Assessing Race and Income Disparities in Crowdsourced Bicycle Safety Data Collection

Jesus M. Barajas*, Kate Beck#, Aditya Medury†

*Safe Transportation Research and Education Center
University of California, Berkeley
2614 Dwight Way #7374, Berkeley, CA 94720, USA
email: jmbarajas@berkeley.edu

Safe Transportation Research and Education Center
University of California, Berkeley
2614 Dwight Way #7374, Berkeley, CA 94720, USA
email: katembeck@berkeley.edu

† Safe Transportation Research and Education Center
University of California, Berkeley
2614 Dwight Way #7374, Berkeley, CA 94720, USA
email: amedury@berkeley.edu

Keywords: crowdsourcing, transportation equity, perceptions, participatory research

1 INTRODUCTION

Planners and public agencies are increasingly using crowdsourcing data apps to gather bicycle travel information from the public. Two examples are San Francisco's CycleTracks and the fitness tracking app Strava, used to collect volume and route choice data, which help enable decisions about where to prioritize cycling improvements [1]. Researchers and transportation agencies have implemented apps to record near misses or other roadway hazards on websites such as BikeMaps.org [2] and phone applications such as Oregon's ORcycle [3]. Crowdsourcing databases provide two advantages: they enable users to document near misses or other locations where they feel unsafe but an incident has not actually occurred, and they are alternative platforms for individuals to report collisions without going through the police. Systematically identifying these types of locations can help cities take a proactive approach to safety, rather than reacting to crashes to identify improvements [4], [5]. Although research on the suitability of using crowdsourced data for bicycle safety improvements is still in its infancy, there is emerging evidence that the data fill reporting gaps relative to official data sources, such as incidents that occur along minor roads and away from intersections [6]. In a study of two crowdsourcing applications, researchers found about three-quarters of the incidents were categorized as near misses [3], [6], which would not have been reported in police, hospital, or insurance databases of bicycle collisions.

However, a major concern about using crowdsourced data to understand safety concerns is one of bias. In some respects, they correct missing information associated with analyzing only police-reported crashes. But apps that work in real time require users to have access to smartphones, a data plan to use them, and the knowledge and desire to participate in data-gathering activities, which may exclude many low-income individuals, Blacks, and Latinos [7]. For example, although the sample size was small, two-thirds of users who submitted data to ORcycle had household incomes greater than $75,000 per year [6]. Crowdsourced data are likely to reflect the characteristics and experiences of people privileged enough to own a smartphone and who opt in to data reporting services, leaving entire neighborhoods unreported and off the map [8]. Therefore, when planners and agencies use crowdsourced data in conjunction with police-reported data, data collection patterns may work to underrepresent low-income communities and communities of color to a greater degree, which could in turn exacerbate disparities in bicyclist safety.

2 PURPOSE AND METHODS

The goal of this project is to understand whether there are biases in reporting through a pilot test of a crowdsourcing data collection app. Specifically, we will analyze whether there are differences in locations of crashes, near misses, perceived unsafe locations, and the types of information people report by different racial/ethnic and income groups. Our study is the first to examine this question.

We will begin by deploying a simple data collection app using a web-based platform. We have preliminarily selected Maptionnaire [9], a geographically-enabled survey platform, as our data collection tool because full-scale development of a uniquely-branded app is beyond the scope of this project. The tool will allow users to provide demographic information, identify points of interest, and associate attributes with each point. We will begin by recruiting about 300 people who live in San Francisco to participate in the study via email and website postings. We will ask users to report all locations where they have gotten into a crash, experienced a near miss, or otherwise felt unsafe. We will then ask participants to use the platform for at least two weeks, identifying additional locations where they have incidents while cycling or where they consider it unsafe. The data attributes we ask users to submit will be broad to identify which characteristics are most useful for safety analysis.

For a second set of participants, we will collaborate with the San Francisco Municipal Transit Agency and a community group involved in safety improvement efforts so we can target outreach to residents in a predominately low-income neighborhood of color in the city that has above average rates of bicycle and pedestrian collisions. We expect to recruit up to 50 individuals whom we will invite to participate in the same data collection exercise as the first group. We will work with the community group to teach participants how to use the platform for people who are not digitally literate. We will also provide mobile devices or an incentive to cover increased mobile phone data use. We anticipate the first group of respondents will be more highly educated and have more technical ability than the population of the city, so the targeted recruitment method will allow the app to reach a larger proportion of the disadvantaged population. We will complete data collection by June 2017.

After data collection, we will analyze the locations, types of incidents, and demographic characteristics of the users for differences by recruiting method. We will analyze how zonal characteristics differ among the datasets (including population, jobs, and other activity locations), and how socioeconomic characteristics of the participants and data collection locations differ. We will also analyze whether there are differences in site-specific characteristics that contribute to elevated risk perceptions, such as roadway features, crosswalk availability, and other safety countermeasures. In this way, we can assess whether the dissemination of a crowdsourcing app without targeted outreach contributes to bias in understanding safety problems. To assess the feasibility of the data as a means for supplementing official data sources, we will also analyze differences in locations and types of incidents reported in the app with police-reported bicycle crash locations in San Francisco's TransBASE database [10] between 2005 and 2012, the most recent data available.

3 EXPECTED FINDINGS AND CONCLUSIONS

While we expect both the reported locations and the types of data collected on the safety app to be different between the two groups, it is difficult to predict how. For example, studies that have analyzed route choice via the Strava fitness app have shown mixed results in the representativeness of the cycling patterns. In Austin, Texas, researchers found recorded routes were primarily recreational in character, occurring in low-density, hillier neighborhoods with fewer bicycle lanes [11]. In contrast, two other studies found the app-based data to have a significant correlation with volumes recorded by manual or automated bicycle counts in more central urban areas, though the strength of association varied by city and context [12], [13]. Higher-income jobs in San Francisco tend to be centralized in the financial district but recreational facilities are located throughout the city, so we may obtain spatially representative data. In other words, incident coverage may be distributed in space in a representative manner even if user demographics are not representative.

Participatory mapping projects have been employed in the environmental justice domain in public health to gather local knowledge about environmental conditions such as pollution hazards [14] and healthy food options [15]. Projects often employ a variety of data collection techniques, including sending teams out on foot to map pertinent locations and taking images and narrating experiences (photo-voice). Our study follows similar models. We will provide a preliminary assessment of which techniques are most effective in engaging diverse community members, and whether they provide actionable information for safety interventions.

REFERENCES

[1] A. Misra, A. Gooze, K. Watkins, M. Asad, and C. Le Dantec, "Crowdsourcing and Its Application to Transportation Data Collection and Management," *Transp. Res. Rec. J. Transp. Res. Board*, vol. 2414, pp. 1–8, Aug. 2014.

[2] T. A. Nelson, T. Denouden, B. Jestico, K. Laberee, and M. Winters, "BikeMaps.org: A Global Tool for Collision and Near Miss Mapping," *Front. Public Health*, vol. 3, Mar. 2015.

[3] B. Blanc and M. Figliozzi, "Safety perceptions, roadway characteristics and cyclists' demographics: a study of the of crowdsourced smartphone bicycle safety data," in *TRB 96th Annual Meeting Compendium of Papers*, Washington, DC, 2017.

[4] R. J. Schneider, R. M. Ryznar, and A. J. Khattak, "An Accident Waiting to Happen: A Spatial Approach to Proactive Pedestrian Planning," *Accid. Anal. Prev.*, vol. 36, no. 2, pp. 193–211, Mar. 2004.

[5] G. Cho, D. A. Rodríguez, and A. J. Khattak, "The Role of the Built Environment in Explaining Relationships Between Perceived and Actual Pedestrian and Bicyclist Safety," *Accid. Anal. Prev.*, vol. 41, no. 4, pp. 692–702, Jul. 2009.

[6] M. Branion-Calles, T. A. Nelson, and M. Winters, "Comparing Crowdsourced Near Miss and Collision Cycling Data and Official Bike Safety Reporting," in *TRB 96th Annual Meeting Compendium of Papers*, Washington, DC, 2017.

[7] A. Smith, "U.S. Smartphone Use in 2015," Pew Research Center, Washington, DC, 2015.

[8] C. A. Le Dantec, C. Appleton, M. Asad, R. Rosenberger, and K. Watkins, "Advocating Through Data: Communitiy Visibility in Crowdsourced Cycling Data," in *Bicycle Justice and Urban Transformation: Biking for All?*, A. Golub, M. L. Hoffman, A. E. Lugo, and G. F. Sandoval, Eds. New York: Routledge, 2016, pp. 70–85.

[9] Mapita Oy, "Map based questionnaire service," *Maptionnaire*. [Online]. Available: https://maptionnaire.com. [Accessed: 24-Feb-2017].

[10] San Francisco Department of Public Health, "TransBase: Linking Transportation Systems to Our Health." [Online]. Available: http://transbasesf.org/transbase/. [Accessed: 25-Feb-2017].

[11] G. P. Griffin and J. Jiao, "Where does bicycling for health happen? Analysing volunteered geographic information through place and plexus," *J. Transp. Health*, vol. 2, no. 2, pp. 238–247, Jun. 2015.

[12] B. Jestico, T. Nelson, and M. Winters, "Mapping ridership using crowdsourced cycling data," *J. Transp. Geogr.*, vol. 52, pp. 90–97, Apr. 2016.

[13] R. L. Sanders, A. Frackelton, S. Gardner, R. Schneider, and M. Hintze, "'Ballpark' Method for Estimating Pedestrian & Bicyclist Exposure in Seattle: A Potential Option for Resource-Constrained Cities in an Age of Big Data," in *TRB 96th Annual Meeting Compendium of Papers*, Washington, DC, 2017.

[14] J. Sadd, R. Morello-Frosch, M. Pastor, M. Matsuoka, M. Prichard, and V. Carter, "The Truth, the Whole Truth, and Nothing but the Ground-Truth: Methods to Advance Environmental Justice and Researcher–Community Partnerships," *Health Educ. Behav.*, vol. 41, no. 3, pp. 281–290, Jun. 2014.

[15] A. Akom, A. Shah, A. Nakai, and T. Cruz, "Youth Participatory Action Research (YPAR) 2.0: How Technological Innovation and Digital Organizing Sparked a Food Revolution in East Oakland," *Int. J. Qual. Stud. Educ.*, vol. 29, no. 10, pp. 1287–1307, Nov. 2016.