

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18 **A Note on Hotspot Identification for Urban Expressways**

19 *Xiaobo Qu^a and Qiang Meng^{1b}*

20 *^aGriffith School of Engineering, Griffith University, Gold Coast, 4222 QLD Australia*

21
22
23 *^bDepartment of Civil and Environmental Engineering, National University of Singapore,*
24
25 *Singapore 117576*

26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

¹ Corresponding author, Tel: +61-7-55529724; fax: +61-7-55528065; E-mail: ceemq@nus.edu.sg (Q. Meng);
61 x.qu@griffith.edu.au (X. Qu);
62
63
64
65

A Note on Hotspot Identification for Urban Expressways

Abstract: Hotspot identification (HSID) is of great importance to land transport authorities (e.g. the Land Transport Authority (LTA) of Singapore) in their efforts to improve the safety of highways. According to a survey of experienced engineers at the LTA of Singapore, we draw the conclusion that the severity of crashes should not be neglected in the HSID process. Accordingly, in this technical note, societal risk-based simple ranking and empirical Bayesian methods are proposed to identify the hotspots in a Singapore expressway on the basis of the detailed three-year casualty data in the Historical Crash Damage (HCD) database. We further conduct a consistency analysis to compare the societal risk-based method and the conventional frequency-based method. The consistency analysis reports that (1) the frequency-based method is more consistent than the societal risk-based method, and (2) the empirical Bayesian method is more consistent than the simple ranking method.

Keywords: Societal risk; hotspots; expressways; crash count

1. Introduction

A crash hotspot, also referred to as a high-risk location or black spot, is a term used in highway safety management systems to denote road sites at which crashes are concentrated (PIARC, 2004). The identification of hotspots is of great importance to land transport authorities and other decision makers in their attempts to enhance the safety of highways. Precise localizations of crashes are essential if decision makers are to implement cost-effective risk reduction solutions. By contrast, inaccurate localizations, namely misidentifications of hotspots, will result in the inefficient use of limited resources (Monsere et al., 2006; Monsere et al., 2008; Cheng and Washington, 2005; Montella, 2010). The identification of crash hotspots is the first step in the highway safety management process (AASHTO, 2010; PIARC, 2004; TRB, 2013).

Several hotspot identification (HSID) approaches have been developed on the basis of accident frequency, accident rate, accident reduction potential, and other variables (e.g. Cheng and Washington, 2005, 2008; Elvik, 2008; Huang et al., 2009; El-Basyouny and Sayed, 2013). In addition to those HSID methods that are based on the crash count without taking into account the severity of the crashes, a few HSID methods involving both crash count and severity have been proposed. The typical equivalent property damage only method is to convert all crashes into the equivalent number of property damage only (PDO) crashes on the basis of their severity or the damage caused (PIARC, 2004). However, these PDO-based approaches are unable to represent the economic impact of hotspots on society. It would be more straightforward to represent the consequences/severity of a crash in terms of the monetary cost.

It is of high priority for land transport authorities to reduce the number of “high-risk” or “high-potential for safety improvement” highway sections, and not necessarily those with high crash frequencies, with a limited budget. That is, the “risk” should be evaluated by taking into account the severity of the crashes. For example, the crash records on two sections of the Central Expressway (CTE) in Singapore (from 2 km to 3 km and from 9 km to 10 km in the southbound direction) in 2007 were 111 and 176, respectively. The latter section can be judged far more dangerous than the former if we only consider the number of crashes. However, there were 6 serious injuries and 59 slight injuries involved in the crashes that occurred on the former section, while only 28 slight injuries were caused by the crashes on the latter. This is because the spot speed in the latter section is significantly lower than that in the former (the latter section is a merging area from an on-ramp to the expressway). Accordingly, on the one hand, there are much fewer traffic conflicts in the former section

than the latter (i.e. more conflicts are likely to occur between vehicles on the on-ramp and those on the main expressway). On the other hand, the severity of the crashes that occur on the former section is significantly higher than those on the latter.

If the severities of crashes are assumed to be randomly distributed along the crash spots on a highway, the “risk” will be appropriately represented by the crash frequency without incorporating the severities of the crashes. This assumption is invalid for the urban expressways in Singapore, however. There are 11 expressways in Singapore, acting as traffic corridors connecting different urban areas, and these expressways are non-homogeneous in terms of the traffic condition as well as the road geometries. For example, the 15-km CTE has 19 on-ramps and 24 off-ramps. More importantly, the expressway connects the Central Business District (CBD) area in the south with the residential area in the north. Accordingly, the traffic volume on the southern part of the expressway is significantly higher than that on the northern part. Correspondingly, the severities of crashes cannot be assumed to be randomly distributed along the expressways due to the non-homogeneous traffic flow and spot speeds. Therefore, a new approach needs to be developed to identify hotspots by taking into account not only the crash frequencies but also the severities on the urban expressways of Singapore.

The objective of this note is to propose a comprehensive HSID method by taking not only crash frequency but also crash severity (evaluated by the monetary losses caused by different crashes) into account from the perspective of the land transport authorities. The Historical Crash Damage (HCD) database (2006-2008) is used in this study. After that, a consistency analysis is used to compare the conventional HSID method to the proposed HSID method. Discussions and conclusions are presented on the basis of the data analyses.

2. Societal Risk-Based Crash Hotspot Identification

Let $\{1, 2, \dots, I\}$ be the set of road sections of a target non-homogeneous urban expressway. Assuming that there are J types of crash denoted by $1, 2, \dots, J$, let f_{ij} and m_{ij} denote the yearly number of crashes of type j occurring on section i and the estimated monetary loss from one crash of type j , respectively. As an example of the sorts of losses that might occur, the National Safety Council (NSC, 2009), the Highway Safety Manual (AASHTO, 2006), and the Federal Highway Administration (FHWA, 1994) have respectively estimated the monetary losses of four types of motor-vehicle crash using insurance and hospital records, and these are presented in Table 1.

Table 1 Comprehensive costs of motor-vehicle crashes with different severities.

Crash category	Loss (USD) (NSC, 2009)	Loss (USD) (HSM, 2010)	Loss (USD) (FHWA, 1994)
Fatal	4,300,000	3,000,000	2,600,000
Serious injury	55,300	42,000	36,000
Possible injury	26,300	22,000	19,000
PDO	2,400	2,300	2,000

The above values are all based on the “willingness-to-pay” approach, estimating the amounts that individuals are prepared to pay to reduce a risk to their lives, which is the value to the individual on an ex-ante basis, or before the fact. In other words, the willingness-to-pay approach attempts to capture trade-offs between wealth and small reductions in risk. People’s preferences (either stated or revealed) demonstrate the value they place on reducing a risk to their life. As can be seen in Table 1, the crashes costs estimated from different studies are quite consistent.

Societal risk has been proposed and used in a number of quantitative risk assessment models to represent risks incorporating both accident frequency and the severity of dangerous scenarios (see Meng et al., 2011a&b). The societal risk is defined as the annual monetary loss due to crashes on one particular section i of an expressway, denoted by R_i , namely,

$$R_i = \sum_{j=1}^J (f_{ij} m_{ij}), \forall i \in \{1, 2, \dots, I\} \quad (1)$$

3. Applications and Discussion

3.1. Data Description

The HCD database of the CTE is used to conduct this study. As shown in Figure 1, the CTE links the north and the south of Singapore via the CBD. There are 19 on-ramps and 24 off-ramps connecting to the expressway, causing non-homogeneous traffic volume among the sections. Clearly, the safety performance of the expressway is homogeneous neither in terms of crash frequency nor crash severity: the CBD area probably has a higher crash frequency but a lower crash severity due to the slower traffic speeds there. The crash records in the HCD database are collected by the Land Transport Authority (LTA) of Singapore. According to the Motor Claims Framework introduced by the General Insurance Association of Singapore (GIA), in the event of a crash on an expressway, everyone involved must inform

the insurance company within one day using the GIA Motor Accident Report form. In addition, in accordance with the Road Traffic Act of Singapore, another report must be made within 24 hours of a crash if an injury has occurred. The HCD database (2006-2008) includes the reported crash records occurring on the CTE from 2006 to 2008, detailing the start time of the crash, its location, the crash type (e.g. rear-end, skidded, chain collision, etc.), the vehicle type (e.g. car-car, car-truck, etc.), the number of slight injuries, number of serious injuries, and number of fatalities. To summarize, there was a total of 6,382 vehicle crashes (3,305 southbound and 3,077 northbound) on the CTE from 2006 to 2008.

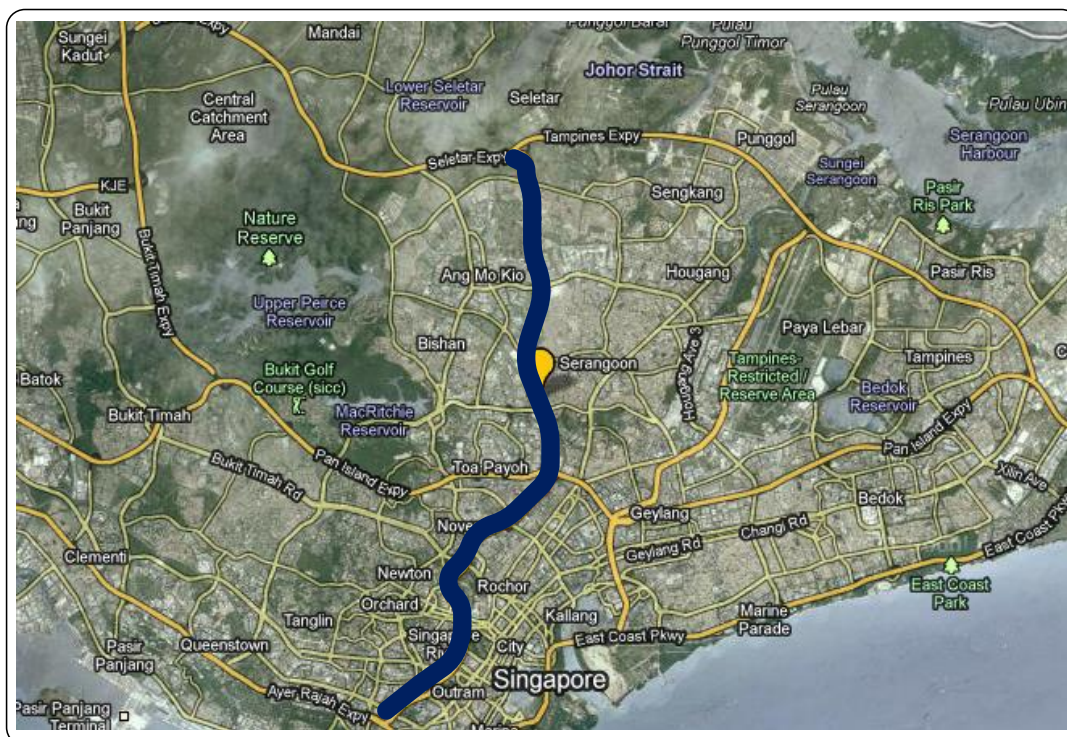


Figure 1 CTE of Singapore

In the HCD database, crashes are classified into four types in accordance with their severity: fatal, serious injury, possible injury and PDO. Similar classifications have been used in a number of previous studies (e.g., Wang and Abdel-Aty, 2008). Recently, some institutions have conducted studies estimating the comprehensive losses due to crashes of different types.

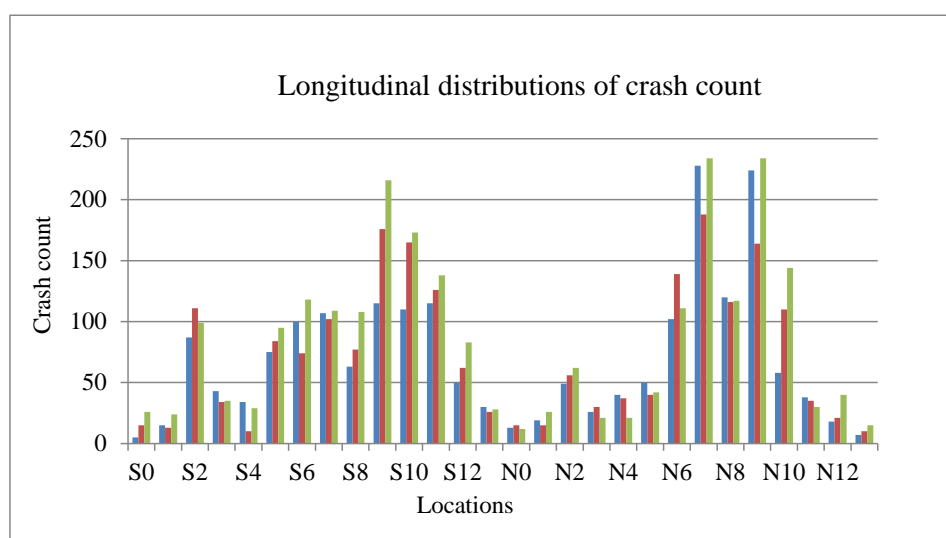
3.2 Results and Analysis

3.2.1 Simple Ranking Approach

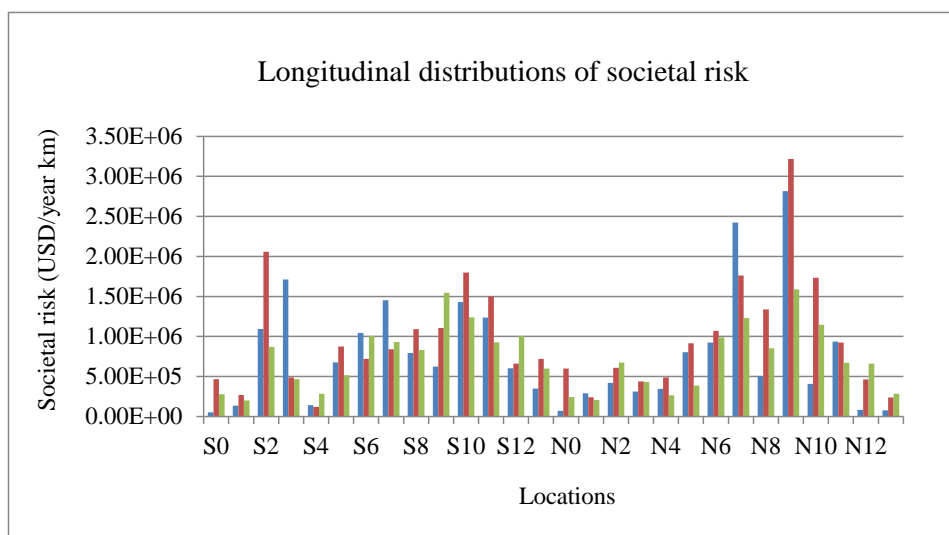
The comprehensive costs of various crash types estimated by the NSC are applied in this study. Horizontal alignment characteristics and traffic conditions are used to divide the

expressway into 26 sections (southbound: S1 to S13; northbound: N1 to N13) with an average length of 1.04 km. Detailed casualty data from crashes occurring on the CTE are obtainable from the HCD database. Figure 2 reports the longitudinal hotspot distributions along the highway.

As can be seen in Figure 2, Site N8 is considered risky according to the simple ranking method (crash count >100 per year per km). However, out of the 353 reported crashes at the site, only 47 involved slight injuries and the remaining 306 were categorized as PDO. The societal risk of the site is thus low (< 1 million USD per year per km) in terms of the monetary loss. Accordingly, it may not be a high priority for the land transport authorities to implement risk reduction solutions at this site.



(a) Crash count distributions from 2006 to 2008



(b) Societal risk distributions from 2006 to 2008

Notations: blue bar – 2006 records, red bar – 2007 records, and green bar – 2008 records

Figure 2 Hotspot identification by simple ranking method.

To summarize, the hotspots identified by the two methods are presented in Table 2. It shows that the results generated by the two methods are not exactly consistent for the two consecutive years.

Table 2 Summary of hotspots identified by the simple ranking method.

	Hotspots (number of crashes > 100 per km)	Hotspots (societal risk > 1 million USD per km)
2006	S7, S9, S10, S11, N6, N7, N8, N9	S2, S3, S6, S7, S10, S11, N7, N9
2007	S2, S7, S9, S10, S11, N6, N7, N8, N9, N10	S2, S8, S9, S10, S11, N6, N7, N8, N9, N10
2008	S6, S7, S8, S9, S10, S11, N6, N7, N8, N9, N10	S9, S10, N7, N9, N10

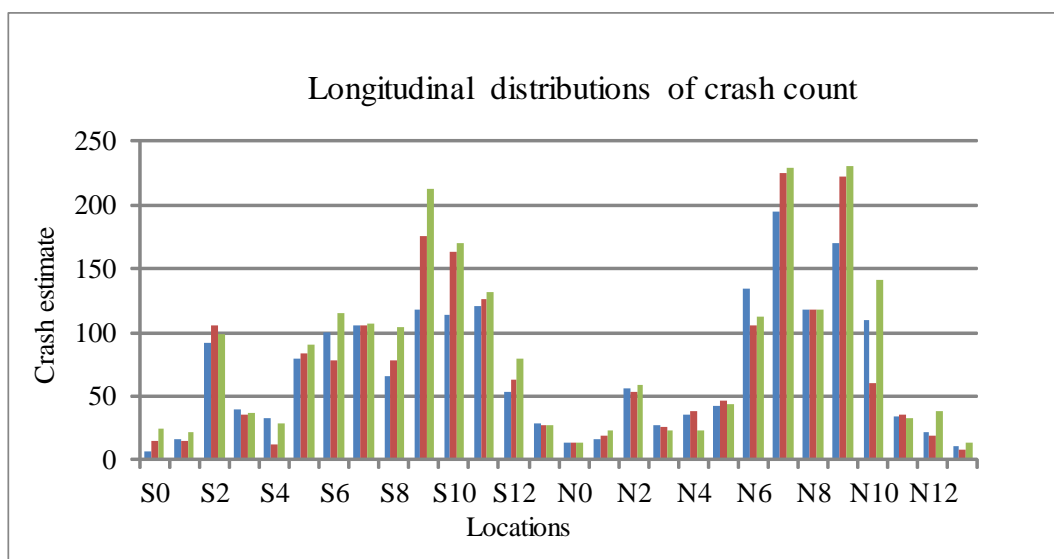
3.2.2 Empirical Bayesian Approach

The highly consistent results of the empirical Bayesian (EB) method can be attributed to its use of a predictive model, typically based upon the road geometry, which addresses the regression-to-the-mean bias present in simpler methods (such as the simple ranking method). This model forms a baseline from which the recorded crash history can be evaluated, leading to a reduction in the impact of randomly high/low crash data on the results of the analysis. In the EB procedure, the crash estimates combine the expected number of crashes and the count of crashes, according to eqns. (2) and (3).

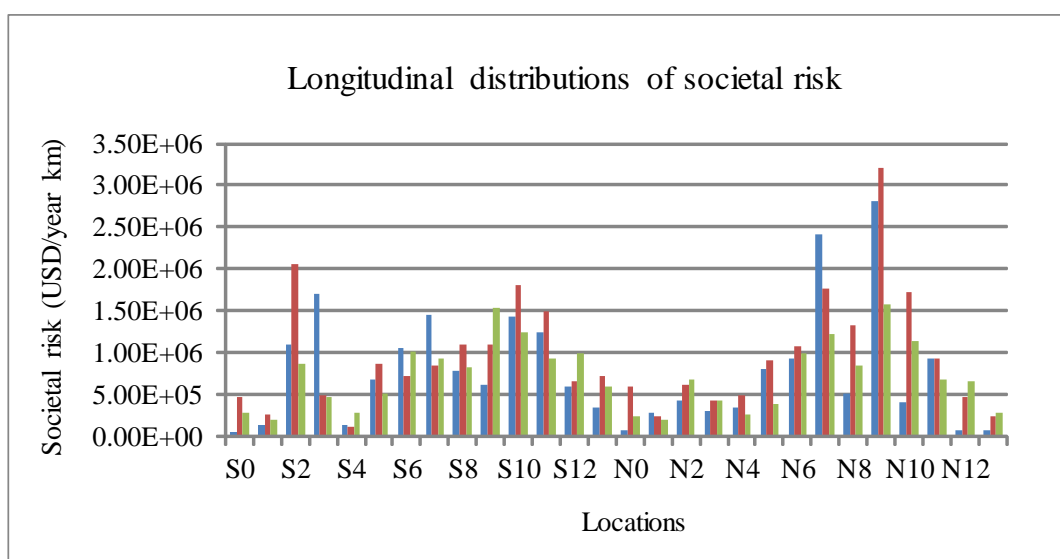
$$EB_{ij} = \varpi E(Y_j) + (1 - \varpi) Y_{ij} \quad (2)$$

$$\varpi = \frac{E(Y_j)}{E(Y_j) + VAR(Y_j)} \quad (3)$$

where EB_{ij} is the estimate of crash or societal risk in year i for segment j ; $E(Y_j)$ is the expected number of crashes or the societal risk of segment j ; $VAR(Y_j)$ is the variance of the number of crashes or societal risk for segment j ; Y_{ij} is the crash count or societal risk in year i for segment j . Based on the crash estimates, the longitudinal hotspot distributions along the highway are presented in Figure 3. The hotspots identified by the EB method are presented in Table 3.



(a) Crash distributions from 2006 to 2008



(b) Societal risk distributions from 2006 to 2008

Notations: blue bar – 2006 records, red bar – 2007 records, and green bar – 2008 records

Figure 3 Hotspot identification by the empirical Bayesian method.

Table 3 Summary of hotspots identified by the empirical Bayesian method.

	Hotspots (number of crashes > 100 per km)	Hotspots (societal risk > 1 million USD per km)
2006	S7, S9, S10, S11, N6, N7, N8, N9, N10	S2, S3, S6, S7, S9, S10, S11, N7, N9
2007	S2, S7, S9, S10, S11, N6, N7, N8,N9	S2, S8, S9, S10, S11, N6, N7, N8, N9, N10

2008 S6, S7, S8, S9, S10, S11, N6, N7, S6, S9, S10, N7, N9, N10
N8, N9, N10

3.2.3 Consistency Analysis

In order to evaluate the consistency of the crash count-based and societal risk-based approaches in the three consecutive years (2006 to 2008), the method consistency test (MCT) and the total rank differences test (TRDT) proposed by Cheng and Washington (2008) are applied in this study.

The MCT evaluates a method's performance by measuring the number of hotspots identified in in all three time periods. It is assumed that road sections are in the same or similar underlying operational state, and that their expected safety performance remains virtually unaltered, over the three periods. With this assumption, the greater the number of hotspots identified in all periods, the more consistent is the performance of the HSID method. The test statistic is given by:

$$MCT_j = \bigcap_{i=1}^I \{S_1, \dots, S_k, \dots, S_K\}_{j,i} \quad (4)$$

The TRDT takes into account the safety performance rankings of the road sections in the three periods. The test is conducted by calculating the sum of the total rank differences of the hotspots identified across the three periods. The smaller is the total rank difference, the more consistent is the HSID method. The test statistic is given by:

$$TRDT_j = \sum_{i=1}^{I-1} \sum_{k=1}^K |R(k_{j,i}) - R(k_{j,i+1})| \quad (5)$$

The consistency test results are reported in Table 4.

Table 4 Consistency test results.

Method	MCT (Test ranking)	TRDT (Test ranking)
Societal risk-based simple ranking	5	45 (3)
Frequency-based simple ranking	8	19 (1)
Societal risk-based EB method	5	38 (3)

Frequency-based EB method	8	17	
		(1)	(1)

As can be seen from Table 4, the frequency-based method is significantly more consistent than the societal risk method according to the MCT and the TRDT. The societal risk-based method is more prone to be affected by a low frequency of fatal or serious crashes, although it would better represent the perspective of the land transport authorities. In fact, the result shows that the occurrence of high-consequence accidents is more random than that of PDO crashes, which might dominate the results. The societal risk-based EB method provides more consistent results than the societal risk-based simple ranking method (TRDT: 38 vs. 45). Therefore, it is highly recommended that the EB method be applied in HSID schemes if the severity of crashes is being taken into account.

4. DISCUSSION

The crashes that occur at sites with lower spot speeds (e.g. congested sites) are mostly low-consequence events. Accordingly, for expressways with non-homogeneous traffic flow, it would not be reasonable to assume that serious crashes are randomly distributed along the expressways. Therefore, societal risk is, from the perspective of the land transport authorities, considered a better representation of the “risk” than the crash count when identifying hotspots on the expressways of Singapore. Where detailed casualty data are available, the societal risk-based method would be more appropriate for HSID. The correlation analysis suggests that the frequency-based method is not a good approximation, but is an acceptable one, to the societal risk-based method if detailed casualty data are unavailable. However, the frequency-based method is more consistent than the societal risk method over consecutive time periods according to the site consistency test, the MCT, and the TRDT. This is caused by the random fluctuations in high-consequence events. Therefore, it is necessary to carry out new statistical methods to eliminate the noise of the random impact of high-consequence events to determine whether high-consequence crashes are random or occur regularly at particular sites? On the basis of two years of casualty data, the EB technique and the full Bayes (FB) technique could be feasible solutions to this issue (Persaud and Lyon, 2007; Persaud et al., 1999; Qu et al., 2012).

5. CONCLUSIONS

The identification of hotspots is of great importance to land transport authorities (e.g. the LTA of Singapore) in their efforts to enhance the safety of highways. According to a survey of experienced personnel at the LTA of Singapore, we concluded that the severity of crashes should be taken into account in the HSID process. Accordingly, a societal risk-based method is proposed here to identify the hotspots on the CTE of Singapore in 2006 and 2007. Correlation analysis and consistency analysis are also carried out to compare the societal risk-based method to the conventional frequency-based method. The correlation analysis suggests that the frequency-based method is an acceptable approximation of the societal risk method if detailed casualty data are unavailable. The consistency analysis reports that the frequency-based method is more consistent than the societal risk-based method. Finally, new research directions are recommended in the discussion section.

ACKNOWLEDGEMENT

This paper is jointly supported by the innovation fund of Land Transport Authority of Singapore (ER 253) and Griffith NRG scheme. Special thanks will be expressed to Mr Kum Thong Yong and Ms Siew Chee Wong from LTA for their supports on data collection for this research.

REFERENCES

- AASHTO, 2010. Highway Safety Manual, First Edition.
- BTE, 2000. Road crash costs in Australia, Report 102. Bureau of Infrastructure, Transport and Regional Economics, Canberra.
- BTRE, 2006. Cost of road crashes in Australia 2006, Report 118. Bureau of Infrastructure, Transport and Regional Economics, Canberra.
- Cheng, W., and S. Washington, 2005. Experimental evaluation of hotspot identification methods. *Accident Analysis and Prevention*. 37, pp. 870-881.
- Cheng, W., and S. Washington, 2008. New criteria for evaluating methods of identifying hot spots. *Transportation Research Record: Journal of the Transportation Research Board*, 2083, pp. 76-85.
- El-Basyouny, K., and T. Sayed, 2006. Comparison of two negative binomial regression techniques in developing accident prediction models. . *Transportation Research Record: Journal of the Transportation Research Board*, 1950, pp. 9-16.

- Elvik, R., 2008. Comparative analysis of techniques for identifying hazardous road locations. *Transportation Research Record: Journal of the Transportation Research Board*, 2083, pp. 72-75.
- Federal Highway Administration, 1994. Motor vehicle accident costs. http://safety.fhwa.dot.gov/facts_stats/t75702.cfm Accessed 8 Dec 2013.
- Highway Safety Manual, 2006. Evaluation of safety alternatives. http://kiewit.oregonstate.edu/pdf/Safety_13.pdf Accessed 8 Jan 2013.
- Huang, H., H.C. Chin, M. Haque. Empirical evaluation of alternative approaches in identifying hot spots. *Transportation Research Record: Journal of the Transportation Research Board*, 2103, pp. 32-41.
- Meng Q., X. Qu, K.T. Yong, and Y.H. Wong, 2011b. QRA model based risk impact analysis of traffic flow in urban road tunnels. *Risk Analysis*. 31, pp. 1872-1882.
- Meng Q., X. Qu, X. Wang, Y. Vivi, and S.C. Wong, 2011a. Quantitative risk assessment modeling for non-homogeneous urban road tunnels. *Risk Analysis*. 31, pp. 382-403.
- Monsere, C., P. Bosa, R.L. Bertini, 2008. Combining Climate, crash, and highway data for improved ranking of speed and winter-weather related crash locations in Oregon. *Journal of Transportation Engineering*, 134, pp. 287-296.
- Monsere, C., R.L. Bertini, P. Bosa, and D. Chi, 2006. Comparison of identification and ranking methodologies for speed-related crash locations. FHWA-OR-RD-06-14, Oregon Department of Transportation, Salem, Ore.
- Montella, A., 2010. A comparative analysis of hotspot identification methods. *Accident Analysis and Prevention*. 42, pp. 571-581.
- National Safety Council. Estimating the costs of motor vehicle injuries. 2009. http://www.nsc.org/news_resources/injury_and_death_statistics/Pages/EstimatingtheCostsofUnintentionalInjuries.aspx Accessed 8 Jan 2013.
- Oh, J., Washington, S., and D., Lee, 2010. Property damage crash equivalency factors to solve crash frequency-severity dilemma: Case study on South Korean rural roads. *Transportation Research Record: Journal of the Transportation Research Board*, 2148, pp. 83-92.
- Persaud, B., and C. Lyon, 2007. Empirical Bayes before-after studies: lessons learned from two decades of experience and future directions. *Accident Analysis and Prevention*. 39, pp. 546-555.

- Persaud, B., C. Lyon, and T. Nguyen., 1999. Empirical Bayes procedure for ranking sites for safety investigation by potential for improvement. *Transportation Research Record: Journal of the Transportation Research Board*, 1665, pp. 7-12.
- PIARC, Technical Committee on Road Safety C13, 2004. Road Safety Manual.
- Qu, X., Meng, Q., and Li, S., 2012. Analyses and implications of the accidents in the Singapore Strait. *Transportation Research Record: Journal of the Transportation Research Board*, 2273, pp. 106-111.
- TRB, 2013. Synthesis Report on safety-related papers presented at the 92nd TRB Annual Meeting
- Wang, X., and M. Abdel-Aty, 2008. Analysis of left-turn crash injury severity by conflicting pattern using partial proportional odds models. *Accident Analysis and Prevention*. 40, pp. 1674-1682.