. Baselines can be selected either through randomisation or matching. Matched baselines have characteristics similar to those of the SCEs under study, ensuring an accurate representation of the crash scenario. While matching does not completely eliminate confounding effects, it helps control for confounders, possibly providing a less biased and more accurate estimation of crash risk (Lash et al., 2021) than if a baseline would be randomly picked. For instance, if most trips in a cyclist population occur in protected bicycle lanes, a randomly selected baseline would also be likely to include more trips in protected bicycle lanes. However, if most SCEs involving cyclists happen on roadways, a randomly selected baseline would not accurately represent the conditions of those SCEs; a matched baseline would therefore be more desirable. Nonetheless, one limitation inherent to matched baselines is that any potential effect on crash causation of any of the variables used for matching will not show in the analysis. For instance, matching based on the time of day would obfuscate any statistically significant differences in crash risk between day and night conditions. Random baselines allow for a broader, more general comparison. They are widely used in crash-causation analysis, not only due to their simplicity but also because they are less resource-intensive (Checkoway et al., 1989).

The odds ratio (OR), a simple metric that enables an understanding of the relationship between exposure and SCEs, is the most common way to compare SCEs with baselines. However, while OR estimates the risk, it does not indicate the extent of the problem within the population. For instance, a factor with a high OR may indicate a significant risk of SCEs but may be extremely rare in the population; thus the overall benefit of addressing that factor is limited. Put simply, factors with similar ORs are not necessarily similar in terms of importance for safety. To compensate for the limitations of OR analysis, researchers often calculate the population attributable risk percentage (PARP) (Herman et al., 2014; Olson et al., 2009), which quantifies the proportion of crashes that can be attributed to specific risk factors (Cole & Macmahon, 1971; Levin, 1953). This metric helps determine which factors, when addressed, offer the highest return in terms of reducing the overall incidence of SCEs. Therefore, by combining OR with PARP we may rank factors according to their actual impact on safety.

While the number of SCEs depends on the dataset, the number of baselines depends on the study design. Hennessy et al. (1999) have shown that for a general case-control study, increasing the controls-to-case ratio for a given OR increases statistical power. However, naturalistic data analysis typically involves video reduction and labelling, so scaling up the number of baselines requires significant time and resources. Previous naturalistic studies have employed a range of ratios, from 1:1 to 14:1 (Dozza & Werneke, 2014; Hankey et al., 2016; Victor et al., 2015; White et al., 2023). Additionally, Hoover et al. (2022) highlighted the importance of considering different baseline-to-SCE ratios. Therefore, understanding how varying the baseline-to-SCE ratio impacts the analysis may further elucidate the risk factors associated with SCEs.

This study used naturalistic riding data to estimate crash risk (OR) and crash prevalence (PARP) for e-scootering. We repeated the analyses with both random and matched baselines while also changing the SCE-to-baseline ratio.

## 2. Methods

We analysed naturalistic data (kinematics and video) collected from instrumented e-scooters. SCEs were identified from the kinematics and subsequently reviewed by watching the videos. We established baselines by identifying trips without SCEs (using both randomised and matched sampling strategies). After labelling SCE and baseline events, we calculated the OR and PARP for each variable, to measure the crash risk and estimate its impact on overall rider safety, respectively. Fig. 1 shows an overview of the methodology used in this study. The study was reviewed and approved by the Swedish Ethical Review Authority (Etikprövningsmyndigheten) (Ref. 2023-04671-01).

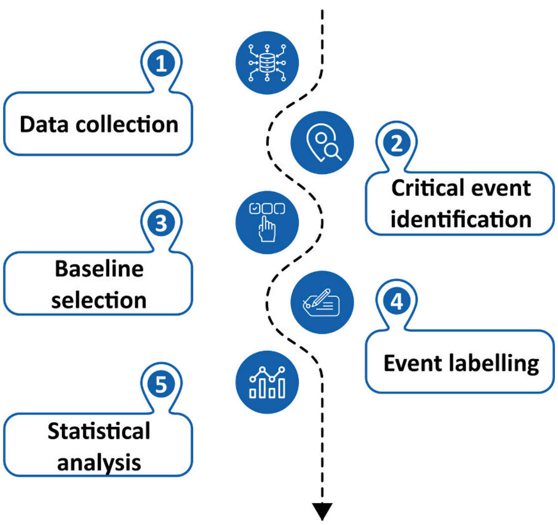


**Fig. 1.** Methodological steps from collection to statistical analysis.

### 2.1. Data collection

Naturalistic riding data were collected from 17 instrumented rental e-scooters in Gothenburg, Sweden (Boda et al., 2023; Pai, 2022). As illustrated in Fig. 2, the instrumentation comprised a data logger that recorded inputs from these onboard sensors: accelerometer, global navigation satellite system (GNSS), wheel speed sensors, and brake lever position sensors. A camera module recording 30 frames per second with 0.4 megapixels complemented the kinematic data with video. Within an 18-month data collection period, a cohort of 4694 distinct participants took 6868 trips, cumulatively covering a distance of 9930 km over 709 h. The e-scooters, part of a commercial rental fleet, were electronically limited to a top speed of 20 km/h, complying with Gothenburg regulations. Due to operational restrictions on the fleet, data were only collected in a four-km$^2$ area within Gothenburg's city centre (a predominantly urban environment), and e-scooters were not used on Friday and Saturday nights between 22:00 and 05:00.

### 2.2. Safety-critical event identification

We adapted the definitions of SCEs from previous naturalistic studies of bicycle riders (Dozza et al., 2016; Dozza & Werneke, 2014). This step was critical to ensure the accuracy of our data analysis, especially considering the dynamics and kinematics of e-scooters are decidedly different (Dozza et al., 2023; Li et al., 2023). An unintended event involving the e-scooter rider dismounting or colliding with surrounding objects or road users while the vehicle was in motion was classified as a crash. A near-crash was defined as an unintended event in which the ego e-scooter rider or the surrounding road users performed a rapid evasive manoeuvre. Candidate events were identified by examining outliers in accelerations, speed, and brake activation signals (Fig. 2). Specifically, we identified three types of candidate events based on kinematic triggers: toppled e-scooters, harsh braking, and sudden swerving manoeuvres (Pai & Dozza, 2025). An event was flagged as a toppled e-scooter if the lateral acceleration exceeded 6 m/s$^2$. A harsh braking event was identified if the longitudinal deceleration exceeded 3 m/s$^2$ and the brake lever was engaged (Li et al., 2023). A sudden swerving manoeuvre was identified if the absolute angular speed along the steering column axis exceeded 100°/s. These thresholds were applied only when the e-scooter speed was greater than 5 km/h. Applying these criteria, we identified 1801 candidate events. Two analysts reviewed each identified event to determine whether it qualified as an SCE.

### 2.3. Baseline selection

#### 2.3.1. Random baseline

After excluding trips that involved SCEs from the dataset, we randomly sampled six trips for each SCE. Each selected trip was associated with only one specific SCE to ensure the independence of baseline data. Subsequently, a random timestamp within each selected trip was determined to serve as the baseline. The selected data point was crucial for the subsequent processes of labelling and analysis. If the e-scooter was stationary at that timestamp, the randomization was repeated to avoid potential bias from stationary periods. Furthermore, to analyse the effect of the baseline-to-SCE ratio, we created two additional baseline sets by randomly subsampling from the initial six baselines per SCE: one set with three baselines per SCE, and one set with a single baseline per SCE.

#### 2.3.2. Matched baseline

For each SCE, GNSS data were used to identify other trips that passed through the same location. Furthermore, other matching criteria (speed, weather condition, lighting condition, time of day, and trip date) were implemented in order of decreasing priority.


**Fig. 2.** Instrumented e-scooter.

Trips matching each SCE with the highest number of criteria were selected as baselines. Because matched baselines had higher requirements than random ones, we limited the selection to three baselines per SCE to ensure that all SCEs had at least three matching baseline events.

### 2.4. Labelling critical and baseline events

Building on the data labelling framework by Pai and Dozza (2025), this study extended the labelling to include random baseline events. While SCEs and matched baselines were previously labelled, the random baselines required extra labelling. To ensure accuracy and reliability, two analysts independently labelled each event in a randomised sequence. For each SCE and baseline event, video segments of 20 s before and ten seconds after the event were analysed. To quantitatively assess the inter-rater reliability of the labels for subjective variables (i.e., those determined by individual analysts based on their perceptions), we calculated Cohen's kappa (Cohen, 1960) for the six subjective variables (out of the 14 analysed in this study; see Table 1). The average Cohen's kappa score was 0.91, indicating a high level of agreement between the analysts; further, they reviewed and discussed all discrepant labels in order to reach a consensus to finalise the labels for each event.

Some of the variable categories in Table 1 bear explaining. The trip purpose was categorised as either *leisure* or *commute* based on k-means clustering, with trip characteristics, including directness, proximity to points of interest, and temporal factors as inputs. The directness factor, categorized as either *detour* or *point-to-point*, was determined by comparing the actual trip distance to recommended route distance (from OpenStreetMap). Trips were classified as *detour* if the travel distance was greater than 1.7 times the recommended distance. Pack riding was defined as the ego e-scooter riding together with another e-scooter or bicycle in close proximity for a sustained period. Rider experience was quantified as the number of trips previously taken on e-scooters by that rider from the same rental provider. The complete codebook containing detailed descriptions of the variables and clustering methods are included in Pai and Dozza (2025).

### 2.5. Statistical analysis

#### 2.5.1. Contingency table

Contingency tables were constructed to organise the data and facilitate statistical analysis. Table 2 shows an example contingency table with the variables $\text{SCE}_{\text{single\_hand}}$ and $\text{Baseline}_{\text{single\_hand}}$ representing the number of single-handed riding SCEs (cases exposed) and baselines (controls exposed), respectively. Similarly, $\text{SCE}_{\text{two\_hands}}$ and $\text{Baseline}_{\text{two\_hands}}$ denote the number of two-handed riding SCEs (cases not exposed) and baselines (controls not exposed), respectively.

$\text{Total}_{\text{single\_hand}}$ and $\text{Total}_{\text{two\_hands}}$ represent the total number of single-handed and two-handed riding events, respectively. $\text{Total}_{\text{SCE}}$ and $\text{Total}_{\text{baseline}}$ indicate the total number of SCEs and baselines respectively. $\text{Total}_{\text{events}}$ refers to the total number of events considered in the analysis.

#### 2.5.2. Odds ratio calculation

We calculated the crude OR to measure the crash risk and corresponding 95 % Confidence Interval (CI) for all the labelled variables (Agresti, 1999; Bishop et al., 2007). Using the contingency table shown in Table 2, the OR and CI for single-handed riding were calculated as

$$OR = \frac{SCE_{single\_hand}}{SCE_{two\_hands}} \times \frac{Baseline_{two\_hands}}{Baseline_{single\_hand}}$$

$$CI = [\exp(ln(OR) - 1.96 \times SE(ln(OR))),\ \exp(ln(OR) + 1.96 \times SE(ln(OR)))]$$

**Table 1**
Variables analysed in this study. # indicates the variables directly influenced by matching criteria. The second column lists subcategories for categorical variables and units of measure for numerical variables.

| Variable | Subcategories/ [Unit of measure] | Variable type |
|---|---|---|
| Intersection# | Signalised, Non-signalised, Roundabout | Subjective, categorical |
| Type of road# | Bicycle lane, Sidewalk, Roadway, Shared (Micromobility users and Pedestrians), Other | Subjective, categorical |
| Trip purpose | Leisure, Commute | Objective, categorical |
| Trip day# | Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, Sunday | Objective, categorical |
| Mean speed | [km/h] | Objective, numerical |
| Directness factor | Detour, Point-to-point | Objective, categorical |
| Time of the event# | [hh:mm:ss] | Objective, numerical |
| Trip distance | [km] | Objective, numerical |
| Trip duration | [s] | Objective, numerical |
| Pack riding | Present, Absent | Subjective, categorical |
| Rider experience | [Trips] | Objective, numerical |
| Phone usage | Handheld, Using the phone holder, No phone usage | Subjective, categorical |
| Object on handlebar | Present, Absent | Subjective, categorical |
| Hand position | Two-handed, Single-handed, No hands on | Subjective, categorical |

**Table 2**
Contingency table for single-handed riding.

| | SCE | Baseline | All events |
|---|---|---|---|
| Single-handed riding | $SCE_{single\_hand}$ | $Baseline_{single\_hand}$ | $Total_{single\_hand}$ |
| Two-handed riding | $SCE_{two\_hands}$ | $Baseline_{two\_hands}$ | $Total_{two\_hands}$ |
| | $Total_{SCE}$ | $Total_{baseline}$ | $Total_{events}$ |

where,

$$SE(ln(OR)) = \sqrt{\frac{1}{SCE_{single\_hand}} + \frac{1}{SCE_{two\_hands}} + \frac{1}{Baseline_{single\_hand}} + \frac{1}{Baseline_{two\_hands}}}$$

Given that the number of SCEs remains constant, varying the baseline-to-SCE ratio will effectively change the proportion of $Baseline_{two\_hands}$ and $Baseline_{single\_hand}$, consequently changing the value of OR. Furthermore, such variations impact the width of the CI, as the value of each variable influences the limits of the CI and, therefore, may also affect the statistical significance of the OR. For a given factor, an OR greater than 1 indicates an increased SCE risk, whereas an OR less than 1 indicates a lowered risk.

### *2.5.3. Population attributable risk percentage*

In addition to OR, we estimated PARP (Walter, 1978) as

$$PARP = \frac{P_e(RR-1)}{P_e(RR-1)+1} \times 100$$

where $P_e$ is the proportion of the events exposed to the risk factor, and $RR$ is the relative risk. The $P_e$ and RR for single-handed riding shown in Table 2 can be calculated as

$$RR = \frac{SCE_{single\_hand} \times Total_{two\_hands}}{SCE_{two\_hands} \times Total_{single\_hand}}$$

and

$$P_e = \frac{Total_{single\_hand}}{Total_{events}}$$

where

$$Total_{single\_hand} = SCE_{single\_hand} + Baseline_{single\_hand}$$
$$Total_{two\_hands} = SCE_{two\_hands} + Baseline_{two\_hand}$$
$$Total_{events} = SCE_{single\_hand} + Baseline_{single\_hand} + SCE_{two\_hands} + Baseline_{two\_hand}$$

The corresponding CI can be calculated as

$$CI = [PARP - 1.96 \times SE_{PARP},\ PARP + 1.96 \times SE_{PARP}]$$

where,

$$SE_{PARP} = RR \times \sqrt{\frac{SCE_{single\_hand}}{Baseline_{single\_hand} \times Total_{two\_hands}} + \frac{SCE_{two\_hands}}{Baseline_{two\_hands} \times Total_{single\_hand}}}$$

A positive PARP indicates the percentage reduction in SCEs if the risk factor is eliminated, while a negative PARP indicates a protective effect, i.e., an increase in SCEs if the factor is hypothetically eliminated.

## 3. Results

The results are presented in three parts: OR, PARP, and the effect of baseline-to-SCE ratio on crash risk. The estimations of OR and PARP were conducted with emphasis on how the baseline sampling strategy affected the results, while the third part is a purely methodological contribution.

### *3.1. Odds ratio*

The crude OR and CI for both the random and matched baselines for a baseline-to-SCE ratio of 3:1 are reported in Table 3. As expected, the analysis revealed that factors unaffected by the matching criteria in the baseline selection exhibited consistent crash risks across both random and matched baselines, with the exception of short-duration and short-distance trips. Overall, the ORs for unmatched factors were lower in matched baselines than in random baselines—with detour trips, short-duration trips, and short-distance trips being notable exceptions. For both baseline types, the OR analysis indicated a statistically significant increase in SCE risk for riders with limited experience (five or fewer trips), phone usage while riding, single-handed riding, riding in a pack, leisure trips, detour trips, and trips with a mean speed of 11 km/h or lower. Similarly, the presence of an object on the handlebar increased the crash risk, although neither OR was statistically significant. While short trips (in terms of both distance and duration) were associated with a lower risk for random baselines, they were associated with an increased risk for matched baselines.

Of the factors influenced by the matching criteria, the presence of intersections increased the crash risk by a threefold for random baselines—but had a negligible effect for matched baselines. Similarly, the analysis revealed a disparity in crash risk between riding in pedestrian versus bicycle lanes; with random baseline sampling, riding in pedestrian lanes was safer (OR = 0.19; CI: [0.04, 0.85]), a trend opposite the one observed with matched sampling. The difference in risk of riding on the roadway compared to riding in the bicycle lane was negligible in both random and matched baseline analyses, with ORs of 0.99 [0.51, 1.94] and 1.2 [0.62, 2.32], respectively; neither finding was statistically significant.

Analyses of the trip characteristics indicated an increased crash risk for trips on Fridays and Saturdays, a pattern consistent across both random and matched baselines, with ORs of 2.75 [1.51, 5.00] and 1.71 [0.95, 3.06], respectively (the latter was not statistically significant, though). The results indicate a lower risk of SCE for trips between 21:00 and 00:00 h, with ORs of 0.22 [0.05, 0.98] and 0.26 [0.06, 1.15] for random and matched baselines, respectively (although the latter was not statistically significant).

### *3.2. Population attributable risk percentage*

For both the random and matched baselines with a baseline-to-SCE ratio of 3:1, Table 4 shows the variables within each category ranked in descending order of their PARP (calculated using random baselines). Factors with an OR less than 1.0 had a negative PARP, indicating a protective effect.

Similar to the results from the OR analysis, the PARP estimation indicated that factors unaffected by the matching criteria in baseline selection are ranked similarly across both random and matched baselines, except for short-duration and short-distance trips. Additionally, the magnitude of the PARP values for random baselines was higher than for matched baselines, suggesting that the PARP estimates for random baselines exacerbate the prevalence of these factors in SCEs. For random baselines, trip characteristics such as leisure trip and rider factors such as pack riding had PARPs of 36.51 % and 17.1 %, respectively. In fact, pack riding suggested a potential impact of 17.1 % for both random and matched baselines.

Among the infrastructure factors that are highly influenced by location, the presence of intersections had the highest PARP at 30.77 % for random baselines but had a minimal effect for matched baselines. Randomly sampled and matched baselines showed PARPs of 26.38 % and 16.67 %, respectively, for trips on Friday and Saturday.

Among the factors exhibiting a protective effect, trips taken between 21:00 and 00:00 were associated with a decrease in SCEs of 8.3 % and 6.8 % for random and matched baselines, respectively. The negative PARP suggests that eliminating trips in the nighttime could potentially increase the overall crash rates.

**Table 3**
Odds ratios of SCEs and confidence intervals for random and matched baseline analyses. (Bold font indicates statistical significance).

| | Random baselines | | Matched baselines | |
|---|---|---|---|---|
| | OR | CI | OR | CI |
| *Infrastructure factors* | | | | |
| Intersection present | **3.01** | [1.66, 5.46] | 0.98 | [0.55, 1.75] |
| Roadway vs bicycle lane | 0.99 | [0.51, 1.94] | 1.20 | [0.62, 2.32] |
| Pedestrian lane vs bicycle lane | **0.19** | [0.04, 0.85] | 1.71 | [0.30, 9.78] |
| *Trip characteristics* | | | | |
| Leisure vs commute trips | **6.85** | [3.54, 13.25] | **2.40** | [1.33, 4.35] |
| Friday + Saturday | **2.75** | [1.51, 5.00] | 1.71 | [0.95, 3.06] |
| Mean speed $\leq$ 11 km/h | **7.83** | [3.39, 18.07] | **3.69** | [1.81, 7.53] |
| Detour vs point-to-point trips | **2.53** | [1.05, 6.12] | **4.93** | [1.79, 13.6] |
| 21:00 to 00:00 vs other | **0.22** | [0.05, 0.98] | 0.26 | [0.06, 1.15] |
| Trip distance $\leq$ 1.45 km | **0.51** | [0.28, 0.93] | 1.25 | [0.68, 2.29] |
| Trip duration $\leq$ 371 s | **0.42** | [0.23, 0.78] | 1.16 | [0.63, 2.17] |
| *Rider factors* | | | | |
| Pack riding | **2.68** | [1.39, 5.17] | **2.68** | [1.39, 5.17] |
| Rider experience $\leq$ 5 trips | **3.15** | [1.52, 6.52] | **2.23** | [1.12, 4.46] |
| Phone usage | **3.35** | [1.26, 8.87] | **2.67** | [1.05, 6.81] |
| Object on handlebar | 2.15 | [0.94, 4.89] | 1.61 | [0.73, 3.55] |
| Single-handed riding | **3.88** | [1.14, 13.22] | **6.51** | [1.58, 26.89] |

**Table 4**
Population attributable risk percentage and confidence intervals for random and matched baseline analyses. (Italic font indicates values where corresponding OR was statistically significant).

| | Random baselines | | Matched baselines | |
|---|---|---|---|---|
| | PARP [%] | CI | PARP [%] | CI |
| *Infrastructure factors* | | | | |
| Intersection present | *30.77* | [30.62, 30.91] | –0.93 | [–1.27, –0.60] |
| Roadway vs bicycle lane | –0.15 | [–0.41, 0.1] | 4.96 | [4.75, 5.18] |
| Pedestrian lane vs bicycle lane | *–18.83* | [–19.08, –18.57] | 1.95 | [1.87, 2.03] |
| *Trip characteristics* | | | | |
| Leisure vs commute trips | *36.51* | [36.41, 36.61] | *24.05* | [23.9, 24.21] |
| Friday + Saturday | *26.38* | [26.24, 26.52] | 16.67 | [16.47, 16.86] |
| Mean speed ≤ 11 km/h | *21.86* | [21.78, 21.94] | *18.05* | [17.95, 18.15] |
| Detour vs point-to-point trips | *7.61* | [7.53, 7.70] | *10.13* | [10.06, 10.20] |
| 21:00 to 00:00 vs other | *–8.26* | [–8.37, –8.14] | –6.79 | [–6.89, –6.68] |
| Trip distance ≤ 1.45 km | *–23.81* | [–24.19, –23.43] | 5.45 | [5.27, 5.64] |
| Trip duration ≤ 371 s | *–30.16* | [–30.58, –29.74] | 3.53 | [3.35, 3.71] |
| *Rider factors* | | | | |
| Pack riding | *17.10* | [16.98, 17.22] | *17.10* | [16.98, 17.22] |
| Rider experience ≤ 5 trips | *14.98* | [14.87, 15.08] | *12.00* | [11.88, 12.12] |
| Phone usage | *7.96* | [7.89, 8.04] | *7.07* | [6.99, 7.15] |
| Object on handlebar | 7.41 | [7.31, 7.50] | 5.21 | [5.11, 5.32] |
| Single-handed riding | *5.58* | [5.52, 5.64] | *6.37* | [6.32, 6.42] |

### *3.3. Baseline-to-SCE ratio*

Table 5 shows the ORs and corresponding CIs for the different baseline-to-SCE ratios. As indicated in Table 5, increasing the baseline-to-SCE ratio generally results in narrower confidence intervals for most variables, with the exception of three (‘Intersection present’, ‘Friday + Saturday’, and ‘21:00 to 00:00′). For ORs, the initial change when increasing from one baseline per SCE to three was greater than the subsequent increase from three baselines to six for all variables except for three (‘Intersection present’, ‘21:00 to 00:00′, and ‘Rider experience’). Specifically, the OR for the variable ‘Intersection present’ shows an increasing trend with the addition of baselines, whereas the variable ‘21:00 to 00:00′ remains fairly constant.

**Table 5**
Odds ratios and confidence intervals for the three different random baseline-to-SCE ratios. (Bold font indicates statistical significance).

| | 6 baselines | | 3 baselines | | 1 baseline | |
|---|---|---|---|---|---|---|
| | OR | CI | OR | CI | OR | CI |
| *Infrastructure factors* | | | | | | |
| Intersection present | **3.75** | [2.15, 6.55] | **3.01** | [1.66, 5.46] | **2.78** | [1.33, 5.84] |
| Roadway vs bicycle lane | 0.85 | [0.46, 1.57] | 0.99 | [0.51, 1.94] | 1.79 | [0.75, 4.26] |
| Pedestrian lane vs bicycle lane | **0.17** | [0.04, 0.71] | **0.19** | [0.04, 0.85] | 0.25 | [0.05, 1.25] |
| *Trip characteristics* | | | | | | |
| Leisure vs commute trips | **6.38** | [3.56, 11.43] | **6.85** | [3.54, 13.25] | **9.47** | [3.55, 25.26] |
| Friday + Saturday | **2.44** | [1.41, 4.22] | **2.75** | [1.51, 5.00] | **1.97** | [0.95, 4.08] |
| Mean speed ≤ 11 km/h | **6.75** | [3.39, 13.41] | **7.83** | [3.39, 18.07] | **6.45** | [2.04, 20.35] |
| Detour vs point-to-point trips | **2.54** | [1.16, 5.58] | **2.53** | [1.05, 6.12] | 2.20 | [0.70, 6.86] |
| 21:00 to 00:00 vs other | 0.24 | [0.06, 1.02] | **0.22** | [0.05, 0.98] | 0.22 | [0.05, 1.10] |
| Trip distance ≤ 1.45 km | **0.44** | [0.25, 0.78] | **0.51** | [0.28, 0.93] | 0.58 | [0.28, 1.20] |
| Trip duration ≤ 371 s | **0.41** | [0.23, 0.73] | **0.42** | [0.23, 0.78] | 0.36 | [0.17, 0.76] |
| *Rider factors* | | | | | | |
| Pack riding | **2.79** | [1.53, 5.07] | **2.68** | [1.39, 5.17] | **4.05** | [1.57, 10.45] |
| Rider experience ≤ 5 trips | **2.83** | [1.49, 5.37] | **3.15** | [1.52, 6.52] | **2.98** | [1.13, 7.83] |
| Phone usage | **2.71** | [1.18, 6.20] | **3.35** | [1.26, 8.87] | **5.11** | [1.05, 24.71] |
| Object on handlebar | **2.22** | [1.05, 4.67] | 2.15 | [0.94, 4.89] | **4.25** | [1.12, 16.11] |
| Single-handed riding | **3.52** | [1.25, 9.91] | **3.88** | [1.14, 13.22] | 6.55 | [0.76, 56.1] |

## 4. Discussion

### 4.1. Crash risk

Factors that were not influenced by the baseline matching criteria exhibited consistent crash risks across sampling strategies. For instance, riders with limited experience consistently showed higher crash risks, which aligns with previous research findings (Austin Public Health, 2019; Cicchino et al., 2021a). Similarly, single-handed riding and phone usage while riding (which also partly contributes to single-handed riding) were significant risk factors. It is worth noting that riding single-handed on an e-scooter is more challenging than on a bicycle. In fact, while the balancing task may feel the same for both e-scooter and bicycles, their manoeuvring differs, especially in crash avoidance (Dozza et al., 2023; Li et al., 2023). Further, e-scooters, due to their design, are inherently unstable at their permitted operational speeds (Paudel & Fah Yap, 2021), requiring riders to exert more effort to maintain balance when perturbed compared to bikes (Kooijman et al., 2011). Riders may also over-rely on their previous experience with bicycles (which is typically much longer than their e-scootering experience), thus underestimating the risk of riding an e-scooter single-handed and being more prone to falling when a perturbation occurs (Li et al., 2023). This study is based on data from 17 instrumented e-scooters in Gothenburg, Sweden, representing an unprecedented naturalistic riding dataset which provides unique insights into e-scooter crash risk factors. A key question for future research is how these risk factors vary across different urban environments; therefore, future studies should aim to incorporate data from multiple cities and potentially larger e-scooter fleets to build upon this foundational understanding of e-scooter crash risk.

As expected, matching on location obscured the effect of location-related factors such as the presence of intersections and different road types, leading to different crash risks for random and matched baselines. For instance, the presence of intersections had a negligible effect on crash risk for matched baselines; however, for randomly sampled baselines, intersection presence increased the crash risk threefold. The increased risk at intersections is consistent with previous research, which showed that intersections are crash-prone areas for all road users, particularly micromobility users (Dozza & Werneke, 2014; Jia et al., 2021; Schepers et al., 2011; Shah et al., 2021; Werneke et al., 2015). The differences in risk estimates based on the sampling strategy align with the findings of Yuan et al. (2021), who also noted that using a matched sampling strategy can obscure certain effects, such as the impact of cycling volume on crash risk. Random baselines also showed a (statistically significant) reduction in crash risk for pedestrian lanes versus bicycle lanes; however, similar to the presence of intersections, this factor was not significant when OR were computed on matched baselines, most likely because we matched by location. Although our findings suggest that pedestrian lanes are safer for e-scooter riders, it is crucial to consider the potential risks for pedestrians. Future studies should evaluate the safety of pedestrian lanes from the perspective of pedestrians to gain a comprehensive understanding of the overall risk for all road users. Additionally, when comparing roadways to bicycle lanes, both the random and matched baselines showed no significant difference in crash risk. The lack of significant difference in both sampling strategies suggests that roadways and bicycle lanes may simply not differ substantially in terms of crash risk for e-scooter riders.

Interestingly, the reduced risk during late evening hours (21:00 to 00:00) was evident in both sampling strategies, with ORs of 0.22 (statistically significant) for the randomly sampled baseline and 0.26 for the matched baseline. It is important to note that in Gothenburg, where our data were collected, there is a ban on e-scooter trips on Friday and Saturday nights, resulting in no data for those time periods. Additionally, our dataset did not include information on alcohol intoxication, a well-known risk factor for crashes. The finding of lower crash risk during night riding appears to contradict previous studies, which reported higher crash rates at night compared to daytime (Shah & Cherry, 2022; Stigson et al., 2021). However, these studies relied on hospital reports and crash databases and—most importantly—included trips on weekends when alcohol may have influenced crash risk. Hence, our study suggests that intoxication may play a larger role than reduced visibility in increasing crash risk during night riding. Future studies should aim to combine crash databases and hospital reports with naturalistic data to provide a more comprehensive understanding of crash risk for intoxicated riding as well as across different times of the day and severity levels.

Trips on Fridays and Saturdays showed an increased risk in both baseline types (OR = 2.75 for random baseline and OR = 1.71 for matched baseline). Notably, the OR from the random baseline is statistically significant, aligning with previous studies (Cicchino et al., 2021b; Stigson et al., 2021) that observed an increased crash rate on those days. It is worth noting that if the Friday and Saturday nights were included, the results could become more pronounced for both matched and random baselines. Shorter trips ($\leq$1.45 km) and shorter trip durations ($\leq$371 s) were associated with lower crash risks in the random baseline analysis (ORs of 0.51 and 0.42, respectively). These short trips might expose riders to fewer risks on average; therefore, exposure alone may explain these results.

Short-duration and short-distance trips present a unique case since the risk patterns differ between the random and matched baselines, although these factors were not controlled in the matching process. In fact, while these trips were associated with lower risk in the random baseline, they showed increased risk in the matched baseline. The discrepancy may still be explained by the matching process, which might have captured trips similar in terms of distance and time within specific parts of the city. For example, trips in densely populated areas might be shorter in distance and duration than trips further away from city centres; thus the matching strategy may indirectly control for population density.

The odds ratios presented in this study are crude estimates, examining the association between individual risk factors and SCEs without controlling for other potential confounders or accounting for repeated measures within riders. This approach was chosen for its simplicity and interpretability in this initial, exploratory analysis of a relatively small dataset. Future research may employ more complex analyses using multivariable models to assess the independent and combined effects of multiple risk factors, and mixed-effects models to address repeated measures. Given that there are 6868 trips from 4694 unique participants, accounting for repeated measures within riders is unlikely to significantly change any of the results. Ideally, future work should be hypothesis-driven and guided by causal inference from directed acyclic graphs to verify how the factors under analysis influence each other—and what such an influence would mean for crash causation and countermeasure development.

### *4.2. Crash prevalence*

Tables 3 and 4 show that although single-handed riding has a high OR, indicating a high crash risk, the low PARP suggests it is not widespread within the population, thus limiting the overall benefit of addressing it. Detour trips and phone usage while riding also have high ORs and low PARPs. In contrast, leisure trips and low mean speed (≤11 km/h) both exhibit high ORs and high PARPs, indicating a significant risk and a high prevalence within the population, for both matched and random baselines. It is important to note that while low mean speed shows high OR and PARP values, experience influences the mean speed (Pai & Dozza, 2025). For random baselines specifically, the presence of intersections has a high OR and PARP, highlighting its significant risk and prevalence. The high-risk and high-prevalence factors should be prioritised for intervention.

The differences between OR and PARP rankings on the importance of a factor in crash causation highlight the limitations of OR alone for capturing the prevalence of risk in the population. It is worth noting that in this study, we calculated the PARP for each factor independently. Our approach does not account for potential interactions between factors; therefore, the results cannot be interpreted as absolute percentages of preventable SCEs. However, it provides a valuable metric to augment the OR by offering supplementary context about the relative contribution of each factor.

The combination of high OR and PARP values reveals that addressing factors such as leisure trips, the presence of intersections, and trips on Fridays and Saturdays should be prioritised. In contrast, factors like single-handed riding, phone usage, and detour trips may be considered lower priority; although they increase the crash risk, they are not very prevalent. Additionally, factors with negative PARP values, such as pedestrian lane usage and trips between 21:00 and 00:00, should be preserved to avoid increasing crash risk. However, as discussed earlier in section 4.1, more research is needed to determine the net safety effect of pedestrian lane usage and the effects of alcohol and traffic density on night trips.

### *4.3. Baseline-to-SCE ratio*

In our study, we observed that increasing the baseline-to-SCE ratio resulted in a higher number of statistically significant variables. This finding is consistent with those of Hoover et al. (2022), who investigated different controls-to-case ratios in their naturalistic data analysis. They observed an increase in the number of statistically significant variables when increasing the controls-to-case ratio from 4:1 to 9:1 and 14:1; however, the statistical significance diminished at higher ratios. In addition, Hennessy et al. (1999) showed that while increasing the baseline-to-SCE ratio enhances statistical power, the incremental gains diminish at higher ratios, resulting in a flattening curve. While our study only tested baseline-to-SCE ratios of 1:1, 3:1, and 6:1, it is possible that increasing the ratio further may reveal a similar trend, though this remains to be tested. Our results, detailed in Table 5, show that increasing the baseline-to-SCE ratio results in narrower confidence intervals for most variables, indicating an increase in the precision of the calculated OR as the number of baseline increases.

While increasing the baseline-to-SCE ratio can give higher-precision estimates, it is important to balance statistical power with computational efficiency. This balance becomes crucial when video annotation is involved, as it requires considerable time and resources.

## 5. Conclusion

This study highlights the importance of prioritizing factors based on a combination of crash risk (OR) and prevalence (PARP). Focusing solely on factors with a high OR, such as single-handed riding, may not be optimal to increase safety; the PARP results may indicate that the underlying behaviours are not widespread among the e-scooter riding population. Further, our findings from the OR analysis challenge previous assumptions based on traditional crash databases and hospital reports: nighttime e-scooter usage might be safer than previously thought, suggesting that nighttime bans may not be an effective safety measure. However, it is important to note that our dataset did not include trips on weekend nights, when alcohol use is often more prevalent. Thus, a possible influence of alcohol on crash risk would not be detected.

Both matched and random baseline sampling techniques offer distinct advantages in crash analysis. Matched baselines can generate more accurate estimates by controlling for confounding variables. However, that sampling technique may not reveal trends and associations attributable to the matched variables. Random baselines, while easier to implement and better suited to uncover the influence of a broader set of variables, are more susceptible to confounding variables—which could potentially exaggerate a factor's influence on SCE risk. Ideally, employing both sampling techniques provides a broader understanding of crash dynamics. However, if only one technique can be chosen, matched baselines may provide a more accurate understanding of SCE risk factors. The study also highlights that the optimal baseline-to-SCE ratio balances statistical precision with computational efficiency. A higher ratio generally leads to more precise crash risk estimations but demands more computational resources, especially for video data labelling.

To conclude, the study shows that prevalence of leisure trips, exposure to intersections, trips on Fridays and Saturdays, pack riding, and inexperienced riding are the most important issues to be solved for safely integrating e-scootering into the transport system.

## Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used Grammarly and Microsoft Copilot to correct grammar and review language. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

## CRediT authorship contribution statement

**Rahul Rajendra Pai:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation. **Marco Dozza:** Writing – review & editing, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization.

## Declaration of Competing Interest

The authors declare no financial interests/personal relationships which may be considered as potential competing interests.

## Acknowledgements

We thank Carol Flannagan, Research Professor at the University of Michigan Transportation Research Institute (UMTRI) for her insightful discussions that greatly contributed to this study. We would also like to thank Dr. K. Mayberry for language revisions and the team at Voi Technology AB for the naturalistic data collection. This work was conducted as part of the project Safe Integration of Micro-mobility in the Transport System (SIMT), funded by Trafikverket through the Forskning och innovation (FoI) program under grant 2022/32014. The data collection was supported by the project e-SAFER funded jointly by Vinnova, the Swedish Energy Agency, the Swedish Transport Administration, and the Swedish vehicle industry through the Strategic Vehicle Research and Innovation (FFI) Program under grant 2022-01641. The work was carried out at Chalmers University of Technology, Gothenburg, Sweden.

## Data availability

The contingency tables will be made available on request.

## References

Agresti, A. (1999). On logit confidence intervals for the odds ratio with small samples. *Biometrics, 55*(2), 597–602. https://doi.org/10.1111/J.0006-341X.1999.00597.X

Austin Public Health. (2019). *Dockless electric scooter-related injuries study*. https://www.austintexas.gov/sites/default/files/files/Health/Epidemiology/APH_Dockless_Electric_Scooter_Study_5-2-19.pdf.

Bishop, Y. M., Holland, P. W., & Fienberg, S. E. (2007). Discrete multivariate analysis theory and practice. In *Discrete multivariate analysis theory and practice*. Springer New York. https://doi.org/10.1007/978-0-387-72806-3.

Boda, C.-N., Ahmed, J., Baluyot, R., Eklöf, K., Pai, R. R., & Dankert, A. (2023). *E-safe pre-study*. https://www.vinnova.se/globalassets/mikrosajter/ffi/dokument/slutrapporter-ffi/trafiksakerhet-och-automatiserade-fordon-rapporter/slutrapport-2021-05060engelska.pdf.

Checkoway, H., Pearce, N. E., & Crawford-Brown, D. J. (1989). *Research methods in occupational epidemiology* (1st ed., Vol. 13). Oxford University Press, Incorporated.

Cicchino, J. B., Kulie, P. E., & McCarthy, M. L. (2021a). Injuries related to electric scooter and bicycle use in a Washington, DC, emergency department. *Traffic Injury Prevention, 22*(5), 401–406. https://doi.org/10.1080/15389588.2021.1913280

Cicchino, J. B., Kulie, P. E., & McCarthy, M. L. (2021b). Severity of e-scooter rider injuries associated with trip characteristics. *Journal of Safety Research, 76*, 256–261. https://doi.org/10.1016/J.JSR.2020.12.016

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37–46. https://doi.org/10.1177/001316446002000104

Cole, P., & Macmahon, B. (1971). Attributable risk percent in case-control studies. *Journal of Preventive and Social Medicine, 25*(4), 242–244. https://www.jstor.org/stable/25565674.

Dozza, M., Bianchi Piccinini, G. F., & Werneke, J. (2016). Using naturalistic data to assess e-cyclist behavior. *Transportation Research Part F: Traffic Psychology and Behaviour, 41*, 217–226. https://doi.org/10.1016/J.TRF.2015.04.003

Dozza, M., Li, T., Billstein, L., Svernlöv, C., & Rasch, A. (2023). How do different micro-mobility vehicles affect longitudinal control? Results from a field experiment. *Journal of Safety Research, 84*, 24–32. https://doi.org/10.1016/J.JSR.2022.10.005

Dozza, M., & Werneke, J. (2014). Introducing naturalistic cycling data: What factors influence bicyclists' safety in the real world? *Transportation Research Part F: Traffic Psychology and Behaviour, 24*, 83–91. https://doi.org/10.1016/J.TRF.2014.04.001

Hankey, J. M., Perez, M. A., & McClafferty, J. A. (2016). *Description of the SHRP 2 Naturalistic Database and the Crash, Near-Crash, and Baseline Data Sets*. http://hdl.handle.net/10919/70850.

Hennessy, S., Bilker, W. B., Berlin, J. A., & Strom, B. L. (1999). Factors influencing the optimal control-to-case ratio in matched case-control studies. *American Journal of Epidemiology, 149*(2), 195–197. https://doi.org/10.1093/oxfordjournals.aje.a009786

Herman, J., Kafoa, B., Wainiqolo, I., Robinson, E., McCaig, E., Connor, J., Jackson, R., & Ameratunga, S. (2014). Driver sleepiness and risk of motor vehicle crash injuries: A population-based case control study in Fiji (TRIP 12). *Injury, 45*(3), 586–591. https://doi.org/10.1016/j.injury.2013.06.007

Hoover, L., Bhowmik, T., Yasmin, S., & Eluru, N. (2022). Understanding crash risk using a multi-level random parameter binary logit model: Application to naturalistic driving study data. *Transportation Research Record: Journal of the Transportation Research Board, 2676*(10), 737–745. https://doi.org/10.1177/03611981221090943

Jia, B., Guenther, D., & Heydinger, G. (2021). Crash factor analysis in intersection-related crashes using SHRP 2 naturalistic driving study data. *SAE International Journal of Advances and Current Practices in Mobility, 3*(5), 2021-01–0872. https://doi.org/10.4271/2021-01-0872.

Kooijman, J. D. G., Meijaard, J. P., Papadopoulos, J. M., Ruina, A., & Schwab, A. L. (2011). A bicycle can be self-stable without gyroscopic or caster effects. *Science, 332*(6027), 339–342. https://doi.org/10.1126/science.1201959

Lash, T. L., VanderWeele, T. J., Haneause, S., & Rothman, K. J. (2021). *Modern epidemiology* ((4th ed.).). Wolters Kluwer Health.

Levin, M. L. (1953). The occurrence of lung cancer in man. *Acta - Unio Internationalis Contra Cancrum, 9*(3), 531–541. http://www.ncbi.nlm.nih.gov/pubmed/13124110.

Li, T., Kovaceva, J., & Dozza, M. (2023). Modeling collision avoidance maneuvers for micromobility vehicles. *Journal of Safety Research, 87*, 232–243. https://doi.org/10.1016/J.JSR.2023.09.019

Olson, R. L., Hanowski, R. J., Hickman, J. S., & Bocanegra, J. (2009). *Driver distraction in commercial vehicle operations*. https://doi.org/10.21949/1502647.

Pai, R. R. (2022). *Logging data from E-scooters to improve traffic safety* [Chalmers University of Technology]. https://hdl.handle.net/20.500.12380/305446.

Pai, R. R., & Dozza, M. (2025). Understanding factors influencing e-scooterist crash risk: A naturalistic study of rental e-scooters in an urban area. *Accident Analysis & Prevention, 209*, Article 107839. https://doi.org/10.1016/J.AAP.2024.107839

Paudel, M., & Fah Yap, F. (2021). Front steering design guidelines formulation for e-scooters considering the influence of sitting and standing riders on self-stability and safety performance. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering, 235*(9), 2551–2567. https://doi.org/10.1177/0954407021992176

Sanders, R. L., & Nelson, T. A. (2023). Results from a campus population survey of near misses, crashes, and falls while E-scooting, walking, and bicycling. *Transportation Research Record, 2677*(2), 479–489. https://doi.org/10.1177/03611981221107010

Schepers, J. P., Kroeze, P. A., Sweers, W., & Wüst, J. C. (2011). Road factors and bicycle–motor vehicle crashes at unsignalized priority intersections. *Accident Analysis & Prevention, 43*(3), 853–861. https://doi.org/10.1016/J.AAP.2010.11.005

Shah, N. R., Aryal, S., Wen, Y., & Cherry, C. R. (2021). Comparison of motor vehicle-involved e-scooter and bicycle crashes using standardized crash typology. *Journal of Safety Research, 77*, 217–228. https://doi.org/10.1016/J.JSR.2021.03.005

Shah, N. R., & Cherry, C. R. (2022). Riding an e-scooter at nighttime is more dangerous than at daytime. In T. Petzoldt, R. Gerike, J. Anke, M. Ringhand, & B. Schröter (Eds.), *Contributions to the 10th International Cycling Safety Conference 2022 (ICSC2022)* (pp. 60–62). Technische Universität Dresden. https://doi.org/10.25368/2022.436.

Stigson, H., Malakuti, I., & Klingegård, M. (2021). Electric scooters accidents: Analyses of two Swedish accident data sets. *Accident Analysis & Prevention, 163*, Article 106466. https://doi.org/10.1016/J.AAP.2021.106466

Trivedi, T. K., Liu, C., Antonio, A. L. M., Wheaton, N., Kreger, V., Yap, A., … Elmore, J. G. (2019). Injuries Associated With Standing Electric Scooter Use. *JAMA Network Open, 2*(1), e187381. https://doi.org/10.1001/JAMANETWORKOPEN.2018.7381

Uluk, D., Lindner, T., Dahne, M., Bickelmayer, J. W., Beyer, K., Slagman, A., Jahn, F., Willy, C., Möckel, M., & Gerlach, U. A. (2022). E-scooter incidents in Berlin: An evaluation of risk factors and injury patterns. *Emergency Medicine Journal, 39*, 295–300. https://doi.org/10.1136/emermed-2020-210268

Victor, T., Dozza, M., Bärgman, J., Boda, C.-N., Engström, J., Flannagan, C., … Markkula, G. (2015). *Analysis of naturalistic driving study data: Safer glances, driver inattention, and crash risk*. *Transportation Research Board*. https://doi.org/10.17226/22297

Walter, S. D. (1978). Calculation of attributable risks from epidemiological data. *International Journal of Epidemiology, 7*(2), 175–182. https://doi.org/10.1093/IJE/7.2.175

Werneke, J., Dozza, M., & Karlsson, M. (2015). Safety–critical events in everyday cycling – Interviews with bicyclists and video annotation of safety–critical events in a naturalistic cycling study. *Transportation Research Part F: Traffic Psychology and Behaviour, 35*, 199–212. https://doi.org/10.1016/J.TRF.2015.10.004

White, E., Guo, F., Han, S., Mollenhauer, M., Broaddus, A., Sweeney, T., Robinson, S., Novotny, A., & Buehler, R. (2023). What factors contribute to e-scooter crashes: A first look using a naturalistic riding approach. *Journal of Safety Research, 85*, 182–191. https://doi.org/10.1016/J.JSR.2023.02.002

Yuan, J., Abdel-Aty, M. A., Yue, L., & Cai, Q. (2021). Modeling real-time cycle-level crash risk at signalized intersections based on high-resolution event-based data. *IEEE Transactions on Intelligent Transportation Systems, 22*(11), 6700–6715. https://doi.org/10.1109/TITS.2020.2994126