Identifying High Collision Locations for Roadway Segments without Traffic Volume Information

Rajib K. Sahaji (Corresponding Author)
Graduate Research Assistant
Department of Civil and Geological Engineering
University of Saskatchewan
57 Campus Drive
Saskatoon, SK S7N 5A9
Phone: (306) 966-8614
Email: rks826@mail.usask.ca

Peter Y. Park
Assistant Professor
Department of Civil and Geological Engineering
University of Saskatchewan
57 Campus Drive
Saskatoon, SK S7N 5A9
Phone: (306) 966-1314
Email: peter.park@usask.ca

George Eguakun
Manager
Traffic Safety Promotion Evaluation
Saskatchewan Government Insurance (SGI)
Regina, SK S4P0J9
Phone: (306) 775-6274
Email: geguakun@sgi.sk.ca

Paper Prepared for Presentation at the Effective Methods for Identification of Potential Sites for Roadway Improvements (A) Session of the 2011 Annual Conference of the Transportation Association of Canada, Edmonton, Alberta

ABSTRACT

Safety network screening is used to select road locations, such as intersections and segments, which are identified based on an abnormally high number of expected collisions or an unusually high proportion of a certain type of collisions. The most commonly used network screening method relies on rigorous safety performance functions (SPFs), which require traffic volume as an input. Many cities in Canada, including Saskatoon, currently do not collect traffic volume for every single segment within the city limits. The lack of traffic volume for a study network severely restricts the applicability of an SPF-based network screening method. On the other hand, the binomial and beta-binomial tests, which may be viewed as formal collision diagnosis tests, can be utilized to screen a roadway network, even for the roadway segments where traffic volume is unavailable. Unfortunately, previous studies that applied these two tests did not explicitly define the circumstances under which we can apply one (or both) of the tests. This study introduced a statistical test known as the 'C(α) test' to determine when we can apply, for instance, the beta-binomial test instead of the binomial test to screen roadway networks. This study used three years' (2007-09) collision data from the City of Saskatoon to demonstrate the potential benefits of applying the collision diagnosis tests as a network screening tool. We also developed collision maps using ArcGIS to visualize the screening results, and to aid governing agencies' decision-making processes in the selection of appropriate safety countermeasures for the screened hotspots.

Keywords: Network Screening, Beta-Binomial test, Overdispersion test, Collision maps, GIS

INTRODUCTION

Road safety is still a major concern for many roadway governing agencies, in spite of significant investment being made in the surface infrastructure system in recent years. In Saskatchewan, for instance, the number of fatalities per 100,000 licensed drivers was recorded as 21.9 in 2008, which is higher than any other province in Canada (Transport Canada, 2011).

One method of mitigating road safety concerns is summarized in the six-step Roadway Safety Management Process (RSMP) in the Highway Safety Manual (HSM) (AASHTO, 2010). The first of the six steps (i.e., 1) Network Screening, 2) Diagnosis, 3) Select Countermeasures, 4) Economic Appraisal, 5) Prioritize Projects and 6) Safety Effectiveness Evaluation), network screening is used to select road locations (e.g., intersections, segments) that show an unusually high number or proportion of collisions in the study network.

One of the most common network screening methods used in practice is the “expected collision frequency with Empirical Bayes (EB) adjustment approach” (AASHTO, 2010), that screens road locations based on the expected number of collisions estimated by safety performance functions (SPFs) coupled with observed collision data. However, the applicability of this method depends heavily on the availability of reliable traffic volume information (e.g., AADT) for roadway segments in the study network. In Canada, many jurisdictions often lack traffic volume information for a substantial number of segments in their governing roadway network. For these segments, SPF-based network screening method cannot be applied, simply due to the lack of input data (i.e., traffic volume).

The City of Saskatoon, for instance, collects daily traffic volume (i.e., AADT) for various classifications of roadways, such as expressways, freeways, arterials, collectors, and local roads. Table 1 shows the number and percentage of segments that contain (or do not contain) traffic volume information for each roadway classification. The results are based on the most recent three years’ traffic volume data (2007-2009) (City of Saskatoon, 2009). The City of Saskatoon has collected daily traffic volume information for more than 50% of road segments classified as expressways, freeways uncontrolled access, controlled major arterial, minor arterial, and ramp (i.e., 61%, 72%, 56%, 74%, and 71%, respectively) at least once in three years. Several other road classifications, however, have recently-collected traffic volume information for less than 50% of segments. Example road classifications include highways, uncontrolled major arterials, major collector, minor collectors, local, and local rural roadways. Notice that even a roadway classification like highway, which is used by a relatively high number of motorists, does not have recently-collected traffic volume information for many segments in its class (i.e., 69%).

Figure 1 further illustrates the lack of traffic volume issue for this study’s target network. The figure displays traffic volume information for road segments consisting of uncontrolled major arterials in the City of Saskatoon. As shown in Table 1, about 56% of road segments in this class of roadway do not have AADT information for the most recent three years (2007-2009), and thus it is not feasible to apply the SPF-based network screening method for the 56% of roadway segments classified as uncontrolled major arterials.

For the roadways that lack traffic volume information, the binomial test and/or the beta-binomial (BB) test, which are known formally as collision diagnosis tests, may be used as a safety network screening tool since these two tests do not require traffic volume information as an input. The two tests use different types of collisions from individual locations as required inputs for screening networks, and are known to be more accurate than a simple collision frequency method. Notice that the HSM referred to the BB test as the "probability of specific collision types exceeding threshold proportion" (AASHTO, 2010).

Several previous studies (Bahar et al., 2007; Bolduc and Bonin, 1995; Heydecker and Wu, 1991; Kononov and Janson, 2002; Kononov, 2002; Lyon et al., 2007; Mollet, 2004; Sayed et al., 1997) have already applied one (or both) of these tests to screen various roadway networks from different jurisdictions, and showed acceptable levels of success.

Unfortunately, however, no previous studies have presented a definitive justification, based on a quantitative analysis, regarding when and why binomial tests or BB tests are applicable for screening a roadway network.

This study has the following three specific objectives:

1. To introduce a formal statistical test to justify the use of an appropriate collision diagnosis test (i.e., binomial test vs. BB test) to screen a roadway network.
2. To develop numerical examples to demonstrate the potential usefulness of applying the proposed two tests as methods of screening collision hotspots, particularly for roadways without traffic volume information.
3. To develop GIS collision maps that visualize the analysis outcome (i.e., probability values) from the proposed network screening method.

The second section (Study Data) describes the study collision database. The third section (Model Development) explains the methodologies used in this study. The fourth section (Analysis Results) presents the outcome of the statistical tests used to select the appropriate collision diagnosis test for this study network, and presents the results of the network screening. The final section (Conclusions) summarizes the major findings.

STUDY DATA

The City of Saskatoon is the target city for this study. We used two different databases from two different agencies:

1. The most recent 3 years' (2007-2009) collision database, supplied by Saskatchewan Government Insurance (SGI).
2. A street network in GIS shape file format, supplied by the City of Saskatoon.

Collision Database

SGI has the responsibility of collecting and maintaining a collision database that records the collisions that have occurred within Saskatchewan. The SGI collision database is called the

"Traffic Accident Information System (TAIS)". Among the available ten years' collision information (2000-2009), this study used only the most recent three years' (2007-2009) collisions in order to demonstrate the potential benefits of using the selected collision diagnosis tests (i.e., binomial test, BB test) as network screening tools. We used three years' collision data since we noticed a substantial amount of changes in configuration, environment, traffic volumes, etc., in the city's roadway network during the ten year period.

The collision database includes collisions that occurred at both intersections and roadway segments. For illustration purposes, this study focuses only on collisions that occurred on roadway segments. In terms of SGI's coding scheme, there are 15 different collision configurations with a total of 6,672 roadway segment collisions, as shown in Table 2. Among the 15 collision configurations, rear end and side swipe same direction (SSSD) collisions were the two most frequent collision configurations, with 36% and 20% of proportions, respectively. This study will use these two collision configurations to demonstrate the potential benefits of applying the proposed methods in safety network screening.

Street Network

To present network screening results spatially, a GIS map presenting Saskatoon's current roadway network is necessary. The City of Saskatoon maintains a GIS base map in a shape file format that contains all of the roadways within the city limits. In the file, each roadway segment has a common location identifier known as a UGRID; this identifier is used to present the location of collisions in SGI's collision database.

ArcGIS Desktop (Version 10) is used to visualize the locations of hotspots identified by the safety network screening procedure.

MODEL DEVELOPMENT

During the last two decades, transportation safety researchers have often used two different collision diagnosis tests to investigate the unusually high proportion of a certain type of collisions for study locations (e.g., intersections, segments). The two diagnosis tests are: 1) the binomial test, and 2) the beta-binomial (BB) test.

Binomial Test

Kononov and Janson (2002) were the first to apply the binomial test (a.k.a., the direct diagnosis method) to investigate unusual collision patterns at locations where the level of safety was a concern. The test has since been applied in a number of roadway safety studies (Kononov, 2002; Kononov and Allery, 2004; Masliah and Bahar, 2006; Masliah et al. 2006; Montella, 2010). The binomial test assumes that a collision occurrence at a particular location follows a binomial distribution with two possible outcomes (i.e., a certain collision configuration or not). This method screens, for instance, a road segment that may show an unusually high proportion of a certain collision configuration, compared to the proportion of the same collision configuration from other locations. The other locations are assumed to be similar to the target location in terms of attributes (i.e., road classification).

Suppose x_i is the number of a certain collision configuration, and n_i is the total number of collisions observed at the same location i. Further suppose that S represents the total number of locations (i = 1, 2, 3, …,S) in our study dataset. The probability of observing x_i out of n_i total collisions can be expressed using the binomial probability mass function:

$$P(x_i|n_i, \bar{p}) = \frac{n_i!}{x_i!(n_i - x_i)!}\bar{p}^{x_i}(1-\bar{p})^{n_i - x_i}, 0 \leq \bar{p} \leq 1 \tag{1}$$

$$\bar{p} = \frac{\sum_{i=1}^{S} x_i}{\sum_{i=1}^{S} n_i} \tag{2}$$

The parameter $\bar{p}$ is the mean proportion of the target collision configuration (x_i) in a total of S similar locations (a.k.a., reference locations) and it is also considered to be the mean probability of observing the target collision configuration (x_i) in a total of S reference locations. The binomial test assumes that this mean probability ($\bar{p}$) remains constant in all reference locations. Thus, the probability of observing x_i or more collisions out of n_i total collisions at location i is given by the following equation:

$$P_i = P(x \geq x_i|n_i, \bar{p}) = 1 - P[x \leq (x_i - 1)] \tag{3}$$

$$= 1 - \sum_{x=0}^{x_i - 1} \frac{n_i}{x!\,(n_i - x)!}(\bar{p})^x(1-\bar{p})^{n_i - x}$$

The smaller the probability for a particular location i, the more likely it is that the expected proportion of the target collision configuration ($p_i = x_i/n_i$) at the location is larger than the estimated mean proportion $\left(\bar{p} = \sum_{i=1}^{S} x_i / \sum_{i=1}^{S} n_i\right)$ from reference locations; therefore, the location can be regarded as having a higher than expected number of collisions, particularly for the target collision configuration.

Although the binomial test has great appeal (mainly because of its simplicity in calculation), this method requires a strong assumption that the mean proportion or probability ($\bar{p}$) of a particular collision configuration remains constant in all locations. In reality, the proportion of a certain collision configuration will likely vary greatly between locations, due to factors such as traffic volume, weather condition, and driver behavior, which can vary greatly between locations; this will, in turn, create a certain amount of variability in $\bar{p}$ between locations. This variability in the proportion of a specific collision configuration is known as 'overdispersion' (Anderson, 1988; Cox, 1983; Liggett and Delwiche, 2005; Pack, 1986; Young-Xu and Chan, 2008), and is difficult to model using a binomial distribution. This means if we apply the binomial test to a collision dataset that shows a significant amount of overdispersion, the estimated probability from this test can reflect a potentially serious bias. In other words, the binomial test is only applicable to collision datasets that do not contain a significant amount of overdispersion. We will introduce a formal statistical test (i.e., C(α) test) that is designed to determine whether a study dataset contains a statistically significant amount of overdispersion.

Beta-Binomial Test

The BB test uses an additional beta distribution to take the overdispersion issues into account in a collision database. Heydecker and Wu (1991) introduced this test in the areas of road safety as a "two stage model of proportions". Since then, a number of researchers have applied this test for various study purposes (Bahar et al., 2007; Bolduc and Bonin, 1995; Kononov and Janson, 2002; Kononov, 2002; Mollet, 2004; Lyon et al., 2007; Sayed et al., 1997).

The BB test uses the binomial distribution to explain the proportion of a certain collision configuration at a target location, and uses the beta distribution to represent the proportion of the same collision configuration between reference locations, which is represented as the prior information from the reference locations. Unlike the binomial test, the BB test assumes that the proportion of a specific collision configuration ($p_i = x_i / n_i$) at a particular location i is not constant, rather it varies between locations. For the BB method, equation (1) will be modified with varying p_i as:

$$P(x_i|n_i, p_i) = \frac{n_i!}{x_i!(n_i - x_i)!} {p_i}^{x_i} (1 - p_i)^{n_i - x_i} \quad (4)$$

As stated, p_i will vary between locations following a beta distribution of p as follows:

$$Beta(p|\alpha, \beta) = \frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha,\beta)}, \quad \text{where, } 0 \le p \le 1 \quad (5)$$

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \quad (6)$$

Here, $Beta(p|\alpha, \beta)$ is the beta prior distribution representing the collision proportion information from reference locations. Γ represents the gamma function and α and β are two positive parameters of the beta prior distribution. The mean, $E(p_i)$, and variance, $Var(p_i)$, of the beta prior distribution are given by:

$$E(p_i) = \frac{\alpha}{\alpha+\beta} \quad (7)$$

$$Var(p_i) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \quad (8)$$

The Empirical Bayes (EB) method is then applied to combine the beta distribution with the binomial distribution and generate the BB distribution. As a result, the integrated BB distribution represents the combination of proportions of a particular collision configuration from a specific location and the proportion of the same collision configuration from the reference locations. The resulting BB distribution can be expressed as follows:

$$P(x_i|n_i, \alpha, \beta) = \frac{n_i!}{x_i!(n_i - x_i)!} \frac{B(\alpha + x_i, \beta + n_i - x_i)}{B(\alpha,\beta)} \quad (9)$$

Two different methods are currently being used to estimate the parameters for the BB test: 1) method of moments (MM) and 2) maximum likelihood (ML) method. Although the current version of the HSM uses the MM to estimate the parameters, many researchers have shown that the ML is the superior parameter estimation method in terms of efficiency, accuracy and consistency (Kottas and Lau, 1978; Ramachandran and Tsokos, 2009; Spanos, 1999). In addition, many automated data analysis programs such as R-language, STATA, and SAS use the ML method as a built-in parameter estimation method. We also note that the ML method is the parameter estimation method suggested by Heydecker and Wu (1991), who introduced this method to road safety area. Therefore, this study uses the ML method to estimate the parameters of the BB test.

If we let L_i be the likelihood of observing x_i specific collision configurations out of n_i total number of collisions at a location i, the likelihood of observing the specific collision configuration at all the reference locations, say S, will be $\sum_{i=1}^{S} L_i$. The log-likelihood function of L_i becomes:

$$Ln(L_i) = Ln\left(\frac{n_i!}{x_i!(n_i - x_i)!}\right) + Ln[B(\alpha + x_i, \beta + n_i - x_i)] - Ln[B(\alpha, \beta)] \quad (10)$$

If equation (10) represents the log-likelihood of observing a specific collision configuration at a particular location i, then the log-likelihood of observing the same collision configuration at all reference locations can be expressed as follows:

$$\sum_{i=1}^{S} Ln(L_i) = \sum_{i=1}^{S}\left(Ln\left(\frac{n_i!}{x_i!(n_i - x_i)!}\right) + Ln[B(\alpha + x_i, \beta + n_i - x_i)] - Ln[B(\alpha, \beta)]\right) \quad (11)$$

Since the term $Ln(n_i!/x_i!\,(n_i - x_i)!)$ is not a function of two parameters α and β, it can be safety eliminated from equation (11) and the log-likelihood function in equation (12) is rewritten as:

$$\sum_{i=1}^{S} Ln(L_i) = \sum_{i=1}^{S}(Ln[B(\alpha + x_i, \beta + n_i - x_i)] - Ln[B(\alpha, \beta)]) \quad (12)$$

Finally, the parameters α and β can be estimated by maximizing the log-likelihood function of equation (12). After applying the EB method to estimate parameters α and β, the posterior parameters α' and β' can be updated as follows:

$$\alpha' = \alpha + x_i \quad (13)$$

$$\beta' = \beta + n_i - x_i \quad (14)$$

To determine the level of safety of a target location, we now need a threshold proportion since the outcome of the BB test simply represents the probability that the estimated proportion of a certain collision configuration at a target location will exceed the (predetermined) threshold proportion of the same collision configuration at the reference locations. The most frequently-chosen threshold proportion in the literature is the median proportion. (Heydecker and Wu, 1991;

Masliah et al., 2006; Masliah and Bahar, 2006; Mollet, 2004).The median proportion (p_m) of a certain collision configuration from the reference locations can be estimated as follows:

$$\int_{p_m}^{1} Beta(p|\alpha,\beta)dp = 0.5 \tag{15}$$

The probability that the proportion (p_i) of a particular collision configuration at a certain location is greater than the median proportion (p_m) in the reference locations is estimated using the posterior parameters α′ and β′ and can be expressed as follows:

$$P(p_i > p_m|x_i,n_i) = 1 - Beta(p_m,\alpha',\beta') \tag{16}$$

As the estimated probability becomes larger, the higher the chance of the target location having a proportion of the target collision configuration beyond the median proportion of the same collision configuration from the reference segments.

Overdispersion Test

As discussed, a binomial method is not the appropriate collision diagnosis test (thus not the proper network screening method) for collision databases that contain overdispersion. To investigate whether a collision dataset contains a statistically significant amount of overdispersion or not, Tarone (1979) introduced the ‘C(α) test’ (a.k.a., Tarone’s Z-statistic). Several studies including Young-Xu and Chan (2008) recommended the ‘C(α) test’ as one of the most reliable overdispersion tests. The test statistic is formulated as follows:

$$Z = \frac{\sum_{i=1}^{S}\left\{\frac{(x_i-n_ip)^2}{p(p-1)}\right\}-\sum_{i=1}^{S}n_i}{\sqrt{2\sum_{i=1}^{S}n_i(n_i-1)}}, p = \frac{\sum_{i=1}^{S}x_i}{\sum_{i=1}^{S}n_i} \tag{17}$$

Tarone (1979) and Young-Xu and Chan (2008) clearly showed that Tarone’s Z-statistic follows an asymptotic normal distribution with the null hypothesis in favor of a binomial distribution. Thus, for a particular collision dataset, if the value of Z-statistics exceeds the critical value at the 95% confidence level, the collision dataset is considered to be overdispersed and the null hypothesis will be rejected. The next section will discuss the outcome of all the analyses, including the ‘C(α) test’.

ANALYSIS RESULTS

This study focuses on roadway segments in uncontrolled major arterials and will make use of the two most frequent collision types (i.e., rear end and side swipe same direction (SSSD)). Before applying the two tests (i.e., binomial and BB tests), we performed the ‘C(α) test’ to check whether the selected collision dataset shows statistically significant overdispersion or not. If overdispersion is statistically significant, the BB test should be applied as an appropriate screening method. Table 3 contains the results of the ‘C(α) test’ (see the column Tarone’s Z-statistic and corresponding p-value). If the p-value is less than 0.05, it means that the study dataset contains a significant amount of overdispersion at the 95% confidence level. Table 3 shows, in general, that the two different types of collision configuration (i.e., rear end and SSSD) contain statistically significant overdispersion for the three year study period (2007-2009), with the exception of the 2008 SSSD collision dataset.

Based on the '$C(\alpha)$ test' results, the BB test was selected as the appropriate network screening method for the majority of our collision datasets. A total of seven collision datasets, including four rear-ends and three SSSDs, were analyzed using the BB test.

To show an example calculation, we estimated probability values using the BB test. Based on the estimated probabilities, Table 4 shows the top 20 riskiest roadway segments classified as uncontrolled major arterials in the City of Saskatoon. As an example, we show the analysis results using the three-year total collisions (2007-2009) for the two target collision configurations (i.e., rear end and SSSD). For rear end collisions, for instance, there is almost a 100% chance that the top five segments (i.e., shaded cells under Rear End Collision in Table 4) will experience a higher than median proportion of this collision configuration, as compared to the reference segments. Similarly, the top five riskiest segments under SSSD collisions have a greater than 88% chance of experiencing a higher than median proportion of this collision configuration, as compared to the reference segments.

To visualize the study results, we used ArcGIS Desktop Version 10.0 to display the estimated probabilities of two different collision configurations spatially on a collision map. Figure 2 and 3 shows the estimated probabilities of rear end collisions and SSSD collisions that exceed the median proportion of the same collision configurations from all reference segments on the uncontrolled major arterials. From both figures, we can observe that red lines represent road segments with high probabilities (i.e., 75% - 100%), and green lines represent road segments with low probabilities. As a result, the red segments may be regarded as potentially hazardous locations in terms of the two target collision configurations.

CONCLUSIONS

The main findings of this study are summarized as follows:

1) As stated by Bahar et al. (2007) and Mollet (2004), the major advantage of the proposed network screening method (e.g., binomial test, BB test) is its capability of screening roadway networks without traffic volume information. The collection of traffic volume (AADT) information for a vast road network often consumes valuable resources and time. Therefore, the network screening method proposed in this study can be viewed as the proper choice, particularly for roadway locations where traffic volume information is not readily available.

2) Previous studies did not clearly explain the circumstances under which we can apply the BB test instead of the binomial test. This study introduced a formal statistical test known as the '$C(\alpha)$ test' to determine the amount of overdispersion in a collision dataset, which determines whether or not the BB test should be applied. All of the collision datasets (except the 2008 SSSD collision dataset) in this study contain a significant amount of overdispersion; in this circumstance, the BB test can be regarded as a more appropriate network screening method than the binomial test.

3) Table 4 and Figures 2 and 3 show roadway segments that have been identified as locations with an unusually high proportion of target collision configurations (i.e., rear end or SSSD). These results could assist public agencies, such as Saskatchewan Government Insurance, to identify roadway segments where safety countermeasures (that target a specific collision configuration) can be implemented.

4) ArcGIS was used in this study to display the analysis results spatially in a collision map. The spatial presentation of the test results (i.e., probability values) for each segment in Figures 2 and 3 provides an efficient way of visualizing the study results. This information can help transportation engineers and public agencies to determine the best locations for implementing specific safety countermeasures more efficiently.

ACKNOWDGEMENT

We acknowledge Saskatchewan Government Insurance (SGI) and the Saskatchewan Center of Excellence for Transportation and Infrastructure (SCETI) for their financial support for this study.

REFERENCES

1. AASHTO, 2010, Highway Safety Manual, 1st Edition. American Association of State Highway and Transportation Officials
2. Anderson D. A., 1988, Some Models for Overdispersed Binomial Data. Australian Journal of Statistics, 30, pp. 125-148
3. Bahar, G., Masliah, M., Philp, C., Shen, K., Park, P., and Bhim, R., 2007, Development of a Traffic Safety Management System: High Proportion Screening Tool: Selection of a High Proportion Testing Tool, State of Utah, US Department of Transportation, Interim Report
4. Bolduc, D., and Bonin, S., 1995, Bayesian Analysis of Road Accidents: Accounting for Deterministic Heterogeneity. Proceedings of the Canadian Multidisciplinary Road Safety Conference IX, Montreal, Canada
5. City of Saskatoon, 2009. Traffic Characteristics Report
6. Cox, D. R., 1983, Some Remarks on Overdispersion. Biometrika, 70, pp. 269-274
7. Heydecker, B. J., and Wu, J., 1991, Using the Information in Road Accident Records. Proc., 19th PTRC Summer Annual Meeting, London
8. Kononov, J., 2002, Identifying Locations with Potential for Accident Reductions: Use of Direct Diagnostics and Pattern Recognition Methodologies. In Transportation Research Record 1784, TRB, National Research Council, Washington, D.C., pp. 153-158
9. Kononov, J., and Allery, B. K., 2004, Explicit Consideration of Safety in Transportation Planning and Project Scoping. In Transportation Research Record 1897, TRB, National Research Council, Washington, D.C., pp. 116–125
10. Kononov, J., and Janson B. N., 2002, Diagnostic Methodology for the Detection of Safety Problems at Intersections. In Transportation Research Record 1784, TRB, National Research Council, Washington, D.C., pp. 51–56
11. Kottas, J. F. and Lau, H. S., 1978, On Estimating Parameters for Beta Distribution. Decision Sciences, 9, pp. 526-531
12. Liggett, R. E., and Delwiche, J. F., 2005, The Beta-Binomial Model: Variability in Overdispersion Across Methods and Over Time. Journal of Sensory Studies, 20, pp. 48-61
13. Lyon, C., Gotts, B., Wong, W. K. F., and Persaud, B., 2007, Comparison of Alternative Methods for Identifying Sites with High Proportion of Specific Accident Types. In Transportation Research Record 2019, TRB, National Research Council, Washington, D.C., pp. 212–218
14. Masliah, M., and Bahar, G., 2006, Using Basic Collision Data to Manage Road Safety. Annual Conference of the Transportation Association of Canada, Charlottetown, PEI
15. Masliah, M., Bahar, G., Clayton, R., and Hall, R., 2006, Managing Road using Basic Crash Data. Traffic Records Forum, Association of Traffic Safety Information Professionals (ATSIP)

16. Mollet, C. J., 2004, Developing a Traffic Safety Improvement Program: A Review and Comparison of Different Screening Network Approaches. Presented at Canadian Multidisciplinary Road Safety Conference XIV, Ottawa, Ontario

17. Montella, A., 2010, A Comparative Analysis of Hotspot Identification Methods. Accident Analysis and Prevention 42, pp. 571-581

18. Pack, S. E., 1986, Hypothesis Testing for Proportions with Overdispersion. Biometrics, 42(4), pp. 967-972

19. Ramachandran, K. M. and Tsokos, C. P., 2009, Mathematical Statistics with Applications. Elsevier Academic Press

20. Sayed, T., Navin, F. and Abdelwahab, W., 1997, A Countermeasure Based Approach for Identifying and Treating Accident Prone Locations. Canadian Journal of Civil Engineering 24, pp. 683-691

21. Spanos, A., 1999, Probability Theory and Statistical Inference: Econometric Modeling with Observational Data. Cambridge University Press

22. Tarone, R. E., 1979, Testing the Goodness of Fit of the Binomial Distribution. Biometrika, 66(3), pp. 585-590

23. Transport Canada, 2011, Canadian Motor Vehicle Traffic Collision Statistics: 2008. TP 3322, Road Safety and Motor Vehicle Regulation Collected in co-operation with the Canadian Council of Motor Transport Administrators

24. Young-Xu, Y., and Chan, K. A., 2008, Pooling Overdispersed Binomial Data to Estimate Event Rate. BMC Medical Research Methodology, 8:58

Table 1: The City of Saskatoon's AADT Information for Individual Road Classifications

Roadway Classification	Total Number of Segments	Number of Segments with AADT	Number of Segments without AADT	Percentage of Segments with AADT	Percentage of Segments without AADT
Expressway	126	77	49	61%	39%
Freeway Controlled Access	54	39	15	72%	28%
Highway	107	33	74	31%	69%
Controlled Major Arterial	140	78	62	56%	44%
Uncontrolled Major Arterial	602	262	340	44%	56%
Minor Arterial	546	402	144	74%	26%
Major Collector	1134	536	598	47%	53%
Minor Collector	567	161	406	28%	72%
Local	7003	137	6866	2%	98%
Local Rural	143	6	137	4%	96%
Ramp	161	115	46	71%	29%

Note: 1) Shaded cells represent the target roadway classification for this study.
2) Segments with AADT represent the segments that have AADT for at least one in three years (2007-2009); segments without AADT do not have any AADT for the study period (2007-2009) (City of Saskatoon, 2009).

Table 2: Observed Collisions by Collision Configuration

Collision Configuration	Collision Frequency				Overall Proportion (2007-2009)
	2007	2008	2009	2007-2009	
Rear End	866	798	706	2370	36%
Side Swipe Same Direction	466	468	391	1325	20%
Fixed or Movable Object	306	287	268	861	13%
Right Angle	154	172	137	463	7%
Lost Control Right Ditch	163	117	127	407	6%
Left Turn Straight Opposite Direction	90	85	107	282	4%
Lost Control Left Ditch	64	68	59	191	3%
Side Swipe Opposite Direction	85	59	38	182	3%
Left Turn Straight	57	54	55	166	2%
Head On	47	36	30	113	2%
Lost Control Right Ditch to Left Ditch	29	31	32	92	1%
Right Turn Same Direction	25	31	24	80	1%
Left Turn Straight Same Direction	24	15	19	58	1%
Right Turn Passing	11	19	12	42	1%
Left Turn Passing	17	8	15	40	1%

Note: Shaded cells represent the two most frequent collision configurations.

Table 3: Overdispersion Test Results for the Study Collision Database

Roadway Classification	Collision Configuration	Collision Year	Number of Collisions	Tarone's Z Statistic	P-value
Uncontrolled Major Arterial	Rear End Collision	2007	285	9.50	0.00
		2008	316	4.25	0.00
		2009	245	11.43	0.00
		2007-2009	846	23.69	0.00
	Side Swipe Same Direction Collision	2007	134	2.75	0.01
		2008	147	-0.24	0.81
		2009	121	3.31	0.00
		2007-2009	402	4.07	0.00

Note: Shaded cells represent the collision dataset that contains statistically significant overdispersion at the 95% confidence level.

Table 4: Network Screening Results for the Top 20 Segments

Collision Year (2007-2009)			
Uncontrolled Major Arterial Segments			
Rear End Collisions		Side Swipe Same Direction Collisions	
Segment ID	Estimated Probability	Segment ID	Estimated Probability
SKH7-1	100.00%	SKH3-3	95.30%
SKB7-1	99.99%	SKG7-93	89.54%
SKC9-9	99.57%	SKE8-49	88.09%
SKG8-35	99.57%	SKF7-103	88.02%
SKH9-50	99.20%	SKL9-4	88.00%
SKJ8-74	97.42%	SKN9-68	85.37%
SKB8-26	95.84%	SKG7-108	84.95%
SKG2-25	95.84%	SKG4-40	83.58%
SKN6-13	95.84%	SKG7-90	82.75%
SKL10-1	95.79%	SKJ9-78	80.23%
SKG5-32	95.04%	SKH1-2	80.23%
SKN6-5	95.04%	SKG1-8	80.23%
SKG5-37	95.04%	SKF5-4	80.23%
SKE7-68	95.04%	SKG2-20	75.08%
SKH7-10	94.19%	SKN5-27	75.08%
SKD5-1	93.45%	SKL11-5	75.08%
SKJ1-4	93.08%	SKJ8-55	72.91%
SKH9-2	91.38%	SKD5-1	72.72%
SKG7-100	91.38%	SKG7-77	71.93%
SKL9-12	90.76%	SKE7-15	71.93%

Note: Shaded cells represent the top five riskiest segments for each collision configuration.

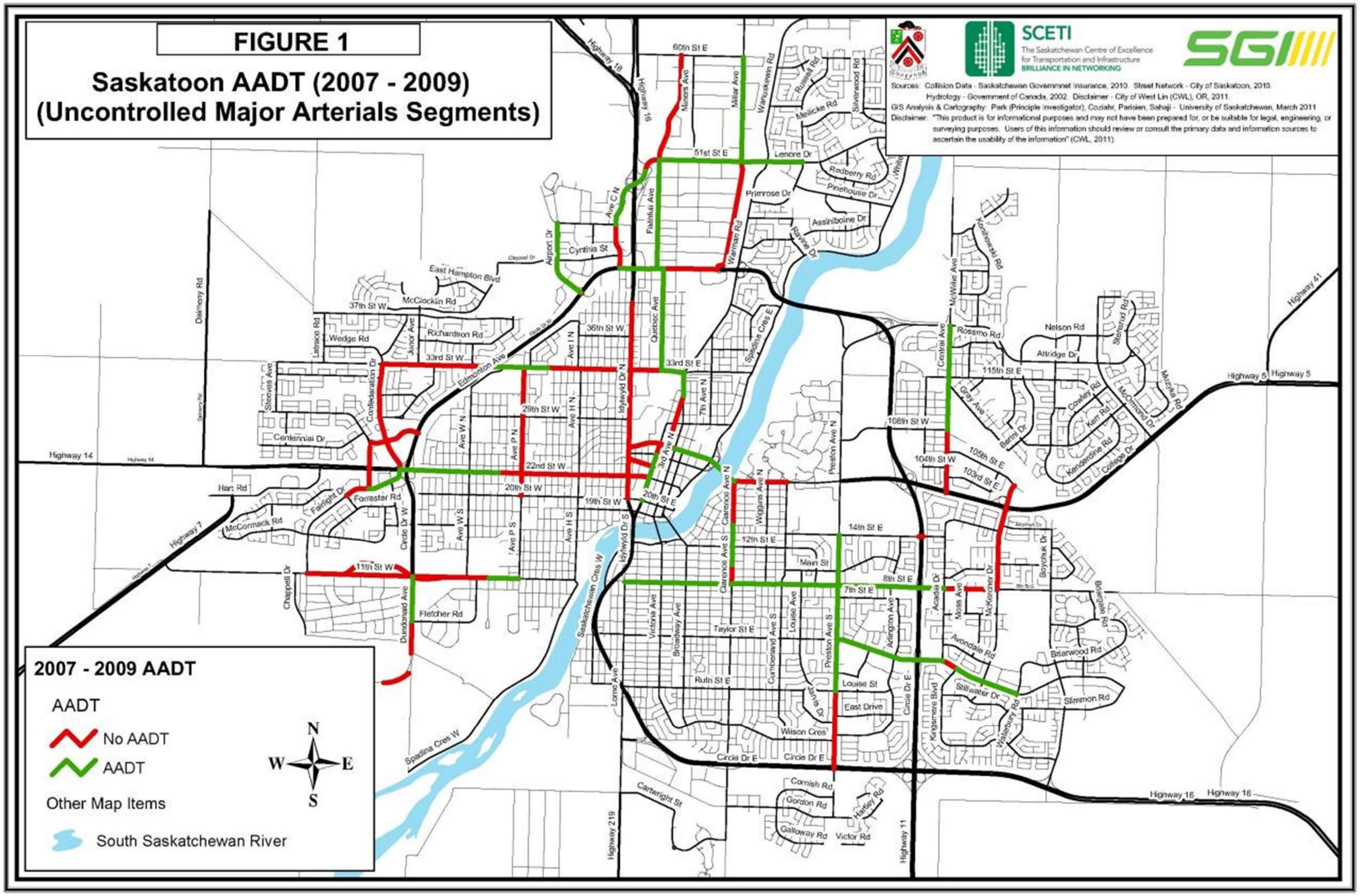

Figure 1: AADT Information for Road Segments on Uncontrolled Major Arterials in the City of Saskatoon

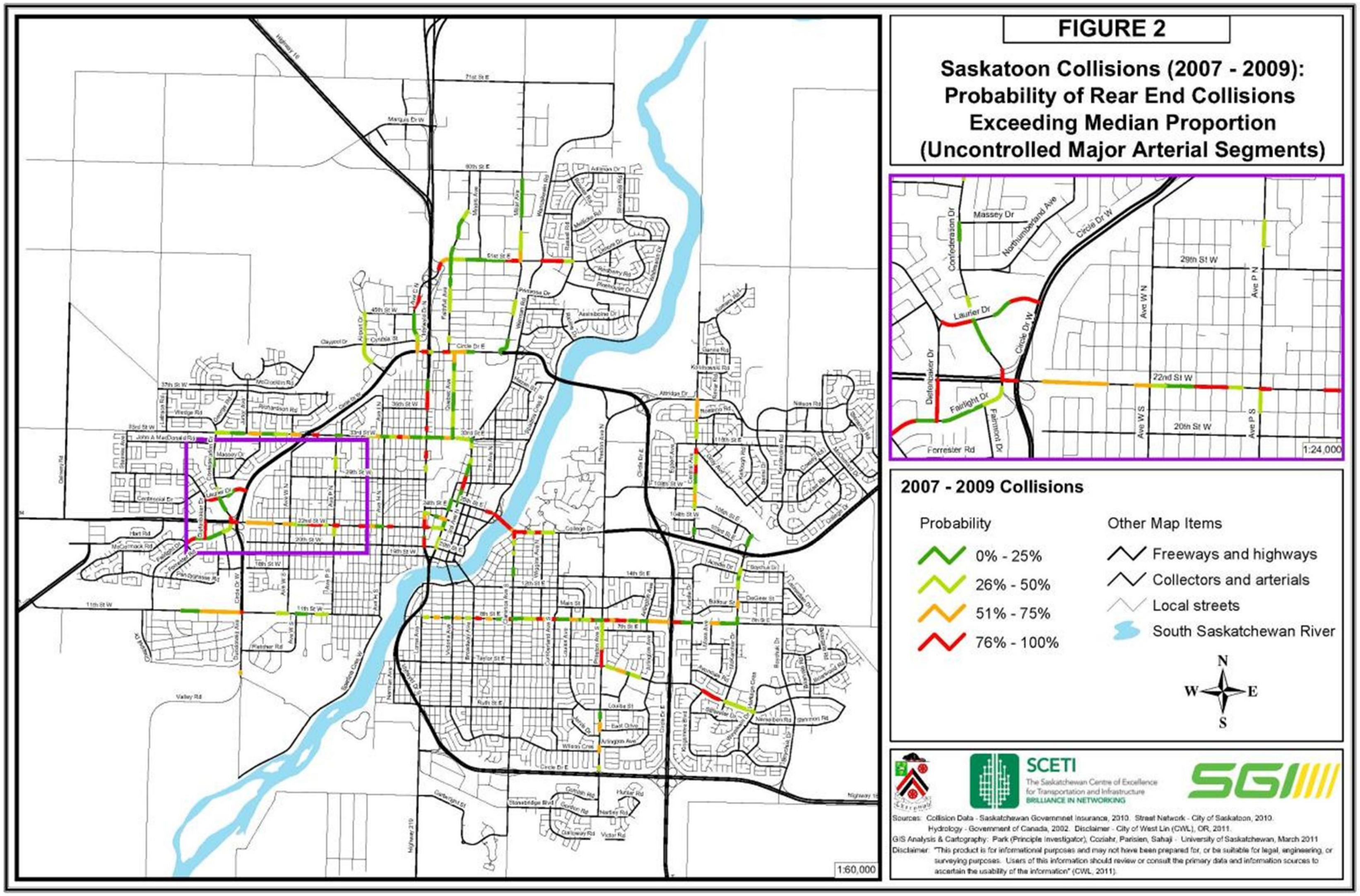

Figure 2: BB Test Results for Rear End Collisions on Uncontrolled Major Arterials

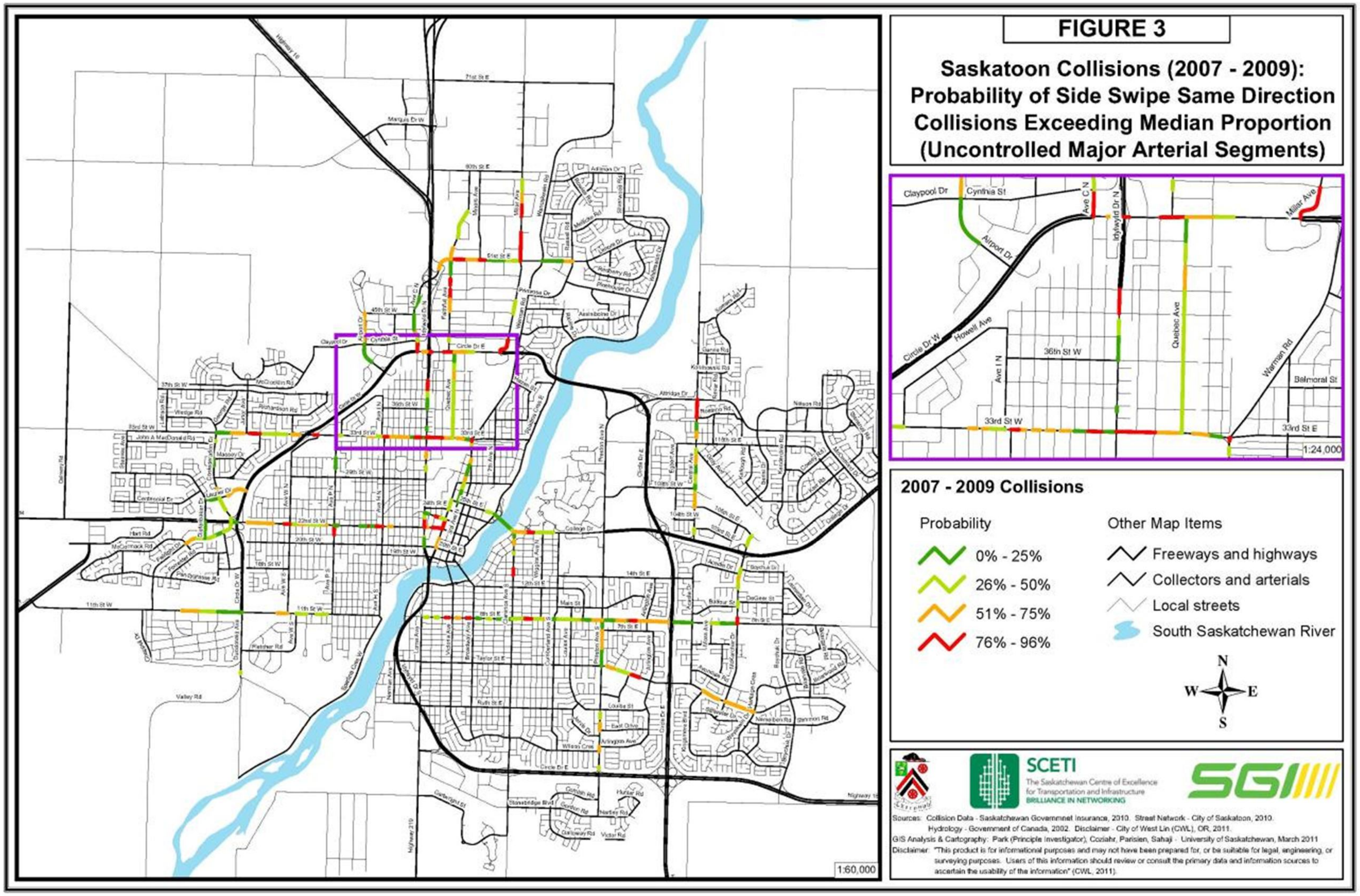

Figure 3: BB Test Results for Side Swipe Same Direction Collisions on Uncontrolled Major Arterials