Accident Analysis and Prevention 61 (2013) 129–137

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

An empirical tool to evaluate the safety of cyclists: Community based, macro-level collision prediction models using negative binomial regression

Feng Wei, Gordon Lovegrove*

School of Engineering, Faculty of Applied Science, University of British Columbia, Kelowna, BC Canada

ARTICLE INFO

Article history:
Received 24 October 2011
Received in revised form 5 March 2012
Accepted 9 May 2012

Key words:
Bicycle safety
Cyclists
Macro-level collision prediction models
Generalized linear regression
Negative binomial regression
Bicycle use levels

ABSTRACT

Today, North American governments are more willing to consider compact neighborhoods with increased use of sustainable transportation modes. Bicycling, one of the most effective modes for short trips with distances less than 5 km is being encouraged. However, as vulnerable road users (VRUs), cyclists are more likely to be injured when involved in collisions. In order to create a safe road environment for them, evaluating cyclists' road safety at a *macro* level in a proactive way is necessary. In this paper, different generalized linear regression methods for collision prediction model (CPM) development are reviewed and previous studies on *micro*-level and *macro*-level bicycle-related CPMs are summarized. On the basis of insights gained in the exploration stage, this paper also reports on efforts to develop negative binomial models for bicycle–auto collisions at a community-based, macro-level. Data came from the Central Okanagan Regional District (CORD), of British Columbia, Canada. The model results revealed two types of statistical associations between collisions and each explanatory variable: (1) An *increase* in bicycle–auto collisions is associated with an *increase* in total lane kilometers (TLKM), bicycle lane kilometers (BLKM), bus stops (BS), traffic signals (SIG), intersection density (INTD), and arterial–local intersection percentage (IALP). (2) A *decrease* in bicycle collisions was found to be associated with an *increase* in the number of drive commuters (DRIVE), and in the percentage of drive commuters (DRP). These results support our hypothesis that in North America, with its current low levels of bicycle use (<4%), we can initially expect to see an increase in bicycle collisions as cycle mode share increases. However, as bicycle mode share increases beyond some unknown 'critical' level, our hypothesis also predicts a net safety improvement. To test this hypothesis and to further explore the statistical relationships between bicycle mode split and overall road safety, future research needs to pursue further development and application of community-based, macro-level CPMs.

1. Introduction

Emerging global problems of climate change, peak oil, traffic congestion, and road safety are forcing governments to consider ways to encourage sustainable transportation systems. Sustainable transportation systems, including walking, bicycling, public transit, green vehicles, and car sharing, make more positive contributions to our society, economy and environment than do automobile-dominated transportation systems. The benefits of cycling are generally understood to include: relatively low costs, emissions, and energy use, together with improved health, and convenient parking. Bicycles traveling at an average speed of 15 km/h, are well suited to the short and medium-distance trips. Bicycling is typically the fastest mode for trips less 5 km (Dejister and Schollaert's, 1999). However, as vulnerable road users (VRUs), cyclists are more likely than auto drivers to be injured in traffic collisions. Research on VRU safety at road intersections has addressed many of the impacts of intersection traffic volume and geometric design on VRUs safety (e.g. Ekman, 1996; Leden et al., 2000; Grey et al., 2010). Although these reactive road safety measures have been effective in improving VRU safety on existing facilities, there is still a lack of planning-level assessment empirical tools that forecast safety effects of VRU mode split on the whole traffic stream (e.g. total collisions over all modes). For example, in many North American (NA) communities, the public perceive that bicycling is dangerous, leading to low bicycle mode share, while in many European Union communities, no such public perception exists and cycling mode share is over 30% with some of the lowest total collisions worldwide. Our hypothesis is that increased VRU mode share will lead to decreased total collisions. However, in order to test this hypothesis, we must address the empirical gap, and develop reliable tools that allow our community planners and engineers to proactively forecast the level of road safety of increased VRU use. As a start toward quantifying the road safety benefits of increased bicycle use, this

* Corresponding author. Tel.: +1 250 807 8717; fax: +1 250 807 9850.
E-mail address: gord.lovegrove@ubc.ca (G. Lovegrove).

0001-4575/$ – see front matter © 2012 Published by Elsevier Ltd.
http://dx.doi.org/10.1016/j.aap.2012.05.018

paper presents a modeling technique to predict bicycle collisions at the community-based, macro-level, which can be used as reliable science-based decision aid tools by community planners and engineers.

The three objectives of the research presented in this paper were to:

(1) Conduct a literature review on micro-level and macro-level bicycle collision prediction models (CPMs).
(2) Use negative binomial (NB) regression to develop community-based, macro-level bicycle CPMs using data from the Central Okanagan Regional District (CORD), in BC, Canada.
(3) Recommend statistical and data issues in this model development for future research.

For some time, it has been generally understood and demonstrated by evidence from previous studies suggests that decreasing auto use, accompanied by increased sustainable transport mode splits (e.g. bicycling, walking, or public transit), is at least statistically (if not causally) associated with lower traffic collisions and road fatalities (Osberg and Stiles, 1998; Newman and Kenworthy, 1999; Marshall and Garrick, 2011). Osberg and Stiles (1998) compared bicycle use and road safety in Boston, Paris and Amsterdam, which have similar social-economics, but very different levels of bicycle use: lowest use in Boston, slightly higher cycle use in Paris; and highest use in Amsterdam. They found that Amsterdam had the lowest total road fatalities rate, at 5.8 fatalities/100,000 population (POP), despite experiencing the highest cyclist fatality rate of these three cities with 1.8 fatalities/100,000 POP. Compare these rates to 0.6 cyclist fatalities/100,000 POP, and 0.3 cyclist fatalities/100,000 POP, and 10 total fatalities/100,000 POP and 8.8 total fatalities/100,000 POP for Paris and Boston, respectively. Newman and Kenworthy (1999) reported that Amsterdam and Copenhagen are considered high bicycle use cities among developed countries, yet they suffer only half the road traffic fatality rate of US cities (5.8 fatalities/100,000 POP for Amsterdam, 7.5/100,000 for Copenhagen and 14.6 fatalities/100,000 POP for US cities on average). Marshall and Garrick (2011) examined the road safety data from 24 California cities and found that cities with higher bicycle use generally showed much lower risks of fatalities for all road users. It is reasonable to expect that, similar automobile transport, as cycling use grows, there will be increased investment in engineering, education and enforcement, in support of mobility, safety, and efficiency cycling needs. Moreover, as those investments occur, cycling-related collisions, injuries, and fatalities would reduce as well. Consequently, it is not unreasonable to expect that as the environmental 'friendliness' toward cycling improves, the NA public perception toward cycling and other vulnerable road users (VRU) would also be expected to improve.

Our hypothesis then follows from these empirical evidences and subsequent reasoned assumptions. The red line in Fig. 1 depicts our hypothesis regarding a causal relationship between overall level of road safety and bicycle mode share that increased bicycle use will lead to a significant reduction in total traffic collisions. The blue line in Fig. 1 depicts an associated hypothesis for NA bicycling levels, that initial increases in bicycle mode share above our currently low levels to some unknown medium level of cycling use will likely cause increased bicycle-related collisions, but as bicycle use increased beyond that unknown medium level there would be a decrease in bicycle-related collisions. Our research seeks to test these hypotheses by development and application of reliable empirical tools. Community-based, macro-level collision prediction models (CPMs) have been shown to provide reliable empirical evidence, and was used in this paper to test our second (blue line in Fig. 1) hypothesis.

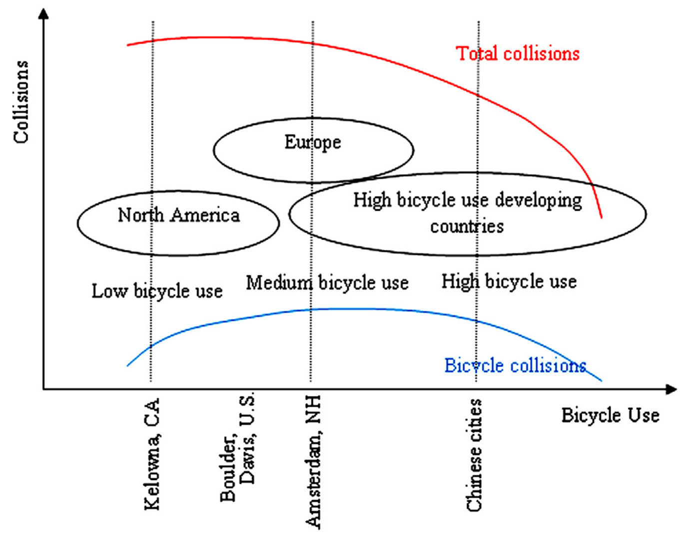

Fig. 1. Hypotheses on bicycle use and overall level of road safety. (For interpretation of the references to color in this figure text, the reader is referred to the web version of the article.)

2. Literature review

2.1. Generalized linear regression approaches for CPMs

In previous CPM studies, generalized linear models (GLMs) are commonly used and proved successfully as they could effectively model the rare, random, sporadic, and non-negative collision data. The generalized linear regression methods for CPM development mainly include Poisson regression and its various extensions such as zero-inflated Poisson regression, negative binomial regression, and Poisson lognormal regression.

2.2. Poisson and zero-inflated poisson regression

Miaou et al. (1992) found that the Poisson regression approach was more effective to predict truck collisions when compared to the regular linear regression techniques. However, Poisson regression assumes that the mean value equals to the variance value, which is not consistent with the over-dispersion of collision data. Therefore, several other regression techniques based on Poisson regression were proposed. The zero-inflated Poisson model (ZIP) is one extension of the Poisson model. It is used to solve the issue of "excess zeros" that can characterize collision data (Shankar et al., 2003; Lord et al., 2005). ZIP models assume a dual-state process which is responsible for generating collision data. The first process generates only zero counts and the second process generates non-zero counts from a Poisson model. The empirical results from related studies show that ZIP regression was more promising for providing explanatory insights into the causality behind collisions than Poisson regression (Qin et al., 2004; Kumara and Chin, 2003; Lee and Mannering, 2002).

2.3. Negative binomial regression

The second extensional approach for developing CPMs uses Poisson–Gamma hierarchy, also called negative binomial (NB) regression. This regression specifically accounts for extra Poisson variation of collisions, and is widely used in many studies for both micro and macro-level CPMs (Miaou and Lord, 2003; Lord, 2006; Sawalha and Sayed, 2006; Hadayeghi et al., 2003; Lovegrove and Sayed, 2006; Ladron de Geuvara et al., 2004). Model results from these studies demonstrated that an NB model was superior to a