Ranking and Selecting Dangerous Accident Locations: Case Study

RA-2005-67

K. Geurts, G. Wets, T. Brijs, K. Vanhoof

Onderzoekslijn Kennis verkeersonveiligheid

Documentbeschrijving

Samenvatting

In Vlaanderen worden momenteel ongeveer 1014 ongevallenlocaties als 'gevaarlijk' beschouwd. Deze gevaarlijke plaatsen of zogenoemde 'gevaarlijke punten' worden geselecteerd op basis van hun historische ongevallendata voor de periode 1997_1999. Meer bepaald wordt een combinatie van gewichten gebruikt om de gevaarlijke ongevallenlocaties te rangschikken en te selecteren: respectievelijk 1 voor elke licht gewonde, 3 voor elke zwaar gewonde en 5 voor elke dode (combinatie 1_3_5). In dit rapport wordt een sensitiviteitsanalyse uitgevoerd om te onderzoeken hoe groot de impact is op de huidige rangschikking van de gevaarlijke locaties wanneer alternatieve rangschikking criteria worden gebruikt. Meer bepaald, nemen we enkel de ernstigste graad van verwonding per ongeval in rekening en gebruiken we gewichten voor deze type verwondingen die gebaseerd zijn op directe kosten, indirecte kosten en de waardering van een mensenleven. Dit resulteert respectievelijk in de gewichten 1_7_33 wanneer het ergste slachtoffer een licht, zwaar of dodelijk gewonde betreft. Daarnaast maken we gebruik van waarschijnlijkheidsgrafieken, geschat uit een hiërarchisch Bayesiaans model, om de kans dat een locatie als gevaarlijk wordt gerangschikt visueel voor te stellen. Resultaten tonen aan dat de combinatie van deze 3 alternatieve rangschikkingcriteria een groot effect heeft op de selectie en rangschikking van gevaarlijke ongevallocaties. Concreet, wanneer we op basis van deze criteria uit de 5326 ongevallocaties met minstens 3 ongevallen de 800 meest gevaarlijke locaties selecteren zullen 40,6% van deze locaties verschillen van de huidige 800 geselecteerde locaties. Op basis van deze resultaten willen we het beleid dan ook sensibiliseren om de criteria om ongevallocaties te rangschikken en selecteren zorgvuldig uit te kiezen.

Summary

In Flanders (Belgium), approximately 1014 accident locations are currently considered as 'dangerous'. These 'dangerous' accident sites are selected by means of historic accident records for the period 1997-1999. More specifically, a combination of weighing values, respectively 1 for each light injury, 3 for each serious injury and 5 for each deadly injury, is used to calculate the priority score for each accident location. In this paper, a sensitivity analysis is performed to investigate how big the impact is on the current ranking of accident sites when alternative ranking criteria are used. More specifically, we only take into account the most serious injury per accident and use a valuation of casualties based on direct costs, indirect costs and validation for human suffering to give weight to the accidents. This valuation results in the weighing values 1_7_33 when the most severe injury respectively concerns a light, serious or deadly injury. Additionally, we generate probability plots, based on estimates from a hierarchical Bayes model, in order to visualize the estimated probability that a location will be ranked as dangerous. Results showed that combining these ranking criteria will have a big impact on the selection and ranking of dangerous accident locations. In particular, when selecting from the 5326 accident locations with minimum 3 accidents, the 800 most dangerous accident sites using the 1_7_33 values, 40,6% of these locations will differ from the current selection. Considering this impact quantity, we want to sensitise government to carefully choose the criteria for ranking and selecting accident locations without stating that the criterion used in this paper should be preferred to the currently used ranking method.

Inhoudsopgave

1. INTRODUCTION

In Flanders (the Flemish speaking community of Belgium), traffic safety is currently one of the highest priorities of the government. Not only does the steady increase in traffic intensity pose a heavy burden on society in terms of the number of casualties, the insecurity on the roads will also have an important effect on the economic costs associated with traffic accidents.

One important group of bottlenecks in traffic safety are the dangerous accident locations. Literature points out that there is no universally accepted definition of what should be considered as 'dangerous' [1]. According to Hauer [2] some researchers rank locations by accident rate, some use accident frequency and some use a combination of the two. Furthermore, there is a wide range of methodologies available, ranging from simple models based on actual accident counts to advanced statistical models based on estimates.

In Flanders, approximately 1014 accident locations are currently considered as 'dangerous' [3]. These 'dangerous' accident sites are selected by means of historic accident records for the period 1997-1999. More specifically, each site where in these three years 3 or more injury accidents have occurred, is selected. Then, a combination of weighing values, respectively 1 for each light injury, 3 for each serious injury (each casualty that is admitted more than 24 hours in hospital) and 5 for each deadly injury (1_3_5), is used to calculate the priority score for each accident location. Finally, a location is considered to be dangerous when its priority score equals 15 or more. To improve the traffic safety on these locations, the Flemish government will each year, starting in 2003 for a period of 5 years, invest 100 million EURO to redesign the infrastructure of the 800 accident locations with the highest priority score.

Previous research [4, 5, 6] has shown that the use of different weighing values on the one hand and giving weigh to the severity of the accident instead of to all the injured occupants of the vehicles on the other hand does have consequences for the selection and ranking of accident locations. Furthermore, using the expected number of accidents, estimated from a hierarchical Bayes model, instead of using historic count data to rank the accident locations can overcome the problem of random variation in accident counts and will also have an effect on the selection of the most dangerous accident locations.

In this paper, a sensitivity analysis is performed to investigate how big the impact would be on the ranking and selection of dangerous accident locations in Flanders when we combine the three different ranking criteria discussed above. More specifically, we will only take into account the most serious injury per accident and use a valuation of casualties based on direct costs, indirect costs and validation for human suffering [7] to rank the accident locations. Next, we will generate probability plots, based on estimates from a hierarchical Bayes model, in order to visualize the estimated probability that a location will be ranked as dangerous.

The remainder of this paper is organized as follows. First, a formal introduction to the techniques that are used in this paper is provided. This will be followed by a description of the dataset. Next, the results of the empirical study are presented. The paper will be completed with a summary of the conclusions and directions for future research.

2. TECHNIQUES

As explained in the introduction of this paper, in this research, we will only take into account the most serious injury per accident and use a valuation of casualties based on direct costs, indirect costs and validation for human suffering to rank the accident locations. Furthermore, two quantitative measures are used in order to examine the ranking and selection of dangerous accident locations: the percentage deviation value and Bayesian ranking plots.

2.1 Valuation of Casualties

The weighing values used in this research are based on accident costs which are often used in cost-benefit analyses to value the impact of road safety measures in Norway [7]. These accident costs were estimated by Elvik in 1993 and are the sum of five main items: medical costs, loss of output, costs of property damage, administrative costs, economic costs and economic valuation of lost quality of life. This sum results in a total accident cost of minimum respectively 16600000, 3780000 and 500000 per respectively fatally injured, seriously injured and slightly injured person (1995-prices, Norwegian kroner). Converting these total costs into cost ratios between the different injury types results in the weighing value combination 1_7_33. These values represent the difference in costs that can be avoided by preventing these injuries from happening. Therefore, we will use these weighing values in our analysis as an alternative for the 1_3_5 weighing values to calculate the priority score for each accident location.

2.2 Percentage Deviation Value

In accordance with our previous research [4, 5, 6], we will use the percentage deviation value (D), eqn (1), to quantify the effects changing the ranking and selection criteria of dangerous accident locations. This measure allows comparing the rankings of two datasets containing different locations by dividing the number of accident locations that do not appear in both data sets by the total number of locations in one dataset.

$$D = 1 - \frac{G}{T} \qquad (1)$$

with G = Number of common elements in both datasets, T = Total number of elements in each dataset.

Note that the percentage deviation only gives information about the number of locations that do not appear in both ranked datasets and does not take into account internal shifts in the ranking position of these common accident locations.

2.3 Bayesian Ranking Plot

A number of statistical models have been used to estimate accident rates and/or accident frequencies at a specific location over a given interval of time (see [1, 2, 8, 9 for a review). The underlying assumption is that road accidents can be treated as random events with an underlying mean accident rate for each accident location. Accordingly, locations that in one period recorded 'x' accidents do not have, on the average, 'x' accidents in the subsequent period. Therefore, the actual count of accidents is not a very good estimate for the expected number of accidents at a location. Consequently, ordered lists constructed by ranking locations according to their empirical accident rate, and thus ignoring the variability associated with each estimate, do not ensure that the worst locations will be identified. To account for this probabilistic nature of accident occurrence

compelling arguments can be found to support the assumption that accident counts follow the Poisson probability law [10].

Recently, Empirical Bayes methods have been used in road safety to identify dangerous locations arguing that adjusting historical data by statistical estimates yields improved predictability (see e.g. [11, 12, 13, 14]). Furthermore, the use of ranking procedures based on a hierarchical Bayes approach has been proposed in literature. These methods can handle the uncertainty and the great variability of accident data and produce a probabilistic ranking of the accident locations [10, 15].

In this research, we will use the observed number of accidents, fatalities, lightly injured and severely injured casualties for a given time period for each site, in combination with 1_7_33 weighing value combination, to estimate by means of a Bayesian estimation procedure the expected ranking order for each accident location. More specifically, we followed the approach of Brijs et al [10], who proposed a multivariate hierarchical Bayes approach for ranking accidents sites. This is done by using a 3-variate Poisson distribution that allows for covariance between the number of fatalities, lightly injured and severely injured casualties. In order to combine all data into a single number that will be used for ranking the sites, a cost function, in this case the values 1_7_33, is being used that measures the expected cost of an accident according to the number of fatalities, heavy and light injured casualties. Based on these expected costs, the posterior density for the rank of each site can be derived. The parameters of the model are estimated via Bayesian estimation facilitated by Markov Chain Monte Carlo (MCMC) methods. A more detailed description of this technique can be found in Brijs et al. [10].

In Geurts et al [5], we elaborated on this technique by developing a method for deriving the estimated probability for each site i of being one of the r worst sites (with l = the total number of locations). This implies that the expected score of location i is among the r highest and hence its rank R is larger than $l - r$ (since in this ranking procedure the larger the value of R, the worst the site). Then the estimated probability $P_r (i)$ is calculated as eqn (2):

$$P_r (i) = \frac{\sum_{j=1}^{N} I\left(R_j^{(i)} > l - r \right)}{N} \qquad\qquad (2)$$

where $I( )$ is the indicator function returning a value of 1 in case that the argument is true and a value of 0 in case that the argument is false. N is the number of MCMC iterations. These probabilities allow for a heuristic rule for selecting worst sites. More specifically, if all sites would have the same characteristics, we expect that for all the sites the required probabilities will be exactly the same as any differences will be merely random perturbations. Accordingly, we expect that this probability will be equal to r/l for each site. Locations with a probability above this limit reveal a deviation from the argument about equal sites. However, note that theoretically, due to random perturbations some probabilities will be larger even in the case of equal sites.

In practice, this technique allows to estimate the probability that a location will be ranked as belonging to the r most dangerous locations. For this research, this implies that we can estimate the probability that a location will be ranked as one of the 800 most dangerous sites. These estimated probabilities will be visualized by means of 'probability plots'.

To facilitate further this approach, we can calculate some sort of confidence intervals for the probabilities, based on the minimum and maximum probability, by repeating the above procedure for a number of times. Indeed, to take into account the instability and variability that characterize the accident counts, the model does not generate exactly one ranking order for each accident site. Instead, for each location, the model produces a series of ranking orders (one for each iteration) that can be expected taking into account that the number of accidents at each location will fluctuate around a mean value that is typical for this location. In practice, this corresponds with splitting up the total number of

MCMC iterations (N) in a number of batches and calculating the estimated probability for each site after each batch. This will allow generating Bayesian confidence intervals for each site. By considering the lower limit of these intervals (the smallest generated value) this will reveal sites with a probability above the limit in a more rigorous basis reducing the effect of random perturbations.

3. DATA

To allow for a sensitivity analysis on the currently used black spot criterion, this study is based on the same data used to select and rank the 1014 currently considered most dangerous accident locations. These data originate from a large data set of traffic accidents obtained from the National Institute of Statistics (NIS) for the region of Flanders (Belgium) for the period 1997-1999. These data are obtained from the Belgian "Analysis Form for Traffic Accidents" that must be filled out by a police officer for each road accident that occurs on a public road (i.e. motorways, national and provincial roads linking towns) involving casualties, since the location of these accidents is accurately known by means of a hectometer stone marker. Hence, the identification of dangerous accident locations is related to roadway segments of numbered roads with a length of 100 meters. Furthermore, each intersection is considered as a possibly dangerous accident site. Accidents occurring in the direct neighborhood of an intersection (within 50 meters) are also incorporated in the calculations of this intersection. This means that the accident sites that are considered as dangerous locations are either roadway segments of 100 meters or intersections. These traffic accident data contain a rich source of information on the different circumstances in which the accidents have occurred: course of the accident, traffic, environmental conditions, road conditions, human conditions and geographical conditions. The accident data needed to perform this sensitivity analysis will be limited to the number of accidents per accident location. Furthermore, these data will only contain the number of fatalities (persons died within 30 days after the accident) and the number of serious casualties (persons hospitalized for more than 24 hours) and light casualties per accident location.

In total, 50961 traffic accidents with casualties are reported in this period. This corresponds with 23184 unique accident locations included in the data set. Analogously with the current selection criterion for dangerous accident locations, we only select the sites where at least 3 accidents occurred between 1997 and 1999. This results in 5326 accident locations that will be analyzed in this research.

4. RESULTS

Using the estimated priority scores from the hierarchical Bayes model, it is possible using the MCMC procedure to estimate the probability for each accident location to belong to the 'r' most dangerous locations.

In figure 1 these results are shown for the 5326 locations where minimum 3 accidents occurred between 1997 and 1999 (horizontal axis) . More specifically, the curved line in figure 1 shows the estimated probability that the location belongs to the 800 most dangerous accident locations (vertical axis), ordered by decreasing probability. The horizontal line in figure 1 represents this probability under the assumption that all sites were equally dangerous and accidents would occur randomly on the different locations. In that case, the probability that a location belongs to the 800 most dangerous accident locations would be equal for all accident locations, namely 800/5326 = 0.15.
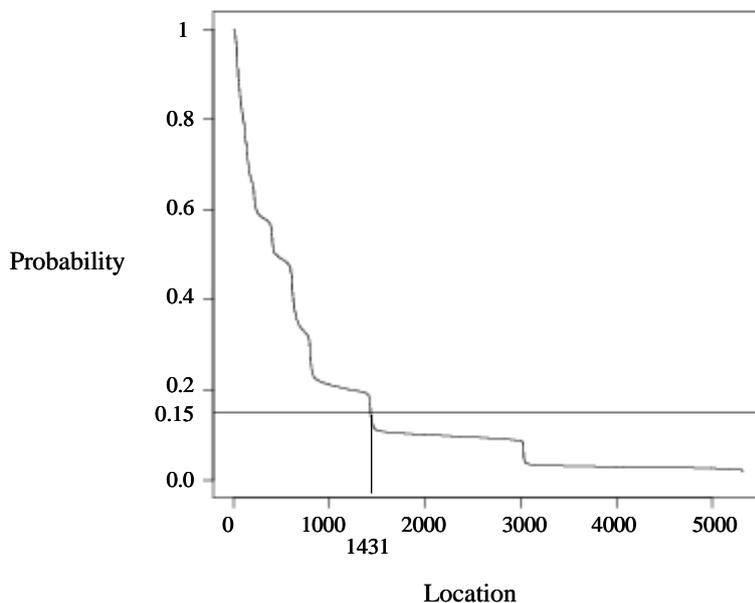

Figure 1: Bayesian ranking plot: Probability of belonging to the 800 most dangerous locations

However, from the curved line in figure 1, it can be seen that the probability of belonging to the 800 most dangerous accident locations is not at all equal for the 5326 locations with minimum 3 accidents. More specifically, 1431 locations have a probability that is larger than 0.15. These locations can be identified in figure 1 as those locations for which the curve is above the horizontal cut-off line. This indicates that these accident locations have a higher probability than expected under random conditions to qualify as one of the 800 most dangerous accident locations. When comparing the 800 accident locations that are currently considered as dangerous with these 1431 locations, it turns out that only 653 of the 800 current dangerous accident locations have a probability that is larger than 0.15. In other words, 147 accident locations are currently considered as belonging to the 800 most dangerous locations while according to the Bayesian ranking technique the probability for these locations is lower than expected under random conditions.

Furthermore, selecting the 800 locations with the highest estimated probabilities based on the results from figure 1 and comparing these sites with the 800 locations that are identified according to the Flemish selection procedure, results in a percentage deviation value of 40.6%. This corresponds with 325 accident locations that are differently selected when targeting the 800 most dangerous accident sites. Translated into costs, this means

that theoretically 205 million EURO of the 500 million EURO investment budget for redesigning these 800 most dangerous accident locations would be differently allocated.

In figure 2, for each accident location the minimum and maximum estimated probability across the different batches of MCMC iterations is shown resulting in confidence intervals.
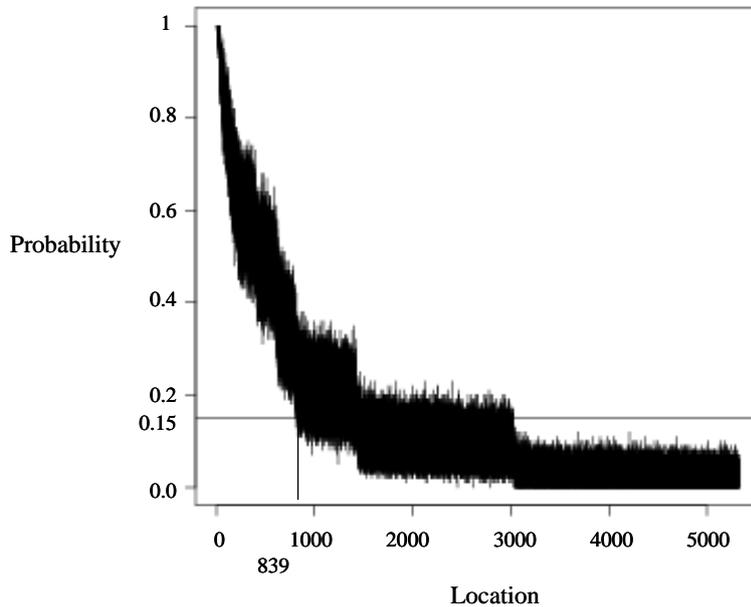

Figure 2: Bayesian ranking plot with minimum and maximum probability of belonging to the 800 most dangerous locations.

More specifically, for each accident location the vertical line in this picture represents the minimum and maximum estimated probability to belong to the 800 most dangerous locations out of the 50 MCMC batches that were included in this analysis. Note that the mean estimated probability for each accident site from the different iterations will equal the estimated probability depicted in figure 1.

These results show that for 839 accident locations the minimum estimated probability value of belonging to the 800 most dangerous accident locations exceeds the limit of 0.15. In other words, by incorporating as much variability as possible and accordingly selecting as strict as possible, 839 accident sites have a probability that is always higher than expected under random conditions to qualify as one of 800 most dangerous accident locations. When comparing these 839 locations with the 800 accident locations that are currently considered as dangerous, results show that only 503 of the 800 current dangerous accident locations have a minimum estimated probability that is larger than 0.15. This indicates that 297 accident locations are currently considered as belonging to the 800 most dangerous locations (current rank between 66 and 800) while according to the Bayesian ranking technique the minimum probability for these locations is lower than expected under random conditions.

Furthermore, results of figure 1 and figure 2 show that using the lower limit of the confidence intervals to select the accident locations with an estimated probability of belonging to the 800 most dangerous accident locations narrows down the number of sites from 1431 to 839. Consequently, the use of confidence intervals results in a more rigorous estimate of the most dangerous accident locations.

5. CONCLUSIONS

In this paper, a sensitivity analysis is performed to investigate how big the impact is on the ranking of accident locations when we use an alternative ranking criterion as the one that is currently used by the Flemish government. More specifically, we used a combination of 3 different criteria, that each were studied in earlier research, to identify and rank the accident locations. First, we only took into account the most serious injury per accident. Secondly, we used a valuation of casualties based on direct costs, indirect costs and validation for human suffering to give weight to the accidents. This valuation resulted in the weighing values 1_7_33 when the most severe injury respectively concerns a light, serious or deadly injury. Finally, we showed that the use of Bayesian estimation values instead of historic count data to rank accident locations can overcome the problem of random variation in accident counts and will also have an important effect on the selection of the most dangerous accident locations. Based on estimates from a hierarchical Bayes model, we generated probability plots, in order to visualize the estimated probability that a location will be ranked as dangerous.

Results showed that combining these ranking criteria will have a big impact on the selection and ranking of dangerous accident locations. In particular, when selecting the 800 most dangerous accident sites of all accident locations, 40.6% of these locations will differ from the current selection. Considering this impact quantity, we want to sensitize government to carefully choose the criteria for ranking and selecting accident locations without stating that the criterion used in this paper should be preferred to the currently used ranking method. It is up to the government to carefully decide which priorities should be stressed in the traffic safety policy. Then, the according weighing value combination can be chosen to rank and select the most dangerous accident locations. Furthermore, giving weight to the severity of the accident corrects for the bias that occurs when the number of occupants of the vehicles are subject to coincidence. However, in some cases (e.g. discotheques, entertainment centers), it can be reasoned that the number of occupants, and accordingly the number of injured persons, is not a coincidence but more likely a trend. For these locations, correcting for the number of passengers (hence taking into account only the most serious injury per accident) would not be advisable since the number of injuries that appear at these locations are inherent to the locations characteristics. Additionally, Bayesian ranking plots can be used to visualize the estimated probability that a location will be ranked as dangerous, based on estimates from a hierarchical Bayes model. These probability plots can provide policy makers with a scientific instrument with intuitive appeal to select dangerous road locations on a statistically sound basis.

Finally, note that one should not only rank the accident locations based on the benefits that can be achieved from tackling these locations. One should also incorporate the costs of infrastructure measures and other actions that these accident sites require in order to enhance the safety on these locations. By balancing these costs and benefits against each other the accident locations can then be ranked according to the order in which they should be prioritized.

6. REFERENCES

[1] Geurts, K. and Wets, G., Black Spot Analysis Methods: Literature Review, Flemish Research Center for Traffic Safety, Diepenbeek, Belgium, 2003.

[2] Hauer, E, Identification of sites with promise, Transportation Research Record: Journal of the Transportation Research Board, No. 1542, TRB. National Research Council, Washington, DC, pp. 54–60, 1996.

[3] Ministry of Flemish Community, Ontwerp-Mobiliteitsplan Vlaanderen, Brussels, Belgium, http://viwc.lin.vlaanderen.be/mobiliteit, Accessed July 2003 (in Dutch).

[4] Geurts, K., Wets, G., Brijs, T. and Vanhoof, K, Identification and Ranking of Black Spots: Sensitivity Analysis, Electronic Proceedings of the 83th Annual Meeting of the Transportation Research Board, Washington, January 11-15, USA, 17 pp, 2004. Also forthcoming in Transportation Research Record.

[5] Geurts, K., Wets, G., Brijs, T., Karlis, D. and Vanhoof, K., Ranking and selecting dangerous accident locations: correcting for the number of passengers and bayesian ranking plots, Flemish Research Center for Traffic Safety, Diepenbeek, Belgium, 2004.

[6] Geurts, K., Wets, G., Brijs, T. and Vanhoof, K., Identifying and ranking dangerous accident locations: Overview sensitivity analysis. Forthcoming in Proceedings of 17th ICTCT Workshop in Tartu, Estonia, 2004.

[7] Elvik, R. and Vaa, T., The handbook of road safety measures, Elsevier, 2004.

[8] Hauer, E. and Persaud, B.N., How to estimate the safety of rail-highway grade crossing and the effects of warning devices, Transportation Research Record, Journal of the Transportation Research Board, No. 1114, TRB. National Research Council, Washington, DC, pp. 131-140, 1987.

[9] Nassar, S. Integrated Road Accident Risk Model, Phd. Thesis, Waterloo, Ontario, Canada, 1996.

[10] Brijs, T., Karlis, D., Van den Bossche, F. and Wets, G. , A Bayesian Model for Ranking Hazardous Sites. In Proceedings of the 11th Symposium Statistical Software, the Royal Dutch Academy of Sciences, Amsterdam, The Netherlands, pp 55-72, 2003.

[11] Elvik, R. , Evaluations of road accident black spot treatment: a case of the iron law of evaluation studies?, Accident Analysis and Prevention 29 (2), pp. 191-199, 1997.

[12] Miaou, S.P, The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions, *Accident Analysis and Prevention 26 (4)*, pp.471- 482, 1994.

[13] Belanger, C., Estimation of Safety of Four-legged Unsignalized Intersections, Transportation Research Record, Journal of the Transportation Research Board, No. 1467, TRB. National Research Council, Washington, DC, pp. 23-29, 1994.

[14] Vogelesang, A.W, Bayesian Methods in Road Safety Research: an Overview. Institute for Road Safety Research (SWOV), Leidschendam, The Netherlands, 1996.

[15] Schlüter, P.J., Deely, J.J. and Nicholson, A.J, Ranking and selecting motor vehicle accident sites by using a hierarchical Bayesian model, The Statistician 46, pp.293-316, 1997.