COMPARATIVE ANALYSIS OF HOTSPOT IDENTIFICATION METHODS IN THE PRESENCE OF LIMITED INFORMATION

Ranja Bandyopadhyaya
Ph.D. Scholar, Department of Civil Engineering, Indian Institute of Technology Kharagpur, Kharagpur - 721302, India, Email: b_ranja@yahoo.co.in

Sudeshna Mitra
Assistant Professor, Department of Civil Engineering, Indian Institute of Technology Kharagpur, Kharagpur - 721302, India, Email: sudeshna@civil.iitkgp.ernet.in

Submitted to the 3rd International Conference on Road Safety and Simulation, September 14-16, 2011, Indianapolis, USA

ABSTRACT

India has the second largest road network in the world. Expansion of road network with simultaneous increase in vehicle ownership in the last decade resulted in increased exposure to road transport, and an increased number of road traffic accidents. Road safety is thus becoming one of the most challenging issues demanding utmost priority in India.

Identification and treatment of unsafe locations (black or hot spots) is the first major step in safety assessment. Many effective Hot Spot identification (HSID) methods are available in literature. However, the more scientific ones like Empirical Bayes (EB) or Safety Performance Functions (SPF), require data such as traffic volume, geometric design details and regulatory information which is rarely available in India. Accident data, with details of number of crashes and nature and number of casualties are only available in most cases. Hence, methods of HSID with Crash frequency (CF), Fatal Crash frequency (FCF), and Equivalent Property Damage Only (EPDO) as well as both crash frequency and severities available from existing literature are considered. In addition, a new metric "Index of frequency severity (I_{FS})" has been proposed by the authors. These four methods are tested with data from NH6 in Howrah district of India with Site consistency, Method consistency and Total rank difference tests. Even though the Crash frequency method seems to outperform other methods, when both frequency and severity are considered the new metric proved to be a better predictor of unsafe locations.

INTRODUCTION

India has the second largest road network in the world. This 33, 000, 00 km road network comprises of the national highways, state highways, connecting expressways, major district roads and rural and other roads. About 65% of freight and 80% passenger traffic is carried by roads. National Highways constitute only about 2% of the road network but carry about 40% of the total road traffic. Number of vehicles in India has been growing at an average pace of 10.16% per annum over the last five years (NHAI). Road accident statistics for last ten years show a steep rise in both number of accidents and number of fatalities. Hence, road safety is a very serious issue in India and needs to be addressed with utmost priority.

The first step for any improvement project is identification of hazardous locations and ranking them in order to prioritize improvement efforts, so that available funds can be optimally allocated. Currently Government of India is focusing on capacity building for better connectivity and mobility but safety aspects in planning and design are not given due priority. There is hardly any centralized effort to systematically collect safety records for the newly constructed highways. As a result it is difficult to conclude if a particular road location is inherently unsafe or it has experienced high crashes by chance.

Some informal initiatives are taken for crash data collection and analysis but there is no systematic method for detecting a high crash location. Experience as well as literature on crash detection indicates that not only the frequencies of accidents but also the exposure is important in quantifying the risk of a location. For this purpose traffic volume and other related information is required. However in India, where data collection and maintenance is not done in organized manner, obtaining information on both exposure and transportation systems, such as geometric design details, traffic volume and regulatory information, are ambitious.

Therefore, many of the available effective methodologies of HSID developed in developed countries, specifically the sophisticated methods such as Empirical Bayes (EB) or methods where Safety Performance Functions (SPF) has to be formulated, cannot be used in India. Hence it is important to properly review and find a suitable method for identification and prioritization of hazardous locations from available limited information.

After a thorough review of existing HSID methods, the appropriate methods that can be used in presence of limited information are identified and tested. The three HSID methods found appropriate are: Crash frequency (CF), Fatal Crash frequency (FCF), and Equivalent Property Damage Only (EPDO). A fourth method, as suggested by Khasnabis et al. (2010), is also appropriate in this context, which uses crash frequency and severity independently to identify hotspots and does not rank the hotspots in order of priority. By following a similar logic a modified index is thus suggested by the authors and termed as "Index of frequency severity (I_{FS})" which considers the combined effect of both frequency and severity to rank the hotspots.

Comparison of the existing methods and the newly suggested I_{FS} method using accident data obtained from a stretch of National Highway-6 (NH6), Howrah district in India is done to find the most efficient HSID method for limited information data. As the tests are done with real data, false identification criteria are not used. Three tests viz. Site consistency test, Method consistency test and Total rank differences test (Cheng and Washington, 2008) are used for comparing the efficiency of HSID methods.

In the present paper, a thorough review of existing literature is done first to find the HSID methods suitable in the Indian context. Next, under methodology, the HSID methods used by the authors are briefly explained as also the tests used to evaluate these methods. A brief description of the data collected and used for identifying hotspots by different methods and evaluation of these methods is given in the subsequent section. This is followed by the results of the data analysis and discussion on the conclusions that may be drawn from it.

The objectives of the present paper are:

1. Identifying and proposing suitable Hot Spot Identification (HSID) methods relevant in Indian context where limited geometric and traffic data are available
2. Identifying hotspots using the above methods in a particular stretch of NH-6 chosen for study
3. Comparing the above methods in terms of their efficiency in identifying hotspots by established tests

LITERATURE REVIEW

The reviewed literature may broadly be classified into two types. The first are those that suggest or validate the different hotspot identification (HSID) methods. This is presented in Table 1 below. The second type, presented in Table 2 below, consist of literature that evaluates the effectiveness of these HSID methods.

Table 1 Literature for HSID methods

Reference	Brief outline/ key points	New Methodology suggested	Data Requirement	Critique/ remarks
Kononov (2002)	Suggested that presence of pattern in occurrence of a particular type of accident along a road segment suggests presence of an element in roadway condition which triggers a deviation from random statistical occurrence.	Proposed a Pattern Recognition algorithm which plots the pattern of a particular accident type along the whole stretch of study area.	Detail of accident type time and exact location etc required	Understanding the patterns of accidents helps to develop effective strategies for addressing safety issues.
Tarko and Kanodia (2004)	Proposed two new HSID methods for ranking hazardous locations. Proposed two criteria for evaluating HSID methods namely efficiency and fairness. are not addressed together by any single method.	Two new HSID methods proposed: "Index of Crash Frequency (I_{CF})" "Index of Crash Cost (I_{cc})",	Traffic Volume	An efficient method ensures that expected total number of crash prevented is maximal (Crash frequency) A fair method takes into account risk faced by all users (Crash Rate). ICF & ICC are observed to satisfy both efficiency and fair criteria.
Elvik (2006)	Proposed a new approach for identifying hotspots based on computation local risk factors for a test site and a reference similar site. The local risk factors for matched pair of locations (i.e. comparative rating to each factor for say road surface friction obstructions etc) are collected. The decision of whether the test site is a hotspot or not is made on the basis of the local risk factors.	Proposed local risk factor based hotspot identification technique.	Site condition details (geometric design etc.) required	It analyzes the factors contributing to accidents, so as to effectively discriminate between true positives and false positives The proposed approach helps in analysis of accidents after hazardous road locations have been identified, rather than helping to identify such locations.
Miranda-Monero et al. (2007)	Proposed multiple testing based approach for hotspot identification. In multiple testing model, multiple hypotheses (one for each site stating weather the site is a	Two Bayesian testing procedures are introduced: i) Bayesian test with weights (BTW). ii) Bayesian test controlling for the false	Requires traffic volume data.	The BTW defines a loss function based hypothesis for each site. The loss function is defined on cost of false identification and the decision of accepting or rejecting a null hypothesis is based on loss function. In Bayesian test controlling for the false

	hotspot or not a hotspot) is tested using a common threshold.	discovery rate (FDR) or false negative rate (FNR).		discovery rate (FDR) or false negative rate (FNR) confidence level of FDR & FNR are specified individually.
Cheng and Wang (2010)	Used combination of two criteria viz. accident count and accident reduction potential (ARP), identify hotspots.	Two evaluation indices are proposed: i) product of accident count and ARP and ii) product of count cost and ARP cost.	Traffic volume data required.	The common sites from independent list generated by count and ARP methods in both lists are designated hazardous under the combined method. The combined method is efficient in identifying sites that have both higher accident counts and ARP.
Khasnabis et al. (2010)	Used crash frequency and severity together for detecting hotspots. Two separate lists from critical frequency criteria and critical severity criteria are generated and the combined list is deemed hazardous.	Joint frequency and severity criteria for identifying hotspots	No traffic volume data required.	Method is based on the assumption that chance of a site having frequent severe is rare. Method do not rank the sites.

Table 2 Literature on HSID evaluation methods:

Reference	Brief outline/ key points	New Methodology suggested	Based on False Discovery?	Critique/ remarks
Hauer and Persaud (1984)	Developed statistical model for evaluation of HSID method Method uses expected accident counts and the expected false identifications given the actual crash counts for evaluation of a HSID method.	Statistical model for evaluation of HSID method.	Yes	Uses False Identification Criteria Requires prior knowledge of weather a site is a hotspot or not.
Hauer et al. (2004)	Developed Profile & Peak (P&P) software, which is used for calibrating SPF function from	Cost effectiveness model for evaluation of HSID methods.	No	The methods compared are based on EB estimate of expected high accident count and

	AADT, and to find the EB estimate of expected accidents as well the highest peak accident count of all sites.			expected severity-weighted accidents. The cost effectiveness of projects are obtained from detailed engineering study by experienced persons.
Cheng and Washington (2008)	Proposed four new quantitative HSID evaluation tests.	The four new evaluation tests described in this paper are: 1. Site consistency test 2. Method consistency test 3. Total rank differences test 4.Poisson mean difference test	No	The new tests suggested by Cheng and Washington (9) can be used for real accident data where the truth (i.e. whether the site is true hotspot or not) is not known.
Montella (2010)	An extension of Cheng and Washington’s (2008) work.	Proposed score test parameter for evaluating HSID methods.	No	Score test parameter is an average of site consistency, method consistency & total rank difference test parameters.

METHODOLOGY

The four HSID methods viz. crash frequency (CF), fatal crash frequency (FCF) equivalent property damage only (EPDO) and Index of frequency severity (I_{FS}) that can be used in presence of limited information are compared to evaluate their performance. Researchers have used the first three methods earlier while the last one is a new method suggested by the authors.

Crash frequency (CF) method uses the crash counts (number of crashes) at a location and ranks sites in descending number of crash counts. The site with maximum number of counts, receive highest rank i.e. Rank 1. The locations with top ten ranks are designated as hotspots. The CF method is useful in detecting sites having high frequency of crashes, but it fails to consider the crash severities. As a result, a site experiencing more frequent non-severe crashes is likely to be indicated as hotspot than a site experiencing relatively less number of crashes, but with higher crash fatalities. Thus a modified CF method called Fatal Crash frequency (FCF) method is used. This method uses the fatal crash counts as opposed to the total crash counts for ranking and the sites are arranged in descending number of fatal crash counts.

The FCF method only takes the fatal crash counts for judging whether a site is critical or not but severity of a crash cannot be only judged by fatality. A severe crash may have many grievous injury crashes but no fatal crashes. The equivalent property damage only (EPDO) considers all levels of severity of crashes occurring at a particular site for deciding a site as critical. In this method total accident cost for each site in terms of its equivalent property damage cost is calculated. The number of property damage only crashes, number of fatalities, number of major injuries and number of minor injuries for each site are multiplied by a their equivalent property damage weight factor and added to calculate the EPDO cost for each site. The costs of various accident types are taken as: property damage cost Rs.16200, fatality cost Rs.535489, major injury cost Rs.242736 and minor injury cost Rs.18855. These costs are taken from a TCS study conducted in 1999 (Sen et al., 2010). The equivalent property damage weight factor is assumed 1 for property damage only crashes, and the equivalent property damage weight factors for other crash types are calculated as: 33.05 for each fatality reported, 14.98 for each major injury reported and 1.16 for each minor injury reported. The sites are ranked in descending order of EDPO values and the top ranked sites are considered as potential hotspots.

Though the EPDO method considers the different levels severity of crashes but very high PDO weight factors are assigned to each fatality, and major injury reported in a particular site. A single accident reporting five fatalities will be indicated as hotspot while five crashes each reporting single grievous injury will be indicated as non-hotspot by this method. Thus it is important to have a method which takes into account the total count crash counts and total crash counts reporting fatal injuries for deciding the safety of a location. Thus the authors propose an index called index of frequency severity (I_{FS}) which takes into consideration the frequency of total crashes and the ratio of fatal to total crashes for a site to decide the safety status of the site. The I_{FS} is defined as:

$$I_{FSi} = \frac{CF_i}{CI_{0.95CF}} + \frac{FCF_i/CF_i}{CI_{0.95R}}$$

Where,

I_{FSi} = Index of frequency severity at location *i*

CF_i = Crash frequency observed at location *i*

FCF_i = Fatal crash frequency observed at location *i*

$CI_{0.95CF}$ = 95% Confidence Interval developed with observed mean crash frequencies in the study locations.

$CI_{0.95R}$ = 95% Confidence Interval developed with observed mean *FCF* to *CF* ratio in the study locations.

The sites ranked in descending order of I_{FS} value. The effectiveness of CF, FCF, EPDO and I_{SF} methods are evaluated using three HSID evaluation tests which do not involve false identification criteria. The tests used are Site consistency test, Method consistency test and Total rank differences test (Cheng and Washington, 2008). The tests are based on consistent identification of a site as hotspot during two study periods given no improvement has been undertaken.

The Site consistency test measures the ability of a HSID method to consistently identify a site as hotspot over subsequent observation periods if no improvement is undertaken. This method compares the sum of observed crashes occurring on n high risk sites (identified in the initial time period) during future time period for all competing HSID methods. The method which yields highest sum of crash counts in future time period is said to perform best. The method consistency test measures a method's performance by finding the number of same sites identified as a hotspot during two subsequent observation periods. A method is more reliable and consistent if greater numbers of same sites are identified as hotspots in consecutive observation periods. The total rank differences test takes the differences in rankings of n hotspots identified in the initial time period and their ranking in the subsequent observation period. The smaller the summation of difference in ranks, the better is the HSID method.

DESCRIPTION OF DATA

The performance of the HSID methods are tested using crash records for three years (2007, 2008 & 2009) over 70 Km stretch, of National Highway (NH-6) in India. NH-6 is a four-lane divided highway with raised median. This is not a limited access facility and functions similar to a multilane suburban highway with no interchanges at intersections. This 70 Km stretch, between

Kolaghat and Bally, fall in Howrah district of West Bengal. The crash records of National Highway -6 (NH-6) are obtained from FIR reports from seven police stations viz. Bagnan, Uluberia, Panchla, Sankrail, Bally, Domjur and Liluah. This stretch was divided into 69 locations. For a particular crash the location of occurrence, the severity (property damage only, minor injury crash, major injury crash, and fatal crash) and the number of persons suffering minor injury, number of persons suffering major injury and the number of fatalities were obtained from the FIR reports. The data was summarized for each intersection and each segment for analysis and is shown in Table 3.

RESULTS & DISCUSSION

The usual practice of evaluating performance of a HSID method is false identification criteria which demands knowledge of truth (i.e. whether a particular site is hotspot or not). But in actual scenario the truth is not known. The other HSID evaluation criteria which do not require knowledge of truth are based on consistent identification of a site as hotspot during two study periods (Cheng & Washington, 2008). The authors adopted Cheng and Washington (2008)'s criteria for evaluating the performance of various HSID methods. Three-year's data available was divided in two time periods: Period 1 consisting of accident records of the year 2007 and Period 2 consisting of accident records for years2008 & 2009. The 69 identified locations were ranked using crash frequency (CF) method, fatal crash frequency (FCF) method, equivalent property damage only (EPDO) method and Index of frequency severity (I_{FS}) method for both time periods and the results are tabulated in Table 3. The top 10 ranked locations are designated as hotspots. The HSID methods are evaluated using Site consistency test, Method consistency test and Total rank difference test and the results are tabulated in Table5.

In Site consistency test (T1) scale the CF method performs best, with total crash count in Period 2 of top ten selected sites in period 1 being maximum (i.e. 175). In both FCF and EDPO methods, the total crash count in Period 2 of top ten selected sites in period 1 is 168. Thus their performance is similar according to the T1 test scale. The proposed I_{FS} method has total crash count in Period 2 of top ten selected sites in period 1 as 159. Thus in T1 test scale the other three methods seem to perform better than the I_{FS} method.

In Method consistency test (T2) the CF method identifies six common locations in the two consecutive periods. The I_{FS} and EPDO methods identify four common locations and the FCF method identifies three common locations in the two consecutive periods. Thus according to the T2 test scale the CF method shows best performance, I_{FS} and EPDO method performs similarly and the FCF method shows worst performance.

In Total rank differences test (T3) CF method shows best performance with minimum rank difference of 110. The FCF and EDPO methods also have rank differences of 148 and 147 respectively. The performance of I_{FS} method according to T3 test scale is the worst with total rank difference of 180.

The consistency tests performed with the HSID methods shows that the CF method performs relatively better than the other HSID methods but none of the HSID methods have very good performance. The Site consistency test T1 shows that crash frequency method has highest total crash count in period II of sites identified in period I. However, the total crash counts, and total fatal crash counts of sites identified as hotspots in period II in the same period (as shown in Table 4) indicates that I_{FS} method actually identifies locations with greater fatal accidents compared to CF method and greater total accidents compared to FCF method. The method consistency test T2 shows that the CF method has only consistently identified six locations as hotspots during the two study periods. There are some sites which experienced high number of crashes during the initial period but experienced no or very few crashes during the second period. Since accidents are random events, a hike in one period followed by reduction in crashes in next period is very natural. Hence, this fluctuation should be attributed to the randomness rather than marking a site as truly a critical site. In the accident record obtained from NH6 the authors have identified a site (location number 130) which has 9 crashes of which only 4 are fatal during period I i.e. year 2007 but during period II the location experienced no crashes. Further investigation showed that location number 130 is actually not hazardous. The Total rank difference test T3 also indicates that the CF method performs best, but the rank difference in all three methods are greater than 100 which is really high.

Table 4: Total CF and Total FCF of top ten sites identified in period II by 4 HSID methods

CF		FCF		EPDO		I_{FS}	
Total CF	Fatal CF	Total CF	Fatal CF	Total CF	Fatal CF	Total CF	Fatal CF
215	104	201	108	204	100	205	107

The HSID evaluation tests which aim to evaluate HSID methods by comparing the sites identified during two study periods do not seem to perform satisfactorily in this situation. The reason may be that the study period I which is of 1 year duration is not actually significant and accident during a short period may occur at a particular location just by chance. Montella (2010) has used five years accident data and divided the two periods as period I 2 years and period II 3 years. From this study it is seen that selection of a 2 year period accident data at least gives more reliable results. .

Table 3 Accident Information of 69 Locations

RANK	PERIOD 1 (2007)								PERIOD 2 (2008 & 2009)							
	CF		FCF		EPDO		IFS		CF		FCF		EPDO		IFS	
	LOC	CF	LOC	FCF	LOC	EPDO	LOC	IFS	LOC	CF	LOC	FCF	LOC	EPDO	LOC	IFS
1	**27**	**19**	**106**	**8**	**121**	**574.35**	**27**	**2.29**	**28**	**30**	**13**	**13**	**106**	**1940.18**	**28**	**1.80**
2	**28**	**15**	**27**	**6**	**27**	**514.04**	**106**	**2.03**	**106**	**26**	**128**	**13**	**28**	**956.28**	**128**	**1.75**
3	**128**	**15**	**124**	**5**	**128**	**396.52**	**128**	**1.92**	**128**	**25**	**106**	**12**	**4**	**908.16**	**106**	**1.74**
4	**121**	**13**	**101**	**5**	**101**	**393.43**	**28**	**1.83**	**25**	**21**	**126**	**12**	**25**	**773.42**	**13**	**1.63**
5	**106**	**11**	**121**	**5**	**28**	**386.86**	**121**	**1.79**	**30**	**20**	**22**	**10**	**128**	**741.62**	**126**	**1.56**
6	**101**	**10**	**128**	**5**	**106**	**326.64**	**101**	**1.64**	**101**	**20**	**25**	**10**	**13**	**666.71**	**25**	**1.51**
7	**130**	**9**	**123**	**4**	**130**	**296.98**	**124**	**1.62**	**4**	**19**	**28**	**10**	**101**	**609.17**	**30**	**1.43**
8	**2**	**8**	**120**	**4**	**120**	**296.98**	**23**	**1.59**	**120**	**19**	**31**	**10**	**126**	**594.43**	**31**	**1.43**
9	**120**	**8**	**130**	**4**	**26**	**263.93**	**130**	**1.47**	**126**	**18**	**2**	**9**	**19**	**580.61**	**22**	**1.41**
10	**31**	**7**	**28**	**4**	**123**	**237.06**	**120**	**1.44**	**31**	**17**	**30**	**9**	**2**	**570.18**	**120**	**1.36**
11	123	7	23	3	124	229.81	123	1.43	19	16	35	9	30	559.84	2	1.34
12	124	7	32	3	2	187.33	1	1.40	123	16	120	8	123	551.34	35	1.34
13	26	6	3	2	32	159.07	13	1.40	2	15	123	8	110	547.32	101	1.33
14	32	6	24	2	23	147.18	18	1.40	22	15	19	7	120	536.36	123	1.29
15	4	5	103	2	31	143.32	105	1.40	13	14	101	7	31	513.74	107	1.26
16	6	5	7	2	119	124.09	32	1.24	35	14	124	7	22	461.55	4	1.25
17	25	5	22	2	6	112.20	3	1.16	27	13	4	6	35	402.31	124	1.23
18	119	5	30	2	126	112.04	24	1.16	3	12	110	6	27	376.33	19	1.23
19	126	5	127	2	4	111.04	103	1.16	110	12	129	6	16	373.90	104	1.16
20	7	4	4	2	7	111.04	31	1.06	127	12	1	5	121	326.94	21	1.11
21	22	4	6	2	22	111.04	7	1.04	10	11	7	5	3	326.17	113	1.11
22	30	4	126	2	30	96.06	22	1.04	125	11	16	5	127	318.53	110	1.09
23	127	4	26	2	103	96.06	30	1.04	129	11	27	5	8	314.28	129	1.09
24	129	4	31	2	127	96.06	127	1.04	1	10	107	5	1	301.07	17	1.07

25	3	3	1	1	3	81.08	26	1.03	5	10	112	5	129	290.50	109	1.07
26	23	3	13	1	24	81.08	4	1.01	121	10	125	5	124	277.45	117	1.07
27	24	3	18	1	25	74.90	6	1.01	7	9	127	5	125	275.91	131	1.07
28	103	3	105	1	132	74.90	126	1.01	8	9	15	4	7	271.27	133	1.07
29	125	3	21	1	105	63.01	2	0.95	16	9	24	4	119	269.64	134	1.07
30	21	2	129	1	125	62.24	21	0.85	24	9	103	4	29	262.00	27	1.03
31	34	2	119	1	35	59.92	119	0.75	26	9	121	4	107	261.31	127	1.01
32	35	2	2	1	129	50.35	129	0.72	112	9	3	3	114	251.27	7	1.00
33	1	1	5	0	21	48.03	25	0.49	114	9	5	3	10	247.02	16	1.00
34	5	1	8	0	1	33.05	125	0.30	124	9	8	3	112	240.15	112	1.00
35	13	1	9	0	13	33.05	34	0.20	18	8	12	3	103	238.06	125	1.00
36	18	1	10	0	18	33.05	35	0.20	103	8	34	3	24	222.08	1	1.00
37	104	1	11	0	34	29.96	5	0.10	116	8	104	3	18	212.81	15	0.97
38	105	1	12	0	5	14.98	104	0.10	9	7	111	3	5	206.33	111	0.96
39	108	1	14	0	104	14.98	108	0.10	23	7	10	2	26	202.08	103	0.90
40	118	1	15	0	108	14.98	118	0.10	119	7	21	2	15	195.21	121	0.90
41	132	1	16	0	118	14.98	132	0.10	15	6	23	2	9	182.85	24	0.89
42	8	0	17	0	8	0.00	8	0.00	32	6	26	2	34	174.05	12	0.85
43	9	0	19	0	9	0.00	9	0.00	34	6	29	2	116	159.46	3	0.84
44	10	0	20	0	10	0.00	10	0.00	12	5	113	2	23	155.98	29	0.82
45	11	0	25	0	11	0.00	11	0.00	107	5	114	2	12	144.09	34	0.80
46	12	0	29	0	12	0.00	12	0.00	111	4	116	2	111	129.11	5	0.79
47	14	0	33	0	14	0.00	14	0.00	6	3	9	1	32	107.95	8	0.78
48	15	0	34	0	15	0.00	15	0.00	29	3	17	1	104	99.15	10	0.72
49	16	0	35	0	16	0.00	16	0.00	104	3	18	1	21	96.06	26	0.67
50	17	0	102	0	17	0.00	17	0.00	105	3	32	1	113	66.10	114	0.67
51	19	0	104	0	19	0.00	19	0.00	108	3	108	1	108	63.01	116	0.64
52	20	0	107	0	20	0.00	20	0.00	21	2	109	1	134	48.03	23	0.63
53	29	0	108	0	29	0.00	29	0.00	113	2	117	1	6	46.10	18	0.52

54	33	0	109	0	33	0.00	33	0.00	11	1	131	1	17	33.05	9	0.49
55	102	0	110	0	102	0.00	102	0.00	17	1	133	1	109	33.05	108	0.49
56	107	0	111	0	107	0.00	107	0.00	102	1	134	1	117	33.05	32	0.46
57	109	0	112	0	109	0.00	109	0.00	109	1	6	0	131	33.05	119	0.34
58	110	0	113	0	110	0.00	110	0.00	117	1	11	0	133	33.05	6	0.15
59	111	0	114	0	111	0.00	111	0.00	118	1	14	0	105	32.28	105	0.15
60	112	0	115	0	112	0.00	112	0.00	131	1	20	0	11	29.96	11	0.05
61	113	0	116	0	113	0.00	113	0.00	133	1	33	0	118	29.96	102	0.05
62	114	0	117	0	114	0.00	114	0.00	134	1	102	0	102	4.64	118	0.05
63	115	0	118	0	115	0.00	115	0.00	14	0	105	0	14	0.00	14	0.00
64	116	0	122	0	116	0.00	116	0.00	20	0	115	0	20	0.00	20	0.00
65	117	0	125	0	117	0.00	117	0.00	33	0	118	0	33	0.00	33	0.00
66	122	0	131	0	122	0.00	122	0.00	115	0	119	0	115	0.00	115	0.00
67	131	0	132	0	131	0.00	131	0.00	122	0	122	0	122	0.00	122	0.00
68	133	0	133	0	133	0.00	133	0.00	130	0	130	0	130	0.00	130	0.00
69	134	0	134	0	134	0.00	134	0.00	132	0	132	0	132	0.00	132	0.00

However, only three years accident records are available which could only be divided in two periods as 1 year period and 2 year periods. The performance evaluation obtained from the Site consistency test, Method consistency test and Total rank difference test are not completely reliable. Thus, the hotspots identified during period II by different methods are compared and their accident history investigated to evaluate the performance of the HSID methods.

Table 5 Test results for evaluation of HSID methods

HSID tests / HSID Methods	Site Consistency Test (T1)	Method Consistency Test (T2)	Total Rank difference Test (T3)
Crash Frequency	175	6	110
Fatal Crash Frequency	168	3	148
Equivalent property damage only	168	4	147
Index of frequency severity	159	4	180

The common sites identified by CF and I_{FS} methods, are site numbers 28, 128, 106, 126, 125, 30, 31 and 120. Sites 13 and 22 are identified by I_{FS} as hotspots but not by CF method while sites 101 and 4 are identified by CF methods as hotspots but not identified by IFS method as hotspots.

The crash counts and fatal crash counts of these four sites (13, 22, 101 and 4) are summarized in table 6.

Table 6 Crash frequency and fatal crash frequency of different hotspots identified by CF and I_{FS}

Site number	Crash Frequency	Fatal Crash Frequency
13	14	13
22	15	10
101	20	7
4	19	6

It is observed that the locations identified by CF method though have very large number of crashes but the crashes often have less percentage of fatality, whereas the sites identified by I_{SF} method detects locations with both high crash frequency and high fatality.

Nine out of top ten accident prone locations by the I_{FS} method are same as those identified by the FCF method. The site (number 120) which is selected by I_{FS} method and not by FCF method has 19 crashes of which 8 are fatal. The site (number 2) which has 15 crashes out which 9 are fatal is selected by FCF method and not by I_{FS} method. Site 120 although has only one less fatality have much more total crash counts.

Thus it is observed that the I_{FS} method identifies locations with high crash frequency with a threshold number of fatality. Thus this method is recommended as a better method when limited information is available on crash, geometric design related variables as well as traffic volumes.

CONCLUSIONS

This study aims to find a suitable method for identifying accident prone locations or hotspots in presence of limited information i.e. when the traffic volume data, geometric design details of locations or regulatory information is not available. The available crash frequency and severity based methods for HSID and the new combined frequency and severity metric proposed by the authors are studied using HSID evaluation tests viz. Site consistency test, Method Consistency test and Total rank difference test available in the literature. The following conclusions can be drawn from the study:

- The HSID evaluation tests available do not perform satisfactorily when any of the study periods are of smaller duration i.e. duration of 1 year. The Crash frequency (CF) method identifies only locations having high number of crashes but fails to identify locations which have relatively high crash fatality. The Fatal crash frequency (FCF) method takes into account only fatal crashes. A site having just one less fatality than a site identified as hotspot but having very high frequency of minor and major injury crashes are often overlooked by this method.
- The I_{FS} method identifies locations with greater fatal accidents compared to CF method and greater total accidents compared to FCF method and thus can be said to identify locations experiencing high number of accidents and high number of fatal accidents more consistently than the CF and FCF methods.

The I_{FS} method is found to be most suitable in identifying high risk locations i.e. locations experiencing fatal and frequent crashes, in absence of exposure data and geometric design information, as it assigns equal weights to accident count and fatality.

ACKNOWLEDGEMENT

The authors would like to thank Mr. Satish Mahato and Mr. Jinu Anchanil, M.Tech students of IIT Kharagpur for their help in data collection.

REFERENCES

Cheng, W., and Washington, S.P. (2008). "New Criteria for Evaluating Methods of Identifying Hot Spots". *Transportation Research Record*, 2083. TRB, National Research Council, Washington, DC, pp. 76-85.

Cheng, W., and Wang, X. (2010). "Evaluation of the Joint Use of Accident Count and Accident Reduction Potential to Identify Hotspot". *Presented at 89th Annual Meeting of the Transportation Research Board*, Washington, DC.

Elvik, R.(2006). "New Approach to Accident Analysis for Hazardous Road Locations". *Transportation Research Record*, 1953. TRB, National Research Council, Washington, DC, pp. 50-55.

Hauer, E., and Persaud, B.N.(1984). "Problem of Identifying Hazardous Locations using Accident Data". *Transportation Research Record,* 975. TRB, National Research Council, Washington, DC, pp. 36-43.

Hauer, E., Allery, B.K., Kononov, J., Griffith, M.S. (2004). "How Best to Rank Sites with Promise". *Transportation Research Record*, 1897. TRB, National Research Council, Washington, DC, pp. 48-54.

Kononov J. (2002). "Identifying Locations with Potential for Accident Reductions". *Transportation Research Record,* 1784. TRB, National Research Council, Washington, DC, pp. 153-158.

Khasnabis, S., Mishra, S., and Safi, C. (2010). "Identification of hazardous locations: An alternative approach". *Presented at Conference on Infrastructure, Sustainable Transportation & Urban Planning,* CiSTUP, IISc Bangalore.

Miranda-Monero, L.F., Labbe, A., and Fu, L. (2007). "Bayesian multiple testing procedures for hotspot identification". *Accident Analysis and Prevention*, 39, 1192-1201.

Montella, A. (2010). "A comparative analysis of hotspot identification methods". *Accident Analysis and Prevention*, 42, 571- 581.

Sen, A. K., Tiwari, G., and Upadhyay, V. (2010). "Estimating marginal external costs of transport in Delhi". *Transport Policy*, 17, 27- 37.

Tarko, A.P., and Kanodia, M. (2004). "Effective and Fair Identification of Hazardous Locations". *Transportation Research Record*, 1897. TRB, National Research Council, Washington, DC, pp. 64-70.

National Highways Authority of India (NHAI). Retrieved on 2nd January 2011 from http://www.nhai.org/roadnetwork.htm