Using Complementary Intersection and Segment Analyses to Identify Crash Hot Spots (Extended Summary)

Grant G. Schultz[1], Camille C. Lunt[2], Timo Pew[3], Richard L. Warr[4]

[1]*Department of Civil & Construction Engineering, Brigham Young University, 430 EB, Provo, UT, USA 84602, gschultz@byu.edu*
[2]*Avenue Consultants, 6605 South Redwood Rd., Taylorsville, UT, USA 84123, lunt.camille@gmail.com*
[3]*Exact Sciences Corporation, 5505 Endeavor Lane, Madison, WI, USA 53719, timopew@gmail.com*
[4]*Department of Statistics, Brigham Young University, 2152 WVB, Provo, UT, USA, warr@stat.byu.edu*

Abstract

Transportation agencies commonly identify and mitigate crash hot spots. Using a full roadway (i.e., including both intersections and segments) crash analysis is beneficial, but the results may be skewed towards locations with high intersection crashes. The purpose of this paper is to demonstrate the use of a combination of two unique, yet complementary crash hot spot analyses, one for intersections and one for roadway segments. The combination of the two allow for an in-depth analysis of causes associated with specific roadway conditions to identify locations that are experiencing more injury-causing crashes than predicted. The crashes on the state route network are separated into one of the two analyses; crashes within the influence area of major intersections become part of the intersection analysis and all other crashes become part of the segment-only analysis. Due to the separation between intersection- and segment-related crashes, safety concerns can be brought to light that might otherwise go unnoticed in a full network analysis. For example, crashes related to driveways or excessive queuing found in the segment analysis indicate a potential need for access management countermeasures. In addition, locations with both adjacent segment and intersection hot spots can be pinpointed as places for more in-depth analysis of the contributing circumstances surrounding the crashes at these locations. Agencies that use an intersection analysis and a segment analysis together to identify crash hot spots can benefit from an increased accuracy of hot spot locations. Furthermore, the approach to mitigating safety concerns can become clearer if hot spots are separated into intersections and segments.

Keywords: Hot spot analysis; segmentation; intersection analysis; safety; UDOT; segment analysis

[1] * Corresponding author. Tel.: +001-801-422-6332;
E-mail address: gschultz@byu.edu

1. Introduction

Finding crash hot spots within a roadway network can be a complex task for agencies globally. Roadways are made up of intersections and segments, and the interplay between them may make it difficult to identify appropriate solutions to existing safety concerns, especially since safety concerns at intersections and segments are typically mitigated by different sets of countermeasures. Agencies and researchers have continually explored ways to prioritize safety projects and reduce the number of fatalities, injuries, and societal costs that occur due to traffic crashes. In the Highway Safety Manual (HSM) the American Association of State Highway and Transportation Officials (AASHTO) has published guidelines for performing hot spot identification analyses on segments of similar number of lanes and functional classification, but little guidance is provided on how to perform an analysis for an entire network of roadways [1].

In coordination with the Utah Department of Transportation (UDOT), the Brigham Young University (BYU) Civil and Construction Engineering and Statistics Departments have conducted a series of safety-focused research. This series included the Roadway Safety Analysis Methodology which analyzed the entire Utah state route network by performing a statistical analysis on highway segments [2, 3]. The endeavor was successful in that it was able to identify segments experiencing higher crash frequencies than predicted, yet the results were concentrated in urban areas with dense, high-volume intersections. With a goal to focus on intersection hot spots, the next research in the series was the Intersection Safety Analysis Methodology which identified crash hot spots among intersections involving two state routes [4, 5]. The most recent endeavor in this series of hot spot analyses is titled the "Segment and Intersection Crash Analysis Methodologies for Utah Highways" which further develops the Intersection Safety Analysis Methodology and introduces the Crash Analysis Methodology for Segments in an effort to separate a segment analysis from any intersection bias [6]. These two most recent crash analysis methodologies form a complementary pair of crash analyses that allows for intersections and segments in the state route network to be analyzed separately without omitting or overlapping any crash data.

The purpose of this paper is to provide a summary of the use of a combination of two unique, yet complementary crash hot spot analyses, one for intersections and one for roadway segments. The combination of the two allow for an in-depth analysis of causes associated with specific roadway conditions to identify crash hot spots and safety concerns otherwise unknown to a single analysis covering the full state-wide network.

2. Methodology

For intersection and segment hot spots to be identified, the required data must first be gathered and then combined into useful data files, one for intersections, and one for segments. The following sections will first breifly sumarrize how the data are prepared for the intersection and segment analyses. The statistical models used to rank the intersections and segments will then be introduced.

2.1 Data Preparation

The data used in this research were obtained through UDOT [7]. Seven sets of roadway data are needed: AADT, speed limit, lanes, urban code, functional classification, intersections, and pavement messages. In addition, crash data for the years 2014-2018 were obtained from the UDOT Traffic & Safety Division and used in the analysis. With the help of code written in Visual Basic for Applications (VBA), the data files were organized into individual intersections and segments.

2.2 Statistical Analysis

The next step after the data preparation is to perform a statistical analysis on the data. The purpose of the statistical analysis is to identify high-priority intersections and segments (i.e., hot spots), defined in this research as those which observe more injury-causing crashes than are predicted. Identifying these locations is done using the data to draw relationships between roadway characteristics and the number of injury crashes, creating distributions of predicted injury crash totals based on roadway characteristics, and flagging segments whose number of observed injury crashes is high compared to the predicted distribution. The following sections will breifly outline the primary statistical model used, the variables used in that model, and a supplemental statistical model.

2.2.1 Primary Statistical Model

Three Bayesian models were applied to the intersection data and segment data: a Negative Binomial-Lindley model, a Zero-Inflated Negative Binomial model, and a Zero-Inflated Poisson model, all of which were built with hierarchical structures. The background and specific development of these models are described in detail in the literature [6, 8]. The models are built with four years of crash data (2014-2017) and are used to create predicted distributions of injury crashes for each intersection or segment. Values of actual (observed) crash data from a fifth year (2018) are compared to the predicted distributions and are assigned a percentile value based on where the observed number of crashes falls within the predicted distribution of crashes. This means that if the statistical model predicted the annual crash frequency at an intersection to be distributed around 6 crashes, but that intersection experienced more than 6 crashes in the comparison year, then the percentile value for that intersection would be some decimal greater than 0.5. Once all locations have been associated with a percentile value, the intersections and segments are ranked separately in order from highest percentile value to lowest.

The results in this paper are from the Zero-Inflated Poisson model which this research terms the prediction model. The probability mass function and regression equation of this model are given in Equation 1 where Y_{ij} denotes the number of crashes at intersection $i = 1,\ldots,1738$ from year $j = 1,\ldots,4$. The variable π can be interpreted as the additional probability of observing a zero, and λ_{ij} represents the expected value of the Poisson component. β_0 represents an overall intercept, β_k is a P-dimension vector of regression coefficients for intersections in urban code k for the explanatory variables and η_i is a random effect which accounts for some of the heterogeneity expected across intersections or segments.

$$P(Y_{ij} = y_{ij} | \pi, \lambda_{ij}) = \begin{cases} \pi + (1-\pi)e^{\lambda_{ij}} & y_{ij}=0 \\ (1-\pi)\frac{e^{\lambda_{ij}}\lambda_{ij}^{y_{ij}}}{y_{ij}!} & y_{ij}=1,2,\ldots \end{cases}$$
$$\ln(\lambda_{ij}) = \beta_0 + x'_{ij}\beta_k + \eta_{ij} \qquad (1)$$

2.2.2 Prediction Model Variables

The prediction model was applied to both the intersection and segment analyses with different variables. There are five variables used in the intersection prediction model: entering vehicles, percent of vehicles that are trucks, maximum number of lanes, maximum roadway width, and maximum approach speed limit. Though the segment prediction model uses the same statistical foundation as the intersection prediction model, the variables it uses are different. There are four variables used in the segment model: speed limit, number of lanes, truck percentage of AADT, and natural log of Vehicle-Miles Traveled (VMT).

2.2.3 Supplemental Statistical Model

In addition to the prediction model as described above, one more model was applied to the intersection and segment analyses. The purpose of this model, termed the severity model, is to identify locations with higher proportions of injury crashes (compared with total crashes) than predicted. The higher a site is ranked, the more likely the number of injury-causing crashes was extreme relative to its predicted proportion of injury-causing crashes; or in other words, percentile values close to 1.0 indicate sites with a larger proportion of crashes being injury-causing than would be expected based on previous years of data. Using the proportions observed in a four-year period (2014-2017), the severity model was used to create predicted distributions of injury crashes for a fifth year (2018). Like the prediction model, the locations are then ranked according to the percentile values of the observed crashes within those distributions. The severity model can be used as a supplement to the prediction model; when a site ranks high in both models it indicates that it is a location that has a high number of crashes compared to similar locations and that the proportion of injury crashes is greater than expected as well.

3. Analysis and Results

The results of the analyses successfully identified intersections and segments with much higher injury-causing crash totals in 2018 than predicted based on the four previous years of crash data (2014-2017). The 20 highest ranking intersections are listed in the literature [6].

The models identified intersections and segments with percentiles above 97.8 percent for the intersections and above 99.4 percent for the segments. Such high percentiles indicate that the number of injury crashes at the top 20 locations are on the extreme (high) end of the predicted crash distributions for those locations. Moreover, it was observed that there was more variability in crash counts at segments versus intersections, which made the average prediction error higher for segment crash counts compared to intersection crash counts.

4. Discussion

The following sections briefly introduce the results of the intersection analysis, the segment analysis, as well as the combination of the two.

4.1 Intersection Analysis Results

The proportions of signalized versus unsignalized intersections in the results (top 20) are approximately the same as the proportions in the overall dataset. The same goes for the respective ratios of state route to state route, state route to federal aid route, and signalized state route to local road intersections. In addition, the representation of the injury crash factors (a list of 31 yes/no fields including intersection-related, speed-related, distracted driving, poor roadway conditions, and night/dark conditions) is very similar to that of the entire dataset. There are, however, differences in the representation of manners of collision. The top 20 crashes have a higher percentage of angle and head-on crashes than do the entirety of the crashes in the dataset. In addition, single vehicle crashes are less represented in the top 20 intersections than they are in the entire dataset. These results indicate that the intersections experiencing higher crashes than other similar intersections may be more likely to have issues with vehicle conflict points; these issues could potentially be solved by incorporating countermeasures such as improving sight distance and increasing protected turning movements.

4.2 Segment Analysis Results

A considerable difference between the segment-only analysis and the full roadway analysis (i.e., an analysis without separation between intersection- and segment-related crashes) is that the top 20 segments were not as concentrated in urban, high-volume areas in the segment-only analysis as they were in the full roadway analysis. No segments in the southern (less populated) half of the state appeared in the top 20 of the full roadway analyses. In the segment-only analysis, however, five segments in that region (Region 4) ranked in the top 20. In addition, only five of the top 20 segments in the segment-only analysis were found in the most populous county, compared to 10 segments in the full roadway analysis. These findings may indicate that removing intersection-bias from the analysis identifies segment crash hot spots that the full roadway analysis could not.

Another interesting thing to note about the segment results is that some intersection-related crashes were present in the top-ranking segments. Considering the fact that the intersection-related crashes near major intersections were not included in the segment analysis, this finding indicates that there may be mid-block driveways and small intersections influencing segment safety. Access management countermeasures may be a strategy to reduce safety problems on these segments. With just the full roadway analysis, the distinction between crashes at busy intersections and crashes at driveways and residential accessways could not be made. The segment-only analysis reveals that a portion of the crashes not at busy intersections are due to "intersection-related" factors.

Overall, the representation of the crash factors (the same 31 fields as described previously, including "intersection-related") and manners of collision in the top 20 segments is very similar to that of the entire dataset. This indicates that the hot spots may not have any factor in particular that is driving up the number of crashes beyond what is predicted. The two most common manners of collision for the segment-related crashes are single vehicle and front to rear; together they make up nearly 75 percent of the crashes.

4.3 Combining the Intersection and Segment Results

When the results of the intersection analysis and the segment analysis are combined, there are locations with both intersection- and segment-related hot spots. Even more clusters like these are evident if the results are expanded to include the top 10 intersections and segments in each of the four UDOT administrative regions. Locations with these clusters would be great candidates for extensive safety reviews such as a Road Safety Audit. These types of safety reviews might reveal safety concerns not otherwise known by looking purely at the data, including inadequate sight distance, roadside distractions, and common unsafe driver behavior.

5. Conclusions

The purpose of this paper was to demonstrate a combination of two unique, yet complementary crash hot spot analyses, one for intersections and one for roadway segments. The combination of the two allow for an in-depth analysis of causes associated with specific roadway conditions to identify crash hot spots and safety concerns otherwise unknown to a single analysis covering the full network (i.e., segments and intersections combined).

The research shows that the simultaneous use of an intersection analysis and a segment analysis can provide insights unavailable to a single roadway analysis. Engineers and agencies can apply this methodology to their networks to improve their crash hot spot analyses. Separating the intersection hot spots from the segment hot spots brings many benefits. In addition to increased accuracy of the locations of current safety concerns, potential countermeasures become clearer. Intersection- and segment-related crashes are typically mitigated by separate techniques; therefore, the separation of the two types increases the ability of an engineer to match a hot spot location with appropriate countermeasures. In addition, removing intersection bias from segment analyses allows for the identification of crash hot spots that would otherwise go unnoticed in the network. In other words, segments that are located in areas without large intersections but that are still experiencing unexpectedly high crash frequencies can be identified. Furthermore, having a matching intersection-only analysis can make it possible for every crash on the state route network to be accounted for in exactly one of two analyses.

Acknowledgment

This research was funded by the Utah Department of Transportation Traffic & Safety Division. The authors acknowledge the guidance of the following individuals on the Technical Advisory Committee for helping to guide this this research: Jeff Lewis, Robert Miles, Ivana Vladisavljevic, Tim Taylor, Travis Jensen, Dallas Wall, Clancy Black, and Charles Allen. This report may be protected under 23 USC 409.

The authors alone are responsible for the preparation and accuracy of the information, data, analysis, discussions, recommendations, and conclusions presented herein. The contents do not necessarily reflect the views, opinions, endorsements, or policies of the Utah Department of Transportation or the US Department of Transportation. The Utah Department of Transportation makes no representation or warranty of any kind, and assumes no liability therefore.

References

1. American Association of State Highway and Transportation Officials (AASHTO). *Highway Safety Manual* (1st ed.), Washington, DC, 2010.
2. Schultz, G.G., S.T. Mineer, M. Saito, J.D. Gibbons, S.A. Seigel, and P.D. MacArthur. *Roadway Safety Analysis Methodology for Utah*. Publication UT-16.13, Utah Department of Transportation Traffic and Safety, Research Divisions, Salt Lake City, UT, 2016.
3. Mineer, S.T., G.G. Schultz, and M. Saito. Roadway Safety Modeling in Utah using the Roadway Safety Analysis Methodology. In Proceedings of *2017 Road Safety & Simulation Conference*, Delft, NE, 2017.
4. Schultz, G.G., J.D. Gibbons, M. Saito, and B.W. Clegg. *Intersection Safety Analysis Methodology for Utah Roadways*. Publication UT-18.06, Utah Department of Transportation Traffic and Safety, Research Divisions, Salt Lake City, UT, 2018.
5. Gibbons, J.D., G.G. Schultz, and M. Saito. A Methodology for Analyzing Intersection Safety. *International Conference on Transportation & Development 2019: Innovation and Sustainability in Smart Mobility and Smart Cities*, American Society of Civil Engineers, Alexandria, VA, Edited by D. A. Noyce, 2019, pp. 143-153 (ISBN 978-0-7844-8257-5 (PDF)).
6. Schultz, G.G., C.C. Lunt, T. Pew, R.L. Warr, and M.J. Heaton. *Segment and Intersection Crash Analysis Methodologies for Utah Highways*. Publication UT-20.04, Utah Department of Transportation Traffic and Safety, Research and Innovation Divisions, Salt Lake City, UT, 2020.
7. Utah Department of Transportation (UDOT). *UDOT Open Data Portal*. http://data-uplan.opendata.arcgis.com/ Accessed May 24, 2021.
8. Pew, T., R.L. Warr, G.G. Schultz, and M. Heaton. Justification for Considering Zero-Inflated Models in Intersection Safety Analysis. *Transportation Research Interdisciplinary Perspectives*, Vol. 8, 2020, 100249, ISSN 2590-1982. https://doi.org/10.1016/j.trip.2020.100249.