G Model
AAP-4137; No. of Pages 9
ARTICLE IN PRESS
Accident Analysis and Prevention xxx (2016) xxx–xxx
Contents lists available at ScienceDirect
Accident Analysis and Prevention
journal homepage: www.elsevier.com/locate/aap

Crash risk: How cycling flow can help explain crash data

Marco Dozza

Chalmers University of Technology, SAFER, Lindholmspiren 3, Floor 2, 417 56 Göteborg, Sweden

ARTICLE INFO

Article history:
Received 18 November 2015
Received in revised form 16 March 2016
Accepted 27 April 2016
Available online xxx

Keywords:
Crash risk
Crash causation
Road user behavior
Exposure
Single-bicycle crashes

ABSTRACT

Crash databases are commonly queried to infer crash causation, prioritize countermeasures to prevent crashes, and evaluate safety systems. However, crash databases, which may be compiled from police and hospital records, alone cannot provide estimates of crash risk. Moreover, they fail to capture road user behavior before the crash. In Sweden, as in many other countries, crash databases are particularly sterile when it comes to bicycle crashes. In fact, not only are bicycle crashes underreported in police reports, they are also poorly documented in hospital reports. Nevertheless, these reports are irreplaceable sources of information, clearly highlighting the surprising prevalence of single-bicycle crashes and hinting at some cyclist behaviors, such as alcohol consumption, that may increase crash risk.

In this study, we used exposure data from 11 roadside stations measuring cyclist flow in Gothenburg to help explain crash data and estimate risk. For instance, our results show that crash risk is greatest at night on weekends, and that this risk is larger for single-bicycle crashes than for crashes between a cyclist and another motorist. This result suggests that the population of night-cyclists on weekend nights is particularly prone to specific crash types, which may be influenced by specific contributing factors (such as alcohol), and may require specific countermeasures. Most importantly, our results demonstrate that detailed exposure data can help select, filter, aggregate, highlight, and normalize crash data to obtain a sharper view of the cycling safety problem, to achieve a more fine-tuned intervention.

1. Introduction

In Western Europe, cycling is increasing, as is the availability of novel technologies to monitor and analyze this popular, healthy, and environmentally friendly activity. These technologies, unavailable just a decade ago, enable collection of real-world data from instrumented bicycles (Mohanty et al., 2014), instrumented cyclists (Gustafsson and Archer, 2013) or instrumented infrastructure (Sayed et al., 2013). These new data promise to inform city planning, road regulations, and bicycle design, so that the cycling community can continue to grow without compromising safety.

The potential of the new data becomes even more exciting when different types of datasets are combined to overcome their intrinsic, individual limitations. Meta-analyses, combining datasets of different natures, have been carried on for a long time (Attewell et al., 2001); however, the availability of novel cycling data from the real world allows new and more complex combinations of datasets. Data can be combined in a rawer format than in traditional meta-analyses and synchronized on different time scales. This paper combines crash databases and cycling flow data from the real world to estimate crash risk, i.e., the ratio between crash frequency and exposure. Crash risk, not available from crash databases alone, can be calculated using different exposure measures (such as the number of miles driven by an average driver) which are often available on a yearly scale. However, more detailed traffic data has greater potential to improve traffic safety analyses (Lord and Mannering, 2010). By using cycling flow data, we could calculate risk on monthly, daily, and hourly time scales.

New combinations of data, such as the one presented in this paper, may help resolve some cycling safety issues which are still open, for instance: single-bicycle crash causation (Niska et al., 2013; Dozza and Werneke, 2014), safety-in-numbers (Bhatia and Wier, 2011; Johnson et al., 2014), helmet use (Attewell et al., 2001; Elvik, 2011), and the relation between crashes and near-crashes in naturalistic data analysis (Dozza et al., 2015) for different crash types. Single-bicycle crashes are particularly important both because they are the largest share of cycling crashes in most countries in Europe (Schepers et al., 2015) and because many human factors may contribute to this wide category of crashes (Schepers and Den Brinker, 2011; Werneke and Dozza, 2013; De Waard and Houwing, 2014). Nevertheless, it is in crashes with motorized vehicles that cyclists are most severely hurt (Bil et al., 2010; Chaurand and Delhomme,

E-mail address: marco.dozza@chalmers.se

http://dx.doi.org/10.1016/j.aap.2016.04.033
0001-4575/

Please cite this article in press as: Dozza, M., Crash risk: How cycling flow can help explain crash data. Accid. Anal. Prev. (2016), http://dx.doi.org/10.1016/j.aap.2016.04.033

2013). This paper considers both single-bicycle crashes and crashes with motorized vehicles.

The purpose of this paper is to show how detailed cyclist flow data (exposure from measuring stations counting each time a cyclist passes) can be combined with crash database data to explain bicycle crash risk, tackle cycling safety issues, and consequently inform the design of countermeasures to bicycle crashes.

2. Methods

Two datasets were used in this study: one including cycling flow data and the other including crash data (from STRADA, the Swedish national accident database). The two datasets were analyzed individually to determine the periodic components of cycling behavior and crash occurrence, and then combined to estimate risk for different crash types over different time windows. Both datasets were constrained in time (years 2012–2014) and space (WGS84 latitude: 57.68–57.735 and longitude: 11.90–12.01, corresponding to Gothenburg city center; Fig. 1).

2.1. Cycling flow data

Cycling flow data was collected from 11 stations in Gothenburg (Fig. 1) and consisted of around 12.6 million passes over three years (Fig. 2; Videos 1–2, Appendix B). The stations continuously record the number of cyclists passing and save this information every 15 min 24/7. The 11 stations are located on the most intensively trafficked bicycle lanes in Gothenburg; however, it is technically possible for a bicycle to cross the area in Fig. 2 without been recorded. A moped passing in front of one of the 11 stations would most likely be counted as a bicycle, but mopeds are not common in Gothenburg and only small mopeds, with a top speed of 25 km/h, are allowed on bicycle lanes.

2.2. Crash data

Bicycle crashes (single-bicycle crashes and crashes between a bicycle and a motorized vehicle) were extracted from STRADA, the Swedish national crash database (Sjöö and Ungerbäck, 2007), for the years 2012–2014 and for the GPS area already mentioned. Single-bicycle crashes are crashes in which the cyclist does not hit any other road user. Crashes between a bicycle and a motorized vehicle occur when the cyclist collides with a car or a truck. Both hospital and police reports were included in the analysis, contributing 481 single-bicycle crashes (468 from hospital reports, 20 from police reports, 7 common to both reports) and 214 crashes between a bicycle and a motorized vehicle (110 from hospital reports, 160 from police reports, 56 common to both reports). Several variables were used for descriptive statistics: gender, injury level, lighting condition, helmet use, and accident type. The purpose of this analysis was to verify previous results and be aware of possible confounders for the temporal and risk analyses that followed. Helmet use and injury reported were only available from hospital reports; the latter included the severity level from the maximum abbreviated injury scale (MAIS). A MAIS3+ (i.e., MAIS equal to, or greater than 3) contains severe to fatal injuries. Only police reports coded weather conditions.

2.3. Temporal analysis

Cyclist flow data were analyzed to identify periodic processes using a fast Fourier transform, which converts time series into the frequency domain and is often used to determine the spectrum of a signal (Rao et al., 2010). This transformation can decompose any time series into simple time series, each with a specific frequency, which can be represented in a spectrum or frequency diagram. Transforming a random time series results in an infinite number of periodic time series, covering all possible frequencies in the whole frequency diagram. The other extreme case is a sine time series, which shows up as a single high value in a frequency diagram. Thus, any high value in the frequency diagram from a Fourier transform identifies a periodic component in the original time series, with the frequency indicated by the diagram. Distributions of cyclist flow were then plotted on yearly, monthly, weekly, and daily scales to determine the origin of the periodic components and relate them to cyclist behavior.

2.4. Risk analysis

Risks for single-bicycle crashes and crashes with motor-vehicles were calculated by dividing the crash rates by number of crashes for the exposure on a monthly, weekly, or hourly scale. The temporal analysis highlighted the fact that cyclist behavior on an hourly scale is similar across all weekdays, and across the days of the weekend, but weekends and weekdays are different from each other. Therefore hourly risk was combined into two categories, weekends and weekdays, for risk analyses. To test statistical significance of the results, the confidence interval around the relative risk was calculated according to Eq. (1) (Rothman 2012) for the relative risk (RR): i.e., the ratio of the crash risk for different time intervals (e.g., weekend vs weekday, or night vs day). In Eq. (1), CI is the confidence interval, RR is the relative crash risk, SE is the standard error, and z is the standard score (z-value). The boundaries of the confidence interval were calculated with Eq. (2), corresponding to a significance level $\alpha = 0.05$. Thus, any time the RR is higher than one and the confidence interval does not include one, there is a statistically significant increase of risk; significant results are marked with '*' in the Results section. It is worth noting that, since the denominators of crash risk are large (because crashes are rare events and cyclists are many), this relative risk approximates an odds ratio.

$$CI = \log(RR) \pm SE \times z \quad (1)$$

$$CI_{\pm} = e^{\log(RR) \pm SE \times 1.96} \quad (2)$$

3. Results

3.1. Descriptive analysis

Single-bicycle crashes comprised 69% of the crashes analyzed, and the other 31% occurred with a motorized vehicle.

Men contributed slightly more than women to the crashes under analysis: 53% vs. 47% for single-bicycle crashes and 57% vs. 43% for crashes between a bicycle and a motor vehicle. Hospitals reported a MAIS3+ injury in 1.7% of the single-bicycle crashes and in 2.3% of the crashes between a bicycle and a motor vehicle. Two cyclists died in crashes with a motor vehicle, but no death resulted from the single-bicycle crashes. From hospital reports, cyclists were reported to be wearing a helmet in 30% of the crashes with motorized vehicles and in 57% of the single-bicycle crashes. From police reports, most of the crashes (80%) happened in dry conditions. Crashes with motorized vehicles happened more often in daylight than did single-bicycle crashes (78% vs 70%).

3.2. Temporal analysis

3.2.1. Cyclist flow

A frequency analysis of the cycling flow data revealed several periodic components (Fig. 3). These components were related to the expected periodicities in cycling behavior due to seasons (over the year) and to commuting times (over the days of the week and the hours of the day). Other faster components were explained by

Please cite this article in press as: Dozza, M., Crash risk: How cycling flow can help explain crash data. Accid. Anal. Prev. (2016), http://dx.doi.org/10.1016/j.aap.2016.04.033