Building and Evaluating a Surveillance System for Bicycle Crashes and Injuries

Building and Evaluating a Surveillance System for Bicycle Crashes and Injuries

A dissertation presented

by

Dahianna Serna Lopez

to

The Committee on Higher Degrees in Health Policy

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Health Policy

Harvard University

Cambridge, Massachusetts

April 2016

Building and Evaluating a Surveillance System for Bicycle Crashes and Injuries

Abstract

For cities aiming to create a useful surveillance system for bicycle injuries, a common challenge is that city crash reporting is scattered, faulty, or non-existent. In chapter 1, I document some of the lessons learned in helping the City of Boston to: 1) create a prototype for a comprehensive police crash database, 2) produce the city's first Cyclist Safety report, 3) make crash data available to the public, and 4) generate policy recommendations for both specific roadside improvements and for sustainable changes to the police department's crash reporting database. Some of the lessons include finding and using committed champions, prioritizing the use of existing data, creating opportunities to bridge divisions between stakeholders, partnering with local universities for assistance with advanced analytics, and using deliverables, such as a Cyclist Safety Report, to advocate for sustainability.

In chapter 2, given that the first step in the public health approach to injury prevention is to identify the problem (Krug et al, 2002), I examine whether police narrative reports cover the information that end-users need to do their part in preventing bicycle injuries. For example, civil engineers can use crash data to identify road conditions that need fixing, such as pavement defects and potholes. Urban planners can use reports to inform their design of the built environment, such as protected bicycle lanes and road diets. Health educators can use the data to plan campaigns. Lastly, police can use the data to determine where and when to focus their enforcement of traffic laws. I used a sample of narrative reports and filled in the fields in a government-recommended bicycle crash form aimed at understanding multiple factors about the crash. I used the percent of missing data across various domains, such as bicyclist information, environmental conditions, road conditions, and others, and found that that the reports did well in

crash typing. Examples of “crash types” are: motorist failed to yield, bicyclist lost control, and bicyclist ran a red light. The percent missingness in the crash-typing domain was, in general, lower and had more variation than percent missingness in other domains. Percent missingness for the crash-typing domain, for example, ranged from just over 40% to 75%. Other domains had little variation, such that missingness was generally over 75%. Police officers generally do not have professional training in road engineering or urban planning or public health and healthcare (which relate to the other domains in the recommended bicycle crash form). In addition, they are not compensated to collect that level of detail. Our results also show that there is less information (more missingness) when police officers take a statement from an involved party either in person or by phone versus when they are onsite. Given that there is a fair amount of missingness in narrative reports, I recommend adopting the Pedestrian Bicycle Crash Analysis Tool (PBCAT) and training officers to use it. The PBCAT software, developed by the US Government, is freely available to anyone and any police department for direct download.

In chapter 3, I identify factors related to a hit-and-run after a vehicle-bicycle collision. Understanding bicycle-vehicle collisions that result in hit-and-run (HAR) behavior is an important concern for law enforcement, public health, and affected individuals. If bicyclists are injured, this issue has implications for expedient access to medical care and for protection from the financial burden of associated medical costs. This study aimed to identify significant predictors of vehicle-bicycle HARs, the results of which can potentially inform preventive interventions for this type of injury and crime. I collected the data from Boston Police Department bicycle crash reports for 2009-2012. The data identified whether a crash was a hit-and-run and other predictor variables including road and bicyclist characteristics. The probability of a HAR was fit to selected variables through logistic regression models. Effects of the

predictors were reported as odds ratios. I found that the probability of a hit-and-run partially depends on time, day of the week, and whether the vehicle type was a taxi. I discuss implications for policies and interventions aimed at preventing this type of collision and crime.

Table of Contents

Acknowledgements

Professors. Dr. Joseph Newhouse: By admitting me to Harvard, you have helped me achieve my American Dream. Dr. David Hemenway, Dr. Mark Glickman, and Dr. Stephen Soumerai: Thank you very much for serving on my dissertation committee and guiding me through the process of conducting my research. ***Support Staff.*** Debbie Whitney, Ayres Heller, and Jessica Livingston: I will forever be indebted to you for your outstanding support in administrative and academic matters (and for your kindness and compassion during challenging times!). Dr. Sheila Thomas and Stephanie Parsons: Thank you for believing in me and for *truly* understanding the difficulties that underserved minorities face while in graduate school. ***Academic Heroes.*** Dr. Shira Mitchell: As a child, my heroine was "She-ra: Princess of Power." She was beautiful, inside and out; she was intelligent, strong, and magical. Little did I know that I would later meet her in the form of a mathematician/statistician at Harvard! Thank you for being so kind and patient with me. Thank you for being able to teach statistics concepts in English! I am incredibly lucky to have met you and I look forward to a lifelong friendship. Dr. Simo Goshev: Have you ever realized that your name contains the word "sigh" in it? You were my sigh of relief from the stress of the dissertation. Thank you for sharing your expertise in analysis and statistics. Thank you for treating me like a colleague and reinvigorating my passion for research. Thank you for being the cheerleader I needed as I was approaching the finish line. ***Lifesavers*****:** Melissa Jacobo: You understood me at the core and you made me stronger. Thank you for keeping me sane. I could not have done this without you. Cristina Contreras: During my years at Harvard, we overcame so many obstacles together. You were there for me; I was there for you… I cannot wait to see what comes next for us both!

I dedicate this doctoral dissertation to:

My mom, Carolina Boe
My dad, Leif Boe
My sister, Sissi Lopez
My brother-in-law, Julio Gonzalez
My nephew, Isaac Gonzalez
My niece, Soleil Gonzalez
&
My husband, Dr. Kenneth Frausto

When I doubted myself, your love gave me the strength to push forward.

Mom, thank you for showing me, through your courageous actions, that:
"In order to attain the impossible, one must attempt the absurd."
-Miguel de Cervantes

Sissi, thank you for all that you sacrificed to make sure I was *always* smiling.
You are my soul mate. I feel so blessed to have you as my sister.

Kenny, let's save the world together! You are my rock.
"And so today, my world it smiles, your hand in mine, we walk the miles. Thanks to you it will be done. For you to me are the only one." - Led Zeppelin

Chapter 1: Generating a city's first report on bicyclist safety: Lessons from the field

Introduction

The first step in the public health approach to reducing injury is to create a data system that practitioners can use to characterize the problem, indicate sensible policies, and evaluate those policies. For cities aiming to create a useful data system for bicycle injuries, a common challenge is that city crash reporting is typically scattered, outdated, or sometimes non-existent. A good crash database is comprehensive and captures relevant characteristics about the vehicles (including bicycles), the parties involved, the built environment, and the sequence of events before, during, and after the crash.[1] Although the United States Department of Transportation (USDOT) has a national template for police reporting, its use is voluntary and often not followed.[2] This is problematic for those charged with improving traffic safety, such as civil engineers, public health practitioners, and law enforcement personnel.

In order to design and implement cost-effective actions to prevent crashes, injuries, and fatalities, city officials need information about the problem, such as the magnitude, the geographic distribution, timing, and probable cause. Surveillance of such injury data is considered an essential component in developing effective injury prevention programs.[3] Unfortunately, creating an ideal data system can take many years. It would require funding, choosing appropriate software, training first responders to use it, and evaluating its effectiveness. So what can cities do in the short term? This was precisely what we, as researchers and participant-observers, asked ourselves as we took on the task of helping the City of Boston (Massachusetts, USA) to: 1) create a prototype for a comprehensive police crash dataset, 2) produce the city's first Cyclist Safety report, 3) make crash data available to the public, and 4) generate policy recommendations for specific roadside improvements and for sustainable changes to the police department's crash reporting database.

In this paper, we document some of the lessons learned in helping the City of Boston. We believe these lessons may be valuable to those working on improving bicycle safety and in reducing injuries generally at the local and state levels.

Method

We used a qualitative research method, commonly used in the field of anthropology, known as Participant Observation in which the researcher observes a group of people over an extended time and participates in their activities.[4] This method often entails building upon prior experience with a similar topic. The lead author had several years of experience working on bicycle and pedestrian safety research and policy in San Francisco, California, and this provided a strong foundation for knowing what types of information to collect and what type of activities in which to engage. Via intensive participation (from 2012 to 2014) with multiple city agencies and advocacy organizations in the City of Boston, she collected the necessary information for this qualitative study.

Lessons

1. Find and use committed champions.

Many organizations have an impact on bicycle safety. These include departments of transportation, urban planning, public health, water and power, and police, as well as private advocacy groups and community advisory committees. Relations among these groups and agencies can become tense. Some of the tension between and within agencies can relate to sharing data. Those who hold the data can be apprehensive about releasing the data because they fear "outsiders" will disregard their expertise or misuse the data.[5] Other reasons for not wanting to share data may just be personal or historical.[6] We found that local leaders acquainted with the political landscape were effective at managing these issues. As academics, any attempt by us to navigate the politics would be ineffective, requiring us to “take sides” rather than to remain impartial. What we could do was emphasize the need for agencies to work together to make the city safer.

We had two champions who helped our research team. Their role was to facilitate the sharing of the data. Both contributed valuable input at meetings and hearings, and advocated for safety. The first was a police captain who was highly invested in the issue and eager to serve as a champion. Uniformed officers and staff members in the police department esteemed him. Citywide agencies, advocates, and the media knew they could rely on him to provide straightforward answers to bicycle safety questions. A second champion was the Director of the Boston Cyclist Union, an advocacy organization focused on encouraging people to use the bicycle as an alternative form of active transportation and promoting bicycle safety.[7]

2. Gain buy-in from agency leaders.

Our champions persuaded agency leaders to stand united in their commitment to reducing bicycle fatalities and injuries. We found that although well-intentioned staff members from each agency (in our case police and transportation) were ready to take necessary actions to improve bicycle safety, they needed the green light from their leaders to move forward. Managing the aftermath of the Boston Marathon bombings was the top priority at the time of this study and the funding and human resources needed to focus on bicycle safety was in a holding pattern. Once the city returned to everyday operations, there was a window of opportunity for prioritizing the bicycle safety issue. The gruesome death of a student who was killed by a large truck while cycling to school served as the impetus for the leadership to allow the team to move forward with the project.

3. Identify existing data.

Examining the police data, we found that there were many key pieces of information that we could gain by merely cleaning up the dataset. For example, while forced-choice (e.g., drop-down or tick box) fields provided some basic information like the names of the parties involved, vehicle types, date and time occurrence, much of the key information was collected in free-text boxes. For example, under the variable “lighting,” the text field was open, which led to more than 160 descriptions of whether it was light or dark. While some wrote “sunny,” others wrote “clear skies” or “no clouds.” We easily recoded these into a few categories. There were other variables, however, like the demographics of the vehicle operators that were unusable due to incoherencies in the database; for these we recommended changes to the database.

4. Ensure that all stakeholders contribute their ideas.

Stakeholders are the ultimate experts, and have been shown to play an important role in the development of effective community-based injury prevention programs.[8] Stakeholder input helped us select the key information from the current data and promote the inclusion of new variables for the future system. We had one-on-one meetings with those who would ultimately collect and use the data, including analysts from road engineering, public health, and law enforcement. We gave them a list of the currently available data elements and the ones in the Pedestrian and Bicycle Crash Analysis Tool (PBCAT),[9] an online tool recommended by the United States Department of Transportation (USDOT) for documenting bicycle and pedestrian crashes, and over a series of working meetings came to a consensus about the most important data elements. They added some elements that were not on PBCAT, but were Boston-specific.

However, as emphasized by our police officer champion, the reporting form needed to be short, simple, and yet meaningful. While more questions about the crash would provide more information, from a practical standpoint, the police officers might not have enough time or inclination to answer everything.

5. Create an opportunity to bridge divisions.

A crucial first step for the bicycle safety advocates was for them to understand the limitations of the existing police crash database. They had assumed that good data were available, but that no one in the police department thought it important enough to analyze them. This assumption created a divide between them and the police. The reality was that the police department had collected a plethora of data, but they were not in a form that could easily be analyzed.

In an effort to bridge this divide, we invited the members of the Boston Cyclist Union, the city's bicycle safety advocacy group, to read the current police reports and to code them against the PBCAT data elements. To address privacy concerns, we redacted police reports with the help of a computer scientist at our university and had participants sign a confidentiality form. After reading the reports, many volunteers stated they were not aware of the scarcity of data. This exercise helped shift the focus away from a divisive relationship and more on the need to build a better data system together. The 30 participants provided feedback to our group about the types of elements that they would like to see based on their experiences in cycling in Boston.

6. Partner with the local university.

A major challenge for Public Health in the 21st century is that its workforce needs more training to reach the level of expertise required to meet surveillance demands.[10] For our project, there were 1800 police reports with personal identifiers. In line with any other research study involving personal information, we needed to remove these to comply with federal privacy regulations and to gain IRB approval. This task could have been carried out by a couple of people over the course of several months, but the window of opportunity would close when another important topic stole the spotlight. To help address the challenge of efficiency, our University was willing to offer the pro bono service of a professional computer scientist. This was because a PhD student, advised by an injury prevention professor, was serving as the lead analyst for the report and would be using the data to complete her dissertation. In addition, the Boston Area Research Initiative, a public policy institute based at the university provided fellowship funds to the student. Ultimately, the computer scientist and PhD student were able to finish the task of de-identifying the data within one week. They were also able to prepare the final dataset for further analysis and for eventual mapping. Partnerships with universities can be especially helpful in communities with limited computational capacity.[11]

7. Work with local graduate students (who have degree requirements to fulfill).

Students are often an untapped human resource. We needed people to help mine and clean the data. Police Department funding was not available for full-time analysts to handle this type of work. Through unprecedented collaboration between the bicycle advocacy organization and the police department, students in an urban planning program received an opportunity to participate in the research process and handle police data as long as they passed a background check and signed a confidentiality form. These students were an outstanding source of competent labor. They 1) had some level of passion for discovery, 2) had technical skills that were more cutting edge than those of the staffers in the police department (e.g., coding and advanced statistical knowledge), 3) wanted to improve their resumes, 4) perceived volunteering as a form of job exploration, and 5) were willing to work for free. However, these students needed a coordinator who could mentor them through the process and serve as a leader. In this case, the person overseeing them was the PhD student on the project. In addition, the head of the Division of Research at the Police Department agreed to be a reference for students in the future.

8. Use deliverables to advocate for funding sustainability.

Similar to a study by Laraque at al. (1995) on preventing injuries in an urban setting, our outcome data took the form of project deliverables (ie. a completed report, publicly available dataset, and interactive online map) that allowed us to both inform stakeholders about the effectiveness of our efforts and advocate for resources that could keep the project going.[12] Prior to the release of the report, NHTSA's federal money for maintaining traffic records was being allocated to other cities in the State of Massachusetts that were able to provide data to show the need for funds to implement highway safety interventions. The City of Boston had not been providing crash data, which made them ineligible to receive funds from the federal government for highway safety improvements. This resulted in the predicament such that without the funds, they could not generate adequate crash data, and without good data, they could not acquire funds for either maintaining their crash records or for implementing highway safety interventions. Shortly after the City released the report, the leadership of the BPD was able to begin negotiations with the National Highway Traffic Safety Administration (NHTSA) to gain federal funds for implementing a new database.

Report Findings and Recommendations

Over a four-year period (2009-2012), there were 1,813 bicycle crashes reported to the Boston police. The number of crashes remained steady over those years, although the City reported an increase in the number of bicyclists over the same four-year period, indicating that the rate of crashes and injuries per bicyclist fell over the time period.[13]

Ninety percent of the bicycle crashes involved a vehicle; the rest were falls from the bicycle. Nearly 80% of cyclists were injured. Men between the ages of 18 and 30 accounted for almost half of the cyclist injuries. The age-adjusted bicycle injury rate in 2012 was 49 per 100,000. This was triple the national rate reported by the National Highway Traffic Safety Administration (NHTSA) in 2012 (i.e., 16.7 per 100,000 population),[14] but 17% lower than that of San Francisco, a city regarded as a fair comparison due to the size of the population and land area and the strong presence of a bicycle culture.

Since Boston has four seasons and its residents take advantage of the summer months for outdoor activities, it was not surprising to find that over half of the crashes occurred in the summer. Crashes spiked during rush hour, particularly between 4:00 PM and 7:00 PM when one third of the crashes occurred.

We found that approximately 60% of the crashes happened at intersections; we were able to identify intersections with high crash numbers using mapping software. Many of the crashes happened along corridors adjacent to major universities. We were also able to find some crash patterns that would not have emerged without the concurrent use of the mapping software and the police narratives. For example, when we saw a cluster of crashes on the map, we zoomed in to read the narratives for those clusters. One important finding from this exercise was that along a small segment of one street, 9 cyclists had their wheels stuck in the trolley tracks, fell and were injured severely enough to be transported to a hospital. Using the narrative reports, we found that bicyclists going against traffic or failing to stop at a red light contributed to many crashes, as did drivers not seeing the bicyclist, especially at night. Finally, a surprising number of the cyclists who crashed were “doored” by passengers emerging from taxicabs.

An examination of the data indicated ways in which the data system needed to be improved. The police narratives provided some clues about driver and bicyclist behavior, but a limitation was that the narrative structure was not standardized and police officers made individual choices as to what to include in the free text box. We provided a detailed template for improvements in data collection. For example, we described how some variables needed to remain open-ended while others needed to be forced-choice. We also urged the City to create an interoperable system in which computer networks from multiple agencies could share data (e.g., Police and Ambulance). Lastly, based on our recommendation, the City made the bicycle crash data available to the public. Users could download the data or view it on an interactive map.

The City's Response

We believe our major contribution was the creation of a tangible document that helped make the case for the need to improve bicycle safety and to prioritize interventions in Boston. The City's response to the report was positive and the findings formed the basis for many new city policies. Shortly after the report releasing the report, outgoing Boston Mayor Thomas M. Menino pledged to decrease the injury rate resulting from bicycle crashes by 50% by 2020.[15]

The City planned to produce public service announcements to play inside of taxis. In fact, to remind passengers to watch for cyclists while exiting, the City furnished over 1,800 taxis with stickers.[16] Given the spike in crashes during rush hour, the City planned to collaborate with companies and businesses to educate employees commuting to work by bicycle about road safety. This plan also included cyclist education, via social media, about how to increase their visibility.[17] The City also sought federal assistance to fund an improved surveillance system by collaborating with Massachusetts Department of Transportation and the National Highway Safety administration. This was key because of its potential for sustainability.

The Boston Police Department (BPD) issued citations and handed out free helmets and blinking bicycle lights in areas with historically high numbers of crashes.[18] Police targeted university areas where high number of crashes involving students historically occurred. In terms of surveillance, the BPD planned to implement software that would capture more details about crashes.

The Boston Transportation Department (BTD) also used the map to complement their ongoing efforts to identify priority intersections and corridors that they needed redesign or retrofit. The Transportation Department added pavement markings at trolley tracks where so many cyclists had crashed when their wheels lodged in the tracks.

Changes have continued since releasing the report in early 2013. Mayor Martin J. Walsh, elected in late 2013, hired a Director of Active Transportation in 2015 to take charge of improving convenience and safety for cyclists and pedestrians.[19] The Boston Police Department hired a transportation safety analyst and adopted Vision Zero, an initiative to take necessary measures to prevent severe injuries and fatalities on the roads.[20,21] The City is working on publishing an updated version with better and more integrated data. In addition, as of February of 2016, nearly three years after its public release, the crash dataset has been viewed over 2,500 times.

Conclusion

Providing an initial report on bicycle crashes in Boston was an important step in reducing bicycle injuries. It helped illuminate specific problem areas (e.g. taxis, certain transit tracks), showed the value of a data system, and provided a blueprint for an even better system. When we talk about evidence-based practice in public health, we often refer to scientific studies. Nevertheless, without good data, good studies are not possible. Building a useful surveillance system depends in no small part on the wise use of advocacy, group dynamics, and politics. Our hope is that the lessons learned from our experience in Boston can help others do even better.

Reference - Chapter 1

1. Bigham, J., & Kang, S. (2013, July). Building a highway linear referencing system from preexisting reference marker measurements for transportation data management. In *URISA* (Vol. 25, No. 1).

2. Governors Highway Safety Association & National Highway Traffic Safety Administration. (2015, July). *Mapping to MMUCC: A process for comparing police crash reports and state crash databases to the Model Minimum Uniform Crash Criteria*. (Report No. DOT HS 812 184). Washington, DC: National Highway Traffic Safety Administration.

3. Mackenzie, S. G., & Pless, I. B. (1999). CHIRPP: Canada's principal injury surveillance program. *Injury prevention*, *5*(3), 208-213.

4. DeWalt, K. M., & DeWalt, B. R. (2010). *Participant observation: A guide for fieldworkers*. Rowman Altamira.

5. Buehler, J. W. (2012). CDC's vision for public health surveillance in the 21st century. Introduction. *Morbidity and mortality weekly report. Surveillance summaries (Washington, DC: 2002)*, *61*, 1-2.

6. Buehler, J. W. (2012). CDC's vision for public health surveillance in the 21st century. Introduction. *Morbidity and mortality weekly report. Surveillance summaries (Washington, DC: 2002)*, *61*, 1-2.

7. Mission, Vision, and Values. (n.d.). Retrieved November 15, 2015, from http://bostoncyclistsunion.org/about_us/mission-vision-and-values/

8. Klassen, T. P., MacKay, J. M., Moher, D., Walker, A., & Jones, A. L. (2000). Community-based injury prevention interventions. *The Future of Children*, 83-110.

9. PBCAT-Pedestrian and Bicycle Crash Analysis Tool. (n.d.). Retrieved December 27, 2015, from http://www.fhwa.dot.gov/publications/research/safety/pedbike/06090/index.cfm

10. Buehler, J. W. (2012). CDC's vision for public health surveillance in the 21st century. Introduction. *Morbidity and mortality weekly report. Surveillance summaries (Washington, DC: 2002)*, *61*, 1-2.

11. Spoth, R., Guyll, M., Lillehoj, C. J., Redmond, C., & Greenberg, M. (2007). Prosper study of evidence-based intervention implementation quality by community–university partnerships. *Journal of community psychology*, *35*(8), 981-999.

12. Laraque, D., Barlow, B., Durkin, M., & Heagarty, M. (1995). Injury prevention in an urban setting: challenges and successes. *Bulletin of the New York Academy of Medicine*, *72*(1), 16.

13. State of the Hub: Boston Bikes 2012 Update. (n.d.). Retrieved December 17, 2015, from https://drive.google.com/file/d/0B5pdqTMhDgHPYkdiZ0hvUUR6djA/view?pref=2

14. National Highway Traffic Safety Administration. (2015). Traffic Safety Facts, 2012 Data: Pedestrians. *Annals of Emergency Medicine*, *65*(4), 452.

15. State of the Hub: Boston Bikes 2012 Update. (n.d.). Retrieved December 17, 2015, from https://drive.google.com/file/d/0B5pdqTMhDgHPYkdiZ0hvUUR6djA/view?pref=2

16. Powers, M. (2013, May 15). With crash data, Boston tries to make bicycling safer. Retrieved December 27, 2015, from https://www.bostonglobe.com/metro/2013/05/14/with-crash-data-boston-tries-make-bicycling-safer/41aQVONIQd8eiiZr2HxaBO/story.html

17. State of the Hub: Boston Bikes 2012 Update. (n.d.). Retrieved December 17, 2015, from https://drive.google.com/file/d/0B5pdqTMhDgHPYkdiZ0hvUUR6djA/view?pref=2

18. Powers, M. (2013, May 15). With crash data, Boston tries to make bicycling safer. Retrieved December 27, 2015, from https://www.bostonglobe.com/metro/2013/05/14/with-crash-data-boston-tries-make-bicycling-safer/41aQVONIQd8eiiZr2HxaBO/story.html

19. Cityofboston.gov - Official Web Site of the City of Boston. (n.d.). Retrieved December 15, 2015, from http://www.cityofboston.gov/news/default.aspx?id=20268

20. Cityofboston.gov - Official Web Site of the City of Boston. (n.d.). Retrieved December 10, 2015, from http://www.cityofboston.gov/news/Default.aspx?id=20044

21. Vision Zero | Just another Business Sweden Sites site. (n.d.). Retrieved December 10, 2015, from http://www.visionzeroinitiative.com/

This page intentionally left blank

Chapter 2: Police narrative reports: Do they cover the information that end-users need to do their part in preventing bicycle injuries?

Introduction

Despite the widely known benefits of cycling to health and the environment, increased bicycle use has not been without adverse consequences. Using a national sample of police crash data the National Highway Traffic Safety Administration estimates that in the United States, 49,000 bicyclists were injured and 726 killed in 2012 as a result of a motor vehicle crash.[1] In that same year, the police department in City of Boston, Massachusetts, reported 493 bicycle crashes of which 489 bicyclists were injured and 4 were killed.[2]

The first step in the public health approach to injury prevention is to identify the problem.[3] A good surveillance system for tracking cyclist-involved crashes can assist those charged with preventing bicycle crashes to take that first step and to then to use the information to intervene according to their scope of practice.[4] For example, civil engineers can use crash data to identify road conditions that need fixing, such as pavement defects and potholes. Urban planners can use reports to inform their design of the built environment, such as protected bicycle lanes and road diets. Health educators can use the data to plan campaigns. And lastly, police can use the data to determine where and when to focus their enforcement of traffic laws. Good surveillance data can provide these “end-users” important information for preventing injuries.[5,6]

The Boston Police Department uses its own crash report form that does not map directly to any crash forms recommended by the US Government, such as the *Model Minimum Uniform Crash Criteria* (MMUCC)[7] or the more bicycle-specific *Pedestrian and Bicycle Analysis Tool* (PBCAT).[8] The latter is comprised of a variety of questions that can be can be categorized into nine domains. One domain, for example, is Bicycle and Facility, which includes the following forced-choice variables: Bicycle Type, Bicycle Defects, Curb Length Width, and Bike Lane Width. The other seven domains are Bicyclist Information, Driver Information, Vehicle Information, Area Characteristics, Roadway Features, Contributing Factors, and Crash Typing. Answers to the questions in these domains could be helpful to end-users.

Boston police crash reports contain a small number of forced-choice variables, such as date, location, and time of occurrence and rely heavily on police officers' accounts of the incident via narrative text. While the former are valuable in identifying the problem and quantifying its magnitude, they are not rich enough to inform specific injury prevention interventions. We use information from the police narratives to fill out the PBCAT form and then use percent of PBCAT variables missing as a proxy for level of detail, such that less missingness suggests more detail. The aims of this study are to examine the extent to which narrative texts cover the information that end-users need in order to do their part in preventing bicycle injuries. In other words, we assess whether average missingness differs by domain across all reports. Because some reports were not filed at the scene, we examine if the percentage of missingness differs between domains when the report was filed by phone/walk-in or at the scene.

Method

Study Design and Data Source

We conducted a 4-year retrospective cohort study of vehicle-vs-bicycle crashes that occurred between January 1, 2009 and December 31, 2012 in Boston, Massachusetts (USA). The data came from the police reports supplied to us by the Boston Police Department (BPD). We dropped cases in which a vehicle served as a weapon to injure a cyclist, as we were only interested in cases that we presumed to be unintentional. The Committee on the Use of Human

Subjects (CUHS), Harvard University's Institutional Review Board, granted permission to conduct this research study.

Variables

Independent Variables. We had 11 categorical independent variables which came from the forced-choice portions of the police report: 1) temperature (0-31 °F**;** 32-59 °F; 60+ °F); 2) rain (yes or no); 3) day of the week (weekend or weekday); 4) time of day (day or night); 5) if the crash occurred at an intersection (yes or no); 6) whether it happened on a main street (yes or no); 5) if the bicyclist was male or female; 6) the cyclist's ethnicity (Black; Hispanic; White; Other); 7) age (0-20 yrs.; 21-40 yrs.; 41-60 yrs.; 60+ yrs.); 8) source (whether the report was taken by phone, by a physical visit to the police department, or at the scene); 9) if the cause of crash involved "dooring" or not; 10) if the motor vehicle was a taxi or not; and 11) if the bicyclists was injured or not. Four variables had missing data, including ethnicity and the bicyclists' gender, whether the bicyclist was injured, and whether the collision occurred on a main street. To account for missing data on these variables in our models, we created an extra category indicating that the information police officers did not enter in these independent forced-choice fields (e.g., gender was recoded as "female", "male" and "missing").

Dependent Variables. To generate the dependent variable, which was percent missingness, 30 individuals read a random sample of the BPD police reports and used the information to fill in the fields in the aforementioned Pedestrian and Bicycle Crash Analysis Tool. We calculated the mean number of empty variables for each report in general and divided by the total number of variables (across all domains) to express overall missingness as a percentage. We also calculated the number of empty variables within each domain and divided by the number of variables within that domain, allowing us to also express missingness as a percentage for each domain. In the example given above, if the "Bicycle Type" and "Bicycle Defects" variables were missing and the other ones in the Bicycle and Facility domain were not, the percent missing in that domain would be 2/4 or 0.5. Each report essentially had nine measures of missingness—each corresponding to one of the nine domains.

Analytic Approach

We fit a linear regression of percent missingness on the explanatory variables to understand their effects and significance. We tested for collinearity among the covariates using variance inflation factors (VIFs). Based on recommendations by Kutner et al (2004), we considered variables with a VIF greater than 10 for removal.[9] We conducted pairwise contrasts between the marginal means of domains and accounted for errors using sequential Bonferroni corrections. We fit a second model with an added an interaction of domain-by-source. These were expressed as percentage point differences. We used an alpha level of 0.05. We used R for our statistical analyses (R Development Core Team, 2013).[10]

Results

Descriptive

The total number of bicycle-related crashes reported to the BPD during the four-year period was 1806. After removing cases with no motor vehicle involvement (i.e., in which the bicyclist fell, hit another bicyclist, or hit a pedestrian), a total of 1646 vehicle-bicycle crashes remained. For the current study, we took a random sample of reports (n=760 or 46% of the larger sample) to minimize coding fatigue among the readers.

See Table 2.1 for descriptive statistics of all covariates. Three quarters (75.4%; n=572) of the bicyclists were male and 56.5% (n= 429) were between the ages of 21-40. Exactly 612 (80.6%) of reported crashes resulted in an injury to the bicyclist. Of the 759 cases, 3.9% (n = 64) had missing data on whether the crash occurred on a main street. Six percent (n= 98) had missing data on injury status. Less than half of a percent (n = 8) of cases had missing data on gender. The variance of percent missingness is not evident when reports are aggregated. When broken down by domain (see Table 2.2 for domain descriptions), the differences in variance are evident between domains and (see Table 2.3).

Table 2.1. *Descriptive statistics with mean and standard deviation of percent missingness across all PBCAT domains by forced-choice independent variables*

Variable	n (%)	All (SD)
Source		
Onsite	618 (81.4)	0.75 (0.07)
Phone or Walk-in	95 (12.5)	0.77 (0.08)
Missing	46 (6.1)	0.77 (0.06)
Cyclist Injury		
Yes	612 (80.6)	0.76 (0.07)
No	100 (13.2)	0.75 (0.09)
Missing	47 (6.2)	0.76 (0.06)
Extended Door		
Yes	83 (10.9)	0.74 (0.07)
No	676 (89.1)	0.76 (0.07)
Intersection		
Yes	441 (58.1)	0.76 (0.07)
No	318 (41.9)	0.75 (0.07)
Time of Day		
Night	309 (40.7)	0.75 (0.07)
Day	450 (59.3)	0.76 (0.07)
Taxi		
Yes	73 (9.6)	0.75 (0.08)
No	686 (90.4)	0.76 (0.07)
Day		
Weekday	592 (78.0)	0.76 (0.07)
Weekend	167 (22.0)	0.75 (0.07)
Precipitation		
Yes	287 (37.8)	0.76 (0.07)
No	472 (65.2)	0.75 (0.07)
Main Street		
Yes	261 (34.4)	0.75 (0.07)
No	468 (61.7)	0.76 (0.07)
Missing	30 (4.0)	0.77 (0.05)
Cyclist Gender		
Female	179 (23.6)	0.76 (0.07)
Male	572 (75.4)	0.75 (0.07)
Missing	8 (1.10)	0.73 (0.08)
Cyclist Age		
0-20	137 (18.1)	0.75 (0.07)
21-40	429 (56.5)	0.76 (0.07)
41-60	144 (19.0)	0.76 (0.07)
61-80	14 (1.8)	0.75 (0.08)
Missing	35 (4.6)	0.76 (0.08)

Table 2.2. *Domain description*

Domain	Description
Domain 1 (or "D1)	Bicyclist
Domain 2 (or "D2")	Bicycle Information
Domain 3 (or "D3)	Environmental Conditions
Domain 4 (or "D4)	Contributing Factors
Domain 5 (or "D5")	Crash Typing
Domain 6 (or "D6")	Roadway Features
Domain 7 (or "D7")	Area Characteristics
Domain 8 (or "D7")	Vehicle Information

Table 2.3. *Mean and SD of % missingness for PBCAT domains (1 - 5) by variables*

	Domain 1		Domain 2		Domain 3		Domain 4		Domain 5	
Source										
Onsite	0.68	0.17	0.74	0.16	1.00	0.03	0.92	0.13	0.12	0.22
Phone/Walk-in	0.70	0.17	0.75	0.16	1.00	0.00	0.93	0.12	0.19	0.25
Missing	0.68	0.22	0.76	0.15	1.00	0.00	0.90	0.14	0.15	0.30
Cyclist Injury										
Yes	0.68	0.18	0.74	0.16	1.00	0.02	0.92	0.13	0.13	0.23
No	0.69	0.17	0.75	0.16	0.99	0.05	0.90	0.14	0.14	0.25
Missing	0.68	0.19	0.77	0.14	1.00	0.00	0.90	0.13	0.14	0.23
Doored										
Yes	0.69	0.16	0.72	0.17	1.00	0.04	0.88	0.13	0.06	0.14
No	0.68	0.18	0.75	0.16	1.00	0.03	0.92	0.13	0.14	0.24
Intersection										
Yes	0.69	0.18	0.75	0.16	1.00	0.02	0.93	0.12	0.13	0.23
No	0.67	0.17	0.73	0.15	1.00	0.03	0.91	0.14	0.13	0.23
Time of Day										
Night	0.69	0.18	0.76	0.15	1.00	0.00	0.92	0.12	0.14	0.24
Day	0.68	0.17	0.73	0.16	1.00	0.03	0.91	0.14	0.12	0.22
Taxi										
Yes	0.67	0.16	0.74	0.14	1.00	0.04	0.89	0.15	0.15	0.26
No	0.68	0.18	0.74	0.16	1.00	0.03	0.92	0.13	0.13	0.23
Day										
Weekday	0.68	0.18	0.74	0.16	1.00	0.03	0.92	0.13	0.13	0.23
Weekend	0.69	0.17	0.75	0.16	1.00	0.00	0.92	0.13	0.13	0.23
Rain										
Yes	0.69	0.16	0.75	0.15	1.00	0.04	0.91	0.14	0.13	0.22
No	0.68	0.18	0.74	0.16	1.00	0.02	0.92	0.13	0.13	0.23
Main Street										
Yes	0.67	0.19	0.75	0.15	1.00	0.04	0.92	0.13	0.12	0.23
No	0.69	0.17	0.74	0.16	1.00	0.02	0.92	0.13	0.13	0.23
Missing	0.71	0.17	0.76	0.15	1.00	0.00	0.91	0.15	0.17	0.25
Cyclist Gender										
Female	0.68	0.18	0.74	0.16	1.00	0.00	0.91	0.15	0.15	0.25
Male	0.68	0.17	0.74	0.16	1.00	0.03	0.92	0.12	0.12	0.22
Missing	0.63	0.22	0.78	0.16	1.00	0.00	0.86	0.17	0.00	0.00
Cyclist Age										
0-20	0.68	0.16	0.74	0.16	1.00	0.00	0.90	0.15	0.11	0.21
21-40	0.69	0.17	0.74	0.16	1.00	0.03	0.92	0.13	0.13	0.22
41-60	0.68	0.17	0.74	0.15	1.00	0.03	0.92	0.12	0.14	0.25
61-80	0.69	0.21	0.73	0.21	0.98	0.09	0.91	0.12	0.17	0.25
Missing	0.67	0.23	0.77	0.15	1.00	0.00	0.94	0.11	0.15	0.30

Table 2.3. *Mean and SD of % missingness for PBCAT domains (6 -9) by variables (continued)*

	Domain 6		Domain 7		Domain 8		Domain 9	
Source								
Onsite	0.81	0.11	0.71	0.25	0.87	0.16	0.99	0.07
Phone/Walk-in	0.82	0.11	0.76	0.25	0.87	0.17	1.00	0.00
Missing	0.83	0.06	0.82	0.24	0.91	0.12	1.00	0.00
Cyclist Injury								
Yes	0.82	0.10	0.72	0.25	0.88	0.16	0.99	0.07
No	0.81	0.13	0.74	0.26	0.86	0.18	1.00	0.05
Missing	0.83	0.10	0.76	0.25	0.87	0.14	0.99	0.07
Doored								
Yes	0.83	0.09	0.76	0.25	0.80	0.23	0.99	0.05
No	0.81	0.11	0.72	0.25	0.88	0.14	0.99	0.07
Intersection								
Yes	0.81	0.11	0.72	0.25	0.88	0.14	0.99	0.06
No	0.82	0.10	0.74	0.25	0.87	0.18	0.99	0.07
Time of Day								
Night	0.81	0.11	0.75	0.25	0.87	0.17	0.99	0.06
Day	0.82	0.10	0.71	0.25	0.87	0.15	0.99	0.07
Taxi								
Yes	0.81	0.11	0.75	0.25	0.86	0.17	0.99	0.08
No	0.82	0.10	0.72	0.25	0.88	0.16	0.99	0.06
Day								
Weekday	0.82	0.10	0.72	0.25	0.88	0.15	0.99	0.07
Weekend	0.80	0.11	0.76	0.26	0.86	0.17	0.99	0.05
Rain								
Yes	0.82	0.11	0.74	0.25	0.87	0.16	0.99	0.08
No	0.81	0.11	0.72	0.25	0.87	0.16	0.99	0.05
Main Street								
Yes	0.81	0.11	0.73	0.25	0.87	0.17	1.00	0.03
No	0.82	0.11	0.73	0.25	0.87	0.15	0.99	0.08
Missing	0.85	0.06	0.70	0.25	0.85	0.12	0.98	0.09
Cyclist Gender								
Female	0.82	0.10	0.72	0.25	0.88	0.14	0.99	0.06
Male	0.81	0.11	0.73	0.25	0.87	0.16	0.99	0.06
Missing	0.81	0.03	0.81	0.26	0.84	0.19	0.94	0.18

Table 2.3. *Mean and SD of % missingness for PBCAT domains (6 -9) by variables (continued)*

	Domain 6		**Domain 7**		**Domain 8**		**Domain 9**	
Cyclist Age								
0-20	0.82	0.11	0.70	0.25	0.88	0.15	0.99	0.08
21-40	0.82	0.10	0.72	0.25	0.87	0.16	0.99	0.05
41-60	0.82	0.12	0.76	0.25	0.88	0.16	0.99	0.07
61-80	0.80	0.14	0.79	0.26	0.84	0.16	1.00	0.00
Missing	0.79	0.12	0.73	0.25	0.89	0.15	0.99	0.08

Inferential

The VIFs for the collection of variables in the first linear model ranged between 1.02 and 1.37 and thus there was no evidence of collinearity among the covariates. Table 2.4 reports our first linear model. Holding all other variables constant, domain 5 (crash typing) had the largest explanatory magnitude. For every unit increase in domain 5, the percent missingness *decreased* by 56 (-0.56; 95% CI: -0.57, -0.54) percentage points when compared to domain 1 (reference domain). Domain 9 (driver information) had the second largest explanatory magnitude, but in the opposite direction, such that for every unit increase in domain 9, the percent missingness *increased* by 31 (0.31; 95% CI: 0.29, 0.32) percentage points when compared to domain 1, after holding all other variables constant. Other variables with a significant effect on percent missingness were day and dooring, but these are likely not practically significant. For example, holding all other variables constant, the percent missingness for the independent variable day was lower by 1 (-0.01; 95% CI: -0.57, -0.54) percentage point when compared to night. In addition, holding all other variables constant, the percent missingness when dooring occurred was lower by 2 (-0.02; 95% CI: -0.03, -0.01) percentage points when compared to when dooring did not occur.

Table 2.4. *Linear Model 1 of Percent Missingness (Note: The reference group for all domain levels is Domain 1; Bicyclist Information)*

Variable	Est.	95% CI		Std. Error	t value	Pr(>|t|)	Sig.
(Intercept)	0.70	0.67	0.72	0.01	61.49	0.00	***
Day (Ref = Night)	-0.01	-0.02	-0.00	0.00	-2.38	0.02	*
Weekday (Ref = Weekend)	-0.00	-0.01	0.01	0.00	-0.65	0.51	
Rain (Ref = No Rain)	0.00	-0.01	0.01	0.00	0.39	0.70	
Main Street (Ref = Not Main Street)	-0.00	-0.01	0.01	0.00	-0.65	0.51	
Main Street Missing (Ref = Not Main Street)	0.00	-0.02	0.02	0.01	0.25	0.80	
Intersection (Ref = No Intersection)	0.00	-0.00	0.01	0.00	0.86	0.39	
Taxi (Ref = Not Taxi)	-0.00	-0.02	0.01	0.01	-0.35	0.73	
Doored (Ref = Not Doored)	-0.02	-0.03	-0.01	0.01	-2.83	0.00	**
Female (Ref = Male)	0.00	-0.01	0.01	0.00	0.75	0.45	
Gender Missing (Ref = Male)	-0.03	-0.08	0.01	0.02	-1.54	0.12	
Age 21 - 40 (Ref = Age 0 – 20)	0.01	-0.00	0.02	0.01	1.14	0.25	
Age 41 - 60 (Ref = Age 0 – 20)	0.01	-0.00	0.02	0.01	1.91	0.06	
Age 61 - 80 (Ref = Age 0 – 20)	0.01	-0.02	0.04	0.02	0.64	0.52	
Age Missing (Ref = Age 0 – 20)	0.02	-0.00	0.04	0.01	1.83	0.07	
Injured (Ref = Not injured)	0.00	-0.01	0.02	0.01	0.70	0.48	
Injured Missing (Ref = Not injured)	0.01	-0.01	0.03	0.01	0.67	0.50	
Source Onsite (Ref = Phone/Walk-in)	-0.02	-0.03	-0.01	0.01	-3.26	0.00	**
Source Missing (Ref = Phone/Walk-in)	0.00	-0.02	0.02	0.01	0.33	0.74	
Domain 2 = Bicycle Information	0.06	0.04	0.08	0.01	7.26	0.00	***
Domain 3 = Environmental Conditions	0.31	0.30	0.33	0.01	38.44	0.00	***
Domain 4 = Contributing Factors	0.23	0.22	0.25	0.01	28.53	0.00	***
Domain 5 = Crash Typing	-0.56	-0.57	-0.54	0.01	-67.86	0.00	***
Domain 6 = Roadway Features	0.13	0.12	0.15	0.01	16.23	0.00	***
Domain 7 = Area Characteristics	0.04	0.03	0.06	0.01	5.13	0.00	***
Domain 8 = Vehicle Information	0.19	0.17	0.21	0.01	23.20	0.00	***
Domain 9 = Driver Information	0.31	0.29	0.32	0.01	37.66	0.00	***

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Adjusted R-squared: 0.7093

Table 2.5 presents the pairwise contrasts between the marginal means of domains. The purpose of making these comparisons is to examine whether there were differences between domains (whereas Model 1 only compared each domain to domain 1). We find that when compared to domain 5, all other domains differ by at least 61 percentage points and at most by 87 percentage points. For example, percent missingness in domain 3 is 61 percentage points greater than domain 5. Nonsignificant contrasts were: domain 2 versus domain 7 and domain 3 versus domain 9. Table 2.6 is the second linear model of percent missingness with a source by domain interaction. It serves to examine whether the source of the report has an effect on missingness by domain. Results show that the only significant interaction is that of domain 5 by source.

Table 2.5. *Pairwise contrasts between marginal means of domains*

Contrast	Estimate	SE	Df	T Ratio	P Value	Sig.
Domain 1 - Domain 2	-0.06	0.01	6804.00	-7.26	0.00	***
Domain 1 - Domain 3	-0.31	0.01	6804.00	-38.44	0.00	***
Domain 1 - Domain 4	-0.23	0.01	6804.00	-28.53	0.00	***
Domain 1 - Domain 5	0.56	0.01	6804.00	67.86	0.00	***
Domain 1 - Domain 6	-0.13	0.01	6804.00	-16.23	0.00	***
Domain1 - Domain 7	-0.04	0.01	6804.00	-5.13	0.00	***
Domain 1 - Domain 8	-0.19	0.01	6804.00	-23.20	0.00	***
Domain 1 - Domain 9	-0.31	0.01	6804.00	-37.66	0.00	***
Domain 2 - Domain 3	-0.26	0.01	6804.00	-31.17	0.00	***
Domain 2 - Domain 4	-0.17	0.01	6804.00	-21.27	0.00	***
Domain 2 - Domain 5	0.61	0.01	6804.00	75.12	0.00	***
Domain 2 - Domain 6	-0.07	0.01	6804.00	-8.97	0.00	***
Domain 2 - Domain 7	0.02	0.01	6804.00	2.13	0.07	
Domain 2 - Domain 8	-0.13	0.01	6804.00	-15.94	0.00	***
Domain 2 - Domain 9	-0.25	0.01	6804.00	-30.39	0.00	***
Domain 3 - Domain 4	0.08	0.01	6804.00	9.90	0.00	***
Domain 3 - Domain 5	0.87	0.01	6804.00	106.30	0.00	***
Domain 3 - Domain 6	0.18	0.01	6804.00	22.20	0.00	***
Domain 3 - Domain 7	0.27	0.01	6804.00	33.31	0.00	***
Domain 3 - Domain 8	0.12	0.01	6804.00	15.23	0.00	***
Domain 3 - Domain 9	0.01	0.01	6804.00	0.78	0.43	
Domain 4 - Domain 5	0.79	0.01	6804.00	96.39	0.00	***
Domain 4 - Domain 6	0.10	0.01	6804.00	12.30	0.00	***
Domain 4 - Domain 7	0.19	0.01	6804.00	23.40	0.00	***
Domain 4 - Domain 8	0.04	0.01	6804.00	5.33	0.00	***
Domain 4 - Domain 9	-0.07	0.01	6804.00	-9.12	0.00	***
Domain 5 - Domain 6	-0.69	0.01	6804.00	-84.09	0.00	***
Domain 5 - Domain 7	-0.60	0.01	6804.00	-72.99	0.00	***
Domain 5 - Domain 8	-0.75	0.01	6804.00	-91.06	0.00	***
Domain 5 - Domain 9	-0.86	0.01	6804.00	-105.52	0.00	***
Domain 6 - Domain 7	0.09	0.01	6804.00	11.10	0.00	***
Domain 6 - Domain 8	-0.06	0.01	6804.00	-6.97	0.00	***
Domain 6 - Domain 9	-0.18	0.01	6804.00	-21.42	0.00	***
Domain 7 - Domain 8	-0.15	0.01	6804.00	-18.07	0.00	***
Domain 7 - Domain 9	-0.27	0.01	6804.00	-32.53	0.00	***
Domain 8 - Domain 9	-0.12	0.01	6804.00	-14.45	0.00	***

Note: Used Sequential Bonferroni Correction Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Table 2.6. *Linear model 2 of percent missingness with a source x domain interaction*

Variable	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.69	0.02	37.94	0.00
Day (Ref = Night)	-0.01	0.00	-2.39	0.02
Weekday (Ref = Weekend)	-0.00	0.00	-0.65	0.51
Rain (Ref = No Rain)	0.00	0.00	0.39	0.70
Main Street (Ref = Not Main Street)	-0.00	0.00	-0.65	0.51
Main Street Missing (Ref = Not Main Street)	0.00	0.01	0.25	0.80
Intersection (Ref = No Intersection)	0.00	0.00	0.86	0.39
Taxi (Ref = Not Taxi)	-0.00	0.01	-0.35	0.73
Doored (Ref = Not Doored)	-0.02	0.01	-2.83	0.00
Female (Ref = Male)	0.00	0.00	0.75	0.45
Gender Missing (Ref = Male)	-0.03	0.02	-1.55	0.12
Age 21 - 40 (Ref = Age 0 – 20)	0.01	0.01	1.14	0.25
Age 41 - 60 (Ref = Age 0 – 20)	0.01	0.01	1.91	0.06
Age 61 - 80 (Ref = Age 0 – 20)	0.01	0.01	0.65	0.52
Age Missing (Ref = Age 0 – 20)	0.02	0.01	1.83	0.07
Injured (Ref = Not injured)	0.00	0.01	0.70	0.48
Injured Missing (Ref = Not injured)	0.01	0.01	0.67	0.50
Source = Onsite (Ref = Phone/Walk-in)	-0.02	0.02	-0.89	0.37
Source Missing (Ref = Phone/Walk-in)	-0.02	0.03	-0.74	0.46
Domain 2 = Bicycle Information	0.05	0.02	2.30	0.02
Domain 3 = Environmental Conditions	0.30	0.02	13.12	0.00
Domain 4 = Contributing Factors	0.23	0.02	9.93	0.00
Domain 5 = Crash Typing	-0.51	0.02	-21.88	0.00
Domain 6 = Roadway Features	0.12	0.02	5.39	0.00
Domain 7 = Area Characteristics	0.06	0.02	2.64	0.01
Domain 8 = Vehicle Information	0.17	0.02	7.42	0.00
Domain 9 = Driver Information	0.30	0.02	13.12	0.00

Significant: $p < 0.05$

Table 2.6. *Linear model 2 of percent missingness with a source x domain interaction (continued)*

Variable	Estimate	Std. Error	t value	Pr(>|t|)
Onsite x Domain 2	0.01	0.02	0.24	0.81
Missing Source x Domain 2	0.03	0.04	0.66	0.51
Onsite x Domain 3	0.01	0.02	0.50	0.61
Missing Source x Domain 3	0.02	0.04	0.53	0.60
Onsite x Domain 4	0.01	0.02	0.22	0.82
Missing Source x Domain 4	-0.01	0.04	-0.20	0.84
Onsite x Domain 5	-0.06	0.02	-2.41	0.02
Missing Source x Domain 5	-0.02	0.04	-0.45	0.65
Onsite x Domain 6	0.01	0.02	0.31	0.76
Missing Source x Domain 6	0.03	0.04	0.78	0.44
Onsite x Domain 7	-0.03	0.02	-1.17	0.24
Missing Source x Domain 7	0.08	0.04	1.94	0.05
Onsite x Domain 8	0.02	0.02	0.71	0.48
Missing Source x Domain 8	0.07	0.04	1.63	0.10
Onsite x Domain 9	0.00	0.02	0.19	0.85
Missing Source x Domain 9	0.02	0.04	0.53	0.60

Significant: $p < 0.05$

Discussion

While all social scientists analyze data, it is rare for them to evaluate the adequacy of the database from which they draw their data. In this study, we examined police reports of bike crashes from the Boston Police Database. Since the reports rely heavily on text narratives, we were interested in knowing whether what they wrote was relevant to those who were interested in using the data to inform their injury prevention interventions. We used a sample of narrative reports and filled in the fields in a government-recommended bicycle crash form aimed at understanding multiple factors about the crash. We used the percent of missing data across various domains, such as bicyclist information, environmental conditions, road conditions, and others, and found that that the reports did well in crash typing. Examples of "crash types" are: motorist failed to yield, bicyclist lost control, and bicyclist ran a red light. The percent missingness in the crash-typing domain was, in general, lower and had more variation than percent missingness in other domains. Percent missingness for the crash-typing domain, for example, ranged from just over 40% to 75%. Other domains had little variation, such that missingness was generally over 75%. Police officers generally do not have professional training in road engineering or urban planning or public health and healthcare (which relate to the other domains in the recommended bicycle crash form). In addition, they do not receive compensation for collecting that level of detail. This might explain why officers generally do not write much about the circumstances of a crash that could help, for example, engineers to fix a pothole that may have been the cause of a crash between a vehicle and a bicycle.

Our results show that there is less information (more missingness) when police officers take a statement from an involved party either in person or by phone versus when they are onsite. A potential explanation is that simply getting a story and having to write it down does not capture the essence of a crash, unless the person filing the report is an excellent communicator and the officer taking the report has the time to write down every piece of information. A report filed in person or over the phone might occur if at the time of the incident, the driver and cyclist agreed that all was fine and thus chose not to call the police. However, what if the cyclist decides to visit the police department later to report an injury with which he or she was diagnosed shortly after the crash (presumably after the rush of adrenaline had subsided)? Ideally, that report should not be different from a report taken at the scene (i.e., onsite) if the form the police officer used was standardized and bicycle-specific.

Injury prevention professionals (end-users of the data) need information that police officers collect routinely.[11] Knowing as much as possible about bicycle crashes (i.e., the parties involved, information about the bicycles and vehicles, environmental conditions, contributing factors, roadway features, and area characteristics) can assist injury prevention professionals to identify high-impact targeted interventions. A good surveillance system should not depend on narrative reports for important crash/injury prevention. Changing the way the data are collected is likely a better option for helping interested end-users. The only note of caution is that police officers may resist filling out a long form, even if the form only consists of forced-choice checkboxes or dropdown menus. While more questions about the crash would provide useful information, from a practical standpoint, police officers might not have enough time or inclination to answer everything.

Recommendations

Given that there is a fair amount of missingness in narrative reports, we recommend adopting the Pedestrian Bicycle Crash Analysis Tool (PBCAT) and training officers to use it. The PBCAT software, developed by the US Government, is freely available to anyone and any police department for direct download. It makes use of pre-drawn diagrams of various situations as well as drop down menus that allow police officers to quickly provide circumstantial information. As emphasized by one of the Captains at BPD, any changes to the current way of reporting bicycle crashes needs to be short, simple, and rich in content.

The fact that injury prevention professionals use administrative datasets from police departments does not mean that the "data collectors" (i.e., police officers) have training in road engineering, urban planning, or public health. Therefore, implementing trainings about the importance of detailed and thorough information gathering of bicycle (and pedestrian) crashes may make them feel more empowered to be involved in the process of preventing injuries, which may result in more information that everyone can use. A sustainable and effective surveillance system can identify when, where, and why interventions should be implemented. However, in order for it to be effective, it must be as detailed as possible without compromising police activities.

Reference – Chapter 2

1. National Highway Traffic Safety Administration. (2014). Traffic Safety Facts: Pedestrians; 2012 data. *Washington DC*.

2. City of Boston Crash Report 2013. (n.d.). Retrieved November 10, 2015, from http://www.cityofboston.gov/images_documents/Crash Report 2013 FINAL reduced 2_tcm3-38304.pdf

3. Krug, E. G., Mercy, J. A., Dahlberg, L. L., & Zwi, A. B. (2002). The world report on violence and health. *The lancet*, *360*(9339), 1083-1088.

4. Klassen, T. P., MacKay, J. M., Moher, D., Walker, A., & Jones, A. L. (2000). Community-based injury prevention interventions. *The Future of Children*, 83-110.

5. Ferrante, A. M., Rosman, D. L., & Knuiman, M. W. (1993). The construction of a road injury database. *Accident Analysis & Prevention*, *25*(6), 659-665.

6. Rosman, D. L., & Knuiman, M. W. (1994). A comparison of hospital and police road injury data. *Accident Analysis & Prevention*, *26*(2), 215-222.

7. Model Minimum Uniform Crash Criteria (n.d.) Retrieved December 28, 2015, from http://mmucc.us/

8. PBCAT-Pedestrian and Bicycle Crash Analysis Tool. (n.d.). Retrieved December 27, 2015, from http://www.fhwa.dot.gov/publications/research/safety/pedbike/06090/index.cfm
9. Kutner, M. H., Nachtsheim, C., & Neter, J. (2004). *Applied linear regression models*. McGraw-Hill/Irwin.

10. R: A language and environment for statistical computing. (2013). Retrieved from https://www.r-project.org/about.html

11. Ferrante, A. M., Rosman, D. L., & Knuiman, M. W. (1993). The construction of a road injury database. *Accident Analysis & Prevention*, *25*(6), 659-665.

This page intentionally left blank

Chapter 3: Identifying factors related to a hit-and-run after a vehicle-bicycle collision

Introduction

The National Highway Traffic Safety Administration estimates that in the United States, 49,000 bicyclists were injured and 726 killed in 2012 as a result of a motor vehicle collision.[1] In that same year, the City of Boston, Massachusetts, reported 493 bicycle collisions of which 489 bicyclists were injured and 4 were killed.[2] Like vehicle-vehicle and vehicle-pedestrian collisions, some fraction of vehicle-bicycle collisions appears to be hit-and-runs (HAR), but the actual incidence and predictors of hit-and-run vehicle-bicycle collisions are not well known. We define a HAR as a vehicle-bicycle collision in which the motorist is reported to have fled the scene after a collision with a bicyclist. In this study, we were not interested in bicyclists who ran because we assumed that if they fled after hitting a vehicle, the outcome would be property damage at most. This issue is of public health concern because HAR collisions can lead to delayed medical care. It is also an important issue to law enforcement because fleeing the scene is a crime for which the offender should be brought to justice. However, ultimately, it is an issue for affected individuals and their families. The "best-case" scenario is that the bicyclist is not harmed and the driver is apprehended. However, it is possible for a bicyclist to sustain injuries that require medical care that is well beyond the liability limits of any party's medical or auto insurance.

Over the last decade, cities have begun to prioritize bicycling as a green and active form of transportation. Developments in bicycling infrastructure such as bicycle lanes and road diets have been, and continue to be, implemented across cities in the United States,[3] including in the City of Boston.[2] Bicyclists commuting to work have doubled from the years 2000 to 2009 in urban and rural areas.[3] In urban areas, overall riding has increased three-fold.[3] As cities become more bicycle-friendly and ridership increases, HARs may proportionately increase, resulting in a greater public health concern.

Cyclists are categorized as vulnerable road users.[4] In the event of a vehicle-bicycle collision, a significant probability exists that the bicyclist will be injured, while the likelihood is small that the vehicle occupants will be injured.[5] The cost of medical care for treating bicyclists' injuries resulting from vehicle-bicycle collisions tends to be higher than for treating the vehicle occupants' injuries resulting from vehicle-vehicle collisions.[5] Bicyclists' vulnerability can be further magnified when drivers commit a HAR. According to Tay et al., the odds of a driver fleeing are 4.67 times as likely in vehicle-bicycle collisions when compared to vehicle-vehicle collisions.[6] If there are no witnesses to report a HAR, medical care for the may be delayed.[6] Delays for treatment of traumatic head injuries may worsen patient outcomes.[7] The aims of this study are to estimate the incidence of HARs among vehicle-bicycle collisions and to identify significant predictors of these incidents in an American urban setting. Predictors that were of particular interest were the bicyclists' gender and their injuries as well as whether the involved vehicle was a taxi.

Previous Work

Several studies have identified a multitude of factors that contribute to a HAR occurrence.[6,8,9,10,11,12,13,14] These studies have generally focused on vehicle-vehicle or vehicle-pedestrian collisions. For example, evidence suggests that drivers are less likely to leave the scene in collisions involving younger or older pedestrians.[9,14] Also, traffic and lighting conditions and day of the week contribute to a drivers' decision to flee.[6,10,11,13] In this study explore whether the factors that predict vehicle-bicycle HARs are similar to those identified in previous HAR studies.

Some HAR studies have selected cases from fatality administrative databases,[9,13,14] while others have used police-reported collision data to conduct their HAR analyses.[6,8,12] In the US, the Fatal Accident Reporting System (FARS) will reliably capture comprehensive fatality data at the national level, making the data appealing. However, the limitation of FARS is that fatalities are relatively rare and arguably a non-representative subset of all collisions. Police-reported collision data, which are considered the national standard for collision analyses in the US,[16] are more representative. However, the problem of missing data is common for many reasons, such as the amount of detail police officers choose to include in the reports. For example, Aidoo et al.[8] conducted their study using 10% of the police-reported collisions in Ghana.[15] Kim et al.[12] only used 25% of HAR cases from police records in Hawaii because the remaining records were incomplete. Although these two examples may be extreme, they are the only articles that provide information concerning missing variables in police reports of HAR events. To our knowledge, there are no estimates from the continental US. Non-random missing data within police reports can lead to biased estimates of the effects of factors predicting HAR occurrences.

Our observational study used an electronic registry of all police-reported bicycle collisions in Boston, Massachusetts (United States). This allowed for a reasonable estimate of the incidence of a HAR and limited the types of biases that occur in studies that select their cases based on the outcome of the event (such as fatality or injury). Our data identified whether a collision was a HAR and included factors that may be predictive of a HAR such as road and characteristics. The registry allowed us to examine variables not generally found in standard police reports (such as whether the vehicle was a taxi).

Hypotheses: Injury, Gender, Taxi

Drivers' motivation to flee the scene after a collision has rarely been explored.[9] However, collision studies on HARs did find that fatal outcomes were associated with a higher likelihood of a HAR.[6,8,11,17] As such, we hypothesized that if the driver perceived the bicyclist to be injured, their likelihood of fleeing would be higher because of the uncertainty of what would happen if they stopped. They might wonder if they could be charged with a crime for severely injuring or killing another person.

Another question was whether the likelihood of a HAR would be different if the bicyclist was observed to be male versus female. Several studies have examined the role of the gender of the offender (i.e., driver) in HARs and found that male drivers are more likely to commit a HAR versus their female counterparts.[6,12] Few studies have examined the role of the gender of the victim (i.e., pedestrian) and found that male and female pedestrians are equally likely to be victims of a HAR.[9,14] Since bicyclists generally share the road with motorists, unlike pedestrians, we hypothesized that HARs would be less likely to occur when the victim was a female bicyclist versus a male bicyclist. The rationale is that female bicyclists may be less susceptible to becoming HAR victims if they take "less risky" side roads that make it difficult for offending drivers to escape after a collision. This is supported by evidence that female bicyclists favor more separation from vehicular traffic than their male counterparts,[18] which could mean traveling on side roads that usually involve lower speeds and are equipped with traffic calming installations such as stop signs. Aidoo et al. also note that drivers are 44% less likely to flee in places with traffic calming installations (e.g., intersections, staggered crossroads, and roundabouts).[8] On the other hand, male bicyclists may be more susceptible to becoming HAR victims if they ride on riskier main roads on which vehicles may travel at higher speeds. This is supported by evidence that male bicyclists are less worried than their female counterparts about sharing the road with motorized vehicles[19] and that vehicles traveling at high speeds prior to a collision may increase the likelihood of a HAR.[6]

Lastly, we were interested in the role of taxis concerning HARs. This was a question of interest among stakeholders in Boston. We hypothesized that taxi drivers in general are more likely to flee the scene of a vehicle-bicycle collision. This is based on some evidence that Boston taxi drivers engage in risky driving practices. Specifically, Fernandez et al. conducted an observational study of seatbelt use among taxi drivers in Boston and found that only 6.8% of drivers wore a seatbelt—compared to the 64% of seatbelt use among all other types of drivers in the same year and in the same city.[20] In addition, it is possible that taxi drivers may fear having their licenses revoked from an at-fault crash—because a revoked license could potentially put them out of work.

Method

Study Design and Data Source

We conducted a 4-year retrospective cohort study of bicycle collisions reported by the Boston Police Department (BPD; Massachusetts, USA) from January 1, 2009 to December 31, 2012. The data included information about vehicle-bicycle, bicycle-bicycle, pedestrian-bicycle collisions, and bicycle falls. Cases involving vehicles used intentionally as weapons were excluded from the dataset because we were only interested in collisions that were otherwise assumed to be unintentional. Some of the factors for this study were coded from police narratives. The Committee on the Use of Human Subjects (CUHS), Harvard University's Institutional Review Board, granted permission to conduct this research study.

It is important to note that the BPD does not use a standard crash report form. The BPD crash report form does include, however, data on collisions between vehicles and bicyclists as well as other variables that are relevant to this study. The variables presented here were taken from both closed-ended and free text fields. The advantage of free text fields is that it provides information about the collision, which may not be captured in a forced-choice type of question (e.g., taxi) on a standard police report form. The disadvantage is that the content and quality of the information in the free text fields may vary by what police officers choose to include.

Variable Construction

The initial set of variables chosen for inclusion in our study was informed by two sources. The first was the findings of studies that had modeled the probability of a HAR studies.[6,8,9,10,11,12,13,14] We note that the cited studies had several variables that were not available at the time of the current study. The second source was interviews with stakeholders about the type of information they would find useful. Stakeholders included bicycle advocates, police officers, injury prevention specialists, urban planners, civil engineers, and experienced data managers. Both sources helped identify the list of variables to extract from the BPD collision database.

Variable Description

The outcome variable was a binary indicator of the occurrence of a HAR event, given a motor vehicle-bicycle collision. Atmospheric covariates were temperature categorized into three levels (0-31 °F**;** 32-59 °F; 60+ °F) and precipitation grouped into two levels (yes or no). Temporal covariates included an indicator of whether collision occurred on a weekend or weekday and an indicator of whether the collision took place during the day or at night. Spatial covariates included an indicator of where the collision occurred (intersection or not), whether it was a main street (yes or no). The demographic covariates for the bicyclist were gender (male or female), ethnicity (Black; Hispanic; White; Other), and age, which were categorized into four groups (0-20 yrs.; 21-40 yrs.; 41-60 yrs.; 60+ yrs.). The one available behavioral covariate was whether the cause of collision involved "dooring" or not. Other variables included the type of auto (taxi or non-taxi) and the bicyclists' injury status at the time of the collision (injured or not).

To define the aforementioned indicator of day or night, we linked historical sunrise/sunset times from the National Oceanic Atmospheric Administration to the date/time of occurrence of each collision. If the date/time of the collision occurred after sunrise, then we coded that as "day"; conversely a collision that occurred after sunset was coded as "night." A main street was defined as an arterial road (otherwise, it was a residential road). "Dooring" is defined as a situation in which a driver or a passenger in a vehicle opens the door in path of, resulting in a fall. We indicated the presence of an immediate injury if at least one of the following was reported: 1) Wounds (e.g., road rash, broken leg, etc.); 2) Complaints of pain made by the bicyclist; and 3) Transportation to a hospital immediately after a collision. Taxis included vehicles that were reported as taxis, private limousines, and other "livery/hackney" vehicles – not just vehicles that were visually identified by witnesses. Four variables had missing data, including the bicyclists' gender and ethnicity, whether the collision occurred on a main street, and whether the bicyclist was injured. To account for missing data on these variables in our models, we created an extra category indicating that the information was not recorded (e.g., gender was recoded as "female", "male" and "missing").

Statistical approach

We computed variable frequencies and chi-squared tests to summarize the unadjusted bivariate relationships between the outcome (HAR) and each explanatory variable. We also fit logistic regressions of HAR on different collections of predictors to understand their effects and significance. Nested models were compared through likelihood ratio chi-squared tests. We examined models including the main effects of our predictors, and models that included potentially important interactions.

We organized our model fitting in the following manner. First, we fit a model including the main effects of variables found to be important by stakeholders and the variables found in other HAR studies available to us. In this model, we checked for collinearity among the covariates through variance inflation factors (VIFs). Variables with a VIF greater than 10, a conventional threshold above which collinearity is considered problematic,[21] were considered for removal. Our second model dropped factors with non-significant p-values in the first model, except for injury and gender, and added the following interactions: Injury-by-Gender, Taxi-by-Gender, and Weekend-by-Night. Effects of the predictors were reported as odds ratios. We conducted our analyses using the statistical software package R.[22]

Results

The total number of bicycle-related collisions reported to the BPD during the four-year period was 1806. For the current study, we excluded cases with no motor vehicle involvement--where the bicyclist fell, hit another bicyclist, or hit a pedestrian--which resulted in a total of 1646 vehicle-bicycle collisions available for analysis.

Collisions and HARs

Table 3.1 summarizes the unadjusted bivariate relationship between collision characteristics and the occurrence of a hit-and-run (HAR). Most collisions occurred in warm weather—59.7% occurred when the temperature was 60 degrees or higher. Over three quarters (76.0%) of the bicyclists were male, 56.5% were between the ages of 21-40, and 56.7% were non-Hispanic Whites. The percentage of collisions that resulted in an injury to the bicyclist was 79.5%. Fewer than six percent (5.7%, n=93) of the total collisions involved a HAR.

Weekends had a higher proportion of HARs than did weekdays (8.0% vs. 5.0%, respectively) and nights had a higher proportion of HARs than did days (7.7% vs. 4.4%). A higher proportion of HARs occurred among collisions with taxis than with non-taxis (10.7% vs. 5.1%). Drivers were less likely to run when the collision resulted from dooring than from other causes (2.0% vs. 6.2%). Other predictors, including gender, did not vary significantly by HAR status in these unadjusted analyses.

Of the 1646 cases, 3.9% (n = 64) had missing data on whether the collision occurred on a main street. Six percent (n= 98) had missing data on injury status. Less than half of a percent (n = 8) of cases had missing data on gender and 9.9% (n = 164) had missing data on ethnicity. What police officers choose to write in the narrative portion of the report may explain the missing data.

Table 3.1. *Unadjusted bivariate relationship between collision characteristics and the occurrence of a hit-and-run (Chi Squared Tests)*

	n (% per category)	**% of collisions in which drivers ran**	**p-value**
Total	1646 (100)	5.7	
Time of Day			<0.01
Night	639 (38.9)	7.7	
Day	1007 (61.1)	4.4	
Day			<0.05
Weekday	1297 (78.8)	5.0	
Weekend	349 (21.2)	8.0	
Temperature (F)			*0.79*
0 - 31	56 (3.4)	3.6	
32 - 59	607(36.9)	5.8	
60 +	983 (59.7)	5.7	
Precipitation			*0.86*
Yes	632 (38.4)	5.6	
No	1014 (61.6)	7.4	
Main Street			*0.79*
Yes	556 (33.8)	5.9	
No	1026 (62.3)	5.2	
Missing	64 (3.9)	4.7	

Table 3.1. *Unadjusted bivariate relationship between collision characteristics and the occurrence of a hit-and-run (Chi Squared Tests; Continued)*

		n (% per category)	% of collisions in which drivers ran	p-value
Intersection				*0.96*
	Yes	960 (58.3)	5.6	
	No	686 (41.7)	5.7	
Taxi				<0.01
	Yes	149 (9.1)	10.7	
	No	1497 (90.9)	5.1	
Extended Door				<0.05
	Yes	199 (12.1)	2.0	
	No	1447 (87.9)	6.2	
Cyclist Gender				*0.76*
	Female	387 (23.5)	5.9	
	Male	1251 (76.0)	5.6	
	Missing	8 (0.5)	0.0	
Cyclist Age				*0.72*
	0-20	313 (19.0)	7.0	
	21-40	930 (56.5)	5.5	
	41-60	299 (18.2)	5.4	
	61-80	37 (2.2)	5.4	
	Missing	67 (4.1)	3.0	
Cyclist Injury				*0.13*
	Yes	1309 (79.5)	5.3	
	No	239 (14.5)	5.4	
	Missing	98 (6.0)	10.2	
Cyclist Ethnicity				*0.88*
	Black	298 (18.1)	6.0	
	Hispanic	189 (11.5)	4.2	
	White	933 (56.7)	5.9	
	Other	62 (3.8)	6.5	
	Missing	164 (9.9)	4.9	

Regression Modeling

The VIFs for the collection of variables in model M1 ranged between 1.09 and 1.99, so that collinearity was not evident in the covariate data. Table 3.2 reports our model for the main effects (M1), and the model that includes significant main effects plus select interactions (M2). Based on Model M1, the odds of a HAR were not greater when the bicyclist was injured versus not injured or when the bicyclist was male versus female, controlling for other variables. The odds of a HAR were 2.40 (95% CI: 1.31, 4.23) times as likely when the vehicle was a taxi. The odds of a HAR were 1.65 (95% CI: 1.08, 2.54) times as likely during night as during daylight hours and 1.74 (95%: 1.07, 2.66) times as likely during the weekend versus during the week. The odds were 3.18 (95%: 1.29, 11.51) times as likely when the incident did not involve a dooring. As is evident in M2, the interactions of male-by-injured, taxi-by-injured, and night-by-weekend were nonsignificant.

Table 3.2. *Logistic regression models predicting the odds of HAR, controlling for other variables (n = 1646)*

	Main Effects Model (M1)		**Significant Main Effects + Interactions (M2)**	
	OR	**95% CI**	**OR**	**95% CI**
(Intercept)	0.01	(0.96, 8.46)	0.02	(0.00, 0.07)
Main Street = Yes	0.86	(0.53, 1.36)		
Intersection = Yes	0.85	(0.55, 1.32)		
Cyclist Age = 21 – 40	0.84	(0.50, 1.45)		
Cyclist Age = 41 – 60	0.80	(0.40, 1.53)		
Cyclist Age = 60+	0.79	(0.12, 2.88)		
Time of Day = Night	1.65	(1.08, 2.54)	1.71	(1.03, 2.85)
Day = Weekend	1.74	(1.07, 2.66)	1.82	(0.87, 3.36)
Taxi = Yes	2.40	(1.31, 4.23)	5.13	(1.23, 18.47)
Extended Door = No	3.18	(1.29, 11.51)	3.06	(1.24, 10.19)
Cyclist Gender = Male	0.96	(0.59, 1.59)	0.40	(0.12, 1.35)
Cyclist Injury = Yes	1.05	(0.58, 1.97)	0.55	(0.18, 1.86)
Male x Injured			3.36	(0.86, 12.8)
Taxi x Injured			0.38	(0.09, 1.86)
Night x Weekend			0.86	(0.05, 4.20)
Residual Deviance	683.26	(df = 1629)	679.17	(df = 1633)
AIC		717.26		705.17

Note: Missing levels for variables that were nonsignificant were excluded from this table, specifically: Injured, Main Street, Gender, and Age

It should be noted that only a small number of cases (n=8) had missing gender data. For the results in Table 3.2, the observations with the unknown gender of the victim were treated as a separate gender category. We performed a sensitivity analysis by investigating two alternative ways to address the missing gender cases. The first was to impute the missing gender data for those cases and the second was to drop the cases with missing gender data altogether. The imputation approach involved performing a logistic regression of gender on all other variables in our model for the 1638 cases in which gender was recorded. The fitted logistic regression model was then applied to the 8 observations with missing gender; if the fitted probability of being female was greater than 50%, the victim was imputed as female, and male otherwise. Our sensitivity analysis (not reported) revealed no change in the significance of the main effects and interaction terms regardless of whether we kept, imputed, or dropped the eight cases with missing gender data.

Discussion

Contrary to our hypotheses, being injured or being a male did not increase the odds of a vehicle-bicycle HAR, controlling for the other factors in our models. The studies by Solnick and Hemenway[13,14] and MacLeod et al.[12] also found that the odds of a HAR did not differ for male versus female pedestrians. One explanation for our null finding is men and women choose to ride on the same types of roads in Boston. As such, both may be equally vulnerable to becoming HAR victims. Another possibility is that the built environment may nudge both gender types to take similar routes (i.e., unprotected bike lanes, such as "sharrows").

In line with one of our hypotheses, taxis were more than twice as likely to commit a HAR when compared to other vehicles. This is not surprising for various reasons. On one hand, there is evidence of risky driving behavior among taxi drivers in Boston; as mentioned earlier, taxi drivers in Boston are much less likely to use their seatbelt when compared to the rates of use by the general population.[20] According to the City's Annual Boston Taxi Report,[23] an estimated 20% of taxis that operate in the city on a given day are "gypsy" taxis (i.e., not officially licensed). Drivers operating an illegal "taxi" may have reason to run since what they are doing is illegal. Like other drivers, they may fear the consequences of being caught, but in their case, they also risk losing their job. On the other hand, taxi drivers may just suffer from fatigued driving—especially when shifts are long or monotonous. Falling asleep at the wheel, they may not be aware that they hit a bicyclist.

Consistent with the HAR literature, we found that drivers were more likely to flee the scene of a vehicle-bicycle collision at night and on weekends. The glare from headlights during the night may reduce the driver's ability to detect bicyclists. Another potential explanation is intoxicated driving. According to NHTSA, in 2014, the alcohol impairment rate was four times higher at night than during the day, among drivers involved in fatal crashes.[1] We were not able to examine the role of alcohol in this study because we did not have those data. However, Solnick and Hemenway (1994) found evidence to suggest that intoxicated drivers are: 1) less inhibited from running, 2) less likely to realize that there was a collision, and 3) less likely to recognize that any penalty for the collision will be greater if it is discovered that their driving was impaired.[14]

We also found that the odds of a HAR were far less likely when a dooring occurred versus when it did not. A motorist who doors a bicyclist has come to a stop or has parked, making it more difficult to leave the scene. In other words, the decision leave would need to be an immediate one—taken before even exiting the car. In addition, the bicyclist's body could land on the road in the way of the vehicle, making it more difficult for the driver to flee.

Our observational study had several advantages. First, we were able to examine the relationship between bicyclists' characteristics (e.g., age, gender, and injury status) and the drivers' decision to run in a representative cohort of police-reported bicycle collisions. Whereas

Aidoo et al.[8] conducted their study using 10% of the police-reported collisions and Kim et al.[12] only used 25% of HAR cases from police records in Hawaii, we were able to use all available collision data and then ascertain the percentage of cases that were HARs. Second, including police narrative data allowed us to code for variables that would likely not be available in typical police reports (e.g., dooring). Third, keeping the analysis to only one city reduced the wide geographic variation present in national-level studies. This was not the case in the MacLeod et al.[9] study that examined the odds of a pedestrian-HAR fatality for Northeast, Midwest, South, and West regions of the United States. Solnick and Hemenway[14] also included a comparison of the South versus other regions in their models. However, even after accounting for population density (i.e., urban versus rural), the rate of cycling use and infrastructure differs greatly from city to city. Since it is difficult to account for these variations without more granularity in the national data, city-level data may have reduced some noise. Lastly, we took into consideration stakeholder input about the research questions and hypotheses, engendering trust and bridging the divide between research and practice—something that is often difficult to do in large-scale studies.

Our study also has limitations. We did not have any information on the demographics or driving history of the motorist or about alcohol involvement. We also did not have data about those drivers who committed a HAR and were subsequently apprehended. As in many collision studies, those resulting in minor injuries or close calls may not have been reported. Lastly, the study findings may not be generalizable to other cities.

Recommendations and Conclusion

With regards to addressing HARs: 1) Holding sobriety checkpoints near busy bars have been found to be effective at getting impaired drivers off of the road, particularly at night and on the weekend.[26] 2) Advocating for a Yellow Alert Law, like the one passed in California in September of 2015 (California Assembly Bill 8, Section 8594.15)[24], which allows law enforcement to inform the public through various media outlets about the HAR with the intent of increasing the likelihood of apprehending the offender. This is similar to an Amber Alert for abducted children, and 3) Encouraging cyclists to ride in well-lit roads at night.

About addressing taxis, we recommend the following: 1) Conducting further research on taxi driver behavior and occupational challenges in that line of work, such as drowsy driving. 2) Encouraging passengers to lookout for bicyclists before opening the door, with stickers/decals pasted inside the taxi (in front of the passenger) and verbal reminders from the driver, 3) Whenever possible, ensuring that taxi lights inform bicyclists whether the taxi is occupied. This may heighten bicyclists' awareness of a potential dooring collision.

The National Highway Traffic Safety Administration in conjunction with the Bureau of Justice Assistance and the National Institute of Justice has recently launched the Data-Driven Approach to Crime and Traffic Safety.[25] This initiative encourages local police departments to deploy officers to areas where traffic collisions and crime occur together. HARs are a perfect example of this overlap and a type of crime and public health issue that the initiative could tackle. This plan may bring the offender to justice and increase the likelihood that affected bicyclists can be treated or compensated for their injuries.

Many factors predict a HAR. While some studies on this topic have focused on road engineering and how that predicts a HAR, others have focused on environmental and behavioral predictors. However, our understanding of the phenomenon is only as sound as the quality and completeness of our data. A study with a more comprehensive database—perhaps linking transportation data to law enforcement and health outcomes data—can help us to better understand and ultimately prevent HARs.

Reference - Chapter 3

1. National Highway Traffic Safety Administration. (2014). Traffic Safety Facts: Pedestrians; 2012 data. *Washington DC*.

2. City of Boston Crash Report 2013. (n.d.). Retrieved November 10, 2015, from http://www.cityofboston.gov/images_documents/Crash Report 2013 FINAL reduced 2_tcm3-38304.pdf
3. Pucher, J., Buehler, R., & Seinen, M. (2011). Bicycling renaissance in North America? An update and re-appraisal of cycling trends and policies. *Transportation research part A: policy and practice*, *45*(6), 451-475.

4. WHO global status report on road safety 2013: Supporting a decade of action. (n.d.). Retrieved November 10, 2015, from http://www.who.int/violence_injury_prevention/road_safety_status/2013/en/

5. Haworth, N., & Debnath, A. K. (2013). How similar are two-unit bicycle and motorcycle crashes? *Accident Analysis & Prevention*, *58*, 15-25.

6. Tay, R., Rifaat, S. M., & Chin, H. C. (2008). A logistic model of the effects of roadway, environmental, vehicle, crash and driver characteristics on hit-and-run crashes. *Accident Analysis & Prevention*, *40*(4), 1330-1336.

7. Dinh, M. M., Bein, K., Roncal, S., Byrne, C. M., Petchell, J., & Brennan, J. (2013). Redefining the golden hour for severe head injury in an urban setting: The effect of prehospital arrival times on patient outcomes. *Injury*, *44*(5), 606-610.

8. Aidoo, E. N., Amoh-Gyimah, R., & Ackaah, W. (2013). The effect of road and environmental characteristics on pedestrian hit-and-run accidents in Ghana. *Accident Analysis & Prevention*, *53*, 23-27.

9. MacLeod, K. E., Griswold, J. B., Arnold, L. S., & Ragland, D. R. (2012). Factors associated with hit-and-run pedestrian fatalities and driver identification. *Accident Analysis & Prevention*, *45*, 366-372.

10. Tay, R., Barua, U., & Kattan, L. (2009). Factors contributing to hit-and-run in fatal crashes. *Accident Analysis & Prevention*, *41*(2), 227-233.

11. Tay, R., Kattan, L., & Sun, H. (2010). Logistic model of hit and run crashes in Calgary. *Canadian Journal of Transportation*, *4*(1).

12. Kim, K., Pant, P., & Yamashita, E. (2008). Hit-and-run crashes: use of rough set analysis with logistic regression to capture critical attributes and determinants. *Transportation Research Record: Journal of the Transportation Research Board*, (2083), 114-121.

13. Solnick, S. J., & Hemenway, D. (1994). Hit the bottle and run: the role of alcohol in hit-and-run pedestrian fatalities. *Journal of studies on alcohol*, *55*(6), 679-684.

14. Solnick, S. J., & Hemenway, D. (1995). The hit-and-run in fatal pedestrian accidents: victims, circumstances and drivers. *Accident Analysis & Prevention*, *27*(5), 643-649.

15. Salifu, M., & Ackaah, W. (2012). Under-reporting of road traffic crash data in Ghana. *International journal of injury control and safety promotion*, *19*(4), 331-339.

16. NASS General Estimates System. (n.d.). Retrieved November 10, 2015, from http://www.nhtsa.gov/Data/National Automotive Sampling System (NASS)/NASS General Estimates System

17. Zhang, G., Yau, K. K., & Zhang, X. (2014). Analyzing fault and severity in pedestrian–motor vehicle accidents in China. *Accident Analysis & Prevention*, *73*, 141-150.

18. Emond, C. R. (2011). Gender Considerations in Performance Measures for Bicycle Infrastructure. In *Transportation Research Board Conference Proceedings* (Vol. 2, No. 46).

19. Garrard, J., Handy, S., & Dill, J. (2012). Women and cycling. *City cycling*, 211-234.

20. Fernandez, W. G., Park, J. L., & Olshaker, J. (2005). An observational study of safety belt use among taxi drivers in Boston. *Annals of emergency medicine*, *45*(6), 626-629.

21. Kutner, M. H., Nachtsheim, C., & Neter, J. (2004). *Applied linear regression models*. McGraw-Hill/Irwin.

22. R: A language and environment for statistical computing. (2013). Retrieved from https://www.r-project.org/about.html

23. Nygaard, N. (2013). Taxi Consultant Report. Retrieved November 10, 2015, from http://www.cityofboston.gov/mayor/pdfs/bostaxiconsultant.pdf

24. California Assembly Bill 8 Emergency services: Hit-and-run incidents. (2015-2016). Retrieved November 10, 2015, from http://leginfo.legislature.ca.gov/faces/billCompareClient.xhtml?bill_id=201520160AB8

25. Data-Driven Approaches to Crime and Traffic Safety. (2014). Retrieved November 10, 2015, from http://www.nhtsa.gov/ddacts

26. Ferguson, S. A. (2012). Alcohol-impaired driving in the United States: contributors to the problem and effective countermeasures. *Traffic injury prevention*, *13*(5), 427-441.