Accident Analysis and Prevention 87 (2016) 8–16

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

A multivariate spatial crash frequency model for identifying sites with promise based on crash types

Jonathan Aguero-Valverde[a], Kun-Feng (Ken) Wu[b,*], Eric T. Donnell[c]

[a] Department of Civil Engineering, University of Costa Rica, Costa Rica
[b] Department of Transportation and Logistics Management, National Chiao Tung University, Taiwan, ROC
[c] Department of Civil and Environmental Engineering, The Pennsylvania State University, United States

ARTICLE INFO

Article history:
Received 16 July 2015
Received in revised form 14 October 2015
Accepted 5 November 2015
Available online 27 November 2015

Keywords:
Multivariate Poisson-lognormal model
Spatial correlation
The systemic approach
Sites with promise

ABSTRACT

Many studies have proposed the use of a systemic approach to identify sites with promise (SWiPs). Proponents of the systemic approach to road safety management suggest that it is more effective in reducing crash frequency than the traditional hot spot approach. The systemic approach aims to identify SWiPs by crash type(s) and, therefore, effectively connects crashes to their corresponding countermeasures. Nevertheless, a major challenge to implementing this approach is the low precision of crash frequency models, which results from the systemic approach considering subsets (crash types) of total crashes leading to higher variability in modeling outcomes. This study responds to the need for more precise statistical output and proposes a multivariate spatial model for simultaneously modeling crash frequencies for different crash types. The multivariate spatial model not only induces a multivariate correlation structure between crash types at the same site, but also spatial correlation among adjacent sites to enhance model precision. This study utilized crash, traffic, and roadway inventory data on rural two-lane highways in Pennsylvania to construct and test the multivariate spatial model. Four models with and without the multivariate and spatial correlations were tested and compared. The results show that the model that considers both multivariate and spatial correlation has the best fit. Moreover, it was found that the multivariate correlation plays a stronger role than the spatial correlation when modeling crash frequencies in terms of different crash types.

1. Introduction

The traditional network screening approach in highway safety relies on the total number of crashes or the expected excess crash frequency to identify sites with promise (AASHTO, 2010). The expected excess crash frequency is the difference between the empirical Bayes adjusted average crash frequency and the predicted average frequency from a Safety Performance Function. It is considered a stable performance measure when compared to the number of reported crashes as it controls for the randomness of crash occurrence (regression-to-the-mean) (Aguero-Valverde and Jovanis, 2006, 2008; AASHTO, 2010). Once high-crash locations are identified, they are studied further to diagnose the problems at locations with high total or excessive crash frequencies. Countermeasures are often selected based on the most commonly-occurring crash types. Because the traditional "hot-spot" approach begins with identifying high-crash locations based on total crashes, and only considers crash type(s) when diagnosing safety problems at specific sites, this approach may not necessarily be the most cost-effective approach to mitigate safety problems at specific sites. Because there is a disconnect between the measures used to identify sites with promise and diagnosing safety problems, the effectiveness of a safety improvement or countermeasure is likely diminished. An alternative approach would be to consider target crashes when identifying sites with promise, and then implement countermeasures that explicitly mitigate those target crashes. This process is often referred to as the systemic approach to managing safety of a roadway network. While a network screening process that specifically targets certain crash types is needed to maximize return on safety improvement, many challenges remain. In order to implement a systemic approach to safety, crash frequency modeling techniques must be able to accurately estimate the expected target crash frequency. This represents a departure from the traditional hot-spot approach to identify sites with promise, which is based on the expected total crash frequency. This study

* Corresponding author.
E-mail addresses: jonathan.aguero@ucr.ac.cr (A.-V. Jonathan), kfwu@nctu.edu.tw (K.-F. Wu), edonnell@engr.psu.edu (E.T. Donnell).

http://dx.doi.org/10.1016/j.aap.2015.11.006

proposes a Full Bayes multivariate crash frequency model with spatial correlation to estimate models of target crash frequency, which could be used in a systemic safety management process.

1.1. Selecting safety performance measures for network screening

The selection of a safety performance measure (e.g., total crashes, excess crash frequency, crash type, or severity outcome) for network screening can significantly influence the return on safety investment. The traditional hot-spot approach usually involves two steps. First, traffic engineers screen a road network to identify sites with the highest expected number of total crashes or excess crashes. Then, crash reports are often reviewed, and perhaps site visits are made to identify candidate countermeasures to mitigate the most common crash type(s) at high-crash locations. Consider a hypothetical example in which 10-years of crash data for 10 roadway segments with the exact same roadway features and traffic volume is included in the network, as shown in Table 1. Suppose one can only select one segment for safety improvement due to budget constraints. Using the traditional hot-spot approach, and the total number of crashes as the safety performance measure, would lead one to select Segment 1 as the "hot-spot" location. Once Segment 1 is selected for safety improvement, a detailed engineering study would then be conducted to identify candidate countermeasures for implementation. The crash data in Table 1 shows that head-on crashes account for the highest proportion of total crashes in Segment 1. A countermeasure targeted to mitigate head-on crashes, such as median barrier or centerline rumble strips, may be implemented. Suppose the countermeasure implemented proves effective in mitigating head-on crashes in Segment 1. A legitimate question can then be asked: why not implement the countermeasure in Segment 5 to mitigate 6 head-on crashes instead of only 4 head-on crashes that were reported in Segment 1?

The hypothetical example above illustrates the implementation procedure of a traditional hot-spot approach using the number of total crashes. For many years, this approach was used to assess highway safety performance, and was based on reducing total crashes. In support of this approach, safety programs were focused on identifying and addressing locations with high crash frequencies, with less attention paid to crash types. Many methods have been developed to assist analysts to more accurately identify locations with high crash frequencies, which are included in publications such as the American Association of State Highway and Transportation Officials' (AASHTO) Highway Safety Manual (AASHTO, 2010). Diagnostic methods, or detailed engineering studies (Hauer et al., 2004), are usually conducted to select appropriate countermeasures in response to common crash types at specific sites. Nevertheless, as illustrated in Table 1, this process is unlikely to effectively identify sites with the highest frequency of a specific crash type and, therefore, the return on the safety investment is not maximized. Moreover, this approach reduces the likelihood that crash types involving severe outcomes will be effectively mitigated.

Table 1
Hypothetical segment-based crash data.

Segment	Total crashes	Head-on	Rear-end	Side-swipe	Fatal
1	10	4	3	3	2
2	0	0	0	0	0
3	0	0	0	0	0
4	0	0	0	0	0
5	8	6	1	1	3
6	4	2	0	2	3
7	3	2	0	1	0
8	1	1	0	0	1
9	0	0	0	0	0
10	9	2	5	2	2
Average	3.5	1.7	0.9	0.9	1.1

As a further example, run-off-road (ROR) crashes often result in severe outcomes and account for a high proportion of total crashes in some segments. However, when network screening is based on the number of total crashes, these segments are unlikely to be chosen for safety improvement even when effective countermeasures are available to directly target these crash types (e.g., shoulder rumble strips). This is mainly because ROR crashes often occur on segments with low traffic volumes, and these segments usually have a lower number of total crashes (due to lower exposure) than those on major arterials with higher traffic volumes. As a result, the locations with a high number of ROR crashes are unlikely to be selected for safety improvements when network screening is based on the number of total crashes. An alternative approach to managing road safety is the systemic method. This method differs from the examples described previously by considering specific, severe crash types and their association with high-risk roadway or roadside features. Safety countermeasures that address the specific, severe crash type are deployed within an entire jurisdiction or region in an effort to mitigate severe crashes.

The systemic approach has been proposed by the Federal Highway Administration (FHWA) to complement the hot-spot approach. The intent is to better address the severe crash types on rural roadways, with the anticipated benefit of reducing fatalities and injuries resulting from traffic crashes. While the traditional hot-spot approach considers total crashes, the systemic approach considers severe crash types. The latter approach generally targets road segments or routes that have an excessive number of ROR crashes in rural areas, and then implements low-cost, proven countermeasures to mitigate ROR crashes. The systemic method has been added by a number of states to their safety planning efforts to better address severe crashes in rural areas (Sawyer et al., 2011; Preston et al., 2010). This approach first identifies a common crash type that a countermeasure will mitigate, and then uses roadway inventory and crash data to identify high-risk features where those countermeasures can be deployed effectively (Sawyer et al., 2011). While the systemic approach is proactive and seemingly cost-effective relative to the hot-spot approach, there are challenges to implement the method. In particular, it is critical to correctly identify sites with a high frequency of severe, target crashes.

1.2. Identification of SWiPs by crash type

To prioritize sites for safety improvement based on ranking crash counts, it is important to accurately estimate the expected number of crashes at a site. One of the challenges in crash frequency modeling is that crash frequencies fluctuate over time at a given site and, because of these annual differences, averaging reported crashes over a few years can be unduly influenced by a single year with an unusually high or low number of crashes (Hauer, 1986; Hauer, 1996; Persaud, 1988; Hauer, 1997; Carriquiry and Pawlovich, 2004). This is known as regression to the mean bias (RTM). Bayesian statistical methods, including Empirical Bayes (EB) and Full Bayesian (FB), have been identified as methods to accurately estimate the expected number of crashes at a site and overcome RTM bias. In general, the EB method is a special case of FB that arises when an FB analysis is simplified by making certain assumptions, but FB is more computationally complex than EB (Carriquiry and Pawlovich, 2004).

Although recent research suggests that EB and FB estimates are comparable (e.g. Carriquiry and Pawlovich, 2004; Persaud et al., 2010), the EB approach is not applicable when trying to compute precise estimates of the expected number of crashes for specific crash types. The EB approach is widely used to identify SWiPs