

SAFETY FINDINGS

Predictors of Biking Behavior Changes after a Safety Incident

Caitlyn Linehan, BS, MS¹, Trisalyn Nelson, Ph.D.¹¹ Geography, University of California, Santa Barbara

Keywords: bicycling safety, random forest, crowdsourced data, injury severity, active transportation behavioral changes

<https://doi.org/10.32866/001c.142431>

Findings

Understanding how a safety incident while biking influences future ridership can inform efforts to increase and retain bicycle ridership. However, there are limited data on how a crash or near miss impacts a person's future riding. Using self-reported incident data, we characterize changes in biking behavior after a safety incident. We predict riding behavior after a safety incident using random forest and input data on demographics, incident types, and injury severity. Our findings show that experiencing a hospitalized injury, involvement in a collision, and female identity are associated with a higher likelihood of changing biking behavior after a safety incident.

1. Questions

After a bicyclist experiences a crash or near miss, they may alter their riding behavior, such as reducing their frequency of biking or replacing bike trips with other forms of transportation that feel safer. For example, Lee, Underwood, and Handy (2015) demonstrate that the severity of injury a bicyclist sustains during a crash is significantly associated with a decline in their bicycling comfort and willingness to continue biking. Identifying the factors that contribute to behavioral change post-incident is essential for improving our understanding of bicycling safety and informing targeted interventions. In this study, we examine how various characteristics: type of incident, whether the respondent was a regular cyclist, whether the respondent was wearing a helmet, terrain, gender identity, age group, who/what the respondent had an incident with, type of injury, and type of road conditions, predict changes in biking behavior after a reported incident. Specifically, we aim to address the following research question:

1. What characteristics are associated with a change in biking behavior following a crash or near miss?

2. Methods

We use self-reported collision and near miss data submitted to [BikeMaps.org](https://www.bikemaps.org), a global crowdsourcing platform where bicyclists can report collisions, near misses, hazards, and thefts. Researchers have shown that BikeMaps data are best used to complement traditional data sources, but it is important to acknowledge the demographic limitations of this data such as a bias towards younger, more technological savvy users (Ferster et al. 2017). Previous

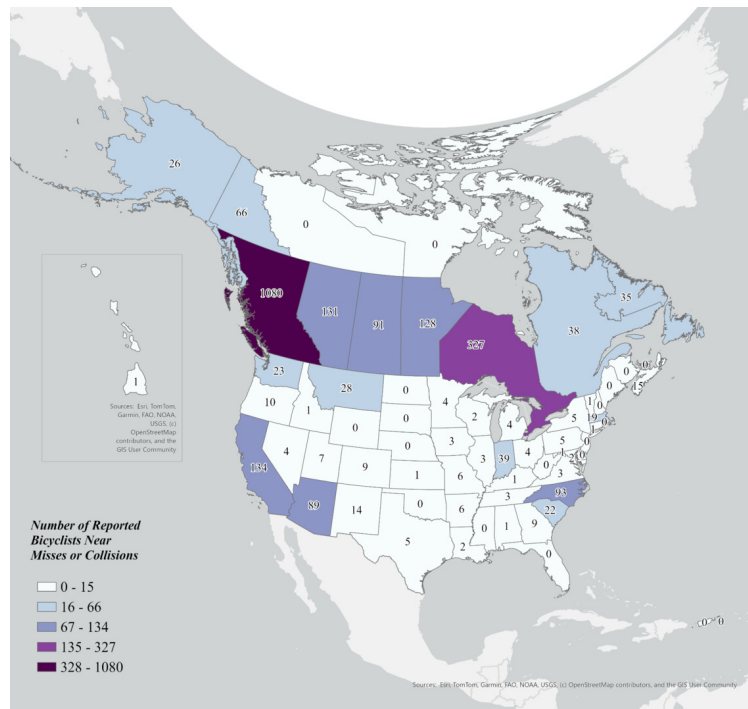


Figure 1. Map of 2502 submitted reports used in analysis.

research in Branion-Calles, Nelson, and Winters (2017) show how BikeMaps crash and near miss data can be used to complement and fill gaps in reports to official collision sources.

For this study, we focus on 2,502 incidents reported in the United States and Canada and limit our dataset to responses with complete records to ensure consistency in modeling ([Figure 1](#)). In our study we include both collisions and near miss reports in our analysis, because prior research emphasizes the importance of near miss events for perceived safety (Branion-Calles, Nelson, and Winters 2017).

We perform an extensive data cleaning process, including filtering incomplete responses, collapsing low-frequency categories to ensure balanced class distributions, and excluding variables with minimal variation. [Table 1](#) shows the different variables of the model as well as their respective values and frequencies.

We analyze predictors of post-incident biking behavior (no change versus some change [which includes biking less, biking more carefully, biking more carefully and biking less, and stopped biking]). We use the following predictors: type of incident, whether the respondent was a regular cyclist, whether the respondent was wearing a helmet, terrain, gender, age group, who/what the respondent had an incident with, type of injury, and type of road conditions. We then utilize a random forest classification model using the tidymodels package in R (Kuhn and Wickham 2020). [Table 1](#) reports the variables used in the development of our model.

Table 1. Included variables in the model

Variable	Value	n	Percentage (%)
Type of incident	Collision	672	26.9
	Near Miss	1830	73.1
Is the respondent a regular cyclist?	No	37	1.5
	Yes	2465	98.5
Was the respondent wearing a helmet?	No	314	12.5
	Yes	2188	87.5
Terrain	Downhill	376	15
	Flat	1880	75.1
	Uphill	246	9.8
Gender Identity	Female	769	30.7
	Male	1733	69.3
Age-Group	5-29	136	5.4
	30-39	785	31.4
	40-49	745	29.8
	50-59	340	13.6
	>= 60	496	19.8
Impact on biking	Bike less; More careful; More careful and bike less; Stopped biking	1426	57
	none	1076	43
Incident with	Animal	21	0.8
	Cyclist or E-scooter	68	2.7
	Infrastructure	96	3.8
	Other	67	2.7
	Pedestrian	48	1.9
	Vehicle, angle; Vehicle, side	769	30.7
	Vehicle, head on; Vehicle, rear end	462	18.5
	Vehicle, open door; Vehicle, passing	361	14.4
	Vehicle, turning left	200	8
	Vehicle, turning right	410	16.4
Type of injury	Injury, hospital emergency visit or hospitalized	182	7.3
	Injury, no treatment, and Injury, saw a family doctor	305	12.2
	No injury	2015	80.5
Road conditions	Dry	2097	83.8
	Other	405	16.2

We trained our model on 70% of data or 1751 reports and tested on the remaining 30% of data or 751 reports. We interpret our model utilizing a variable importance plot and partial dependence plots. We utilize variable importance plots to see what variables contributed more to a decision made in a random forest tree. We perform partial dependence plots in order to show the effect of a variable on the predicted outcome, while maintaining the other values of the feature.

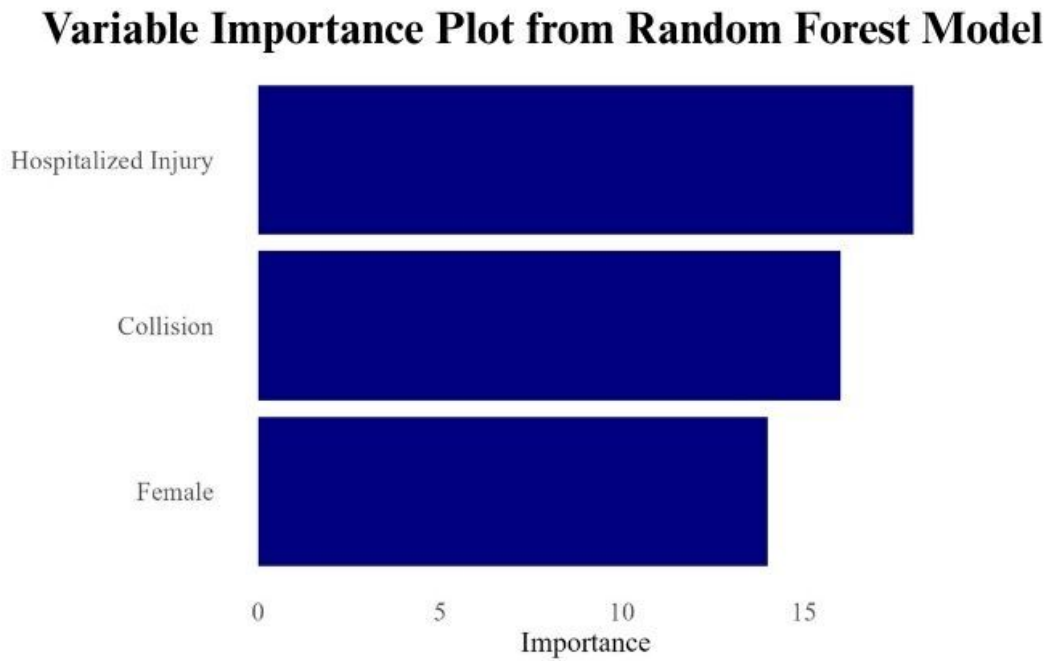


Figure 2. Variable importance plots in RF model for predicting bicycle behavior changes post safety incident

3. Findings

Our random forest model achieved a ROC AUC of 0.64, indicating modest discriminatory ability which leaves room for improvement. One contributing factor may be the class imbalance of our variables as shown in [Table 1](#). While we mitigated this by combining classes when appropriate, the class imbalance may affect the model's ability to predict. We also acknowledge the spatial imbalance in our dataset, with most reports in Canada, which may limit our ability to capture more regional-specific behavioral responses ([Figure 1](#)). Variable importance plots ([Figure 2](#)) were used to interpret which predictors were most vital in the model making the decision in the random forest classification for determining biking behavior following an incident. The x-axis displays the Gini importance, a relative measure of how impactful each variable is in making decisions in the model. The most influential predictors contributing to node splitting in the random forest classification model for predicting biking behavior were hospitalized injury followed by collision classification and then identification as a female.

While variable importance plots indicate the relative contribution of each predictor to the model's decision-making process, they do not convey the direction of a variable's influence on biking behavior. We therefore use partial dependence plots to examine the marginal effect of individual predictors on the probability of a given biking behavior outcome ([Figure 3](#)).

The partial dependence plots show that individuals who experienced a hospitalized injury, reported a collision rather than a near miss, or identify as female are more likely to have a change in biking (which includes biking

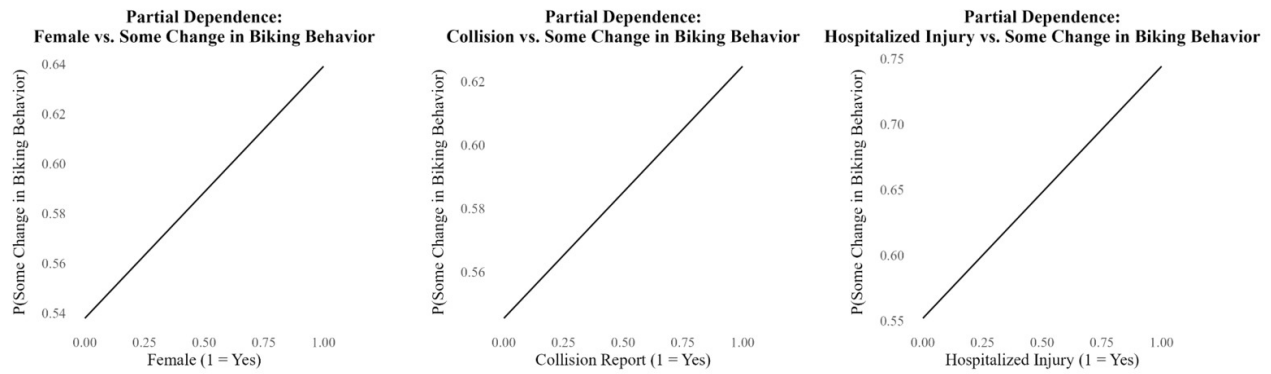


Figure 3. Partial dependence plots for variables in the model for predicting bicycle behavior changes post safety incident

less, biking more carefully, biking more carefully and biking less, and stopped biking) behavior following a crash or near miss. These results are consistent with previous findings showing that women are more likely than men to reduce their cycling exposure following a crash (Fraser and Meuleners 2020). These predictors are positively associated with the probability of some type of biking behavior change while holding the other variables constant, as illustrated by the positive slopes in the partial dependence plots ([Figure 3](#)).

Our models can be used to help identify bicyclists who are more likely to reduce bicycling following a safety incident. This can inform the allocation of post-crash resources or targeted interventions. Future work should investigate whether cities offering support services see improved retention of bicyclists after crashes. The reports combined alongside qualitative interviews may provide further insight into the types of support most valued by affected populations. Our work can be situated alongside other research such as Fraser and Meuleners (2020), which identifies protective factors such as group riding and those who had a full recovery post safety incident being associated with lower odds of reducing bicycling exposure. Our work can help shape data-based interventions at improving post biking safety incident recovery and retention.

Submitted: June 15, 2025 AEST. Accepted: July 22, 2025 AEST. Published: August 01, 2025 AEST.



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-SA-4.0). View this license's legal deed at <https://creativecommons.org/licenses/by-sa/4.0> and legal code at <https://creativecommons.org/licenses/by-sa/4.0/legalcode> for more information.

REFERENCES

- Branion-Calles, M., T. Nelson, and M. Winters. 2017. "Comparing Crowdsourced Near-Miss and Collision Cycling Data and Official Bike Safety Reporting." *Transportation Research Record* 2662 (1): 1–11. <https://doi.org/10.3141/2662-01>.
- Ferster, C. J., T. Nelson, M. Winters, and K. Laberee. 2017. "Geographic Age and Gender Representation in Volunteered Cycling Safety Data: A Case Study of BikeMaps.Org." *Applied Geography* 88:144–50. <https://doi.org/10.1016/j.apgeog.2017.09.007>.
- Fraser, L., and L. B. Meuleners. 2020. "Getting Back on the Bike: Participation in Cycling after a Hospitalisation Crash." *Accident Analysis & Prevention* 146:105726. <https://doi.org/10.1016/j.aap.2020.105726>.
- Kuhn, M., and H. Wickham. 2020. "Tidymodels: A Collection of Packages for Modeling and Machine Learning Using Tidyverse Principles." Computer software. 2020. <https://www.tidymodels.org>.
- Lee, A. E., S. Underwood, and S. Handy. 2015. "Crashes and Other Safety-Related Incidents in the Formation of Attitudes toward Bicycling." *Transportation Research Part F: Traffic Psychology and Behaviour* 28:14–24. <https://doi.org/10.1016/j.trf.2014.11.001>.