Journal of Safety Research xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Safety Research

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

Wider view over bicycle crashes: Complementing and extending bicycle crash statistics in urban areas using surveys

Laura Ringel [a,*], Clemens Kielhauser [b], Bryan T. Adey [c]

[a] Fachstelle Verkehrssicherheit, Strasseninspektorat Tiefbauamt, Baudirektion Kanton Zürich, Walcheplatz 2, 8090 Zurich, Switzerland
[b] Transport Infrastructure Group Berne University of Applied Sciences, Pestalozzistr. 20, 3400 Burgdorf, Switzerland
[c] Chair of Infrastructure Management, ETH Zurich, Stefano-Franscini-Platz 5, 8093 Zurich, Switzerland

ARTICLE INFO

Article history:
Received 18 November 2022
Received in revised form 8 August 2023
Accepted 20 September 2023
Available online xxxx

Keywords:
Crash reports
Unreported crashes
Crash types
Crash statistics
Bicycle crashes
Tram tracks

ABSTRACT

Introduction: In pursuit of sustainability goals, many cities are introducing measures to increase the usage of bicycles as a means of transportation. City planners aim to ensure that this increase does not lead to an increase in crashes, but must make corresponding infrastructure decisions with limited information. Sufficient data to perform a statistical analysis of location-specific crash frequencies is rarely available. For example, only approximately 10% of all bicycle crashes are reported to the police (Shinar et al., 2018). Therefore, urban planners often rely on expert opinion, which may lead to suboptimal prioritization and realization of infrastructure improvements. Method: This paper demonstrates how surveys on bicycle crashes can be used to aid urban planners in making infrastructure decisions. In addition to confirming the location and characteristics of reported crashes, surveys can uncover characteristics of crashes that are not reported to the police, situations in which a crash almost occurred, and locations perceived by cyclists to be dangerous. Surveys also allow urban planners to investigate non-infrastructure related causes of crashes, such as the frequency with which individual cyclists use other modes of transportation. Practical Applications: The usefulness of surveys in the determination of urban cycling safety is demonstrated in this paper through analysis of survey results from the city of Zurich in 2018.

1. Introduction

Efforts are being made to improve quality of life in the world's cities by reducing pollution, congestion, and noise. This can be achieved by increasing the number of cyclists. Unfortunately, increasing bicycle traffic might also increase the number of crashes involving cyclists. Therefore, urban transport planners strive to adapt infrastructure to make cycling in the city both enjoyable and safe. Decisions on where and how to improve the safety of infrastructure should be supported by extensive amounts of crash data. In most cases, however, these data does not exist. In Switzerland (BFU, 2020, p. 8) and in many other countries, it is estimated that only 1 in 10 bicycle crashes are reported to the police (Shinar, et al., 2018). Several researchers have confirmed that crashes reported to the police account for only a small proportion of the total number of crashes (Juhra, et al., 2012; Isaksson-Hellman & Werneke, 2017; Meuleners, et al., 2019; Meuleners, et al., 2020). Being aware of only about 10% of all crashes heavily reduces the information basis for improving cycling infrastructure. This dearth of data also impedes quantification of the effectiveness of imposed policies.

To mitigate the scarcity of data, this paper proposes conducting surveys among cyclists to complement police records. This can provide a comprehensive picture of all cycling crashes and locations being perceived as dangerous by cyclists.

The benefits of this approach are demonstrated in the City of Zurich, Switzerland. In 2018, an online survey on bicycle safety was conducted in which 2076 participants reported a total of 6,997 dangerous locations ("Gefährliche Stellen"), 1,452 near crashes ("Beinahe-Unfälle"), 1,014 personally experienced crashes ("selbst gehabte Unfälle"), and 416 observed bicycle crashes ("gesehene Unfälle"). The current analysis consists of three components: a descriptive and multivariate analysis, an assessment of survey usage in conjunction with police reports, and a discussion of demographic and behavioral data in context of bicycle crashes.

When analyzing cyclist road safety, road safety departments and researchers frequently rely on police crash reports (Pai, 2011; Pokorny, Drescher, Pitera, & Jonsson, 2017; Jensen, 2017; Akgün, et al., 2021) although they are aware of the underrepresentation of cyclist crashes in statistics derived from police reports

* Corresponding author.
E-mail addresses: laura.ringel@bd.zh.ch (L. Ringel), clemens.kielhauser@bfh.ch (C. Kielhauser), adey@ibi.baug.ethz.ch (B.T. Adey).

https://doi.org/10.1016/j.jsr.2023.09.018
0022-4375/

Please cite this article as: L. Ringel, C. Kielhauser and B.T. Adey, Wider view over bicycle crashes: Complementing and extending bicycle crash statistics in urban areas using surveys, Journal of Safety Research, https://doi.org/10.1016/j.jsr.2023.09.018

(Stutts, Williamson, Whitley, & Sheldon, 1990). Some researchers try to compensate by adding insurance or hospital crash reports to the data (Hels & Orozova-Bekkevold, 2007; Sakshaug, Laureshyn, Svensson, & Hydén, 2010; Juhra, et al., 2012; Schepers, et al., 2015). Others use video recordings (Buch & Jensen, 2017; Meuleners, et al., 2019; Wu & Lin, 2020) or attempt simulations of certain scenarios using driving simulators (Warner, Hurwitz, Monsere, & Fleskes, 2017; Jannat, Hurwitz, Monsere, & Funk, 2018). Despite having access to incomplete datasets, common approaches such as the development of safety performance functions (SPFs)[1] to predict the safety of road segments are often based solely on police report data (Chen, Wang, Roll, Nordback, & Wang, 2020; Nordback, Marshall, & Janson, 2014; Nordback, Kothuri, Marshall, Gibson, & Ferenchak, 2017).

To complement existing crash data, researchers have attempted to collect observations on cyclist stressors and near crashes to predict where crashes are likely to occur. Such work includes that of Gustafsson and Archer (2013) who recorded cyclists in Stockholm on camera during their commutes. Sayed et al. (2013) recorded cyclists passing through a specific intersection. Luan et al. (2020) analyzed GPS data of trips using a public bike sharing service to find factors influencing riders to travel against the specified direction on one-way roads. Blanc and Figliozzi (2016) went a step further, using the app *Orcycle* to enable 170 cyclists to rank the comfort level of each trip over a period of seven months on a five-point ordinal scale. Additionally, users could categorize factors that lead to stress as "Not concerned," "Car traffic," "Large commercial vehicles (trucks)," "Public transport (buses, light rail, streetcar)," "Parked vehicles (being doored)," "Other cyclists," "Pedestrians," and "Other." Karakaya et al. (2020) used data from the platform *SimRa*, which collects motion-related data through the smartphone GPS, accelerometer, and gyroscope sensors. It identifies acceleration spikes in the data and asks the user after a trip to label those incidents, with the goal to identify characteristics and locations of near crashes. Nelson et al. (2015) has set up an open-source website called BikeMaps.org to allow riders to map near crashes, crashes, or dangerous locations. When mapping, the user is given different options to further classify the incident. Similarly, the Swiss website *bikeable.ch*, though not used for research, allows users to map cycling hazards with a photo and description.

The studies conducted so far have mostly been based on police data (Kielhauser et al., 2020; Akgün et al., 2021), police and hospital data (Meuleners et al., 2020; Juhra et al., 2012), insurance data (Isaksson-Hellman & Werneke, 2017), video analysis (Wu & Lin, 2020), or a combination of the above (Sakshaug et al., 2010). Each approach has certain limitations. For example, video analysis covers a limited area of observation. The use of hospital data alone focuses less on the details of the crash and more on the injuries sustained; combining this dataset with insurance data can create further inconsistencies due to limited standardized reporting. Surveys have been used to complete mixed study designs (Meuleners, et al., 2019) and to understand the extent of underreporting in specific regions (Shinar, et al., 2018; Winters & Branion-Calles, 2017) without looking at crash types or infrastructure features. This study suggests to consider crash constellation, crash type, infrastructure characteristics, and unreported cases in further research. Surveys are a useful complementary tool to combine all of the mentioned crash characteristics. The underreporting rate in the different samples of other studies varied between 50% (Isaksson-Hellman & Werneke, 2017) over 60% (Meuleners et al., 2019; Meuleners et al., 2020), 68% (Juhra et al., 2012), 75% (Hels & Orozova-Bekkevold, 2007), 88% (Winters & Branion-Calles, 2017) to 90% (Shinar et al., 2018).

The proposed use of surveys to improve cycling safety in urban areas allows all of these characteristics to be considered, in addition to addressing issues of perceived crash potential. Furthermore, the use of surveys allows urban planners to investigate non-infrastructure related correlations with crash frequency, such as the frequency with which cyclists use other vehicular modes of transportation. Examples of the use of surveys in cycling safety are not entirely new. Surveys have been conducted to compare crash rates between countries (Branion-Calles et al., 2020) and to investigate how cyclist behaviours lead to increased risk. The latter includes studies on red light running (Richardson & Caulfield, 2015), mobile phone use (de Waard, Westerhuis, & Lewis-Evans, 2015), the correlation between personal physical aggression and bicycle crash frequency (Stephens, O'Hern, Trawley, Young, & Koppel, 2019), differences in crash risk between e-bike users and conventional bicyclists (Fyhri, Johansson, & Bjørnskau, 2019) and the influence of vigilance, predictability, and cautiousness on crash risk reduction (Hoglund, 2018). However, surveys have not yet been used in investigations as to where crashes have occurred or dangerous situations in urban areas are thought to exist.

The combination of surveys and police reports gives a more nuanced and comprehensive picture of where potential exists for improving infrastructure for cycling than if only police-reported crashes were used. The following chapters explain how the survey responses on dangerous locations and crashes that were not reported to police provide an insight into the crash potential of the existing cycling network, and how such an analysis can provide an enhanced information base for planning infrastructure improvements.

The remainder of this paper is structured as follows. Section 2 provides an overview of the conducted survey and crash data and describes the analytical approach. Section 3 demonstrates the additional benefit of including the survey data into the analysis. Section 4 contains a discussion of the limitations of the explanatory power of the research, summarizes the conclusions of the paper, and provides a critical reflection on how infrastructure managers should conduct such analyses in the future. Finally, Section 5 provides suggestions for future work.

2. Data collection

2.1. Overview

This paper demonstrates the usefulness of surveys in the analysis of cycling crashes through the presentation of an analysis of cycling safety in Zurich in 2018. Two datasets are used for this case study: On-site crash data collected by the police, and results from a survey of cyclists in the City of Zurich to collect self-reported information about crashes. The survey was conducted to obtain a representative sample of cyclists' experiences, perceptions, and behaviors that may influence crash statistics. In this section, an overview of the data is provided. The next section explains the survey methodology. In the subsequent section, the two datasets are compared.

The police crash data were collected between 01 Jan 2014 and 31 Dec 2020. The required data on police crash report forms is listed in Table 2.1. The police report distinguishes between 11 crash type groups: singles-vehicle crashes, overtaking or lane-changing crashes, rear-end crashes, turning crashes, lane-crossing crashes, head-on collisions, parking crashes, pedestrian crashes, animal crashes, and other. Each group is further subdivided into different types of crashes. A total of 66 different crash types and 117 main causes of crash are present in the crash report form.

[1] A safety performance function relates the expected crash frequency of a certain road element to measurements of traffic volume and selected geometric and/or operational characteristics.

Table 2.1
Attributes on bicycle crashes collected by the police.

Person involved	Object	Crash
Gender	Number of persons involved	Date and time
Age	Leading cause of involvement	Number of objects involved
Injured / uninjured	Causes 1 – 3	Crash severity
Protective equipment	Trailer? (yes/no)	Crash type group
	Possession of driver's license (if applicable)	Crash type
	Purpose of travel	Leading cause of crash
	Breath alcohol test executed? (yes/no)	Map coordinates
	Breath alcohol concentration	Type of road
	Blood alcohol test executed? (yes/no)	Traffic volume
	Blood alcohol concentration	Speed limit
	Blood/urine drug test executed? (yes/no)	Right-of-way characteristics
	Blood/urine drug concentration	Road characteristics at crash site (straight, curve, intersection…)
	Blood/urine narcotic test executed? (yes/no)	Road condition
	Blood/urine narcotic concentration	Weather
		Lighting condition

The survey was conducted in the last quarter of 2018. Participants were not asked about the date and time of the relevant crashes, as it was assumed that they would not be able to provide accurate information in most cases. The survey data requested a classification of the crash type and a personal description of incident events.

2.2. How the survey was conducted

The survey, entitled "Bicycle Safety in the City of Zurich," was conducted on the website *Maptionnaire*. It was promoted online, in print media, and via word-of-mouth. In addition, a posting promoting the survey was published on RonOrp, a city news website, and was also included in the RonOrp newsletter. Promotional Facebook posts were published by *Velo Plus* (a popular bicycle shop chain), by *Pro Velo Zürich* (an association that advocates for the interests of cyclists in Zurich) and by *bikeable.ch* (an online platform to report dangerous spots for cyclists). Pro Velo also published an article in its print medium *Velojournal*. Leaflets were distributed at the bicycle market and in 54 bicycle shops in and around the City of Zurich. Participants could reach the website either via a QR code or via a link. The language of the survey was German as Zurich is situated in the German speaking part of Switzerland. Since the questions were easy to understand, some answers were given in English as well.

2.3. Comparison of data composition

The police crash data consist of three broad data categories: *Mitfahrende* (persons involved), which contains the persons involved in the incident and their personal characteristics; *Objekt* (object), which contains the vehicles involved in the crash and their characteristics; and *Unfall* (crash), which contains other information about the crash. The three data pools are linked through a set of unique identifiers (UID). The standardized data collected are shown in Table 2.1. The data set contains any bicycle crash that was registered by the police in the considered timeframe. There is no threshold that has to be met so that the police include a crash into the data base. A crash gets included if police presence was necessary of any sort (e.g., to manage the crash site) or if someone reports the crash to the police.

The survey data also consists of three data categories: demographic and behavioral characteristics of the participants (age, frequency of cycling, lifetime cycling crashes, etc.); the coordinates of the commuter cyclists' route to work; and the crash characteristics (coordinates and description of crashes, near crashes, and danger spots). All provided data were linked to the ID of the respondent. Table 2.2 and Table 2.3 show the attributes collected about the characteristics of survey respondents and the characteristics of crashes, respectively. For a complete list of all attributes, see Appendices Table 7.1 and Table 7.2.

Since typically the ways people travel differ between workdays and weekend days, two different categories were given for these purposes. The goal was to see if the practice in kilometers shows any correlation with the number of accidents. To find a meaningful classification the categories were built in the style of a logarithmic scale. As calculation basis the Renard series R 10/3 was chosen. The beginning and the end of the scale deviates from the series since the goal was to not create extremely small or big categories.

Between 2014 and 2020, 4,305 bicycle crashes were registered by the police. Following the exclusion of 25 responses that contained fewer than 7 of the total 34 attributes queried, the survey data consisted of 2,076 entries on respondents' characteristics, 266 entries on routes to work, and 9,879 entries on crashes, near crashes, or danger spots. Respondents mapped locations of 1,014 personally experienced crashes, 416 observed crashes, 1,452 near crashes and 6,997 danger spots. Descriptions in writing were given for 89% of all personally experienced crashes, 89% of all observed crashes, 86% of all near crashes and 89% of all danger spots. For 95% of the personally experienced crashes, an answer was given to the question of whether or not the crash was reported to the police. The crash constellation was stated in 94% of the cases.

The crashes, near crashes, and dangerous spots reported in the survey were classified based on the written descriptions of the reported circumstances. The text was cleaned up with a dictionary created using the software TreeTagger (Schmid, 1995). Subsequently, each description was assigned a crash type group, a crash type, and a leading cause where possible. In order to compare both data sets, the same crash type groups, crash types, and leading cause categories were used that are used in the standardized police form. For 78% of the descriptions, a crash type group could be assigned, for 88%, a crash type and for 80%, a leading cause. The two data sources can be correlated by these methods, the analysis of which is discussed in the next chapter.

2.4. Analytical approach

Due to extensive advertising channels, a semi-representative sample was achieved in the survey. Comparing the age distribution of Zurich's overall population with the ages of survey participants, the data show an underrepresentation of residents under the age of 21 and over the age of 60 (Fig. 1), but the desired distribution is otherwise representative. The average age of the 2,076 survey participants was 37.63 years, while the average age of the population

Table 2.2
Attributes gathered about the survey participants.

Demographic and behavioral characteristics		
Respondent-ID	Typical weekend day cycling distance	Regular commuting route
Screen width of personal device	Purpose of the trip	Number of cycling crashes
Age	Weather conditions in which the participant does not like to travel	Number of downhill or mountain bike crashes
Bicycle type	Light conditions in which the participant does not like to travel	Number of cycling crashes in the City of Zurich
Frequency of bicycle trips	Self-assessment of driving ability while tired or drunk	Number of near crashes in Zurich
Cycling motivation	Music listening habits while cycling	Location of spots perceived as dangerous
Age of learning to cycle	Smartphone usage listening habits while cycling	Opinion about Dutch Reach
Driver's license? (no/if yes, which)	Helmet usage habits	Opinion about integration of Dutch Reach in driving school lessons
Frequency of trips via car	Reason(s) to not wear a helmet	Duration of response
Typical weekday cycling distance	Area of residence	Date and time of response

Table 2.3
Attributes gathered about the crash / near crash / danger spot location.

Crash information		
Respondent-ID	Crash constellation	Crash severity
ID of location	Crash description	Familiarity of the route
Type of location	Crash registration	Improvement suggestions
Coordinates	Assignment of blame	
Description of dangerous spots	Change in cycling frequency	

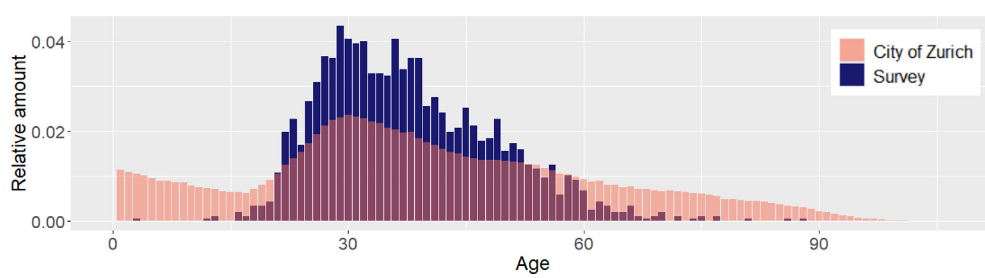

Fig. 1. Comparison of age distribution of the overall population of Zurich with survey participants (n (survey) = 2′076, n (City of Zurich) = 428′737).

of Zurich in 2018 was 39.49 years. Most survey participants reported that they frequently use bicycles and rarely use cars (Fig. 2). Thus, the sample overestimates characteristics of the middle-aged population and frequent cyclists.

63% of the participants stated they see the bicycle both as means of transport and for leisure purpose. 33% saw it only as means of transport and 4% used their bicycle only for leisure purposes. 89% of the participants owned a drivers' license, 11% did not. 81% of the participants learned how to ride a bicycle at the age of 6, 16% between 7 and 12 years old, 2% between 12 and 50 years old, and 1 person with more than 50 years. 30% of the participants rode 8–16 km on a typical weekday, 28% 4–8 km, 16% 2–4 km and 11% rode 16–30 km. 22% of the participants rode 4–8 km on a typical weekend day, 18% 2–4 km, 17% 8–16 km and 10% rode 16–30 km. The participants were asked for which purposes they generally choose to take the bike and, when they choose to take the bike, if they cycled the whole way or just a part of the way. 74% of the participants stated they use the bicycle for the complete trip to work, 14% for a part of the commute, 6% did not use it for the commute. 71% of the participants used the bicycle for the complete trip to their leisure activity, 21% for a part of the way, 2% did not use it for leisure activities. 65% of the participants used the bicycle for the complete shopping trip, 9% for a part of the shopping trip, 15% did not use it for the shopping. 27% of the participants used the bicycle for the complete trip to education, 6% for a part of the education trip, and 28% did not use it for the education trip.

The advantage of using survey data manifests in knowing about details of unreported cycling crashes. For 95 % of the 1,014 personally experienced cycling crashes stated in the survey, respondents gave the information if the crash got registered by the police or not. Thus, the dark figure of crashes can be estimated at 86%. The next section presents the advantages of using survey data together with police-reported crash data over using the latter alone. The presentation is built on three stages:

- The comparison of results based on police-reported data alone versus a combined dataset of police and survey data (Section 3.1). In order to compare the police data with the combined dataset (police and survey data), an aggregation was conducted using the proportions estimated in the survey, i.e., a ratio of 14%

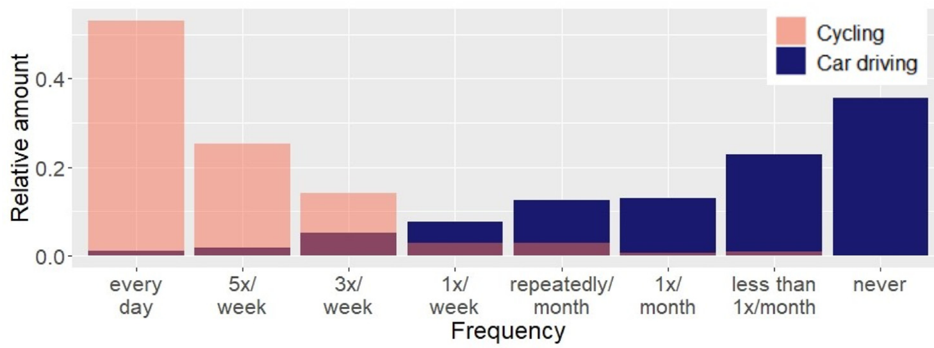

Fig. 2. Bicycle and car use frequency of the survey participants (n (Cycling) = 2064, n (Car driving) = 733).

to 86% between reported and unreported cases. Furthermore, the relative spatial variation between police-reported and non-police-reported crashes was illustrated on a map.

- Testing hypotheses on possibly existing correlations between the number of urban cycling crashes and certain demographic or behavioral characteristics, which is infeasible with only police-reported crash data (Section 3.2). Hypotheses were tested, two-dimensional relationships were examined, and a Random Forest approach was selected to determine which characteristics were more likely to serve as explanatory variables for the crash frequency. Then, the two-dimensional relationships of the variables that were more likely to be an explanatory variable for the number of crashes were examined and different combinations of these variables were used as input variables for cluster analysis.
- The comparison of the frequency of crash types between near crashes, perceived danger spots, and actual crashes (Section 3.3).

The comparisons of crash severity at different locations (Section 3.1.3) were made using police data from 2014 to 2018, as the survey was conducted in 2018. All other analyses were carried out with the entire dataset, that is, crashes that occurred from 2014 to 2020.

3. Results: Comparison of bicycle crash research based on only police data versus survey and police data

3.1. Differences between an analysis using only police reported data, and a combined data set of police and survey data

A comparison of the two data sets shows that 86% of the crashes reported in the survey were not reported to the police. While 15.8% of police-reported crashes resulted in serious injuries, only 2.7% of the survey-reported crashes did the same. Table 3.1 shows the proportions of crash severity for reported and unreported crash data.

There is considerable spatial variation between police-reported and non-police-reported crashes. Fig. 3 shows this difference for the districts of the City of Zurich, using both color and transparency. The color scale gives information about how reported and unreported crashes are distributed across the city in terms of differences in percentages. A red district has more unreported crashes, and a blue district more reported crashes. For example, a district where 8% of the police-reported crashes but 10% of the unreported crashes occurred, would be colored medium-red, as there is an absolute difference of (8% − 10%) = -2%. The transparency scale gives information about the total number of crashes in a district; the less transparent, the more crashes. Fig. 3 illustrates that, relatively speaking, far more unreported crashes occurred in the center of Zurich and fewer occurred in the surrounding areas. This indicates that the spatial distribution of bicycle crashes varies considerably if unreported bicycle crashes are taken into account.

In order to compare the police data with the combined dataset (police and survey data), an aggregation was conducted using the proportions estimated in the survey (i.e., a ratio of 14% to 86% between reported and unreported cases). For example, if the police data has a 28.3% share of single vehicle crashes and the unreported survey crash data has a 43.5% share, the aggregated share is 28.3% *0.14 + 43.5%*0.86 = 41.4%.

3.1.1. Crash type

Fig. 4 shows the breakdown of crashes by crash type for police-reported crashes and survey-augmented crashes. Considering only police-reported data appears to overestimate the number of collisions with oncoming traffic when taking left turns (4.9% vs. 1.0%), collisions with crossing traffic (10.5% vs. 6.2%), and collisions with stationary vehicles (3.8% vs. 1.2%). On the other hand, the number of single-vehicle crashes without collisions (28.3% vs. 41.4%) and the number of dooring crashes (3.0% vs. 4.8%) are underestimated.

3.1.2. Leading cause

Fig. 5 shows a comparison of leading causes of police-reported crashes compared to those of survey-augmented crashes. Both datasets contain a high number of crashes involving a single vehicle without a collision (see Section 3.1.1). The combined dataset showed a top leading cause of "acute angle tram track crossing" (26%), and the police-reported dataset showed this leading cause for only 5.5% of crashes. On the other hand, "other influence due

Table 3.1
Proportions of crash severity for reported and unreported crashes.

Crash severity	Police data(n = 4,305)	Survey data(police reported, n = 133)	Survey data(unreported, n = 834)
Seriously injured (including fatally)	16%	11%	1%
Slightly injured	65%	70%	58%
Property damage	19%	19%	26%
No damage			15%

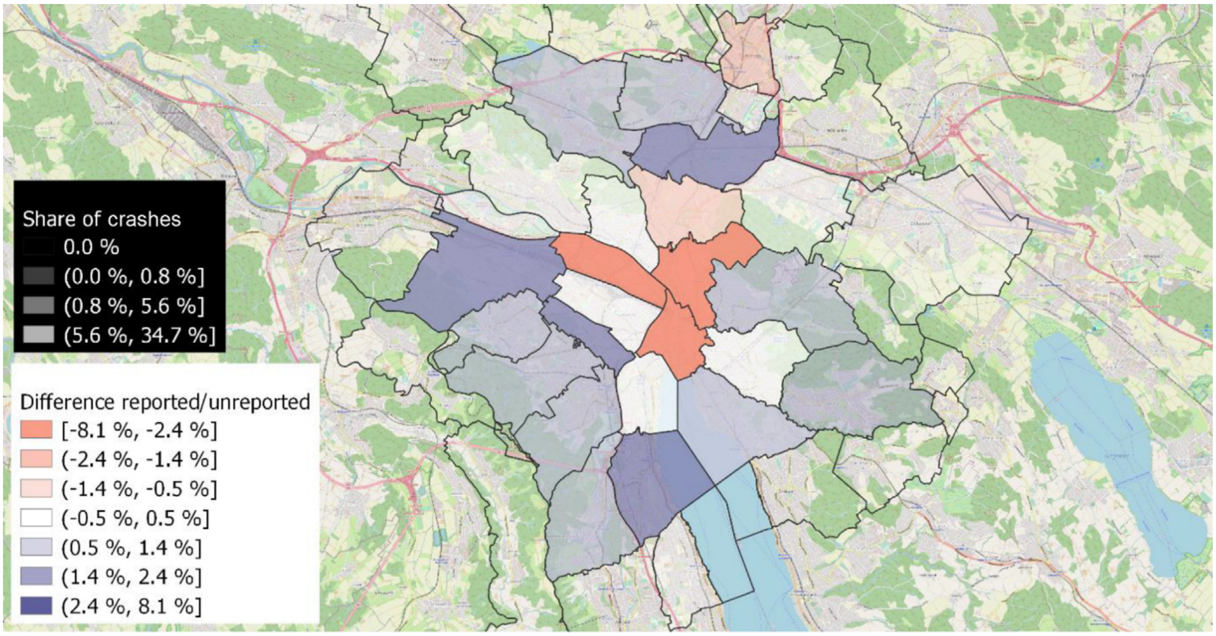

Fig. 3. Difference of the relative proportions of police reported and unreported survey bicycle crashes in the districts of the City of Zurich (n (unreported) = 834, n (reported) = 2850).

to inattentiveness or distraction" represented the most frequent leading cause within the police-reported crashes (19.1%), while the combined data set shows that leading cause in only 4.3% of the crashes. Furthermore, the influence of alcohol was significantly more common in the reported crashes (7.9%) than in the combined data set (2%).

"Lack of adaptation to road conditions (wet, icy, gravel, leaves, etc.)" was much more prevalent in the combined data set (6.9%) than in the police-reported data set (0.4%). In summary, the combined crash data show that the most important leading crash causes are largely underestimated in the police-only data, while the leading crash cause identified in the police-only data is relatively uncommon and overestimated by a factor of almost 4.

3.1.3. Crash location and economic cost

To identify crash hotspots, the spatial distribution of crash costs is often used, as this allows to combine different crash severities. Fig. 6 shows core density estimates of crash costs from police-reported crashes (blue) and the combination of crash data sources (red). The economic costs were calculated using crash cost figures from the Swiss standard VSS 41 824. For the purpose of information privacy, the categories "serious injury" and "fatal" were combined beforehand in the police data. The set "serious injury" contains very few fatal crashes (about 1 to 2%). In the survey dataset fatal crashes are by definition not included. For the estimation, the following values were used for estimation in accordance with VSS 41 824: CHF 498,700 for a crash with serious injuries, CHF

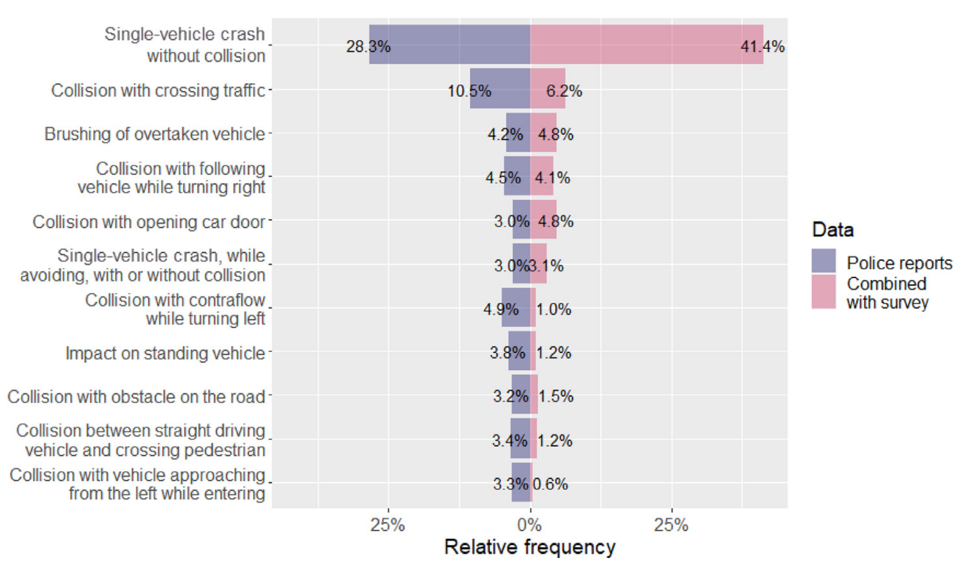

Fig. 4. Comparison of crash types of only police reported crashes and the proportional mix of police reported and not police reported crashes (n (police data) = 4'305, n (survey data) = 832).

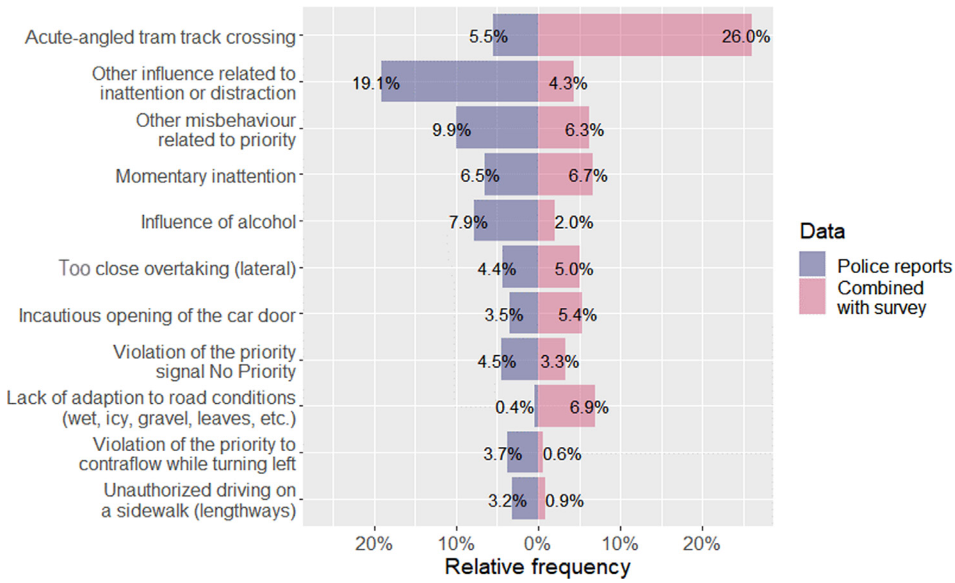

Fig. 5. Comparison of leading causes of only police reported crashes and the proportional mix of police reported and not police reported crashes (n (police data) = 4′305, n (survey data) = 739).

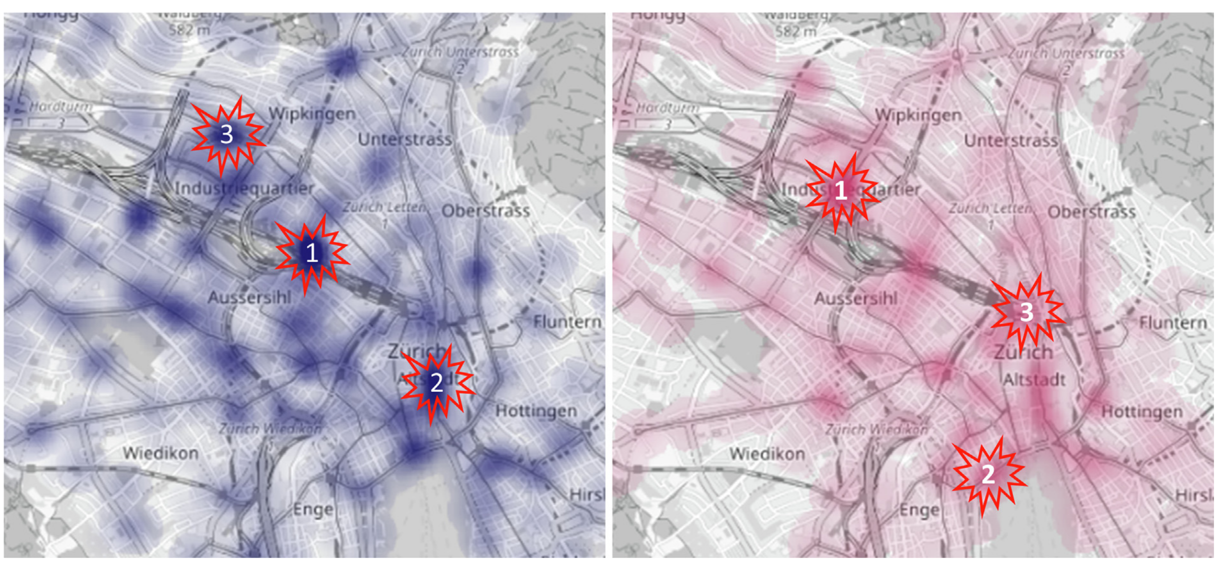

Fig. 6. Density of economic costs of police reported crashes (blue) and the extrapolated combination of police reported and not police reported crashes (red) (n (police data) = 4′305, n (survey data) = 967).

33,500 for a crash with minor injuries, and CHF 7,011 for a crash with property damage. The Swiss norm does not differentiate between motorized traffic and bicycle traffic.

In Fig. 6, the three peak crash cost density locations are marked. To create the combined data visualization, the underreporting percentages were applied separately for each category and adjusted with police-reported crash costs such that the relative distribution of costs is comparable. The figure shows that the economic costs of police-reported crashes peak in other areas of the city than in those of the combined data.

Table 3.2
Estimated crash costs due to police reported crash and the underreporting rate of the survey.

	Number of police reported bicycle crashes (2014–2018)	Economic cost of police reported crashes	Survey underreporting rate	Economic cost including underreporting rate
Slight injuries	1822	61,037,000 CHF	6.19	377,819,030 CHF
Serious injuries	488	243,365,600 CHF	1.73	421,022,488 CHF
Property damage	540	3,785,940 CHF	8.64	32,710,522 CHF
Sum		**308,188,540 CHF**		**831,552,040 CHF**
				− 308,188,540 CHF
Cost without known spatial allocation				*523,363,500* CHF

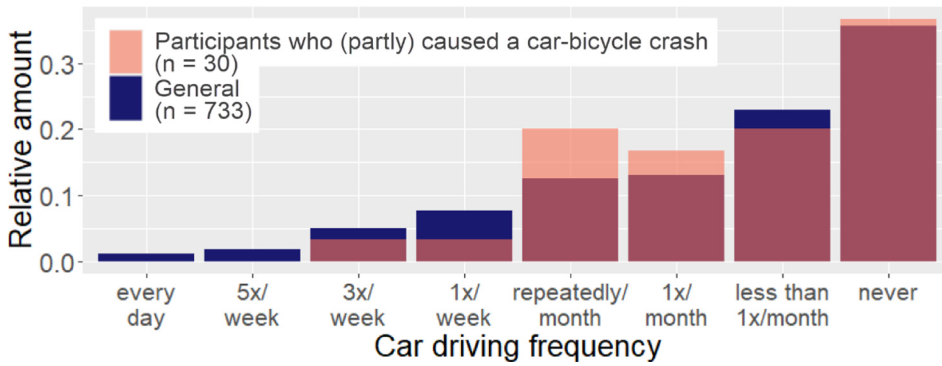

Fig. 7. Comparison of the car driving frequency of participants in general and that of cyclists who have once (at least partially) caused a car-bicycle crash (the dark red part shows the overlap in the shares).

Using the underreporting rates determined in the survey, the total police-reported crash costs with known location can be compared to those unknown. Table 3.2 shows the known crash costs of police-reported crashes and a projection of the actual total crash costs.

The spatial allocation remains unknown for approximately CHF 520 M (1.7 times the value of the police-reported crashes if solely the police-reported crash data are analyzed. The economic crash costs of non-police-reported crashes amounted to CHF 22,845,603 in the 2018 survey. Thus, in this case, an increase in economic costs corresponding to one third (34%) of the police-reported crash costs in one year (2018: CHF 67,548,074) can be mapped with the help of the survey.

These results suggest that if more information resources (e.g., survey data) were included in the crash analysis, the priorities for infrastructure improvements with respect to cycling would be significantly impacted.

3.2. Demographic and behavioral factors

With crash data reported by the police alone, it is difficult to analyze correlations between the frequency of crashes and the demographic or behavioral factors of cyclists. By conducting a survey, however, it is possible to analyze some of these aspects.

Fournier et al. (2020) found in their research that car drivers who cycle more frequently or are more familiar with certain cycling infrastructure pay more attention to cyclists. It is therefore reasonable to hypothesize that a certain percentage of cyclists who blame themselves for a collision with a car have little to no driving experience. However, the survey data did not confirm this hypothesis; there was no significant discrepancy. The average car driving frequencies of the participants[2] and those of participants who (at least partly) caused a car-bicycle crash show a high similarity (Fig. 7).

In the survey, the respondents were asked to estimate how frequently they have cycled since the crash on a slider that could be moved from "I no longer cycle" to "I cycle as often as I used to." The slider was mapped to a scale from 1 to 101. Of the participants, 85% answered that they had not changed their cycling habits after the crash in question (Fig. 8).

The characteristics of the cycling crashes after which the cyclists rode less frequently did not differ significantly from those after which the cyclists did not change their riding habits. The same applies to the characteristics of the cyclists themselves.

The participants were also asked whether or not they were familiar with the route on which the crash occurred or not. Fig. 9 illustrates that the crashes in which the participants stated that they had not frequently ridden the route (orange) contained significantly more single-vehicle crashes than those in which the cyclist was familiar with the route (blue).

The survey data contained various attributes describing personal characteristics and habits of the participants. Finding correlations between these attributes and the number of cycling crashes could be valuable for creating and directing preventive measures. Therefore, a Random Forest approach was selected to determine which characteristics were more likely to serve as explanatory variables for the crash frequency. Then, the two-dimensional relationships of the variables that were more likely to be an explanatory variable for the number of crashes were examined. Random Forest was chosen for its property having a built-in feature selection process, to handle nonlinear relationships and interactions between variables and thus being well-suited for identifying relevant variables in a complex environment like urban cycling. Especially the variable importance measures generated by Random Forest provide a more comprehensive understanding of the relative importance of the predictors, enabling us to focus on the most critical variables that contribute to the outcome of interest. As input variables almost all demographic and behavioral attributes collected about the participants were used (see Appendix Table 7.1) (excluded were respondent ID, duration of response, date of response, responses regarding Dutch Reach, numbers of cycling crashes), as Random Forest is robust to noisy or irrelevant variables due to the property of using only a subset when building the decision trees and thus irrelevant variables averaging themselves out automatically. The Random Forest approach was conducted twice. Once including all of the attributes as they were collected (see Appendix Table 7.1), and once trying the approach with categorical subgroups for the attributes "Condition in which the respondent does still feel capable of cycling," "Frequency of listening to music while cycling," and "Frequency of using the smartphone while cycling," which were collected on a continuous scale from 1 to 11. For the second approach they were categorized in three groups (1 to 4, 5 to 7 and 8 to 11).

In addition, different combinations of these variables were used as input variables for cluster analysis. To find clusters between different mixed input variables (one of which is always the number of crashes either in the urban area or in Zurich), the clustering algorithm PAM (Partitioning Around Medoids) was employed, using the R code of Filaire (2018) as a basis. The analysis did not reveal any clearly separated clusters nor other significant relationships. The results hence indicate that the number of crashes involving

[2] Due to a technical error, not all participants were enabled to answer this question. The question was only activated during the last third of the survey period.

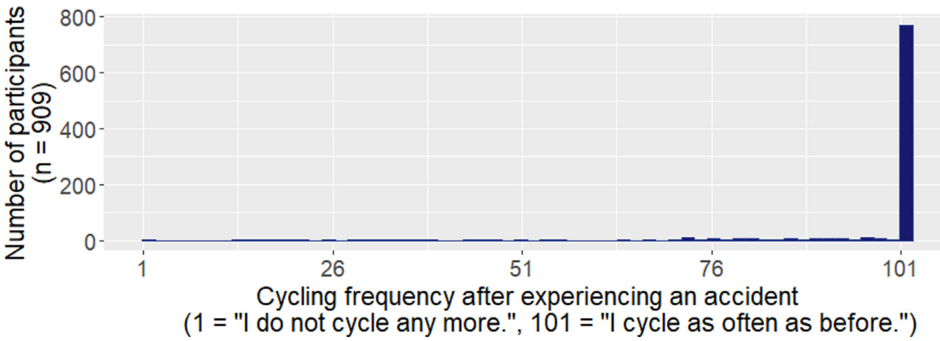

Fig. 8. Frequency of cycling after having experienced a bicycle crash.

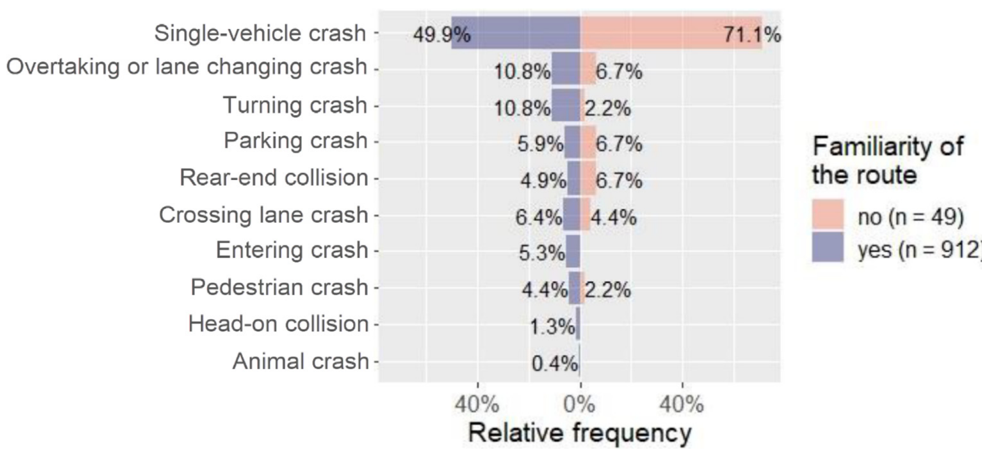

Fig. 9. Relative frequency of crash type groups with respect to familiarity of the route.

cyclists in cities does not significantly correlate with any one or any combination of the recorded characteristics of cyclists.

Another aspect that can be achieved only through a combination of surveys and police-reported data is the comparison of places that are perceived to be dangerous with places that with increased crash frequency in reality. This analysis is described in the next section.

3.3. Risk perception and actual crashes

Understanding the risk perception of cyclists is an indicator that urban planners can use to understand which configurations should be avoided in order to make cycling more attractive. A comparison between perceived and actual risk gives one more insight into the psychology of cyclists' risk perception.

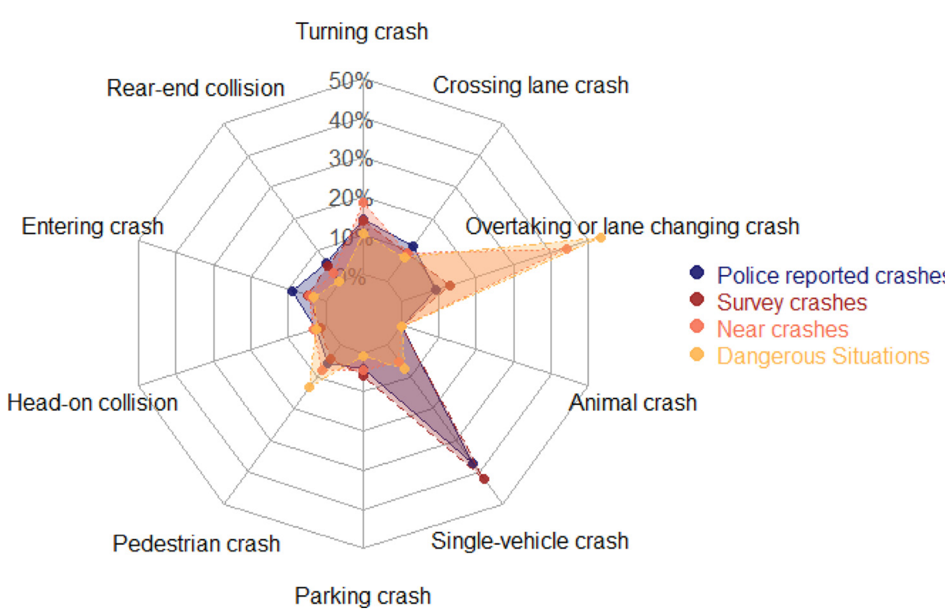

Fig. 10. Comparison of the frequency of the crash type groups (n (police data) = 4′305, n (survey crashes) = 1′290, n (almost crashes) = 1′247, n (dangerous situations) = 6′131).

The comparison of the crash type groups of actual crashes, near crashes and situations perceived as dangerous shows significant discrepancies (Fig. 10). Actual crashes peak at single-vehicle crashes in 37% (police reports) and 42% (survey) of the cases. Near crashes and perceived dangerous situations, however, peak in the category "overtaking or lane change crash" in 44% and 53%, respectively. These results indicate that there seems to be a large difference between perceived risk and actual risk for these two types of crashes. Pedestrian and turning crashes also show a tendency to greater perceived discomfort.

3.4. Summary

Using surveys in addition to police reports provides a more complete picture of bicycle crashes. The proportions of crash types, main crash causes, and crash locations differ significantly from considering only police-reported crashes. Surveys also enable testing of hypotheses related to demographic or behavioral characteristics of cyclists and inclusion of risk perception in crash analysis.

For example, in the city of Zurich, the number of cycling crashes not reported to the police was estimated at 86%. Using a combined data set, the three primary crash locations are completely different from the three primary crash locations of the police-reported data. Further, crashes with main causes "acute-angled crossing of tram tracks" and "failure to adapt to road conditions (wet, icy, gravel, leaves, etc.)" were severely underestimated when solely the police-reported bicycle crashes were used.

In the analysis, the hypothesis that cyclists who never or rarely drive a car are more likely to cause a cycling crash with a car could not be confirmed. Additionally, no significant correlation between demographic or behavioral characteristics of cyclists and their crash frequency could be identified. The situations most frequently perceived as dangerous by cyclists were being passed by a car or being followed by one while changing lanes.

4. Discussion

Collecting information on cycling crashes with surveys helps to obtain a more holistic picture of cycling crash occurrence and also enables specific research questions to be answered. The following three sections go into detail about the findings of this study and its limitations, and draws a conclusion.

4.1. Findings

Research questions examined in the present study included correlations between crash occurrence and certain demographic or behavioral characteristics as well as perceptions of danger. Specifically, in the city of Zurich, the number of crashes with the main causes "acute-angled crossing of tram tracks" and "failure to adapt to road conditions (wet, icy, gravel, foliage, etc.)" was greatly underestimated in the police-reported data. Combining non-police-reported and police-reported crashes highlighted key crash locations that appeared to be less important when only police-reported crashes were considered.

Statistically, 42% of survey crashes that resulted in serious injuries were not reported to the police. This result is in line with other studies that compared hospital crash records with police crash records, which found similarly high numbers of unreported bicycle crashes (Hels & Orozova-Bekkevold, 2007; Juhra, et al., 2012; Meuleners, et al., 2019; Meuleners, et al., 2020). As the prioritization of infrastructure renewal is partially based on the results of crash analysis, more studies should be done to gain insights into non-police-reported crashes. This study demonstrated that a combined dataset of police and survey crash data reveals a significantly different distribution of crash-related economic costs.

The situations most frequently perceived as dangerous were overtaking and lane changing; however, the type of crash that occurs most frequently in Zurich is a single-vehicle crash. One attempt to explain this seemingly paradoxical phenomenon lies in the psychological effect of expectation. When a cyclist is closely overtaken or followed by another vehicle, the potential danger to them is obvious. Therefore, higher vigilance due to conscious knowledge of a danger could lead to a lower probability of a crash in this case. Conversely, single-vehicle crashes could more often represent situations in which the cyclist does not anticipate a hazard and is therefore less alert, which could consequently lead to a higher probability of a crash. Further research is needed to confirm this theory.

Regarding the ambition to increase the use of cycling as a mode of transport in cities, the results suggest that reducing the likelihood of overtaking and lane-changing crashes would increase the appeal of cycling. To reduce the number of single-vehicle crashes, research should focus on better understanding the reasons for the main cause "other influences related to inattention and distraction." In addition, the analysis suggests that prevention measures should target driving under the influence of alcohol, the improvement of cycling infrastructure on roads with tram tracks, and the alleviation of slippery conditions for cyclists. The study furthermore shows, that significantly more single vehicle crashes occurred on routes cyclists were not familiar with, which indicates that traffic guidance for cyclists in the city of Zurich has potential for improvement.

4.2. Limitations

While surveys can provide information on a more representative sample of crashes, police-reported data provide information on a sample of cycling crashes for which the police were present in some form. Therefore, the data sets are inherently different. Due to inherent subjectivity in survey data, the evaluation of the police dataset seems to promise to be more objective but is limited to the information that police officers are able to collect at the scene of a crash.

The biggest shortcoming of the crash data survey presented is that it does not contain dates of the crashes; therefore, accurate comparisons over time were not possible. Furthermore, the panel of respondents to a survey is influenced by the channels through which they are reached. In this study, the channels used included Pro Velo, Velo Plus and bikeable.ch, which reach an audience of cyclists more actively engaged in cycling related topics. Additionally, cyclists under 20 and over 60 years old are underrepresented. It remains unclear whether or not the results would have been significantly impacted if these age groups had been better represented. Furthermore, since the survey questions were quite scarce concerning socioeconomic details of the participants (in order to get a lot of people answering about the crash locations), it is not possible to judge the representation of different demographic groups.

One significant difference between the data sets that should be kept in mind for all analyses is that a police accident data set can contain fatal crashes, while a survey data set cannot include fatal crashes. Only survivors would be able to respond. In the used police crash data set fatal and serious crashes were already combined beforehand under the category "serious injury" for purpose of information privacy. There were very few fatal crashes in the set of crashes that caused serious injury (about 1 to 2%).

It should also be noted that in the extrapolation in 3.1.3, it was assumed that the non-police-reported crashes for which the location was not given occurred with the same spatial distribution as the non-police-reported crashes reported in the survey for which a location was given.

The sample size of 834 non-police-reported bicycle crashes provided a substantial basis for comparing non-police-reported and police-reported bicycle crashes. To determine whether or not there might be any bias in the types of crashes that participants would be more willing to report in a survey, a comparison between police-reported crashes in the survey and police-reported crashes in general would be interesting. However, in this study, it was assumed that with only 129 police-reported crashes in the survey, profound conclusions in this regard could not be drawn.

4.3. Conclusion

The results of this study align very much with other studies that show that there is a high number of unreported cycling crashes, and that not only slight injury crashes fail to be reported, but also quite a high share of crashes that result in serious injury (Isaksson-Hellman & Werneke, 2017; Meuleners et al., 2019; Meuleners et al., 2020; Juhra et al., 2012; Hels & Orozova-Bekkevold, 2007; Winters & Branion-Calles, 2017; Shinar et al., 2018). It adds to that knowledge that unreported cycling crashes appear to not have the same spatial distribution nor leading causes as the police reported crashes. Thus, the belief of traffic safety officers, that despite only knowing about 10% of the accidents at least through knowing details about those the most dangerous spots and causes are known, appears to be very wrong. This study shows that infrastructure safety officers would focus their work significantly different - spatially and with respect to general prevention measures - had they more information about unreported cycling crashes. The results thus also align with the conclusions of other studies that insurance and hospital data can complement police crash data with a significant value (Juhra et al., 2012; Isaksson-Hellman & Werneke, 2017; Shinar et al., 2018; Hels & Orozova-Bekkevold, 2007; Meuleners et al., 2020). However, solutions still have to be found for the lack of detailed information about crash sites and circumstances as well as for the difficulties in connecting the data due to considerable differences in the data sets of hospitals and insurances. This study adds the valuable approach to collect detailed information about crash sites and circumstances with surveys. The results add an insight into the juxtaposition of risk perception and actual crashes.

5. Further research potential

The presented method enables the estimation of the number of unknown crashes, and shows their characteristics, which provides a better understanding on where to best focus preventative measures. In addition, the method has shown promise in gathering more general information about crash sites and situations that cause cyclists to feel uncomfortable. However, the approach only includes police and survey data. To get an even broader overview, it is recommended to combine other data sets that are available in this area for further research, including police data, hospital data, insurance data, and data from bicycle companies (e.g. bike sharing) (Fig. 11.).

A further approach that seems promising for additional investigations is the use of social media to mine crash data. Milusheva *et al.* (2021) analyzed posts on Twitter in Nairobi to collect crash locations, as Twitter is used extensively by road users in the region to communicate traffic information due to poor road infrastructure. To determine if a similar approach is potentially applicable, social media channels should be scanned to understand where, how, and to what extent bicycle crash information is available.

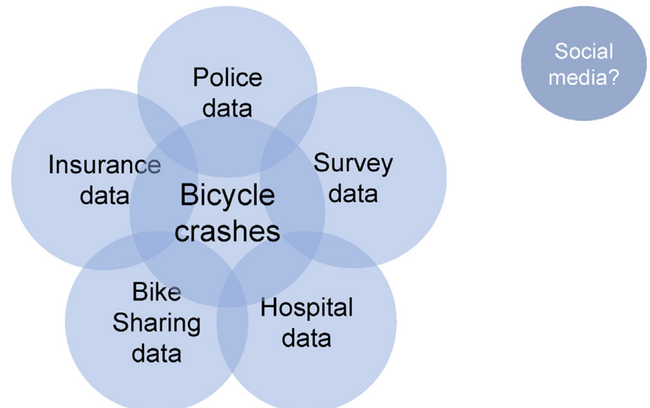

Fig. 11. Future vision for bicycle crash data collection.

For future surveys, to improve analysis efficiency and precision, it is recommended to use radio buttons that allow participants to select the crash type (group) and the main cause of the crash from a range of options similar to those used by the Federal Road Authority. The presented results show an example of the distribution of crash types and their main causes. It is therefore advisable to use this knowledge to design a differentiated questionnaire on crash type groups, crash types, and main crash causes. It is also recommended to ask for the year of the crash occurrence. A further interesting factor would be knowing the purpose of the trip on which the crash happened (work, kindergarten, shopping, etc.). This could also help the memory.

When advertising the survey, the goal was to get as many participants as possible. Limitations imposed were the timing at the beginning of winter (due to the start of the Master's thesis) and the available time span of 3 months. Due to the advertising channels chosen, participants were mainly cyclists who cycle very frequently, while the younger and older generations were underrepresented. For future surveys, it is recommended to focus on a more representative age distribution and greater diversity in terms of average frequency of cycling. The latter can be particularly interesting when analyzing demographic or behavioral trends. The authors believe that greater diversity can be achieved by advertising for an entire year on bicycle stands, in children's cycling classes (where parents and children of different ages are present), at bicycle markets, and in bicycle shops.

It would be worthwhile for future studies to compare crash locations having similar characteristics, for example the geometry of curbs, markings, signs, roads, buildings, traffic density, traffic flow, and HGV traffic. One aspiration of transport planning is to design infrastructure in a way that is intuitively understandable to the user. Thus, it would be valuable information for crash prevention to find reasons for unfavorable behavior patterns that are triggered by certain design features.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank the Head of Traffic Safety of the City of Zurich, Dr. Wernher Brucks for his support during the research project.

Appendix

One unidentified factor in the reported bicycle crashes in the city of Zurich are single bicycle crashes with the main cause "other influence due to inattention or distraction." In order to meaningfully target prevention measures, further research should be conducted to better subdivide this group. Interviews with police officers who regularly collect information on crash sites, examination of crash sketches, and comparisons with other cities are possible approaches.

In this study, a clear difference between risk perception and actual risk was found across the different types of crashes in Zurich. The results raise the question of whether or not cities with different cycling infrastructure exhibit the same phenomenon. In general, comparisons of different cities in terms of the number of unknown crashes and cyclists' perception of risk could provide a basis for measuring the success of crash prevention strategies and cycling infrastructure planning.

Finally, a further question of interest is whether or not the number of actual crashes would decline were the number of locations perceived to be dangerous to decrease, because bicyclists would spread their vigilance more evenly over the entire route (rather than increasing their vigilance at the locations they perceive to be most dangerous).

Attributes gathered about the participants

(See Table 7.1).

Table 7.1
Demographic and behavioural attributes and entries of the questionnaire used in the survey conducted in Zurich.

Attribute	Entry
Respondent-ID	*numerical*
Screen width	*numerical*
Age	*numerical*
Type of bicycle(s) the respondent drives regularly	■ Racing bicycle ■ Mountain bike ■ City bike ■ E-bike ($\leq$ 25 km/h) ■ E-bike ($\leq$ 45 km/h) ■ Trekking bicycle ■ Cargo bike ■ Tandem ■ Other *(text)*
Frequency of cycling	■ Every day ■ 5x/week ■ 3x/week ■ 1x/week ■ Repeatedly/month ■ Less than 1x/month
Question if respondents view the bicycle as:	■ Means of transportation ■ Recreational means ■ Both
Age of learning how to ride a bike	■ $\leq$6 years old ■ 7–12 years old ■ 12–20 years old ■ 20–50 years old ■ >50 years old
Response to the question if respondents own a driver's license	■ Yes ■ No
Frequency of car driving	■ Every day ■ 5x/week ■ 3x/week ■ 1x/week ■ Repeatedly/month ■ Less than 1x/month ■ Never
Cycling distance on a typical weekend day	■ none ■ <1 km ■ 1–2 km ■ 2–4 km ■ 4–8 km ■ 8–16 km ■ 16–30 km ■ 30–48 km ■ 48–64 km ■ >64 km

Table 7.1 (*continued*)

Attribute	Entry
Cycling distance on a typical weekday	■ none ■ <1 km ■ 1–2 km ■ 2–4 km ■ 4–8 km ■ 8–16 km ■ 16–30 km ■ 30–48 km ■ 48–64 km ■ > 64 km
Response to the question for which purposes the respondent chooses to take the bike	■ Commute (complete, halfway, not) ■ Leisure trip (complete, halfway, not) ■ Way to school/university (complete, halfway, not) ■ Way to the supermarket (complete, halfway, not)
Weather conditions in which the respondent does not like to or does not cycle	■ Windy ■ Wet streets ■ Icy streets ■ Snowy streets ■ Shower ■ Rain ■ Snow ■ Thunderstorm ■ Hot temperatures ($\geq$ 27°C) ■ Cold temperatures ($\leq$ 5°C) ■ Commuter traffic
Light conditions in which respondent does not like to or does not cycle	■ Twilight ■ Darkness, with streetlights ■ Darkness, without streetlights
Condition in which the respondent does still feel capable of cycling	■ *numerical (1: "Only fresh as a daisy" - 11: "Very tired")* ■ *numerical (1: "Only sober" - 11: "Very drunk")*
Frequency of listening to music while cycling	■ *numerical (1: "Never"- 11: "Always")*
Frequency of using the smartphone while cycling	■ *numerical (1: "Never"- 11: "Always")*
Response to the question if the respondent wears a bicycle helmet or not, and how often	■ Yes, always ■ Yes, often ■ Yes, occasionally ■ Yes, seldom ■ No, never
Situations in which the respondent does not wear a helmet while cycling	*text*
Response to the question if the respondent lives in the City of Zurich	■ Yes ■ No(If not, where? *text*)
Response to the question if the respondent commutes regularly to Zurich	■ Yes ■ No
Number of cycling crashes	*numerical*
Number of cycling crashes downhill or mountain bike trails	*numerical*
Number of cycling crashes in the City of Zurich	*numerical*
Number of near cycling crashes in the City of Zurich	■ Every day ■ Repeatedly/week ■ 1x/week ■ Repeatedly/month ■ 1x/month ■ Repeatedly/year ■ 1x/year or less ■ Never before
Response to the question if there are places where the respondent perceives danger when cycling in Zurich	■ Yes ■ No
Response to the question if the respondent believes that the Dutch Reach is a good idea	■ Yes ■ No(If not, why? *text*)
Response to the question if respondent believes the Dutch Reach should be integrated in driving school lessons	■ Yes ■ No(If not, why? *text*)
Duration of response in seconds	*numerical*
Date and time of response	*numerical*

Attributes gathered about the spots

(See Table 7.2).

Table 7.2
Spatial attributes and entries of the questionnaire used in the survey conducted in Zurich.

Attribute	Entry
Respondent-ID	*numerical*
ID of location	*numerical*
Type of location	■ Dangerous spot ■ Observed crash ■ Near crash ■ Personally experienced crash
Coordinates	*numerical*
Description of dangerous spots	*text*
Constellation of personally experienced crashes	■ Fall ■ Collision with a car ■ Collision with a pedestrian ■ Collision with another cyclist ■ Collision with a motorcyclist ■ Collision with a tram ■ Collision with a bus ■ Collision with a truck
Crash description	*text*
Response to the question if the crash was registered by the police or not	■ Yes ■ No
Assignment of blame	■ Personal negligence ■ Route guidance ■ Signage ■ Third party negligence
Assessment of how much cycling frequency changed after the crash	■ *Numerical (1: "I do not cycle anymore" - 101: "I cycle as often as before")*
Crash severity	■ Slightly injured ■ Seriously injured ■ Killed ■ Only property damage ■ No harm or damage
Response to the question if the respondent drove the way on which he or she experienced the crash already repeatedly before or not	■ Yes ■ No
Response to the question what the respondent would like to change on the spot if he or she could	*text*

References

Akgün, N., Daniels, S., Bell, M. C., Nuyttens, N., Thorpe, N., & Dissanayake, D. (2021). Exploring regional differences in cyclist safety at roundabouts: A comparative study between the UK (based on Northumbria data) and Belgium. *Accident Analysis & Prevention, 150*. https://doi.org/10.1016/j.aap.2020.105902 105902.

BFU. (2020). Sinus 2020, Sicherheitsniveau und Unfallgeschehen im Strassenverkehr 2019.

Blanc, B., & Figliozzi, M. (2016). Modeling the impacts of facility type, trip characteristics, and trip stressors on cyclist_comfort levels utilizing crowdsourced data. *Transportation Research Record, 2587*, 100–108.

Branion-Calles, M., Götschi, T., Nelson, T., Anaya-Boig, E., Avila-Palencia, I., Castro, A., ... Winters, M. (2020). Cyclist crash rates and risk factors in a prospective cohort in seven European cities. *Accident Analysis & Prevention, 141*. https://doi.org/10.1016/j.aap.2020.105540 105540.

Buch, T. S., & Jensen, S. U. (2017). Incidents between straight-ahead cyclists and right-turning motor vehicles at signalised junctions. *Accident Analysis & Prevention, 105*, 44–51. https://doi.org/10.1016/j.aap.2016.07.035.

Chen, C., Wang, H., Roll, J., Nordback, K., & Wang, Y. (2020). Using bicycle app data to develop Safety Performance Functions (SPFs) for bicyclists at intersections: A generic framework. *Transportation Research Part A: Policy and Practice, 132*, 1034–1052. https://doi.org/10.1016/j.tra.2019.12.034.

de Waard, D., Westerhuis, F., & Lewis-Evans, B. (2015). More screen operation than calling: The results of observing cyclists' behaviour while using mobile phones. *Accident Analysis & Prevention, 76*, 42–48. https://doi.org/10.1016/j.aap.2015.01.004.

Filaire, T. (2018). Clustering on mixed type data – A proposed approach using R. Retrieved from https://towardsdatascience.com/clustering-on-mixed-type-data-8bbd0a2569c3.

Fournier, N., Bakhtiari, S., Valluru, K. D., Campbell, N., Christofa, E., Roberts, S., & Knodler, M. (2020). Accounting for drivers' bicycling frequency and familiarity with bicycle infrastructure treatments when evaluating safety. *Accident Analysis & Prevention, 137*. https://doi.org/10.1016/j.aap.2019.105410 105410.

Fyhri, A., Johansson, O., & Bjørnskau, T. (2019). Gender differences in accident risk with e-bikes—Survey data from Norway. *Accident Analysis & Prevention, 132*. https://doi.org/10.1016/j.aap.2019.07.024 105248.

Gustafsson, L., & Archer, J. (2013). A naturalistic study of commuter cyclists in the greater Stockholm area. *Accident Analysis & Prevention, 58*, 286–298. https://doi.org/10.1016/j.aap.2012.06.004.

Hels, T., & Orozova-Bekkevold, I. (2007). The effect of roundabout design features on cyclist accident rate. *Accident Analysis & Prevention, 39*, 300–307. https://doi.org/10.1016/j.aap.2006.07.008.

Hoglund, M. W. (2018). Safety-oriented bicycling and traffic accident involvement. *IATSS Research, 42*, 152–162. https://doi.org/10.1016/j.iatssr.2017.10.004.

Isaksson-Hellman, I., & Werneke, J. (2017). Detailed description of bicycle and passenger car collisions based on insurance claims. *Safety Science, 92*, 330–337. https://doi.org/10.1016/j.ssci.2016.02.008.

Jannat, M., Hurwitz, D. S., Monsere, C., & Funk, K. H. (2018). The role of driver's situational awareness on right-hook bicycle-motor vehicle crashes. *Safety Science, 110*, 92–101. https://doi.org/10.1016/j.ssci.2018.07.025.

Jensen, S. U. (2017). Safe roundabouts for cyclists. *Accident Analysis & Prevention, 105*, 30–37. https://doi.org/10.1016/j.aap.2016.09.005.

Juhra, C., Wieskötter, B., Chu, K., Trost, L., Weiss, U., Messerschmidt, M., ... Raschke, M. (2012). Bicycle accidents – Do we only see the tip of the iceberg? A prospective multi-centre study in a large German city combining medical and police data. *Injury, 43*, 2026–2034. https://doi.org/10.1016/j.injury.2011.10.016.

Karakaya, A.-S., Hasenburg, J., & Bermbach, D. (2020). SimRa: Using crowdsourcing to identify near miss hotspots in bicycle traffic. *Pervasive and Mobile Computing, 67*. https://doi.org/10.1016/j.pmcj.2020.101197 101197.

Kielhauser, C., Herrmann, J. S., & Adey, B. T. (2020). Identifying changes in bicycle accident trends using GIS and time series information in the city of Zürich. *Frontiers in Sustainable Cities, 2*, 21. https://doi.org/10.3389/frsc.2020.00021.

Luan, S., Li, M., Li, X., & Ma, X. (2020). Effects of built environment on bicycle wrong Way riding behavior: A data-driven approach. *Accident Analysis & Prevention, 144*. https://doi.org/10.1016/j.aap.2020.105613 105613.

Meuleners, L. B., Fraser, M., Johnson, M., Stevenson, M., Rose, G., & Oxley, J. (2020). Characteristics of the road infrastructure and injurious cyclist crashes resulting in a hospitalisation. *Accident Analysis & Prevention, 136*. https://doi.org/10.1016/j.aap.2019.105407 105407.

Meuleners, L. B., Stevenson, M., Fraser, M., Oxley, J., Rose, G., & Johnson, M. (2019). Safer cycling and the urban road environment: A case control study. *Accident Analysis & Prevention, 129*, 342–349. https://doi.org/10.1016/j.aap.2019.05.032.

Nelson, T. A., Denouden, T., Jestico, B., Laberee, K., & Winters, M. (2015). BikeMaps.org: A global tool for collision and near miss mapping. *Frontiers Public Health, 3*, 53. https://doi.org/10.3389/fpubh.2015.00053.

Nordback, K., Kothuri, S., Marshall, W., Gibson, G., & Ferenchak, N. (2017). Improving Bicycle Crash Prediction for Urban Road Segments.

Nordback, K., Marshall, W. E., & Janson, B. N. (2014). Bicyclist safety performance functions for a U.S. city. *Accident Analysis & Prevention, 65*, 114–122. https://doi.org/10.1016/j.aap.2013.12.016.

Pai, C.-W. (2011). Overtaking, rear-end, and door crashes involving bicycles: An empirical investigation. *Accident Analysis & Prevention, 43*, 1228–1235. https://doi.org/10.1016/j.aap.2011.01.004.

Pokorny, P., Drescher, J., Pitera, K., & Jonsson, T. (2017). Accidents between freight vehicles and bicycles, with a focus on urban areas. *Transportation Research Procedia, 25*, 999–1007. https://doi.org/10.1016/j.trpro.2017.05.474.

Richardson, M., & Caulfield, B. (2015). Investigating traffic light violations by cyclists in Dublin City Centre. *Accident Analysis & Prevention, 84*, 65–73. https://doi.org/10.1016/j.aap.2015.08.011.

Sakshaug, L., Laureshyn, A., Svensson, Å., & Hydén, C. (2010). Cyclists in roundabouts—Different design solutions. *Accident Analysis & Prevention, 42*, 1338–1351. https://doi.org/10.1016/j.aap.2010.02.015.

Sayed, T., Zaki, M. H., & Autey, J. (2013). Automated safety diagnosis of vehicle–bicycle interactions using computer vision analysis. *Safety Science, 59*, 163–172. https://doi.org/10.1016/j.ssci.2013.05.009.

Schmid, H. (1995). TreeTagger - a part-of-speech tagger for many languages. Retrieved from https://www.cis.uni-muenchen.de/ schmid/tools/TreeTagger/.

Schepers, P., Agerholm, N., Amoros, E., Benington, R., Bjørnskau, T., Dhondt, S., ... Niska, A. (2015). An international review of the frequency of single-bicycle crashes (SBCs) and their relation to bicycle modal share. *Injury Prevention, 21* (e1), e138–e143. https://doi.org/10.1136/injuryprev-2013-040964.

Shinar, D., Valero-Mora, P., van Strijp-Houtenbos, M., Haworth, N., Schramm, A., Bruyne, G. D., ... Tzamalouka, G. (2018). Under-reporting bicycle accidents to police in the COST TU1101 international survey: Cross-country comparisons and associated factors. *Accident Analysis & Prevention, 110*, 177–186. https://doi.org/10.1016/j.aap.2017.09.018.

Stephens, A. N., O'Hern, S., Trawley, S., Young, K. L., & Koppel, S. (2019). Self-reported aggression amongst active cyclists. *Accident Analysis & Prevention, 128*, 46–52. https://doi.org/10.1016/j.aap.2019.04.004.

Stutts, J. C., Williamson, J. E., Whitley, T., & Sheldon, F. C. (1990). Bicycle accidents and injuries: A pilot study comparing hospital- and police-reported data. *Accident Analysis & Prevention, 22*, 67–78. https://doi.org/10.1016/0001-4575(90)90008-9.

Warner, J., Hurwitz, D. S., Monsere, C. M., & Fleskes, K. (2017). A simulator-based analysis of engineering treatments for right-hook bicycle crashes at signalized intersections. *Accident Analysis & Prevention, 104*, 46–57. https://doi.org/10.1016/j.aap.2017.04.021.

Winters, M., & Branion-Calles, M. (2017). Cycling safety: Quantifying the under reporting of cycling incidents in Vancouver, British Columbia. *Journal of Transport & Health, 7*, 48–53. https://doi.org/10.1016/j.jth.2017.02.010.

Wu, K.-F., & Lin, P.-J. (2020). Using on-board video data for safety analysis – An analysis of right hook crashes involving large buses at signalized intersections. *IATSS Research*. https://doi.org/10.1016/j.iatssr.2020.06.005.

Laura Ringel MSc. Laura Ringel is a traffic safety specialist at the Cantonal Office for Civil Engineering in Zurich. She holds a Master's Degree in Economics of Transport and researched at the Swiss Federal Institute of Technology (ETHZ) in the field of cycling safety. Her research is based on various statistical analyses such as text analysis, multivariate analysis and spatial analysis in GIS.

Dr. Clemens Kielhauser. Dr. Clemens Kielhauser is a Postdoctoral Researcher with demonstrated experience working in industrial and academic environments, such as Carmentae Infrastructure Management or at the Swiss Federal Institute of Technology (ETHZ). He also advises the Global Infrastructure Hub (a G20 Initiative) as a member of the Academic Advisory Panel. His current research areas include Bicycle Safety, Data-driven Infrastructure Management, Data Mining and Machine Learning.

Prof. Dr. Bryan T. Adey. Prof. Dr. Bryan T. Adey is the Professor for Infrastructure Management, in the Institute for Construction and Infrastructure Management of the Department of Civil, Environmental and Geomatic Engineering, at the Swiss Federal Institute of Technology in Zürich (ETHZ), Switzerland. Prof. Adey's research is focused on improving the effectiveness and efficiency of infrastructure management. This includes the definition and standardization of the infrastructure management process, from setting goals to determining optimal monitoring and intervention programs, and evaluating the performance of infrastructure management organisations.