ESTIMATING THE SAFETY PERFORMANCE OF URBAN ROAD TRANSPORTATION NETWORKS

by

**Dominique Lord***
Center for Transportation Safety
Texas Transportation Institute
Texas A&M University
3135 TAMU
College Station, TX
77843-3135
Tel. (979) 458-1218
E-mail: d-lord@tamu.edu

**Bhagwant N. Persaud**
Department of Civil Engineering
Ryerson University
350 Victoria Ave.
Toronto, Ontario
M5B 2K3
Tel. (416) 979-5000, ext. 6464
E-mail: bpersaud@ryerson.ca

AA&P Paper A-2160

March 1st, 2003

* Corresponding author
ABSTRACT

Transportation planning models are typically used to estimate future traffic patterns, peak period traffic, travel time, and various environmental or other related traffic flow characteristics. Unfortunately, traffic safety is seldom, if ever, explicitly considered proactively during the transportation planning process. This omission is attributed to various factors, including the lack of available tools needed to estimate the number of crashes during this process. To help fill this void, the research on which this paper is based aimed, as a primary objective, to develop a tool that would allow the estimation of crashes on digital or coded urban transportation networks during the planning process. The secondary objective of the research was to describe how the predictive models should be applied on these networks and explain the important issues and limitations surrounding their application. To accomplish these objectives, safety performance functions specifically created for this work were applied to two sample digital networks created with the help of EMME/2, a software package widely used in transportation planning. The results showed that it is possible to predict crashes on digital transportation networks, but confirmed the reality that the accuracy of the predictions is directly related to the precision of the traffic flow estimates. The crash predictions are also sensitive to how the digital network is coded, and it is shown how appropriate adjustments can be made.

Key words: safety, crash prediction models, safety performance functions, urban network, transportation planning
INTRODUCTION

Programs aimed at reducing collision frequency and severity in urban areas have traditionally been reactive in that they are geared at roads in service. Hence, by the time a location has been flagged as dangerous and treatable, many people may have already been killed or injured. Ideally, the anticipated number of crashes should be estimated before a facility is built or upgraded so that potentially hazardous elements can be identified and corrected accordingly, before the new facility is used. Although a thorough planning analysis is always conducted when a transportation agency is planning to build, modify or upgrade its facilities, traffic safety is rarely, if ever, explicitly accounted for during this process, which typically is concerned with impacts such as travel time, air pollution and fuel consumption.

The lack of available tools is a deterrent to quantifying the safety of a transportation facility during the planning process. While the tools to identify the potentially hazardous locations for roads in service are well developed, these tools are difficult to use with current urban transportation planning models, an issue that has been raised by various transportation organizations (Roberts, 2001; TRB, 2002). In a recent study, De Leur and Sayed (2002) have lamented the lack of available tools to evaluate the safety of a transportation network and have proposed a framework for proactive road safety planning, arguing that such an approach to road safety not only complements the traditional or reactive method but also prevents unsafe situations from arising in the first place. Further evidence of the recognition of the relevance of safety in the planning process and of the need for tools to do so is the commissioning of a new project under the National Cooperative Highway Research Program (NCHRP Project 8-44: Incorporating Safety into Long-Range Transportation Planning). The objective, according to the project statement, is to “develop a guidebook for practitioners that identifies and evaluates alternative ways to more effectively incorporate and integrate safety considerations in long-range state-wide and metropolitan transportation planning and decision-making processes.”

The primary objective of this paper is to demonstrate a tool that could be used in the planning process to estimate the number of crashes on digital or coded transportation networks for which the most likely flows can be estimated outside of the tool. The focus of this research is on the macroscopic representation of physical networks and on the application of safety performance functions (SPFs), or crash prediction models (CPMs), that describe the relationship between the number of crashes and the most important independent variables such as traffic flow or the number of lanes on a road segment. The SPFs developed and applied are simplistic in the sense that they tend to contain predictive rather than causal factors and since they are not intended to estimate changes in safety for minor upgrades in geometric design (e.g., lane widening, etc.) (Such a tool has been developed by Harwood et al., 2000). The focus is on important issues and limitations regarding the application of SPFs rather than on their development. The paper is divided into four sections. The first section presents a review of the coding process for physical networks. The second section describes the characteristics of available SPFs that could be applied for this research. The third section shows how the models are applied. The last section presents a discussion on the important issues and limitations related to the application of SPFs on digital transportation networks.

DIGITAL REPRESENTATION OF PHYSICAL NETWORKS

The term network may refer either to physical structures (such as streets, railroads or airline paths) or conceptual constructions (such as the assignment of personnel inside an organization or affiliations between corporations). In either case, a network is characterized by a series of nodes (i.e., vertices or points), and links (i.e., arcs or edges) that connect these nodes. In digital networks, the nodes and links usually represent physical structures such as streets, intersections and interchanges.

The digital representation of urban transportation infrastructure components is very important to analysts in that, as used by transportation planning software programs, they characterize, in a
mathematical form, the movement of goods and people within a physical network. As explained by Potts and Oliver (1972), a regularity exists in the habits of an urban population that establishes certain patterns of movements that can be described by mathematical models. Moreover, the goal of such models is to predict or estimate, as accurately as possible, future traffic patterns, peak periods, travel time, and various environmental or other related impacts. In the end, digital networks enable transportation analysts both to estimate the flow (vehicles, passengers, pedestrians) that travels through each component of the physical network and, given this estimated flow pattern, evaluate several measures of interest such as travel costs, operation characteristics (revenues, etc.), or flow by-products (e.g., pollution).

Figure 1 illustrates a typical macroscopic digital representation of a physical urban transportation infrastructure. The actual physical network is shown in Figure 1a. The network consists of four major arterial roads and several local streets. The digital representation of the physical network is illustrated in Figure 1b. This network has five nodes and four links. Four nodes (one in each corner) are used to represent the intersection of the four arterial roads. The fifth node (located to the left) is used to connect the centroid to the network through one of the three centroid connectors. This node does not represent an actual physical intersection. The four links represent the four arterial roads. The centroid and the centroid connectors are utilized to simulate the number of trips that occur on the local streets (i.e., within the zone). The intersection of local streets with the major arterial roads is not depicted in the digital representation. Local street networks can also be digitized but are not the focus of the paper since they usually fall under the category of microscopic digital networks. The reader is referred to Potts and Oliver (1972), Sheffi (1985) and Easa (1991) for a more detailed discussion concerning the digital representation of urban physical transportation networks.

**FIGURE 1 HERE**

### ISSUES IN ESTIMATING APPROPRIATE SAFETY PERFORMANCE FUNCTIONS (SPFs)

#### Previous Work

SPFs are used to estimate the number of collisions on transportation facilities such as intersections or road segments. How to estimate SPFs is not the primary focus of this paper, however, since such details are adequately covered in sources such as Hauer et al. (1988), Kulmala (1995), Miaou (1996), Poch and Mannering (1996), Sawalha and Sayed (2001), and Lord and Persaud (2000). The focus here is on the application of SPFs to transportation networks, a topic for which the literature is quite sparse. One of the rare papers on this subject, by Al-Deek et al. (1993), evaluated the safety of an urban highway corridor in Orlando, Florida for which a risk matrix that used queuing and traffic density was applied as input to predict the number of crashes. Other work of relevance includes Maher et al. (1993), who developed a function to simultaneously optimize collisions and traffic flows on urban networks, and Burrow and Taylor (1995), who used a series of SPFs to compute the expected number of crashes on two urban networks in England in order to assess network crashes for different scenarios (e.g., re-routing of traffic) and different traffic calming measures. Similarly, Nicholson and Turner (1996) and Turner and Nicholson (1998) also applied different SPFs on three urban networks New Zealand. Finally, Chatterjee and McDonald (1998), Lord et al. (1999), and Look and Abdulhai (2001) examined the effect of Dynamic Route Guidance (DRG) on traffic accidents on digital networks in the U.K. and Canada.

These studies on the application of SPFs on digital networks had a few limitations. First, some researchers applied models developed from a reference population located outside the study area. Thus, in many instances, the model details were not known and therefore the models could not be properly calibrated or validated for the study network. Indeed, the number of crashes was under-estimated by up to 34% in one study (Nicholson and Turner, 1996). Second, it was unclear in some of the studies whether the digital network was a macroscopic or microscopic representation, a detail that is important in the choice of an appropriate SPF.
A thorough literature search was conducted on models used to predict crashes at intersections and for urban corridors. The full results of the review can be found in Lord (2000). The review found that many model forms exist to predict crashes at intersections. Various model forms are summarized in Maycock and Hall (1984), Maher and Summersgill (1996), Mountain and Fawaz (1996), Brüde et al. (1998), and Miaou and Lord (2003). One characteristic for urban corridor models is related to the non-linear relationship between segment length and crashes (Mountain et al., 1996; Sawalha and Sayed, 2001). This characteristic has a direct effect on the prediction of crashes on links used on digital transportation networks. Hence, many current predictive models, which typically assume that crashes are proportional to segment length, may not be appropriate for predict crashes on links in digital networks.

**Approach for this Research**

A series of predictive models (or SPFs) was created for predicting the number of crashes at nodes and on links using data from Toronto, Ontario for 1990 to 1995. The data included 868 signalized 4-legged intersections, 200 signalized 3-legged intersections, 59 unsignalized 4-legged intersections, 117 unsignalized 3-legged intersections, and 284 urban arterial segments. The detailed description on how these models were estimated is the subject of three other publications (Lord, 2000; Lord and Persaud, 2000; Persaud et al., 2001). Three different model types were developed and applied; although models were developed for different crash severities, only the models that include all severities, including property damage only (PDO), are employed for the purposes of this paper. The first model type was used for predicting crashes at nodes. The second and third types were utilized for predicting crashes on links. As suggested by Jackett (1993), the prediction of collisions on links could be separated into two components: mid-block and intersection. The mid-block component models predict the number of crashes between minor intersections that are located on the physical network but which are not coded as nodes on the digital network. The intersection component models estimate the number of crashes at these minor intersections. This type of model only uses the link flow as input, making them more compatible with transportation planning software programs that do not utilize flows on minor intersecting approaches. In Toronto, an intersection-related crash is defined as one occurring within 15 m from the center of the intersection. Figure 2 illustrates the above description.

The predictive models for each type were separated into different categories as shown in Table 1. According to Hauer and Persaud (1996), given that enough data points exist, it is preferable to separate the data in different categories and develop a model for each category rather than to create one model with many categorical covariates.

The modeling effort recognized what has been confirmed by several researchers -- that traffic flow explains more than 50% of the systematic variation in crash occurrence, particularly for urban networks (Kulmala, 1995). This recognition, taken together with the difficulty of obtaining data on other possible predictors, and the fact that the models would be used to estimate the effects of traffic flow changes, necessitated a focus on developing models with traffic flow as the only variable of consequence. It was, however, important in developing these models to ensure that the best possible model form was used, that the proper error structure for crash count data was applied, and that account was taken of the temporal correlation in data available on an annual basis. These three issues of consequence are briefly discussed below with reference made to sources that provide more detail on each.

The first issue of consequence is the need to specify the most appropriate model form. Results of the exploratory analysis showed that the relationship between crashes and traffic volume follows either the Power function or the Gamma function. To determine which is the case, the Integral-Differentiate (ID) method proposed by Hauer and Bamfo (1997) was utilized. An example of the
selected model forms is presented in equations (1) to (3) for each of the three model types. Although the relationships presented here are of all of the Power form, many covariates incorporated into the other models developed for this work followed a Gamma relationship.

Models for nodes:

\[ E[k] = \alpha F_1^{\beta_1} F_2^{\beta_2} \]  
(1)

Models for link mid-block:

\[ E[k] = \alpha L^{\beta_1} F_L^{\beta_2} \]  
(2)

Models for link intersection:

\[ E[k] = \alpha F_L^{\beta_1} \]  
(3)

where,
- \( E[k] \) = the expected number of crashes per year;
- \( F_1, F_2 \) = entering flows of the major and minor roads in AADT;
- \( F_L \) = link flow in AADT (both directions);
- \( L \) = length of section in kilometers (less 30 m per minor intersection on the physical network but not on the digital network);
- \( \alpha, \beta_1, \beta_2 \) = coefficients to be estimated.

The second issue of consequence is the specification of the error structure. The SPFs were estimated with Genstat 5, Version 4.1 (Payne et al., 1993) which allows for the specification of a Negative Binomial (NB) error structure. The NB (also known as a Poisson-Gamma model) has been shown to more accurately describe the distribution of traffic crashes between sites than the normal distribution assumed in conventional regression modeling. The overdispersion parameter \( \alpha \) of the NB distribution, which is estimated in the model calibration process, is such that

\[ VAR[k] = E[k] + \alpha E[k]^2 \]  
(4)

where,
- \( VAR[k] \) = the variance of the expected number of crashes;
- \( E[k] \) = the expected number of crashes per year;
- \( \alpha \) = overdispersion parameter of the NB distribution.

This size of the overdispersion parameter was used to assess which models best fit a given set of data, in that, as seen from Equation 4, the smaller the value of \( \alpha \) the better a model is. In addition use was made of the more traditional goodness of fit measures.

The third issue of consequence is the temporal correlation in the annual data. Since several years of data were used for the model estimation, the temporal correlation in the data needed to be accounted for, especially in the light of the year to year variation in crash counts that is caused by factors that may change every year (e.g., economic conditions, weather patterns, crash reporting practices, etc.). Specifics of the model calibration, in which the built-in Generalized Estimating Equations (GEE) procedure of Genstat was utilized to account for temporal correlation, can be found in Lord and Persaud (2000). Further details on the development of SPFs are described in Lord (2000).

Figure 3 illustrates how the models are applied on computerized transportation networks. The same network shown in Figure 2 is used in the illustration.
APPLICATION OF MODELS

The models described in the previous section were applied on two sample networks. Sample network 1 was utilized to describe the basics of how the models should be applied to estimate the safety of a computerized road network. Sample network 2 was used to illustrate the different planning applications once the safety of a network could be estimated.

Sample Network 1

Sample network 1 is a digital representation of a small street system located in Toronto, Canada. This digital network has six links, six nodes, one centroid (labeled 7), and four centroid connectors (or dummy links) connecting the centroid to four nodes (Figure 4). The street system in this figure is laid out in a shape of a rectangle and is about one kilometer wide by two kilometers long. This network is a small subset of the Greater Toronto Area (GTA) EMME/2 (INRO, 1996) network currently used for conducting transportation planning analyses by the Joint Program in Transportation (JPINT) at the University of Toronto. All the aspects related to the coding of the network and the simulation of traffic for sample network 1 were carried out by the JPINT.

The characteristics of the links are presented in Table 2. This table includes a description of the attributes of minor intersections not coded as nodes, the number of lanes, and the measured distances of links. The 4 nodes illustrated in Figure 4 are signalized 4-legged intersections.

The traffic flow simulation, provided by the JPINT, was performed for the entire GTA area for a typical morning peak hour. Unfortunately, the predicted flows did not correspond accurately to the estimated traffic flow conditions recorded on the physical network. With the aim of improving the predicted flows, a second simulation run was carried out with Paramics (Quadstone, 1998), a microsimulation transportation software frequently used for intelligent transportation systems (ITS) applications. Again, the simulation results for the second run showed no improvements on the accuracy of the predicted flows.

After consulting with the staff at the JPINT, it was discovered that the inaccurate flow estimates were caused by the input O-D matrix and the volume-delay functions currently used by EMME/2. Since this software is normally applied at the regional level, traffic tends to be assigned to more important links such as freeways and large arterial roads, both of which have considerably higher speed limits than local roads. It was brought to the researchers’ attention that software programs such as EMME/2 are sometimes used to screen the total traffic counts traveling between the predefined regions defined in the analysis, rather than to conduct a link-by-link assessment of the traffic flow. According to the JPINT, it would be possible to substantially increase the prediction accuracy of traffic on the network by improving the volume-delay functions and by including turning penalties at nodes. Unfortunately, the JPINT was unable, due to financial and time constraints, to provide better functions and additional tools to improve the prediction of traffic. In the end, changes or modifications related to the EMME/2 output were beyond the control of this research project.

As a last measure, the predicted flows had to be substituted with actual traffic counts recorded on the physical network. The peak hour flows were expanded to AADT values from expansion factors computed from traffic counts on the physical network given that the total annual number of crashes for the entire network is of interest. Since the values are estimates computed from actual counts, the sum of inflows at intersections does not exactly equal the sum of outflows for each
intersection. Nonetheless, they were deemed to be close enough for the purpose of this
demonstration. The AADT values are shown in Figure 5.

FIGURE 5 HERE

The expanded AADT output flows of Figure 5 were used as input in the three predictive model
types. The number of crashes at nodes was estimated with the signalized 4-legged intersection
model, since all the nodes are signalized with four approaches. The predicted crashes were
estimated at nodes 1, 2, 4 and 5, all of which represent actual physical intersections, as opposed
to the nodes 3 and 6, neither of which represent a physical intersection. The number of crashes
on the links connecting nodes 1, 2, 4 and 5 was calculated by the link models. Moreover, the
number of crashes for the mid-block sections was estimated with the 4-lane non-central business
district (CBD) and 6-lane link models.

The predicted number of mid-block crashes on links located on the North-South axis was
multiplied by the adjustment factors calculated with equation (5). The mid-block models described
above were originally estimated from a sample of urban arterial road sections located between
nodes representing major physical intersections. As is, the mid-block models cannot be used
directly on the North-South links since nodes 3 and 6 do not represent physical intersections
(note: crashes are not estimated at these two nodes). Since a non-linear relationship (usually with
0.5<β<=1.0) exists between the segment lengths and the number of mid-block collisions, the
models predict more crashes for each section individually, than when applied on the entire link,
as is their actual function. The non-linear relationship can be explained by the influence of the
land-use (strip malls, gas stations, etc.) adjacent to nodes located at both ends of the link on the
number of crashes. Since between 30% and 70% of all mid-block crashes on a given link occur
near these nodes, shorter road sections, proportionally, would tend to have a greater number of
accidents than longer road sections, everything else being equal (see Lord, 2000 for additional
information on the non-linear relationship). Thus, adjustment factors are used to make predicted
number of crashes for both sections equal to the predicted number of crashes for the entire link,
as if the model of equation (2) was applied directly for the entire link. The factors are computed by
the following equation:

\[
AF = \frac{P}{P^\beta} = P^{(1\cdot\beta)} \tag{5}
\]

where,
- \(AF\) = adjustment factor;
- \(P\) = proportion of original link (0 < P ≤ 1), where P = length of section
divided by the total length of original link (as described above);
- \(\beta\) = power coefficient of the parameter L in equation (2).

The application results of the models on sample network 1 are presented in Figure 6 and Table 3.
The number inside the parentheses in this figure is the actual number of crashes that occurred on
the physical network for 1995. The figure reveals that link 1-2 has the highest number of
predicted crashes, while links 2-3 and 4-5 have the lowest number of predicted crashes. In
general, the models predict values that are very similar to the true accident counts that occurred
on the physical network, as depicted in Table 3. Almost all the crash counts were encompassed
within two standard deviations of the predicted crashes, which are deemed adequate in statistical
analysis (Hauer and Bamfo, 1997).

TABLE 3 HERE

FIGURE 6 HERE
Sample Network 2

Within the urban planning process, one could evaluate different scenarios and estimate the number of crashes that are likely to occur on various components of the network and hence add another measure in the evaluation of different scenarios. An example of such an application is shown below.

The hypothetical network was built arbitrarily with EMME/2. Sample network 2 has 15 nodes, 6 centroids, 18 links, and 6 centroid connectors. Nine nodes are used to represent intersections, while the other six nodes are employed as gateway nodes located between the centroids at the outskirts of the network and the street system. The hypothetical physical network is illustrated in Figure 7a and its digital representation appears in Figure 7b.

The sample network is four kilometers wide by two kilometers long. The attributes of the links are presented in Table 4. The prediction of crashes was performed on links and nodes located inside the rectangle bounded by nodes 11, 13, 15, and 17. The other links were used only to connect the centroids with the rectangular network.

Under the assumption that the estimated flows were predicted accurately by EMME/2, the expected number of crashes was initially computed for the original layout, as it appears in Figure 7. An origin-destination matrix consisting of 9,800 vehicle-trips was assigned to this network. The product of this assignment was expanded to AADT flows with an average factor computed from the Toronto data. The predicted number of crashes from the 2-lane, 4-lane, 6-lane road and signalized 3- and 4-legged intersection models were applied on sample network 2.

In the proposed scenario, a traffic generator was added near the center of the network. The new centroid is connected to link 16-19, half-way between both nodes. The proposed traffic generator could represent a casino, a grocery store, a new subdivision, etc. In this scenario, 1,800 vehicle-trips were added to those of the original scenario. The new flows were assigned to the network and crashes were predicted with the models described above. Table 5 presents the difference in the predicted number of crashes between the original layout and the proposed scenario used for sample network 2, for injury and all crashes combined.

Table 5 reveals that an increase of 1,800 vehicle-trips (~16%) leads to an increase of about 13% in the predicted crashes on the entire network. The biggest increases occur on every link that connects to node 19 in the center of the network. For link 16-19, the increase in crashes is attributed primarily to the new node 26.

Based on the results shown in Table 5, it would be possible to plan for the implementation of additional measures to minimize the increase in crashes for the critically affected links and nodes. One could, for example, plan measures that would redistribute the flow of traffic within the transportation network. Alternately, one could plan various localized interventions (site specific) on the physical network to improve safety, such as at nodes 19 and 26.

The SPF's can also be used to quantify safety at the planning stage for forecasted traffic flows. The aim is to estimate, in general terms, the safety of different scenarios without knowing all the detailed characteristics of the physical network that, more often than not, are unavailable for long-term planning projects. With the use of SPF's that account for time trend in accident occurrence the safety implications of different scenarios could also be estimated for short to mid-range planning projects.
DISCUSSION OF ISSUES

This paper has shown that SPFs can be used to estimate crashes on digital transportation networks. However, this process should not be undertaken without an appreciation for a few notable issues. It is important to note that some of the issues can be easily overcome if one has access to additional information and better output that may be provided by more sophisticated planning processes. Six such issues are addressed below.

The first issue is related to the prediction of traffic flow by transportation planning models. As described extensively in the previous section, the output flows of transportation planning software programs can be inaccurate and this, of course, will lead to an inaccurate appraisal of network safety. Therefore, if several alternative scenarios are being evaluated, and the number of predicted collisions is one of the evaluation criteria, it is advisable to conduct a sensitivity analysis that reflects the uncertainty in the estimates of predicted flow and, therefore, predicted collisions.

The second issue concerns the applicability and accuracy of SPFs. The models applied for this work were developed and validated with traffic flows, crash and road characteristics data from Toronto. Thus, their use to estimate crashes on a digital representation of the Toronto arterial network is appropriate. However, they can be used for other jurisdictions using a recalibration procedure in Harwood et al. (2000) that was tested by Persaud et al. (2001). Regarding the accuracy of SPFs it should not be surprising that the predicted number of crashes will always be associated with a certain degree of uncertainty, hence the importance of estimating the variance in this kind of analysis. Ideally, the crash counts should lie within two standard deviations of the predicted crashes; values outside this boundary are usually considered as outliers.

The third issue is related to how the network is coded. The link-based models proposed here and in the literature indicate a non-linear relationship between segment length and accident counts. This non-linearity has important implications for the prediction of crashes on computerized networks. As described above, there exist many links that are subdivided into smaller sections in order to connect zonal flows to the network. In some networks, every street and intersection can be digitally coded on the transportation network. The proposed link-based models would overestimate the number of crashes on the various segments on the network since each segment is not bounded by a major arterial road. Hence, it is necessary for the modeller to adjust the predicted flow with equation (5) described in this paper. On the other hand, the adjustment would not be necessary if crashes were proportional to segment length.

The fourth issue is that the application of SPFs for the evaluation of the safety impact of changes in a network as was illustrated with sample network 2 can be problematic. The predictive models used in this work were applied with the assumption that all the characteristics that impact on safety with the exception of the modified ones remained unchanged. It is unreasonable to expect that a modification for one road characteristic will not affect other factors that influence safety. For instance, increasing the capacity of a road may influence the land-use located along the modified links or nodes (private entrances, parking, etc.). In other instances, transportation agencies may automatically conduct minor road improvements within the vicinity of an added traffic generator. Unfortunately, the models used in this research and others that are available are unlikely to capture the safety impacts of such changes. More detailed models could be used to capture the safety effect of changes such as the ones described above. However, the lack of readily available information about the physical characteristics of the network presents a formidable obstacle that a new NCHRP project (17-26: Methodology to Predict the Safety Performance of Urban and Suburban Arterials) is trying to overcome. Even so, it may be difficult to know, at the planning stage, the exact physical characteristics of roads and intersections or the type of land-use located within different sections of a future network being evaluated. Consequently, it may be difficult to apply detailed models, even if these were available. In addition, many characteristics that influence crashes (economic conditions, introduction of new laws and city by-laws, etc.) may change over time. Thus, the models applied in this research may not be able to reasonably predict the number of crashes when used for long-term forecasting analysis. Nonetheless, they
do provide a good base since traffic flow is by far the most important exploratory variable that is associated with accident occurrence.

The fifth issue is related to the amount of information required to perform the analysis. For instance, the location and the characteristics of each minor intersection on links must be known before the link-based models can be applied to digital networks as illustrated in this paper. Since it may not be always possible to know the exact location of these minor intersections on networks that are yet to be built, it would be desirable for general-purpose models to be created and evaluated.

The final issue is that the models developed for this research could not be used directly within transportation planning software programs. Instead, the traffic flow output of these programs was used in another computer process (i.e., a spreadsheet) to calculate the predicted crashes on the digital network. This computation process turned out to be very cumbersome. It would be desirable, therefore, to include the accident prediction facility directly inside the planning software, as is done in MicroBENCOST (TTI, 1993). Before this step can be taken, however, it would be necessary to adequately develop and test the required tools. This research has provided a flying start in this development.

CONCLUSION

This research has shown that it could be very informative, and that it is certainly possible, to estimate the safety of transportation networks at the planning stage. The tools developed and demonstrated in this research would allow planners to add a very useful criterion to the evaluation of different alternatives. Most important, the results demonstrated some important pitfalls and limitations in the application of SPF s on computerized transportation networks. Given the increased liability of transportation agencies for traffic safety problems, it is important that they take action at the planning stage, which will allow them to show that proper steps were taken to reduce the number and severity of crashes. As stated by a speaker at a recent seminar on road safety audits (MTO, 1998, pg. 40): "...an organization that audits, that recognizes that audits are a process intended to help road safety, that gets an audit done and then follows through and reports and documents what it will and will not do with respect to the audit report, suffers a better chance in a court of law than an organization that buries it's head in the sand and says, 'We don't want to find out about these things, we're better off to be blind to them'.” Although additional work needs to be done regarding the integration of traffic safety models into the urban transportation planning process, the product of this research should help other researchers pursue interesting and useful areas of research that would help, at the planning stage, to reduce the number and the severity of collisions on urban roads.

FURTHER WORK

There are many avenues for further research. For instance, additional models for nodes should be developed and applied. These models should be separated by crash type and for different time periods (e.g., peak periods, nighttime, etc.), as proposed by Turner and Nicholson (1998). To simplify computation of crashes on links, SPF s that combine segment- and intersection-based models should be investigated. As it stands now, the analyst has to use a series of three models that may be cumbersome to handle if the analysis is conducted outside computer transportation program such as in a spreadsheet. The SPF s, once finalized, should be directly incorporated as a separate module inside transportation planning software programs, possibly with a facility to recalibrate them for different jurisdictions and time periods using the procedure in Harwood et al. (2000). By including the models in the program, it would be easier to visually assess the predicted crashes on the network with the graphical tool of the software. The use of crash modification factors (CMFs) to estimate minor changes on the network should be explored within the context of this research (Hauer and Persaud, 1996; Harwood et al., 2000). Network-based predictive models should also be explored, as proposed by Miaou et al. (2002). These models consider a network as one entity and attempt to incorporate spatial statistics attributes, restricted to the
physical network, within the modeling process. Finally, it would be desirable to estimate the safety within zones in order to predict the total number of crashes in the study area by including all local streets not shown in the computerized road network.

ACKNOWLEDGEMENT

The research was partly supported by an operating grant from the Natural Sciences and Engineering Research Council of Canada (NSERC). This support is greatly appreciated, as is the assistance of the City of Toronto Transportation Department who provided the data for calibrating the safety performance functions used in the research.

REFERENCES


Maycock, G, Hall, R.D., 1984. Accidents at 4-arm Rondabouts. TRRL Laboratory Report 1120, Transportation and Road Research Laboratory, Crowthorne.


a) Physical network                                b) Digital network

FIGURE 1 Digital representation of networks
FIGURE 2 Illustration for three model types \((E(k)_N, E(k)_{LM}, E(k)_{LI})\)

\(E(k)_N\) = expected number of accidents at nodes
\(E(k)_L = E(k)_{LM} + E(k)_{LI}\) = expected number of accidents on links
\(E(k)_{LM}\) = expected number of accidents for mid-block component (link flow)
\(E(k)_{LI}\) = expected number of accidents for intersection component (link flow)
FIGURE 3 Illustration of how the SPF's are applied on digital transportation networks

\[ E(k)_n = \alpha_F \cdot F_1^{\beta_1} \cdot F_2^{\beta_2} \]

\[ E(k)_c = \alpha_3 \cdot F_1^{\beta_1} + \alpha_3 \cdot F_2^{\beta_2} + \alpha_c \cdot L^{\beta_3} \cdot F^{\beta_4} \]

\[ E(k)_l = \alpha \cdot F_1^{\beta_1} \cdot F_2^{\beta_2} \]

- \( E(k)_n \) = expected number of accidents at nodes
- \( E(k)_c \) = expected number of accidents on links
- \( \alpha_3 \) and \( \alpha_c \) are for the intersection component models; \( \alpha_l \) is for the mid-block component model

Note: the model forms illustrated are for a power relationship.
FIGURE 4 Sample network 1
FIGURE 5 Expanded AADT counts
FIGURE 6 Predicted crashes in one year on sample network 1

Nodes not representing intersections

Note: ( ) = actual number of collisions
a) Physical network

b) Digital transportation network

FIGURE 7 Sample network 2
<table>
<thead>
<tr>
<th>Nodes</th>
<th>Link Mid-Block</th>
<th>Link Intersection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signalized 4-legged</td>
<td>2-lane roads</td>
<td>Signalized 4-legged</td>
</tr>
<tr>
<td>Signalized 3-legged</td>
<td>4-lane roads (CBD and non-CBD)</td>
<td>Signalized 3-legged</td>
</tr>
<tr>
<td>Unsignalized 4-legged</td>
<td>4-lane roads CBD</td>
<td>Unsignalized 4-legged</td>
</tr>
<tr>
<td>Unsignalized 3-legged</td>
<td>4-lane roads non-CBD</td>
<td>Unsignalized 3-legged</td>
</tr>
<tr>
<td></td>
<td>6-lane roads</td>
<td>All grouped together</td>
</tr>
<tr>
<td>Link</td>
<td>lanes</td>
<td>^1Sig4L</td>
</tr>
<tr>
<td>------</td>
<td>-------</td>
<td>---------</td>
</tr>
<tr>
<td>2-3</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>3-4</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>1-2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>4-5</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>6-1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>5-6</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

^1 Sig=signalized, Unsig=unsignalized, 4L=4 legs, 3L=3 legs
^2 length less 0.030 km for each minor intersection on link + 0.015 for the node at each end, if it represents an intersection
<table>
<thead>
<tr>
<th>Link/Node</th>
<th>Intersection Component</th>
<th>Mid-block Component</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>E(k)(^1)</td>
<td>K(^2)</td>
<td>E(k)</td>
</tr>
<tr>
<td>2-3</td>
<td>5.2 (2.44)(^3)</td>
<td>7</td>
<td>8.6 (4.53)</td>
</tr>
<tr>
<td>3-4</td>
<td>5.4 (3.41)(^3)</td>
<td>4</td>
<td>8.4 (4.42)</td>
</tr>
<tr>
<td>1-2</td>
<td>6.9 (4.38)(^3)</td>
<td>17</td>
<td>27.2 (12.6)</td>
</tr>
<tr>
<td>4-5</td>
<td>4.9 (3.08)(^3)</td>
<td>0</td>
<td>23.0 (12.1)</td>
</tr>
<tr>
<td>1-6</td>
<td>11.6 (4.83)(^3)</td>
<td>12</td>
<td>13.2 (6.96)</td>
</tr>
<tr>
<td>5-6</td>
<td>10.7 (3.43)(^3)</td>
<td>3</td>
<td>13.0 (6.87)</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total on network</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^1\) Estimated crashes calculated from SPFs, \(^2\) Actual crash count, \(^3\) Standard deviation
### TABLE 4 Characteristics of links in sample network 3

<table>
<thead>
<tr>
<th>Link (node to node)</th>
<th>Lanes</th>
<th>Sig4L</th>
<th>Sig3L</th>
<th>Unsig4L</th>
<th>Unsig3L</th>
<th>length (km)</th>
<th>length - intersections</th>
</tr>
</thead>
<tbody>
<tr>
<td>11-12</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1.00</td>
<td>0.94</td>
</tr>
<tr>
<td>12-13</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1.00</td>
<td>0.94</td>
</tr>
<tr>
<td>13-14</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2.00</td>
<td>1.94</td>
</tr>
<tr>
<td>14-15</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2.00</td>
<td>1.94</td>
</tr>
<tr>
<td>15-16</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1.00</td>
<td>0.97</td>
</tr>
<tr>
<td>16-17</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1.00</td>
<td>0.97</td>
</tr>
<tr>
<td>17-18</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2.00</td>
<td>1.94</td>
</tr>
<tr>
<td>18-11</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2.00</td>
<td>1.94</td>
</tr>
<tr>
<td>12-19</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2.00</td>
<td>1.97</td>
</tr>
<tr>
<td>14-19</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1.00</td>
<td>0.97</td>
</tr>
<tr>
<td>16-19</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2.00</td>
<td>1.97</td>
</tr>
<tr>
<td>18-19</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1.00</td>
<td>0.97</td>
</tr>
</tbody>
</table>
## TABLE 5 Difference in the expected number of crashes per year
(proposed scenario - original layout)

<table>
<thead>
<tr>
<th>Link</th>
<th>Mid-Block Component</th>
<th>Intersection Component</th>
<th>Total Columns 1 + 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Injury</td>
<td>All crashes</td>
<td>Injury</td>
</tr>
<tr>
<td>11-12</td>
<td>0.34</td>
<td>1.27</td>
<td>0.34</td>
</tr>
<tr>
<td>12-13</td>
<td>0.35</td>
<td>1.32</td>
<td>0.21</td>
</tr>
<tr>
<td>13-14</td>
<td>0.03</td>
<td>0.10</td>
<td>0.00</td>
</tr>
<tr>
<td>14-15</td>
<td>1.61</td>
<td>5.44</td>
<td>0.48</td>
</tr>
<tr>
<td>15-16</td>
<td>0.27</td>
<td>1.02</td>
<td>0.08</td>
</tr>
<tr>
<td>16-17</td>
<td>0.17</td>
<td>0.63</td>
<td>0.03</td>
</tr>
<tr>
<td>17-18</td>
<td>1.49</td>
<td>4.78</td>
<td>0.40</td>
</tr>
<tr>
<td>18-11</td>
<td>0.06</td>
<td>0.19</td>
<td>0.01</td>
</tr>
<tr>
<td>12-19</td>
<td>1.49</td>
<td>5.87</td>
<td>0.32</td>
</tr>
<tr>
<td>14-19</td>
<td>1.84</td>
<td>7.25</td>
<td>0.07</td>
</tr>
<tr>
<td>16-19</td>
<td>0.68</td>
<td>3.06</td>
<td>0.03</td>
</tr>
<tr>
<td>18-19</td>
<td>1.11</td>
<td>4.27</td>
<td>0.11</td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* a positive number indicates an increase in the expected number of crashes.
Underline: change in predicted crashes significant (5%)