Another Look at the Relationship Between Accident- and Encroachment-Based Approaches to Run-Off-the-Road Accidents Modeling

Shaw-Pin Miaou
Research Staff Member
Center for Transportation Analysis, Energy Division
Oak Ridge National Laboratory
P.O.Box 2008, MS 6073, Building 3156
Oak Ridge, TN 37831, USA
Phone: (423) 574-6933; Fax: (423) 574-3851
E-mail Address: PNI@ORNL.GOV

August 1997

To be submitted for review for PUBLICATION in Accident Analysis & Prevention

Prepared by the
Oak Ridge National Laboratory
Oak Ridge, Tennessee 37831-6366
managed by
LOCKHEED-MARTIN ENERGY RESEARCH CORP. for the
U.S. DEPARTMENT OF ENERGY
under contract DE-AC05-96OR22464

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED
DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.
Another Look at the Relationship Between Accident- and Encroachment-Based Approaches to Run-Off-the-Road Accidents Modeling

Shaw-Pin Miaou
Research Staff Member
Center for Transportation Analysis, Energy Division
Oak Ridge National Laboratory
P.O. Box 2008, MS 6073, Building 3156, Oak Ridge, TN 37831, USA
Phone: (423) 574-6933; Fax: (423) 574-3851; E-Mail: PNI@ORNL.GOV

ABSTRACT

Understanding the relationships between roadside accidents and roadside design is imperative to developing cost-effective, road-related countermeasures to improve roadside safety. Much of what is known today about the relationships remains to be qualitative in nature. Recent studies have suggested that new, cost-effective analysis approaches and data collection efforts are essential if a more quantitative basis of such relationships is to be developed. Historically, models used in previous studies to develop the relationships have been categorized as using either an accident-based approach or an encroachment-based approach. The former has a solid statistical ground, but has been criticized as being overly empirical and lacking engineering basis. The latter, on the other hand, has analytical and engineering strengths, but has been described as being full of subjective assumptions and lack of sufficient supporting data. In addition, these two approaches seem to have been treated as two competing, disconnected approaches, and very few attempts have been made by roadside safety researchers to combine the strengths of both. The purpose of this study was to look for ways to combine the strengths of both approaches. The specific objectives were (1) to present the encroachment-based approach in a more systematic and coherent way so that its limitations and strengths can be better understood from both the statistical and engineering standpoints, and (2) to apply the analytical and engineering strengths of the encroachment-based approach to the formulation of mean functions in accident-based models. To demonstrate the strength of mean functions so obtained, accident-based models were developed using such mean functions for guardrail and utility pole accidents. Furthermore, to show how the accident-based model can be useful to the encroachment-based model, the developed accident models were used to estimate the roadside encroachment rate—a basic input parameter that is required by the encroachment-based model and is expensive and technically difficult to collect. The estimated rates were found to be consistent with those obtained in earlier encroachment-based studies. This is an indication that estimating basic encroachment parameters using accident-based models can be a viable approach to reducing encroachment data collection cost. In addition, unlike estimating encroachment parameters from the field-collected encroachment data, the use of accident-based models to estimate encroachment parameters does not require the development of a procedure to distinguish between controlled and uncontrolled encroachments, which can be subjective and technically difficult to do in practice. This paper concludes with a discussion on future research.

Key Words: Run-Off-the-Road Accident, Roadside Design, Vehicle Roadside Encroachment, Encroachment-Based Model, Accident-Based Model
Another Look at the Relationship Between Accident- and Encroachment-Based Approaches to Run-Off-the-Road Accidents Modeling

Shaw-Pin Miaou
Research Staff Member
Center for Transportation Analysis, Energy Division
Oak Ridge National Laboratory
P.O. Box 2008, MS 6073, Building 3156, Oak Ridge, TN 37831, USA
Phone: (615) 574-6933; Fax: (615) 574-3851; E-Mail: PNI@ORNL.GOV

1. INTRODUCTION

Understanding the relationships between roadside accidents and roadside design is imperative to developing cost-effective, road-related countermeasures to improve roadside safety. To date, much of what is known about the relationships remains to be qualitative in nature or dependent on subjective engineering guesses [Ray et al., 1995; Daily et al., 1997]. Recent studies suggested that new and cost-effective analysis approaches and data collection efforts are essential if a more quantitative basis of such relationships is to be developed [Mak and Sicking, 1992; Viner, 1995; Mak and Bligh, 1996]. Models used in previous studies to develop the relationships between run-off-the-road accidents (RORA) and roadside hazards, such as utility poles, trees, guardrail, median barriers, and embankments, have been categorized as using either an accident-based approach or an encroachment-based approach [Transportation Research Board (TRB), 1987; Daily et al., 1997].

The accident-based approach uses statistical regression models to develop the relationships in which the RORA frequency of hitting a particular or a combination of roadside hazards is the dependent variable, and traffic flows, roadway mainline designs, roadside designs, and other variables are the explanatory variables (or covariates) [Zegeer et al., 1987; Zegeer et al., 1990; Miaou, 1996]. In the last decade or so, there has been a steady realization of the statistical advantages of using the Poisson and negative binomial (NB) regression models over the conventional normal distribution-based regression models when this approach is used to model road accidents [Maycock and Hall, 1984; Miaou and Lum, 1993]. The theory behind the Poisson and NB regression accident-based models have been discussed quite extensively in many recent publications [e.g., Miaou, 1994; Maher and Summersgill, 1996; Miaou, 1996]. The goal of these accident-based models is not only to estimate the expected number of accidents and its association with key covariates, but also to estimate the statistical uncertainty associated with the estimates. In general, these accident-based models have been developed with a solid statistical ground.

Under the Poisson and NB regression models, a mean function, which is a function that relates the mean (or expected) number of accidents to the covariates, is typically assumed to have an exponential form. This functional form has several desirable mathematical properties and has been widely accepted in other research areas, such as biostatistics and econometrics [Miaou, 1996]. Some of the desirable properties include: (1) it is a multiplicative function that allows interactive effects of covariates on accidents to be easily represented; (2) it ensures that the mean accident rate is always nonnegative; and (3) it is mathematically convenient to obtain standard statistical inferences for the model. However, the use of such a functional form has been criticized as being overly empirical and lack of engineering basis.

The encroachment-based approach uses a series of conditional probabilities to describe the sequence of events resulting in a roadside accident [Glennon, 1974; TRB, 1987; Daily et al., 1997].
A typical sequence of events considered by this approach is: (1) an errant vehicle leaves the traveled way and encroaches on the shoulder; (2) the location of encroachment is such that the path of travel is directed towards a potentially hazardous roadside object; (3) the hazardous object is sufficiently close to the travel lanes that control is not regained before encounter or collision between vehicle and object; and (4) the collision is sufficiently severe enough to result in an accident of some level of severity. The idea of the encroachment-based approach was to formulate and estimate these conditional probabilities based on a combination of traffic, vehicle dynamics, and driver behavior theories. Appendix F of Transportation Research Board (TRB) Special Report 214 (SR214) provides a good description of the concept behind the encroachment-based approach and its application on two-lane undivided roads [TRB, 1987].

Over the last 30 years, there has been a constant effort to develop and refine the encroachment-based models. Despite these efforts, the encroachment-based approach is still being criticized as being full of subjective assumptions and lacking sufficient supporting data [Daily et al., 1997]. In addition, available vehicle encroachment data, including encroachment rates, were collected on a small number of road sections and are largely outdated. The Federal Highway Administration (FHWA) and TRB have been addressing the requirements and collection of such data through their sponsorship of several roadside safety projects. As a result, rather comprehensive data collection plans have been proposed, and results of pilot data collection efforts have been reported [Mak and Sicking, 1992; Mak and Bligh, 1996; Daily et al., 1997]. A review of these plans and pilot data collection results suggests that the cost of collecting the required roadside field data will be very high. Furthermore, the validity of any field-collected encroachment data may be questionable because of the technical difficulty in distinguishing between controlled (or intentional) and uncontrolled (or unintentional) encroachments.

A recent review of the encroachment-based approach and its relationship with the accident-based approach is given in Miaou [1996]. The encroachment-based approach is appealing because of its analytical and engineering strengths. It allows useful results from other studies, especially those in the areas of driving behavior and vehicle dynamics, to be directly incorporated into the model in a sensible way. In addition, systematic exploration and assessment of different road- and vehicle-based countermeasures, which have the potential of reducing the probability of the occurrence of each encroachment event described above, can be conducted with such an approach. The study by Fancher et al. [1994] is an example of using the encroachment-based approach to assess the potential benefits of using Intelligent Transportation Systems (ITS) technologies to improve road safety.

Historically, researchers on roadside safety seem to have treated these two approaches as two competing, disconnected approaches and seldom or never attempt to seek the opportunity to combine the strengths of both [Mak and Sicking, 1992; Miaou, 1996]. In a recent study, Miaou attempted to point out the complementary nature of the two approaches and suggested that the accident-based approach can benefit from the encroachment-based thinking in obtaining a mean function that has better engineering basis and interpretation [Miaou, 1996; Miaou, forthcoming]. Furthermore, because the data required to estimate basic encroachment parameters, such as encroachment rates, for use in the encroachment-based model are expensive and difficult to collect in practice, Miaou proposed a method to estimate some basic encroachment parameters using accident-based models. The method was developed based on an exploration of the probabilistic relationship between a roadside encroachment event and an RORA event. Miaou illustrated the concept and use of such a method by first using data from three States, which were contained in an FHWA Seven States Cross-Section Data Base, to develop a RORA prediction model for rural two-lane undivided roads. The model was then used to estimate encroachment rates by setting the clear zone width to zero and the sideslope ratio to 1:1 in the RORA prediction model.
The study presented in this paper is an extension of Miaou’s study [Miaou, 1996; Miaou, forthcoming]. The purpose of this study was to look for ways to combine the strengths of both approaches in roadside safety research. The specific objectives were (1) to present the encroachment-based approach in a more systematic and coherent way so that its limitations and strengths can be better understood from both statistical and engineering standpoints, and (2) to apply the analytical and engineering strengths of the encroachment-based thinking to the formulation of mean functions in accident-based models. To demonstrate the use of mean functions so obtained, an accident-based model was developed using such mean functions for guardrail and utility pole accidents from the same data base as that used by Miaou [1996]. Furthermore, to show how the accident-based model can be useful to the encroachment-based model, similar to the approach proposed by Miaou, encroachment rates are estimated using the developed guardrail and utility pole accident models. The rates were compared with those estimated by Miaou [1996] and other encroachment-based studies, such as Hutchinson and Kennedy [1966], Cooper [1980], and SR214 [1987].

This paper is organized as follows: Section 2 provides a systematic examination of the encroachment-based thinking, addressing its limitations and strengths from both the statistical and engineering standpoints. Section 3 describes a way to formulate mean functions for accident-based models using encroachment-based thinking described in Section 2. Section 4 uses the formulation suggested in Section 3 to develop accident-based prediction models for guardrail and utility pole accidents and then estimates roadside encroachment rates from these models. This paper concludes with a discussion of future work.

The following discussion focuses on two-lane, undivided roads. However, the extension to other roadway types should be straightforward. Also, in the discussion, a "roadside encroachment" is said to occur when an errant vehicle crosses the outside edges of the travelway and encroaches on either the inside or outside shoulder. Thus, for a two-lane, undivided road which has no inside shoulder, the total number of roadside encroachments includes departures of vehicles from near-side and far-side edges of the travelway in both directions. It is also important to note that roadside encroachments refer only to uncontrolled (or unintentional) encroachments. In other words, the "controlled or intentional encroachments" resulting from vehicles intentionally driven outside of the travel lane (e.g., onto shoulders and traversable medians) are not counted as encroachments.

---

2. ENCROACHMENT-BASED THINKING

Consider a vehicle traveling through a road section of length $L$. The road section is characterized by its mainline and roadside conditions. On the mainline, the condition is characterized by its key design attributes, including lane width, horizontal curvature, and vertical grade, and by its traffic conditions, such as traffic density and car-truck mix percentages. On the roadside, the number of various types and sizes of roadside objects, their locations along the road section, and their lateral offsets from the edge of travelway are some of the main safety-related characteristics of interest.

For clarity, in the following presentation, it may be necessary in some instances to indicate which road section within the sample sections is being considered. Under such instances, we will assume, without loss of generality, that the $i$th sample road section is under consideration.
Overall Concept

For a particular type of roadside object, such as guardrails or utility poles (made of the same material), the passage that leads the subject vehicle to hit one of these objects and results in a reportable accident is modeled as four sequential stochastic processes. Each of the four processes is modeled by a conditional probability. Figure 1 gives an overview of these processes and the key determinants that affect the outcome of each process. Process 1 determines the probability that the vehicle will encroach on the roadside. If the vehicle encroaches, the encroachment is characterized by its encroachment speed and angle, which are used as input to Process 2. Given an encroachment speed and angle, Process 2 determines the probability that the encroachment location of the vehicle is in a potentially hazardous envelope associated with one of the roadside objects under consideration, which makes the impact with the object possible. If the encroaching vehicle is in a hazard envelope, Process 3 determines whether the vehicle will encroach far enough to collide with the object. The key output of Process 3 is an impact speed, which can be zero (i.e., no collision). Provided that the vehicle collides with the object, the impact speed is used as input to Process 4 to determine whether the impact will result in a reportable accident.

For ease of exposition, this paper uses the following notations: $\xi_q$ symbolizes the occurrence of a vehicle roadside encroachment; $H_q$, $C_q$, and $A_q$ represent, respectively, the hazard envelope, collision event, and accident event that are associated with the $q$th roadside object on the road section, where $q = 1, 2, ..., Q$, and $Q$ is the total number of objects on both sides of the road section; $X^{(1)}$, $X^{(2)}$, $X^{(3)}$, and $X^{(4)}$ denote key determinants associated with Processes 1, 2, 3, and 4, respectively, that are observable and available; and $Z^{(1)}$ represents those unobservable and unavailable variables associated with Process 1. As noted from the accident-based literature, variables in $Z^{(1)}$ are mainly driver- and vehicle-related variables which are usually unavailable by individual site [Miao, 1996]. Using these notations, the conditional probabilities associated with these four processes, which pertain to the $q$th object on the road section, are further denoted by $P(\xi_q | X^{(1)}, Z^{(1)})$, $P(In_H_q | \xi_q, X^{(2)})$, $P(C_q | In_H_q, X^{(3)})$, and $P(A_q | C_q, X^{(4)})$. Note that (1) “In $H_q$” symbolizes that the vehicle is in the hazard envelope and implies that a roadside encroachment has occurred; (2) for the models to be discussed later in this paper on $P(C_q | In_H_q, X^{(3)})$, collision $C_q$ implies that the encroachment is in hazard envelope $H_q$; and (3) accident $A_q$ implies the occurrence of collision $C_q$.

Mathematically, the encroachment-based approach separates the probability of a vehicle being involved in $A_q$, given determinants $X^{(1)}$, $X^{(2)}$, $X^{(3)}$, $X^{(4)}$, and $Z^{(1)}$, into a series of conditional probabilities as follows:

$$P(A_q | X^{(1)}, X^{(2)}, X^{(3)}, X^{(4)}, Z^{(1)})$$

$$= P(\xi_q | X^{(1)}, Z^{(1)}) P(A_q | \xi_q, X^{(2)}, X^{(3)}, X^{(4)})$$

$$= P(\xi_q | X^{(1)}, Z^{(1)}) P(In_H_q | \xi_q, X^{(2)}) P(C_q | In_H_q, X^{(3)}) P(A_q | C_q, X^{(4)})$$

At the microscopic level, the mechanics and driver behavior that determine the outcome of each of the four processes can be quite complex. To make the concept workable and useable, there has been a conscientious effort attempting to simplify each of these processes based on engineering judgment and limited, available data. How each process has been and can potentially be treated in the encroachment-based study is described in more detail in the rest of the section.
Process 1

One of the efforts made to simplify the analytical procedure is to separate Process 1 from the rest of the processes. This is done by introducing the assumption that roadside design has a negligible effect on roadside encroachment probability. In other words, this assumption says that the probability of encroachments can be solely determined from mainline conditions. This assumption simplifies the analytical procedure significantly. Specifically, it allows the analysts to stay focused on the analysis of the last three processes, which pertain to the effect of alternative roadside designs on the accident probability, without worrying about the roadside encroachment rates and characteristics being altered as a result of different roadside designs. However, the validity of this assumption has not been formally challenged and may be debatable. For example, some may believe that some roadside conditions, such as a sharp sideslope, can significantly affect a driver’s driving behavior (e.g., driver’s attentiveness) and therefore affect roadside encroachment rates.

Using this assumption, previous encroachment-based studies typically started the analysis with the premise that data on encroachment rates and two key encroachment characteristics, namely, encroachment speed and angle, are available either from some data collection efforts or can be pre-estimated. Encroachment rates, angles, and speeds are, however, expected to vary by weather condition, roadway functional class, average annual daily traffic (AADT), lane width, horizontal curvature, vertical grade, and even car-truck mix percentage [Mak and Sicking, 1992]. Thus, besides the potential technical difficulties to be discussed next, the effort required to collect a comprehensive set of encroachment data under all these conditions will be tremendous.

Many data collection instruments have been used to obtain roadside encroachment data, including observing tire-tracks, monitoring maintenance records of roadside objects, and using electronic monitoring equipment such as video cameras [Mak and Sicking, 1992; Daily et al., 1997]. However, so far, most of the data collection efforts have been of small-scale and experimental in nature and were not considered successful. As indicated earlier, even if the controlled encroachments can be distinguished from the uncontrolled encroachments, the effort required to collect all of the necessary data is expected to be expensive. One additional note regarding controlled and uncontrolled encroachments is that no formal, quantitative definitions of them exist. Recognizing that oftentimes there is a fine line between a controlled and an uncontrolled encroachment, the procedure used to remove controlled encroachments from the collected encroachment data are bound to be subjective.

For simplicity, the discussion in this paper does not distinguish far-side and near-side (i.e., left- and right side) encroachments. When accidents are available by side of road and direction of travel and when roadside objects, and perhaps traffic, are inventoried by side of road, there may be some incentive to consider them separately, which is conceptually not difficult to do.

Output of Process 1

The output of this process is the occurrence or non-occurrence of an encroachment event. The probability of the occurrence of the event is represented by a conditional probability $P(\xi | X(I), Z(I))$. If an encroachment occurs, the encroachment is further characterized by its encroachment angle and speed, two key parameters which will be used in the subsequent processes to determine whether the encroachment can result in a collision with a roadside object. As indicated before, the assumption is that these basic encroachment parameters, including $P(\xi | X(I), Z(I))$ and encroachment speed and angle, are not affected by roadside conditions.
**Input to Processes 2-4**

The input to the rest of the processes is an encroachment event which is characterized by its speed and angle. The speed and angle are treated as random variables generated from a joint probability density function, \( f(v, \phi) \), where \( v \) and \( \phi \) represent encroachment speed and angle, respectively. The conditional probability for Processes 2-4 can now be re-expressed in terms of \( v \) and \( \phi \) as

\[
P(A_q \mid \xi, X^{(2)}, X^{(3)}, X^{(4)}) = \int_{v_0}^{v_{ref}} \int_{\phi_0}^{\phi_{max}} P(A_q \mid \xi, v, \phi, X^{(2)}, X^{(3)}, X^{(4)}) f(v, \phi) \, dv \, d\phi
\]

Given \( f(v, \phi) \), the focus of the encroachment-based studies has been on the three conditional probabilities: \( P(In H_q \mid \xi, v, \phi, X^{(2)}) \), \( P(C_q \mid \xi, v, \phi, X^{(3)}) \), and \( P(A_q \mid C_q, v, \phi, X^{(4)}) \).

Due to lack of data on encroachment speed and angle, the density functions that have been used in the literature and current roadside analysis software are largely chosen using engineering judgment [Mak and Sicking, 1992; Daily et al., 1997]. To facilitate the following discussion, Figure 2 gives an example joint probability density of encroachment speed and angle. Without real data, this example density function is considered to be as plausible and complex as any function that has been used or proposed in the encroachment-based studies by TRB [1987], Mak and Sicking [1992], Daily et al. [1997].

In Figure 2, the encroachment speed density is first represented by a triangular function as

\[
f(v) = \begin{cases} 
\frac{2(v - v_{min})}{(v_{ref} - v_{min})(v_{max} - v_{min})}, & \text{if } 0 \leq v \leq v_{ref} \\
0 & \text{otherwise}
\end{cases}
\]

where \( v_{ref} \) is a reference speed that relates to the posted speed limit and design speed, and \( v_{min} \) and \( v_{max} \) are the minimum and maximum operating speeds, respectively. Second, the maximum possible encroachment angle, \( \phi_{max} \), is postulated to be dependent of the encroachment speed, \( v \), in a linear fashion as follows:

\[
\phi_{max}(v) = \phi_{max}(v_{min}) - \frac{\phi_{max}(v_{max}) - \phi_{max}(v_{min})}{v_{max} - v_{min}}(v - v_{min}), \quad \text{for } v_{min} \leq v \leq v_{max}
\]

where \( \phi_{max}(v_{min}) \) and \( \phi_{max}(v_{max}) \) are, respectively, the maximum possible encroachment angles when encroaching at speeds of \( v_{min} \) and \( v_{max} \). That is, higher encroachment speeds are expected to be associated with lower encroachment angles (for not exceeding the roadholding capacity that a vehicle can realistically provide). Third, the probability density of the encroachment angle \( \phi \), given \( \phi_{max} \), is represented by another triangular function as
\[ f(\phi | \phi_{\text{max}}) = \frac{2}{\phi_{\text{max}} - \phi_{\text{min}}} \left[ 1 - \frac{\phi - \phi_{\text{min}}}{\phi_{\text{max}} - \phi_{\text{min}}} \right], \quad \text{if} \quad \phi_{\text{min}} \leq \phi \leq \phi_{\text{max}} \]  
\[ = 0 \quad \text{otherwise} \]

An obvious deficiency of this density function is that it ignores the effect of horizontal curvatures on \( \phi \). Note that, mathematically, it is straightforward to combine all three equations into a joint density function by using the relationship

\[ f(v, \phi) = f(\phi | v)f(v) \]

\[ = f(\phi | \phi_{\text{max}})f(v) \]

However, this mathematical expression is quite complicated and will not be presented here.

To have some understanding of the statistical properties of the exemplified joint density function, a Monte-Carlo simulation was performed with the following parameter values: \( v_{\text{min}} = 0 \), \( v_{\text{ref}} = 55 \) mi/hr, \( v_{\text{max}} = 70 \) mi/hr, \( \phi_{\text{max}}(v_{\text{min}}) = 40^\circ \), \( \phi_{\text{max}}(v_{\text{max}}) = 15^\circ \), and \( \phi_{\text{min}} = 0.25^\circ \). The reason why \( \phi_{\text{min}} = 0.25^\circ \) is used, instead of 0°, will be given later. The simulation results show that the expected value of encroachment speed, expressed as \( E[v] \), is 41.7 mi/hr, and the unconditional expectations \( E[\phi] \) and \( E[\phi_{\text{max}}] \) are, respectively, equal to 8.5° and 25.1°. Note that the same set of parameter values for the joint density function, \( f(v, \phi) \), was used for all of the Monte-Carlo simulations performed in this study.

**Process 2**

Process 2 determines \( P(\text{In } H_q | \xi, v, \phi, X^{(2)}) \) for each of the \( Q \) objects. Figure 3 shows a commonly used formula to compute the size of the hazard envelope, given \( \phi \), vehicle swath width \( (W_{\text{veh}}) \), and length and width of the considered object \( (l_{\text{obj},q} \) and \( W_{\text{obj},q}) \). This is expressed as

\[ l_{\text{env},q} = l_{\text{obj},q} + W_{\text{obj},q} \cot(\phi) + W_{\text{veh}} \csc(\phi) \]

This formula has been widely used and discussed [TRB, 1987; Daily et al., 1997]. It is based on the assumption that the road section is a straight section with no horizontal curvature and that the trajectory of the encroaching vehicle is a straight line.

Hazard envelopes can, of course, overlap when multiple objects exist and are located closely to one another. This is especially true when \( \phi \) is small. By considering the potential for overlapping of \( H_q \), the conditional probability can be expressed as

\[ P(\text{In } H_q | \xi, v, \phi, X^{(2)}) = P(\text{In } H_q | \xi, \phi, X^{(2)}) = \frac{l_{\text{env},q} - OL_q(\phi)}{2L} \]

\[ = \frac{l_{\text{obj},q} + W_{\text{obj},q} \cot(\phi) + W_{\text{veh}} \csc(\phi) - OL_q(\phi)}{2L} \]

where \( L \) is the length of the road section, and \( OL_q(\phi) \) represents an adjustment of the size of \( H_q \) that is overlapped by other objects and is a function of \( \phi \). The use of \( 2L \) in the denominator
requires the assumption that encroachments are equally likely to occur on the right and left sides of the road, and that the encroachment is equally likely to occur at any location within the road section. These assumptions are good only for a straight and level road section that is homogeneous in traffic and design variables such as AADT, lane width, within the road section.

Estimating the size of overlapping envelopes can be a tedious bookkeeping process and is usually avoided in the modeling stage by either selecting objects that are typically far away from one another or by treating closely located objects as one continuous object [TRB, 1987; Daily et al., 1997; Rodgman et al., 1989]. Another potential problem with estimating the $H_q$, which is usually ignored, is that the envelope of those objects that are located close to the end points of the road section can have significant part of the envelope falling on adjacent sections. Of course, adjacent sections may have envelopes falling on the section under consideration. If the objects are evenly located across the section, this “boundary problem” will roughly cancel out. Otherwise, more tedious bookkeeping procedures may be required.

In practice, when a type of point object is considered, $l_{obj,q}$ and $W_{obj,q}$ are usually set to constants for all $Q$ objects. In addition, when a round-shaped object is considered, it is typically treated as a squared object and $l_{obj,q}$ and $W_{obj,q}$ are set approximately equal to the diameter of the object. For example, for utility poles, $l_{obj,q} = W_{obj,q} = 8$ inches have been used for all poles [TRB, 1987]. For a particular type of continuous object considered, such as guardrails, the length $l_{obj,q}$ varies over $q$, but $W_{obj,q}$ remains fairly constant for all objects. Also, $W_{veh}$ is fixed and determined with a design vehicle (e.g., a mid-sized car) in mind. For example, $W_{veh}$ was set to 6 ft in SR214 and to 12 ft in Roadside Design Guide [American Association of State Highway and Transportation Officials (AASHTO), 1989]. Note that, for the point object, the size of the hazard envelope, as calculated in Eq. (7), is dominated by the term $W_{veh} \csc(\phi)$ because $W_{veh}$ is considerably larger than $l_{obj,q}$ and $W_{obj,q}$. On the other hand, for the continuous object, $l_{obj,q}$ is usually the dominant term in Eq. (7).

To gain some understanding of the size of $H_q$, two Monte-Carlo simulations were conducted with the same joint density $f(v, \phi)$ as in the previous simulation and with $W_{veh}$ set to 9 ft. The first simulation is intended to represent the size of $H_q$ for a long guardrail with $l_{obj,q} = 1,320$ ft (or 0.25 mi) and $W_{obj,q} = 1$ ft. The simulation results show that the minimum, average, and maximum sizes of the envelopes are about 1,336 ft, 1,493 ft, and 3,610 ft, respectively. (Note that the average size of the envelope is a simulation-based estimate of the integral $\int_0^{\pi} \int_{-\phi}^{\phi} \left[ l_{obj,q} + W_{obj,q} \cot(\phi) + W_{veh} \csc(\phi) \right] f(v, \phi) \, dv \, d\phi$.) In this simulation, on average, $W_{veh} \csc(\phi) + W_{obj} \cot(\phi) = 173$ ft, in which $W_{veh} \csc(\phi) = 156$ ft and $W_{obj} \cot(\phi) = 17$ ft. This suggests that, for a short continuous object, e.g., $l_{obj,q} < 500$ ft, the proportion of the envelope that is attributed to vehicle swath width and object width may not be ignored. This is mainly because $\phi$ is typically small (for straight road sections), which makes $\csc(\phi)$ and $\cot(\phi)$ large.

The second simulation run was performed to represent the envelope of a utility pole which was assumed to have an 8 inch diameter. The corresponding minimum, average, and maximum sizes of the envelope are 16 ft, 168 ft, and 2,214 ft, respectively. In this simulation, on average, $W_{veh} \csc(\phi) = 156$ ft and $W_{obj} \cot(\phi) = 11$ ft. This suggests that, even with a relatively small point object like a utility pole, the size of the envelope can be very wide due, again, to vehicle swath width and object width, as a result of small encroachment angles. These two simulations indicate that...
\( \frac{\ell_{obj}}{2L} \) will not be a good approximation of the conditional probability in Eq. (8) in practice when point objects are considered and may not be good for continuous objects when the length of the object is short.

Note that the average sizes of \( W_{veh} \csc(\phi) \) and \( W_{obj} \cot(\phi) \) obtained from this utility pole simulation is considerably larger than that obtained by Hutchinson and Kennedy [1966] and that used in Roadside Design Guide [AASHTO, 1989]. For example, in Hutchinson and Kennedy, a near-side encroachment angle of 6.1° and a far-side encroachment angle of 11.5° were estimated from a limited encroachment data. These angles were used to compute the average size of the envelope using Eq. (7). Assuming that two-thirds of the encroachments are near-side encroachments and one-third are far-side encroachments, this gives an average encroachment angle of about 8°. Now, using the same vehicle swath width of 9 ft as in the simulation, we have
\[
W_{veh} \csc(\phi) + W_{obj} \cot(\phi) = 64.7 \text{ ft} + 4.7 \text{ ft} = 69.4 \text{ ft}.
\]
In Roadside Design Guide, an encroachment angle of 15.2° is recommended for both near-side and far-side encroachments. The basis of this recommendation is, however, not clear. Nevertheless, under this recommendation and using the same vehicle swath width, \( W_{veh} \csc(\phi) + W_{obj} \cot(\phi) = 34.3 \text{ ft} + 2.5 \text{ ft} = 36.8 \text{ ft} \). Thus, assuming that the vehicle swath width can be agreed on, the size of \( W_{veh} \csc(\phi) + W_{obj} \cot(\phi) \) obtained from the utility pole simulation in this study is, respectively, about 2.4 and 4.5 times that used by Hutchinson and Kennedy [1966] and Roadside Design Guide [1989]. Unfortunately, there is no good supporting data to judge the validity of any of these estimates.

One may notice that, in the two examples above, a single encroachment angle is used to estimate the expected size of the hazard envelope with Eq. (7), instead of using the entire distribution of the encroachment angle. This has been practiced by many studies, e.g., SR214 [TRB, 1987] and Daily et al. [1997]. This poses an important question as to what a good choice of \( \phi \) value would be if one is to choose one value to estimate the average or expected size of \( H_q \) using Eq. (7). One observation made from the simulations above was that, when \( \ell_{obj,q} \) is short, plugging in the average encroachment angle in Eq. (7) is not a good estimate of the average size of envelope. For example, in the guardrail example, if one uses the average angle 8.5° in Eq. (7), the average size of the envelope would be estimated as 1,387 ft, which is about 7% lower than the actual average envelope (1,493 ft). However, in the utility pole example, the estimate would be 66 ft, which is about 60% lower than the actual average envelope of 168 ft. The reason for the underestimation is that the distribution of \( \phi \) is skewed to the right (with a positive coefficient of skewness) and the size of the envelope is a nonlinear function of \( \phi \). If the average encroachment angle is not a good choice, what \( \phi \) value should be used to estimate the expected size of \( H_q \) using Eq. (7)? This particular question did not seem to be addressed by earlier studies when estimating the size of the envelope for point objects such as sign posts and utility poles [TRB, 1987; Daily et al., 1997]. The validity of these practices is, therefore, questionable.

Note that the expected size of \( H_q \), which is represented analytically as
\[
\int_{\phi_{min}}^{\phi_{max}} \int_{v_{max}}^{v_{min}} \left[ \ell_{obj,q} + W_{obj,q} \cot(\phi) + W_{veh} \csc(\phi) \right] f(v, \phi) dv d\phi,
\]
does not exist if \( \phi_{min} = 0 \). This is the reason \( \phi_{min} = 0.25° \) was used instead of 0° in previous simulations. The choice of \( \phi_{min} = 0.25° \) is arbitrary; it is simply chosen to represent a very small encroachment angle.
Process 3

Process 3 determines $P(C_q | In H_q, v, \phi, X^{(3)})$, the probability that, if the vehicle encroaches and is in $H_q$, the vehicle will hit the object. As indicated in Figure 1, many factors are involved in this process, including roadside, vehicle, and driver conditions. The main roadside conditions that have been considered in the encroachment-based studies are lateral offset of the object, $D_{obj}$, and surface type, slope, and wetness. These determinants are also recognized as the key road variables by the accident-based studies [Zegeer et al., 1987; Zegeer et al., 1990; Miaou, 1996]. It is obvious that $D_{obj}$ can vary from object to object, which will be denoted by $D_{obj}$. The main vehicle conditions identified by the encroachment-based studies include vehicle braking system and tire condition, while the main driver variables identified include the driver's response delay (to the occurrence of roadside encroachment) and the driver's braking and steering behavior after the vehicle encroaches. As mentioned earlier, this is the process where research results in vehicle dynamics and driver behavior (including maneuvering characteristics and in-vehicle habits) can potentially be incorporated. For example, the Highway Vehicle Object Simulation Model is a sophisticated vehicle-handling simulation model that has been used in some encroachment-based studies [Mak and Sicking, 1992; Mak and Bligh, 1996]. Note that, depending on the focus of the study, the choice of the level of analytical complexity for modeling this process is oftentimes at the discretion of the analysts.

To illustrate the concept, simple kinematic equations will be presented here. These equations have been used to estimate time to collision and impact speed for a given encroachment speed and angle with the assumptions that encroachment trajectory is a straight line and deceleration rate is a constant [Mak and Sicking, 1992]. Table 1 shows these equations under different encroachment conditions. In these equations, the lateral offset, $D_{obj}$, is the key roadside variable that is explicitly modeled; while surface type, slope, and wetness are implicitly modeled through the choice of vehicle deceleration rate ($\Delta$). Vehicle braking system and tire condition are also implicitly represented via the choice of $\Delta$. Driver response delay, $t_r$, is explicitly considered, while braking behavior is remotely implied in the choice of $\Delta$. Steering behavior, on the other hand, is ignored completely. Despite their simplicity, these equations do capture a simple crash-avoidance maneuver very well at the conceptual level and can serve as a basis for considering more complicated maneuvers.

In theory, one can choose $t_r$ from a distribution function which represents the probability of a driver's response delay in typical roadside encroachment circumstances due to the driver's temporary inattentiveness, as well as in extreme cases where a driver falls asleep completely. The data to calibrate this distribution function is, however, not available and has not been considered in the data collection plans mentioned earlier.

It is worth pointing out that ignoring a driver's ability to make steering correction is a critical limitation, since most of the encroachments are expected to have small encroachment angles and, under such encroachments, the alerted driver can usually apply a combination of braking and steering operations to swerve the vehicle back into the travel lane before hitting any object.

Clearly, the effects of many factors are being lumped together and implied in a simple variable, $\Delta$. Without additional modeling effort, it will be difficult, and possibly meaningless, to suggest a probability distribution function for $\Delta$. Intuitively, $\Delta$ is highly dependent on the roadside surface condition, especially sideslope ratio, which is of great interest to this study. By limiting the choice of $\Delta$ to a constant, one can at best assume some sort of ideal braking and roadside surface conditions.
Since a straight-line trajectory is assumed above, the impact angle is the same as the encroachment angle. More complicated trajectories and vehicle yaw rates can, of course, be considered in the process so that a collision is characterized not only by its impact speed, but also by its angle and position of impact. Again, the data to support these more complicated scenarios are either extremely limited or not available.

Given that an encroached vehicle is in the envelope of an object, additional simulations were conducted to illustrate the conditional probability using the kinematic equations in Table 1. In these simulations, new parameters were set: \( t_r = 1 \) sec and \( \Delta = 0.5 \) g = 16.1 ft/sec\(^2\). These parameters represent an alerted driver and a car encroaching on a relatively flat surface with a good tire-surface friction coefficient. Lateral offsets, \( D_{\text{obj}} \), are set for 0, 5, 10, 15, 20, ..., 50 ft at different simulation runs. Given \( v, \phi, t_r \), and \( \Delta \), and the encroachment location within \( H_q \), the impact speed associated with the \( q \)th object, \( V_{c,q} \), can be computed deterministically using the equations in Table 1. Figure 4 shows the distribution of impact speeds as a result of different encroaching speeds, angles, and locations within \( H_q \). Analytically, Figure 4 shows the collision probability

\[
P(C_q \mid H_q, X^{(3)}) = \int \int P(C_q \mid H_q, v, \phi, X^{(3)}) f(v, \phi) \, dv \, d\phi.
\]

Figure 4(a) shows that the percentage of vehicles that hit the object decreases as the lateral offset of the object increases. For those that impact the object, Figure 4(b) further shows the distributions of their impact speed for different lateral offsets of the object. It was noted that the rate of decrease in collision probability in Figure 4(a) is almost a constant, but it is not as pronounced as those used in the current Roadside Design Guide and that estimated in Miaou [AASHTO, 1989; Miaou, 1996; Miaou, forthcoming]. Note that Miaou assumed an exponential rate of decrease for a roadside surface that is relatively flat and showed that the results are quite consistent with those used in the Roadside Design Guide [AASHTO, 1989; Miaou, 1996]. Of course, there are many possible explanations for their discrepancies, such as the simplification of behavior models in the simple kinematic equations.

Combining Processes 2 and 3

Both Processes 2 and 3 are conditional on \( v \) and \( \phi \). To understand the analytical property of these two processes jointly, Monte-Carlo simulations were conducted for different sizes of point objects and for various lengths of continuous objects. Mathematically, the simulation seeks to understand the following process:

\[
P(C_q \mid \xi, X^{(2)}, X^{(3)}) = \int \int P(C_q \mid \xi, v, \phi, X^{(2)}, X^{(3)}) f(v, \phi) \, dv \, d\phi
\]

\[
= \int \int P(In H_q \mid \xi, \phi, X^{(2)}) P(C_q \mid In H_q, v, \phi, X^{(3)}) f(v, \phi) \, dv \, d\phi
\]

Eq. (8) and the simple kinematic equations in Table 1 are used in the simulation. Again, the simulations were conducted for different lateral offsets of the object. The length of the road section is fixed and set to \( L = 1 \) mi (5,280 ft).

For the simulations of point objects, two round-shaped objects of the same lateral offset, one on each side of the road, are considered. The simulations were conducted for various sizes of the object, ranging from 4 inches to 12 inches in diameter. Figure 5 shows the collision probabilities by lateral offsets and by size of the object. Figure 5(a) shows that the probability decreases monotonously as the lateral offset increases. Additional computations revealed that the percentage decrease in collision probability is not a constant; it decreases drastically initially and becomes a constant as the lateral offset increases. This suggests that, for a particular type of point object considered (of the same size), an exponential function of the lateral offset is a good candidate
function to represent the collision probability. However, within the exponential function, the coefficient associated with the lateral offset may have to change as the lateral offset increases. Thus, within the exponential function, either a second or higher order polynomial function or a step function of the lateral offset (which will be described in the next section) may have to be considered. Figure 5(b) indicates that, for a fixed lateral offset, the probability of a collision increases in a linear fashion as the size of the object increases. More specifically, the probability of a collision increases in a linear fashion as the average size of the hazard envelope increases (not shown in the figure).

For the simulations of continuous objects, two objects of the same lateral offset, one on each side of the road, are again considered. The width of the objects are set to a constant ($W_{obj,q} = 1$ ft) and length $l_{obj,q}$ varies from 100 ft to 2,500 ft. Figure 6 shows the same probabilities as those presented for the point objects. Figure 6(a) shows that the probability of a collision decreases as the lateral offset increases. As in the point object simulations, an exponential function of the lateral offset is a good candidate function to represent the collision probability. An interesting observation is that, for $l_{obj,q} > 750$ ft, the rate of decrease in collision probability appears to be fairly constant as the lateral offset increases. This suggests that, within the exponential function, a simple linear function of the lateral offset with one coefficient will be sufficient for representing the collision probability when $l_{obj,q} > 750$ ft. As the length of the object becomes shorter (relative to 750 ft), the percentage decrease in collision probability deviates further from a constant. And, as the length becomes very small, the percentage decrease in collision probability decreases drastically initially and becomes a constant as the lateral offset increases. As in the point object, a second or higher order polynomial function or a step function of the lateral offset becomes necessary within the exponential function. Again, Figure 6(b) indicates that, for a fixed lateral offset, the probability of a collision increases in a linear fashion as the size of the object (or, more precisely, the size of hazard envelope) increases.

**Output of Process 3**

One of the key outputs from Process 3 is an impact speed. If the impact speed is greater than zero (i.e., an object is hit), then it serves as input to Process 4.

**Process 4**

Process 4 determines $P(A_q \mid C_q, \phi, v, X^{(4)})$. As mentioned earlier, in a more sophisticated vehicle trajectory model, the collision is characterized by impact speed, impact angle, and the position of impact. However, in previous studies, very simple conditional probabilities have been assumed based on limited data and some engineering judgment. For example, a constant probability of 0.9 was used for hitting utility poles in SR214. A more sensible way would be to determine the probability based on the impact speed (or the lateral offsets of objects since the impact speed is a function of lateral offsets). For simplicity, we consider the impact speed only in this study. To make the impact speed explicit in the conditional probability, we can write $P(A_q \mid C_q, V_{cs}, X^{(4)})$, instead of $P(A_q \mid C_q, \phi, v, X^{(4)})$. This probability has to be determined for each type of object. For example, it may be decided that if the impact speed with a utility pole is greater than 5 mph, then it will result in a reportable accident. That is,
\[ P(A_q \mid C_q, V_{c,q}, X^{(4)}) = 1 \quad \text{if } V_{c,q} > 5 \text{ mi/hr} \]
\[ = 0 \quad \text{otherwise} \tag{10} \]

The choice of the threshold impact speed is, of course, dependent on the nature of the object considered (e.g., the material of which the object is made).

Similar to the setup of the last simulations, a simulation was conducted to gain some understanding of the probability when different thresholds of \( V_{c,q} \) are chosen. Figure 7 shows the probabilities for an 8-inch diameter point object and a 0.25-mile continuous object. Naturally, the probability decreases as a higher threshold value of \( V_{c,q} \) is chosen. However, the simulation results suggest that the decrease in probability is very small when \( V_{c,q} \) increases from 0 to 10 mi/hr. Another interesting observation is that the decreasing rates of the probability as the lateral offset increases remain very much the same under different threshold values of \( V_{c,q} \).

The severity of accidents is expected to be some function of impact speed. Therefore, this is the process where the conditional probability by accident severity could potentially be considered. Conceivably, crash test results, which can be computer simulated, can potentially be used to determine the relationship between severity and impact speed. Also, this is the process where the changes of safety features in vehicle population can be reflected, such as the percentages of vehicles equipped with airbags and dynamic side-impact protection and the rate of seat-belt use.

**Overall Probability and Expected Number of Accidents**

Provided that the overlapping of \( H_q \) can be properly accounted for (or avoided), the probability of an encroached vehicle to be involved in an accident with one of the objects on the road, symbolized by \( A_q \), is the sum of the probability over all objects:

\[ P(A_q \mid X^{(2)}, X^{(3)}, X^{(4)}) = \sum_{q=1}^{Q} \left[ \int \int P(A_q \mid \xi, v, \phi, X^{(2)}, X^{(3)}, X^{(4)}) f(v, \phi) dv d\phi \right] \]
\[ = \sum_{q=1}^{Q} \left[ \int \int P(\text{In } H_q \mid \xi, v, \phi, X^{(2)}) P(C_q \mid \text{In } H_q, v, \phi, X^{(3)}) P(A_q \mid C_q, v, \phi, X^{(4)}) f(v, \phi) dv d\phi \right] \tag{11} \]

As stated earlier, in the encroachment-based studies, it is expected that the (expected) encroachment rate can be pre-determined based on mainline conditions. The encroachment rate, denoted by \( R_\xi \), is typically given as the number of encroachments per million vehicle-miles traveled. Given \( R_\xi \) and all determinants, the expected number of accidents on the road section in \( N \) years is calculated as

\[ E[Y \mid X^{(1)}, Z^{(1)}, X^{(2)}, X^{(3)}, X^{(4)}] = \left( 365 \times N \times AADT \times L / 10^6 \right) \times R_\xi \times P(A \mid \xi, X^{(2)}, X^{(3)}, X^{(4)}) \tag{12} \]

where \( Y \) is the number of reportable accidents involving the type of objects, \( L \) is in miles, \( AADT \) is in number of vehicles, and \( \left( 365 \times N \times AADT \times L / 10^6 \right) \times R_\xi \) is the expected total number of encroachments during the period. One of the basic encroachment parameters of interest is the encroachment rate per mile per year. If \( R_\xi \) is estimated and symbolized as \( \hat{R}_\xi \), then the
encroachment rate can be obtained as \( \left( 365 \times \text{AADT} / 10^6 \right) \times \hat{R}_i \). An implicit assumption used here is that the RORA experiences of individual drivers traveling through the section are independent from one another (which is a reasonable one).

### 3. FORM OF MEAN FUNCTIONS

Unlike the encroachment-based approach, the accident-based approach deals with one conditional probability only: the probability of having \( y \) accidents, given the available determinants, or mathematically, \( P(Y = y | X^{(1)}, Z^{(1)}, X^{(2)}, X^{(3)}, X^{(4)}) \), where \( y = 1, 2, 3, \ldots, \infty \). There are basically five major tasks in developing accident-based models: (1) find a good probability (mass) function to describe the random variation of accident frequency; (2) determine an appropriate functional form and parameterization for the mean function which describes the effect of key variables on accident frequency; (3) select the variables that have statistically significant effects on accident frequency for inclusion in the mean function; (4) estimate the regression parameters in the mean function and obtain good statistical inferences for the estimated parameters based on available data; and (5) assess the quality of the model; judge whether the developed model makes good engineering sense; decide whether the developed model meets the planning and design requirements; and identify cost-effective ways to improve the model. The work involved in each of these tasks has been described in Miaou [1996]. Since the theory behind the Poisson and NB regression accident-based models have been discussed quite extensively in many recent publications [Maycock and Hall, 1984; Miaou et al., 1993; Miaou and Lum, 1993; Miaou, 1994; Maher and Summersgill, 1996; Miaou, 19961, the readers are referred to these publications for a review of these models.

As seen in the last section, the focus of the encroachment-based model has been on task 2, determining the appropriate functional form and parameterization for the mean function and on identifying key determinants. In this section, several mean functions will be formulated for use in the accident-based models, following the encroachment-based thinking described in the last section. Specifically, the objective is to determine mean functions based on Eqs. (11) and (12) and simulation experience gained in the last section.

As stated earlier, the encroachment-based thinking described above does not provide any engineering insights on what the plausible functional forms of \( P(X^{(1)}, Z^{(1)}) \) might be. Thus, it does not provide any clue as to the functional form of \( R_i \) in relation to \( X^{(1)} \) and \( Z^{(1)} \). It has been argued in the accident-based literature that a plausible functional form should be multiplicative in nature and represent the interactive effects of various mainline design and traffic variables on accident frequency. The exponential form has been suggested to be a good candidate because it is simple and it ensures that the expected value will always be non-negative [Miaou, 1996]. By applying the same argument to the encroachment frequency, we can express the encroachment rate as

\[
R_i = \exp(\beta_i^* + \sum_{j=2}^{m} \beta_j x_{ij} + \sum_k \gamma_k z_{ik})
\]

(13)

where, for clarity, the subscript \( i \) is introduced to represent the road section being considered is the \( i \)th section in the sample road sections; \( x_{ij} \) is the values of the available mainline design and traffic variables for the \( i \)th road section; \( z_{ik} \) is the value of the unobservable variables for the section; and \( \beta_i^* \), \( \beta_j \)'s, and \( \gamma \)'s are unknown model parameters.
The encroachment-based thinking described in the last section does provide some good ideas on the choice of the functional form in Eq. (11). One such choice, which is consistent with the simulation results shown in Figures 5-7, would be

\[ P(Al|\ell_{obj},W_{obj},D_{obj}) = \sum_{q=1}^{Q} \left[ \frac{\ell_{obj,q} + \delta_q}{2L} \right] \exp(\alpha_1 D_{obj,q} + \alpha_2 D_{obj,q}^{(1)} + \alpha_3 D_{obj,q}^{(2)} + \ldots) \eta_q \]  

(14)

where \( \delta_q \) is the part of the hazard envelope associated with vehicle swath width, object width, and adjustments for the overlapping of envelopes; a polynomial or a step function of \( D_{obj,q} \) is used within the exponential function as suggested from the previous simulations; \( \alpha \)'s are unknown parameters associated with the polynomial function; and \( \eta_q \) is associated with the threshold impact speed discussed earlier and is a constant between 0 and 1. When \( \eta_q = 1 \), it indicates that any impact speed greater than 0 will result in an accident. According to the simulation results shown in Figure 7, \( \eta_q \) should be fairly close to 1 if the threshold impact speed is less than 10 mi/hr. Thus, for typical non-breakaway, fixed objects, if we are interested in all reportable accidents (regardless of their severity) a reasonably good estimate of \( \eta_q \) would be close to 1 and can be determined outside of the accident model.

Note that, to use a polynomial function of \( D_{obj,q} \) in Eq. (14), \( D_{obj,q}^{(1)} = D_{obj,q}^2, D_{obj,q}^{(2)} = D_{obj,q}^3, \) and \( D_{obj,q}^{(3)} = D_{obj,q}^4, \) etc., and for a step function, \( D_{obj,q}^{(1)} = D_{obj,q}, \) if \( D_{obj,q} \) is greater than a predetermined distance, e.g., 5 ft, otherwise a zero is assigned; \( D_{obj,q}^{(2)} = D_{obj,q}, \) if \( D_{obj,q} \) is greater than a larger predetermined distance, e.g., 10 ft, otherwise a zero is assigned; and \( D_{obj,q}^{(3)} = D_{obj,q}, \) if \( D_{obj,q} \) is greater than another larger predetermined distance, e.g., 15 ft, otherwise a zero is assigned, etc. The step function so arranged would allow the rate of decrease in collision probability to change step-wise as the lateral offset increases. For the example step function above, the percent decrease in collision probability is approximately \(-100\alpha_1\%\) per ft increase of \( D_{obj,q} \) for \( D_{obj,q} \) between 0 and 5 ft, \(-100(\alpha_1 + \alpha_2)\%\) for \( D_{obj,q} \) between 5 and 10 ft, and \(-100(\alpha_1 + \alpha_2 + \alpha_3)\%\) for \( D_{obj,q} \) between 10 and 15 ft, etc.

In Eq. (14), \( D_{obj,q} \) is the only roadside design variable considered in determining collision probability after roadside encroachments occur. As indicated earlier, sideslope and surface types are two other key variables which were implicitly considered in the simulations presented earlier. Thus, a plausible extension of Eq. (14) is to add sideslope as the second roadside design variable within the exponential function on the right-hand side of Eq. (14). However, a representative sideslope may be difficult to obtain for two reasons: (1) the sideslope can vary significantly within a road section, and (2) for a particular object, the sideslopes of interest are the slopes that are located within the hazard envelope of the object, the location and size of which also need to be estimated. To reflect the difference in friction coefficient of different surface types, such as paved and unpaved shoulders, in the model, the lateral offset, \( D_{obj,q} \), can be provided by surface type and the model parameter associated with the lateral offset can vary from one surface type to another.

Substituting Eqs. (14) and (15) into Eq. (12), we have a general mean function for the accident-based models for both point and continuous objects.
\[ E[Y|X^{(1)}, X^{(2)}, X^{(3)}, X^{(4)}] = \]
\[ \left(365 \times N \times AADT \times L / 10^4\right) \times \exp(\beta_1 + \sum_{j=2} \beta_j x_{ij}) \]
\[ \times \left\{ \sum_{q=1}^Q \left[ \frac{\ell_{obj,q} + \delta_q}{2L} \right] \exp(\alpha_1 D_{obj,q}^{(1)} + \alpha_2 D_{obj,q}^{(2)} + \alpha_3 D_{obj,q}^{(3)} + \ldots) \right\} \eta_i \}
\]

where \( \alpha \)'s, \( \beta \)'s and \( \gamma \)'s are unknown model parameters to be estimated from the data. It should be noted that the term \( \exp(\sum_k \gamma_k z_{ik}) \) has been taken out of the expectation since the variable \( Z^{(1)} \) is not available in practice. Using the Poisson assumption for the randomness of accident frequency, together with the assumption that the exponential function of the unobservable variables (i.e., \( \exp(\sum_k \gamma_k z_{ik}) \)) is gamma distributed with an expected value of one, a negative binomial regression model can be derived with Eq. (16) as its mean function [Miaou, 1996]. These two assumptions are quite plausible and flexible for count data analysis. They have been used in accident-based models and widely accepted in other fields such as biostatistics and econometrics.

In Eq. (16), \( \delta_q \) is a variable which varies over individual object. If overlapping of hazard envelopes, i.e., the size of \( OL_q(\phi) \), is judged to be small, then \( \delta_q \) can be treated as a constant, say \( \delta \), and it may be estimated within or outside of the accident-based model. More discussion on the estimation of \( \delta_q \) will be provided in the following paragraphs. Also, as indicated earlier, the simulation results presented in Figure 7 suggested that, when non-breakaway, fixed objects are considered, \( \eta_i = 1 \) is a good estimate and can be determined outside of the accident-based model if all reportable accidents are of interest. Note that, after the model parameters in Eq. (16) are estimated, the expected encroachment rate per mile per year can be obtained for a given set of covariate values, \( x_{ij} \), as:

\[ \left(365 \times AADT / 10^4\right) \exp(\hat{\beta}_1 + \sum_{j=2} \hat{\beta}_j x_{ij}) \]

For point objects where the overlapping of hazard envelopes is small, \( \delta_q \) is approximately a constant, represented by \( \delta \). Also, \( \ell_{obj,q} \) is considerably smaller than \( \delta \) and can be ignored. Under these conditions, Eq. (16) becomes

\[ E[Y|X^{(1)}, X^{(2)}, X^{(3)}, X^{(4)}] = \]
\[ \left(365 \times N \times AADT \times L / 10^4\right) \times \exp(\beta_1 + \sum_{j=2} \beta_j x_{ij}) \]
\[ \times \left\{ \frac{Q \delta}{2L} \sum_{q=1}^Q \omega_q \left[ \exp(\alpha_1 D_{obj,q}^{(1)} + \alpha_2 D_{obj,q}^{(2)} + \alpha_3 D_{obj,q}^{(3)} + \ldots) \right] \eta_i \right\} \]

where \( Q \delta \) is the total length of the hazard envelopes associated with the \( Q \) objects along the road section, \( \omega_q = 1/Q \) represents the fraction of the total hazard envelopes that is associated with the \( q \)th object and \( \sum_{q=1}^Q \omega_q = 1 \). The two parameters \( \delta \) and \( \beta_1 \) in the equation cannot be uniquely determined from the estimation procedure of the accident-based model. In order to estimate \( \beta_1 \), which is required if the encroachment rate is to be estimated from the model, \( \delta \) has to be determined outside of the accident-based model. However, recall that, even if the vehicle swath width is set to be a constant of 9 ft, \( \delta \) can still vary significantly over different choices of \( \phi = 168 \) ft according to the previous utility pole simulation, about 69.4 ft when using Hutchinson and Kennedy's [1966] encroachment angle, and about 36.8 ft when using the encroachment angle
recommended by *Roadside Design Guide* [1989]. (Note that \( \eta \) is again assumed to be determined outside of the accident-based model.) One important observation can be made from Eq. (17) is that a good estimate of \( \beta_1 \) not only requires a good estimate of \( \delta \), but also requires the functional form of the effect of \( D_{obj,q} \) on collision probability be appropriately specified from the data, especially for the range where \( D_{obj,q} \) is close to zero.

For relatively long continuous objects, \( \hat{\ell}_{obj,q} \) (say, greater than 750 ft) will be considerably larger than \( \delta_q \). Thus, Eq. (16) can be approximated by

\[
E[Y|X^{(1)}, X^{(2)}, X^{(3)}, X^{(4)}] = \left(365 \times N \times AADT \times L / 10^6 \right) \times \exp(\beta_1 + \sum_{j=2}^{q} \beta_j x_{ij}) \times \left\{ \frac{L_{obj}}{2L} \sum_{q=1}^{Q} \omega_q \left[ \exp(\alpha_1 D_{obj,q}) \right] \hat{\eta}_r \right\}
\]

where \( L_{obj} = \sum_{q=1}^{Q} \hat{\ell}_{obj,q} \) is the total length of the object on both sides of the road section and \( \omega_q = \hat{\ell}_{obj,q} / L_{obj} \) represents the fraction of the total hazard envelopes that is associated with the \( q \)th object and \( \sum_{q=1}^{Q} \omega_q = 1 \). Note that in Eq. (18), based on the simulation results presented earlier, a constant rate of decrease for the collision probability as \( D_{obj,q} \) increases is adopted and the parameter \( \alpha_1 < 0 \). Unlike the point object, \( \beta_1 \) can be determined with good accuracy without requiring a good estimation of \( \delta \) to be obtained (if \( \eta_1 \) is determined externally as stated earlier).

The discussion above suggested that there are several advantages of using long continuous objects to develop accident-based models for estimating encroachment rates over the use of point or short continuous objects:

1. For long continuous objects, good estimates of encroachment rates do not require good estimation of \( \delta \). For short continuous and point objects, the accuracy of the estimation of encroachment rates is directly dependent on the estimation of \( \delta \), which could be off by a factor of 4.5, depending on which encroachment angle is assumed. Also, a good estimate of the encroachment rate will require the functional form of the effect of \( D_{obj,q} \) on collision probability be appropriately specified from the data, especially for the range where \( D_{obj,q} \) is close to zero.

2. For the same reason stated above, the overlapping of hazard envelopes is less of a problem when long continuous objects is considered (when compared to short continuous and point objects).

3. The simulation in this study suggested that, for long continuous objects (> 750 ft), a simple exponential function of the lateral offset is suffice to represent the collision probability, and more complicated functions of the lateral offset will be required if short continuous or point objects are considered.

4. If the sideslope is to be considered as a determinant of the collision probability, given an encroachment, it will be a lot more difficult to obtain a representative sideslope for point and short continuous objects than for long continuous objects. This is because the sideslopes of interest are the slopes that are located within the hazard envelope of the object, the location and size of which need to be estimated. And, as indicated earlier, better estimates of the size of
the hazard envelopes can be obtained for long continuous objects than for short continuous or point objects.

### 4. ESTIMATING ENCROACHMENT RATES

This section is intended to show how accident-based model can be useful to the encroachment-based models. Specifically, roadside encroachment rates for rural, two-lane, undivided roads will be estimated using the accident-based model. The candidate mean functions suggested in the last section will be employed in developing the accident-based models for accidents involving guardrails and utility poles.

Accidents and roadway data for rural, two-lane, undivided roads from a roadway cross-section design data base [Rodgman et al., 1989] administered by FHWA were used to develop an accident-based model. This data base contains 1,944 road sections, most of which are located in rural areas. Specifically, out of the 1,944 sections, about three-quarter of them can be considered to be truly rural, undivided roads. One of the important features of this particular data base is that it contains a rather detailed description of key design elements of various roadside obstacles. This data base has been used in previous studies to develop accident-based models for rural, two-lane, undivided roads, such as Zegeer et al. [1987], Zegeer et al. [1990], and Miaou [1996]. A good description of the data collection process and general statistics of the road sections included in this data base can be found in Rodgman et al. [1989] and Zegeer et al. [1990]. The road sections contained in the data base represent a stratified random sample from seven States: Alabama, Michigan, Montana, North Carolina, Utah, Washington, and West Virginia. Except for Alabama, which has about 2.5 years worth of data, five years of accident data from 1980 to 1984 were available for analysis. Note, however, that accident data were not broken down by year. For each road section, the data base has the inventory of roadside objects within 30 ft of the edge line of travel lanes.

**Guardrail Accidents**

Only those road sections with guardrails in the data base are of interest. For each road section in the data base, the total number of miles of guardrails (or guardrail-miles) is recorded according to the clear zone width. However, in the data base, clear zone widths are grouped into eight categories: < 1.5 ft, 1.5 to 3.5 ft, 3.5 to 6.5 ft, 6.5 to 10.5 ft, 10.5 to 15.5 ft, 15.5 to 20.5 ft, 20.5 to 25.5 ft, and 25.5 to 30 ft. That is, for each clear zone width category within each road section, we have the total guardrail-miles, but we don’t know how many disjointed guardrails have been included. We also do not know their relative positions within the road section and how many are on the left and right sides of the road. Note that, for those guardrails that fall into different clear zone width categories, we do know for sure that they are disjointed. Also, for the modeling purpose, the mid-point of each clear zone width category is used to represent the clear zone width for all guardrails located in this category.

Recall that in order to use Eq. (18) as the mean function, we need to be able to select road sections with relatively long guardrails. Specifically, we need \( \ell_{obj,q} \) to be considerably larger than \( \delta_i \). Also, recall that \( \delta_q \) is about 173 ft (if hazard envelopes are not overlapped), according to the simulation presented earlier, and is much smaller if the encroachment angle from Roadside Design Guide [AASHTO, 1989] or Hutchinson and Kennedy [1966] are used. For the reasons stated above, in this data base, we will never be certain that we have included only road sections with long
guardrails. However, the probability that most of the road sections that are included have long guardrails should increase if we choose to remove more road sections with short guardrail-miles. On the other hand, we clearly need to have a reasonably large sample size to develop meaningful statistical models.

To compromise, road sections with guardrail miles less than 0.1 mi (or about 528 ft) per side of road were first removed, leaving 272 road sections for analysis. The length of these sections ranges from 1 to 9.37 mi, with an average length of about 3 mi. The total length of these road sections is 841 mi, while the total guardrail-miles is about 109 mi per side of road. During the period considered, there were 450 recorded guardrail accidents on these road sections, regardless of vehicle and accident severity type. With the total vehicle-miles estimated to be 3,486 million and vehicle guardrail-miles traveled estimated to be 471 million, the overall guardrail accident rate was 0.12 accidents per million vehicle-miles traveled and 0.96 accidents per million guardrail-miles traveled. Note that the rate is calculated assuming that guardrail-miles are equally distributed on both sides of the road. Of the 272 road sections, about 43% of them (118 sections) had no recorded guardrail accident. The maximum number of guardrail accidents recorded for an individual road section was 19 during the 5 year period.

Of the 109 guardrail-miles per side of the road, their distribution across the eight clear zone width categories were: 1.37%, 14.31%, 36.88%, 34.74%, 11.96%, 0.60%, 0.06%, and 0.08%, respectively. That is, the majority of the guardrail-miles are located in categories 2 to 5 or between 1.5 to 15.5 ft. For these 272 road sections, in addition to vehicle-miles traveled (in millions), other covariates considered and their associated ranges are as follows:

- **AADT per lane**, used as a surrogate measure for traffic density; AADT is between 160 and 10,000 vehicles per day
- **Lane width**: Between 9 and 13 ft
- **Horizontal curvature**: Non-homogeneous within a section, i.e., each section may contain multiple curves; length-weighted horizontal curvature is between 0 and 21 degrees/100 ft arc
- **Vertical grade**: Non-homogeneous within a section, i.e., each section may contain multiple grades; length-weighted vertical grade is between 0 and 8 percent
- **Clear zone width** (or lateral offset), measured from the outside edges of travelway to the guardrail, which includes:
  - **paved shoulder width**, Between 0 and 12 ft
  - **unpaved shoulder width** (i.e., earth, grass, gravel, or other stabilized shoulder width), Between 0 and 10 ft
  - **additional clear zone width beyond shoulders**, Between 0 and 6.2 ft

About 90% of the road sections have shoulders that are either paved or unpaved, i.e., only about 10% of the road sections have a mixed shoulder type. Also, about 90% of them have a posted speed limit of 55 mi/hr. In addition, about 12% of the sample road sections do not have horizontal curvature data, and about 21% do not have vertical grade data. Furthermore, most of the road sections have 11 ft as the lane width.

Guardrail accident models were developed using these 272 road sections, as well as two subsets of these sections, in which road sections with longer guardrail-miles (> 0.15 mi or 792 ft and > 0.2 mi or 1,056 ft per side of road) were selected. The total numbers of road sections available for developing models for the two subsets were 217 and 188 sections, respectively. An extended NB regression model, as described in Miaou [1997], was employed in this study. The model is a general-purpose model which allows mean functions to have the form as shown in Eqs. (17) and (18).
To be more specific, for each road section \( i \), the conventional NB regression has a multiplicative mean function of the following form:

\[
E[Y_i|v_i, x_{ij}, j = 1, 2, \ldots, J] = v_i \exp\left(\sum_{j=1}^{J} \beta_j x_{ij}\right)
\]

\[
= v_i \exp(\beta_1 x_{i1}) \exp(\beta_2 x_{i2}) \cdots \exp(\beta_J x_{iJ})
\]

\[
= v_i \left[ \prod_{j=1}^{J} \exp(\beta_j x_{ij}) \right]
\]

where \( v_i \) is typically called an offset of the model (which usually represents an exposure measure in the accident-based model), \( x_{ij} \) is the value associated with the \( j \)th covariate of the road section \( i \), and \( \beta_j \) is a regression parameter associated with the \( j \)th covariate. That is, in the conventional NB regression model, the effect of covariate \( j \) on the expected number of accidents is modeled as a simple exponential function of the form: \( \exp(\beta_j x_{ij}) \). In the extended NB model, the effect of the \( j \)th covariate on the mean function is allowed to have the following form:

\[
\sum_{s=1}^{S} w_{is} \exp\left(\sum_{k=1}^{K} \beta_{jk} x_{jsk}\right)
\]

where \( \omega_{is} \) represents a predetermined weight associated with a subsection \( s (=1,2,\ldots,S_p) \) of road section \( i \), \( x_{jsk}, k=1,\ldots, K_j \), are the covariate values that characterize the subsection \( s \), and \( \beta_{jk} \) are unknown model parameters that need to be estimated from the data. Typically, for each road section, we have \( \sum_{s=1}^{S_p} \omega_{is} = 1 \), i.e., the sum of the predetermined weight over all subsections is equal to 1. The maximum likelihood method is used to estimate unknown model parameters, and the observed Fisher Information Matrix is used to obtain statistical inferences for the estimated parameters.

The mean function in Eq. (20) was formulated specifically to deal with road sections (or sites) that have non-homogeneous attributes along the road section (or within the site). For example, in this study, each road section may have multiple horizontal curvatures and vertical grades, and roadside objects within each road section may have different offsets. To illustrate, here we use the lateral offset of roadside objects in Eq. (17) as an example. That is, the \( j \)th covariate is now the lateral offset. Let's say that road section \( i \) has \( S_{ij} \) roadside objects of interest and their lateral offsets are denoted by \( D_{obj,s} \), where \( s = 1,2,\ldots,S_p \). In this example, \( \omega_{is} \) is the fraction of the total hazard envelope that is associated with the \( s \)th object. Also, if we use the step function described under Eq. (14) to model the effect of lateral offsets, then \( x_{js1} = D_{obj,s}^{(1)}, x_{js2} = D_{obj,s}^{(2)}, x_{js3} = D_{obj,s}^{(3)}, \text{ etc.} \)

Furthermore, in Eq. (17), \( v_i = (365 \times N \times AADT \times L / 10^6 \times (Q \delta / 2L)) \). Note that here, without confusion, the subscript \( J \) has been omitted from \( N, AADT, L, Q, \) and the lateral offset.

The following variable selection procedure was adopted in selecting the final model: Initially, all covariates listed earlier were included in the extended NB regression model and their parameters estimated. Then, the variable that had the least absolute t-statistic value which is less than 1.9 (i.e., not significant at about a 5% significance level) was removed from the model. The parameters of the smaller model were re-estimated and their t-statistics reassessed. The procedure continued until the t-statistics of all parameters in the model are greater than about 1.9. Lane width, horizontal curvature, and vertical grade were removed from the model at different stages of the variable
selection process. The estimated parameters, as well as their associated standard deviations and t-statistics, of the final selected models are presented in Table 1. Note that AADT per lane is not statistically significant in the last model presented in the table (for sections with guardrail-miles > 0.2 mi per side of road). This model is presented for comparison purpose only. Also, even though the second and higher order terms (or the step function) of the later offset, $D_{obj,q}$, were not required in Eq. (18), as suggested by the simulation, they were still tested in the variable selection procedure and were not found to be statistically significant.

From the three models presented in columns 2-4, it is clear that clear zones have very significant effects on guardrail accident rates. The second and third models (for road sections with guardrail-miles > 0.15 mi per side of road) is preferred because it has a higher probability of excluding road sections with short guardrails. The second model (in column 3) does indicate that different roadside surface types may have different effects on the collision probability. For this particular data set, the difference in their effects is, however, not statistically different. All three models give similar encroachment rates (assuming $q_v = 1$ in all three models). Assuming $q_v = 1$, the expected numbers of encroachments per mile per year are estimated by AADT using the third model and are shown in Figure 8. The estimates can be seen to be slightly higher than those presented by Miaou [1996], which are also accident-based estimates. The estimates can also be seen to be compatible with those obtained by Hutchinson and Kennedy [1966] and Cooper [1980] which were estimated from field-collected data.

The consistency of these estimates is an indication that estimating basic encroachment parameters using accident-based models, as proposed by Miaou [1996], can be a viable approach to reducing encroachment data collection cost. Most importantly, it is straightforward to use such an approach to estimating basic encroachment parameters for various mainline traffic and design conditions, such as AADT, lane width, horizontal curvature, and vertical grade. The only premise is that a sound accident-based model be developed. Another strength of using accident-based models to estimate basic encroachment parameter is that there is no need to procedures to distinguish between controlled and uncontrolled encroachments, which, as indicated earlier, can be subjective and technically difficult.

**Utility Pole Accidents**

This subsection presents the use of utility pole accidents to develop accident-based models and estimate roadside encroachment rates. The limitations described earlier regarding the use of point objects to estimate encroachment rates are then illustrated. Note that due to a serious limitation of the utility pole data contained in the data base, which will be described later, the utility pole accident-based models presented in this subsection are not considered to be reliable and are for illustration purpose only.

Only those road sections with utility poles in the data base are of interest. As in the guardrail data, the total number of utility poles is inventoried by eight clear zone width categories. For each clear zone width category within each road section, we have the total number of utility poles. However, we do not know their relative positions within the road section and how many are on the left and right sides of the road.

Recall that, in order to use Eq. (17) as the mean function for modeling point object accidents, the overlapping of hazard envelopes has to be small. To reduce the probability of overlapping in hazard envelopes, only road sections with an average pole spacing greater than 0.1 mi (528 ft) were considered. Note that the average pole spacing is calculated as two times the section length divided by the total number of utility poles on both sides of the road. Since we do not know the position of these poles, the choice of 0.1 mi is admittedly arbitrary and is mainly based on the simulation
results presented earlier where the size of the hazard envelope associated with a utility pole was estimated to be about 168 ft on average.

There were 855 road sections that met the selection criteria, including the criterion of having an average pole spacing of at least 0.1 mi. The length of these sections ranges from 0.92 to 9.37 mi, with an average length of 3.02 mi. The total length of these road sections is 2,586 mi, while the total number of utility poles is 20,424 (or 10,212 per side of road). During the period considered, there were 293 recorded utility pole accidents on these road sections, regardless of vehicle and accident severity type. With the total vehicle-miles estimated to be 9,931 million, the overall utility pole accident rate was about 0.03 accidents per million vehicle-miles traveled. Also, there were about 42,211 million vehicle-poles, which gave 0.69 accidents per 100 million vehicle-poles. Note that here a vehicle-pole is defined as a vehicle passing two poles, one on each side of the road, and the rate is calculated assuming that utility poles are equally distributed on both sides of the road. Of the 855 road sections, about 79% of them (673 sections) had no recorded utility pole accident. The maximum number of utility pole accidents recorded for an individual road section was 12 during the 5 year period.

Of the 20,424 utility poles, their distribution across the eight clear zone width categories were: only 9 poles (0.05%) in categories 1 and 2, and 313 (1.53%), 1,574 (7.71%), 3,668 (17.96%), 4,412 (21.60%), 4,044 (19.80%), 6,402 (31.35%), respectively, in categories 3 to 8. Thus, very few utility poles (less than 10% of the total) are located within 10.5 ft of the travel lane, about 40% are between 10.5 and 20.5 ft, and over 51% of the poles are located between 20.5 and 30.5 ft. As indicated earlier, a good estimate of $\beta_1$ in Eq. (17) not only requires a good estimate of $\delta$, but also requires the functional form of the effect of $D_{obj,q}$ on collision probability be appropriately specified from the data, especially for the range where $D_{obj,q}$ is close to zero. With very few utility poles located within 5 ft (or 10 ft) of the travel lane, the model development process has a low probability of success in identifying the appropriate functional form for $D_{obj,q}$ from the data set which, as stated at the outset, is considered a very serious limitation of this data set.

In addition to vehicle-miles traveled (in millions), the following covariates have been considered for individual road sections:

- **AADT per lane**, used as a surrogate measure for traffic density: AADT is between 160 and 10,000 vehicles per day
- **Lane width**: Between 9 and 13 ft
- **Horizontal curvature**: Non-homogeneous within a section, i.e., each section may contain multiple curves; length-weighted horizontal curvature is between 0 and 21 degrees/100 ft arc
- **Vertical grade**: Non-homogeneous within a section, i.e., each section may contain multiple grades; length-weighted vertical grade is between 0 and 11 percent
- **Sideslope** (e.g., 3:1 and 7:1 slopes are recorded as $1/3=0.33$ & $1/7=0.14$, respectively.): Non-homogeneous within a section; length-weighted sideslope is between 0 and 0.91
- **Clear zone width** (or lateral offset), measured from the outside edges of travelway to the guardrail, which includes:
  - paved shoulder width, Between 0 and 11 ft
  - unpaved shoulder width (i.e., earth, grass, gravel, or other stabilized shoulder width), Between 0 and 12 ft
  - additional clear zone width beyond shoulders, Between 0 and 28 ft
About 93% of the road sections have shoulders that are either paved or unpaved, i.e., only about 7% of the road sections have a mixed shoulder type. Also, about 93% of them have a posted speed limit of 55 mi/hr. In addition, about 22% of the sample road sections do not have horizontal curvature data, and about 39% do not have vertical grade data. Furthermore, most of the road sections have 11 ft as the lane width.

Several step functions of the clear zone, as described earlier, were also tested in the variable selection process. The Column 1 of Table 3 contains a complete list of the covariates considered. Note that, since the position of the poles are unknown, the sideslope measure considered here is more of a characterization of the roadside slope for the entire road section than that for the hazard envelopes encompassed by utility poles.

Recall that a good estimate of the encroachment rate will require the functional form of the effect of $D_{obj,q}$ on collision probability be correctly identified, at least for $D_{obj,q}$ close to zero. To illustrate the difficulty of developing the accident-based model the this data set, Table 3 shows 3 estimated extended NB regression models (in Columns 2 to 4). The first model (Column 2) uses a simple exponential function to model the effect of the clear zone width on collision probability, with different effects for different surface types; the second model (Column 3) uses a simple exponential function of the clear zone width without considering surface types; and the third model (Column 4) uses the step function described earlier. Note that many other models were evaluated and are not presented here. Our overall experience is that when more detailed step functions or higher order functions of the clear zone width were considered, the estimation results tended to become unstable, even with a slight change in the function form or variable definition.

Comparing the estimated regression parameters associated with the clear zone width among the three models in Table 3 suggests that a simple exponential function of the clear zone width will not suffice to model the collision probability and a more complicated function form will be required. For example, the first model indicates that the estimated parameters are statistically different for the shoulder ($0.15\pm0.035$ for paved shoulders and $0.18\pm0.027$ for unpaved shoulders and recalled that shoulders are between 0 and 12 ft) and for additional clear zone width beyond the shoulder ($0.05\pm0.015$) at a 5% significance level. Also, the parameters of the step function used in the third model, though not well determined (as seen by their low t-statistics), indicate that the percentage change in collision probability does decrease as the clear zone width increases.

Figure 8 shows the estimated roadside encroachment rates from the first and third models under three different assumptions of the size of $\delta$. A wide variation of the estimates can be observed from the figure (even without considering the sampling variations). Because of the limitation of the data set stated earlier, it would not be worth it to dwell on the implications of these models.

---

5. DISCUSSION

The effect of underreported minor injury and property-damage-only accidents on model parameter estimation and, in turn, on encroachment rate estimation has not been discussed in this paper. However, it is generally expected that the underreporting probability increases as the severity of accidents decreases. Because the severity of accidents is highly related to the impact speed and, therefore, to the lateral offset of roadside objects, the underreporting probability of accidents is expected to increase as the impact speed decreases or as the lateral offset of objects increases. Under these notations, the parameter $\alpha_1$ in Eqs. (17) and (18) reflects not only the collision probability, but also the underreporting probability of accidents, as the lateral offset increases.
Thus, a higher underreporting rate in the accident data would result in a higher $|\alpha_1|$ value being estimated.

The effect of underreported accidents on parameter $\beta_1$ is, however, expected to be relatively small. The reason is that the severity of accidents is expected to be high when vehicles collide with fixed objects that are fairly close to the travel lane (or, more precisely, that have $D_{obj,4} = 0$) and, therefore, very few unreported accidents are expected under such collisions. This small effect on $\beta_1$ suggests that underreporting of minor accidents should have a small effect on encroachment rate estimation when Eqs. (17) and (18) are used.

The research presented in this paper can potentially be extended in several directions:

(1) Current encroachment-based studies do not provide any clue on what the plausible functional forms for modeling Process 1 (i.e., encroachment probability) might be. Thus, an interesting extension would be to apply traffic flow, driver behavior, and vehicle dynamics theories to gain some engineering insights on this functional form.

(2) This study used the same data base as in Miaou [1996]. It would be interesting to see if consistent encroachment rates can be estimated when other independent data sets are used.

(3) In theory, the proposed method for estimating encroachment rates can be extended to consider road sections of roadway classes other than the rural, two-lane, undivided roads.

ACKNOWLEDGEMENTS

This paper is based on the results of a research project sponsored by the Office of Safety and Traffic Operations R&D, Federal Highway Administration (FHWA). The opinions expressed in this paper are, however, solely those of the author. The author's research presented in this paper was originally motivated by the late Mr. Harry Lum.
REFERENCES


LIST OF TABLES

Table 1. Simple kinematic equations used to estimate time to collision and impact speed for a given encroachment angle when encroachment trajectory is a straight line and deceleration rate is a constant.

Table 2. Estimated regression coefficients of extended negative binomial regression models and associated statistics for guardrail accidents.

Table 3. Estimated regression parameters of extended negative binomial regression models and associated statistics for utility pole accidents.

LIST OF FIGURES

Figure 1. An overview of the encroachment-based approach.

Figure 2. An example joint probability density function of encroachment speed and angle.

Figure 3. Impact envelope of a roadside object for a given vehicle encroachment angle.

Figure 4. Illustrated conditional probability distributions of impact speeds as a function of the lateral offset of the object when an encroached vehicle is in the hazard envelope.

Figure 5. Probabilities of an encroached vehicle to hit point objects of different sizes at various lateral offsets.

Figure 6. Probabilities of an encroached vehicle to hit continuous objects of different lengths at various lateral offsets.

Figure 7. Probabilities of an encroached vehicle to collide with a point object and a continuous object at a speed greater than v.

Figure 8. Comparison of the roadside encroachment frequency estimated from the accident prediction model developed in this study and observed frequencies from earlier studies.

Figure 9. Roadside encroachment frequencies estimated from two utility pole accident models under three different sizes of hazard envelopes.
Table 1. Simple kinematic equations used to estimate time to collision and impact speed for a given encroachment angle when encroachment trajectory is a straight line and deceleration rate is a constant.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Time to Collision, ( t_c )</th>
<th>Impact Speed, ( v_e )</th>
</tr>
</thead>
<tbody>
<tr>
<td>if ( d_c \leq v_o t_c )</td>
<td>( t_c = \frac{d_c}{v_0} )</td>
<td>( v_e = v_0 )</td>
</tr>
<tr>
<td>if ( v_0 t_c &lt; d_c \leq v_0 t_c + \left( v_0^2 / 2\Delta \right) )</td>
<td>( t_c = t_c + \frac{v_0}{\Delta} ) ( - \left{ \frac{v_0^2}{2\Delta} - v_0 t_c \right}^{1/2} / \Delta )</td>
<td>( v_e = \left{ \frac{v_0^2}{2\Delta} - 2\Delta [d_c - v_0 t_c] \right}^{1/2} )</td>
</tr>
<tr>
<td>if ( d_c &gt; v_0 t_c + \left( v_0^2 / 2\Delta \right) )</td>
<td>( t_c = \infty ) (i.e., no collision)</td>
<td>( v_e = 0 )</td>
</tr>
</tbody>
</table>

Notations:
- \( d_c \): encroachment distance to collision (mi or km) (see the notes below)
- \( D_{obj} \): lateral offset of the object (mi or km), measured from the outside edge of travelway (see Figure 3)
- \( v_0 \): encroachment speed (mi/hr or km/hr)
- \( v_e \): impact speed (mi/hr or km/hr)
- \( \phi \): encroachment angle (degree)
- \( t_r \): driver's response delay (sec), measured from the time encroachment occurs until the driver applies the brake
- \( \Delta \): constant deceleration rate (mi/sec\(^2\) or km/sec\(^2\)), e.g., 0.5g (=0.5x32.2/5280 mi/sec\(^2\)=0.00304 mi/sec\(^2\)=0.00304x3600\(^2\) mi/hr\(^2\)=39.518 mi/hr\(^2\)); \( \Delta \) actually depends on surface type, sideslope, and wetness.

Notes:
From Figure 3, the distance that a vehicle needs to travel to collide with the object after encroachment can be estimated as
\[
d_c = \frac{D_{obj}}{\sin(\phi)}
\]
if the encroachment occurs in zones 1 and 2;
and, depending on where the encroachment occurs within zone 3, the distance is between
\[
d_c = \frac{D_{obj}}{\sin(\phi)} \quad \text{and} \quad d_c = \frac{D_{obj} + W_{obj}}{\sin(\phi)}
\]
Strictly speaking, it also depends on vehicle orientation at the time of collision. The two equations above suggest that \( D_{obj} / \sin(\phi) \) is a good estimate of the distance only when \( D_{obj} \gg W_{obj} \) or when continuous objects are considered where \( l_{obj} \gg W_{obj} \) and \( \phi \) is not too small (i.e., the hazard envelope of zone 3 is considerably smaller than that of zones 1 and 2 combined).

Unit conversions: 1 mi = 5,280 ft = 1.6 km = 1,600 m
Table 2. Estimated regression parameters of extended negative binomial regression models and associated statistics for guide rail accidents.

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Estimated NB Model Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Guardrail Length &gt; 0.1 mi</td>
</tr>
<tr>
<td></td>
<td>(n=272 Sections)</td>
</tr>
<tr>
<td>Dummy intercept (=1): the associated parameter is ( \beta_1 ) as described in the text</td>
<td>1.66290 (±0.229; 7.26)</td>
</tr>
<tr>
<td>AADT per lane ( (10^3 ) vehicles per day per lane)</td>
<td>-0.20515 (±0.095; -2.16)</td>
</tr>
<tr>
<td>Lane width (ft)</td>
<td></td>
</tr>
<tr>
<td>Horizontal curvature (degrees/100 ft arc)</td>
<td></td>
</tr>
<tr>
<td>Vertical grade (percent)</td>
<td></td>
</tr>
<tr>
<td>Clear zone width (ft): Paved shoulder width</td>
<td>-0.18391 (±0.034; -5.42)</td>
</tr>
<tr>
<td>Clear zone width (ft): Unpaved shoulder width (i.e., Earth, grass, gravel, or stabilized shoulder width)</td>
<td>-0.17013 (±0.040; -4.26)</td>
</tr>
<tr>
<td>Clear zone width (ft): Additional clear zone width beyond shoulders</td>
<td>-0.27137 (±0.077; -3.53)</td>
</tr>
<tr>
<td>Clear zone width (ft): Total available width</td>
<td></td>
</tr>
<tr>
<td>Dispersion parameter of the NB model</td>
<td>0.89325 (±0.155; 5.77)</td>
</tr>
<tr>
<td>Log-likelihood function/n</td>
<td>-1.64</td>
</tr>
<tr>
<td>Akaike Information Criterion Value/n</td>
<td>3.33</td>
</tr>
<tr>
<td>Expected vs. Observed total number of accidents</td>
<td>508 vs. 450</td>
</tr>
</tbody>
</table>

Notes: (1) About five years of accident data (1980-1984) were available. (2) Values in parentheses are asymptotic standard deviation and t-statistics of the estimated parameters above. (3) ----- indicates "not selected in the final model.” (4) The exposure measure, vehicle-miles traveled, in the model is in million vehicle-miles. (5) 1 mi = 1.61 km and 1 ft = 0.3048 m.
Table 3. Estimated regression parameters of extended negative binomial regression models and associated statistics for utility pole accidents.

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Estimated NB Model Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dummy intercept (=1): the associated parameter is $\beta_1$ and, as described in the text, $\delta = 168$ ft</td>
<td>0.68440 (±0.264; 2.59) 0.68438 (±0.266; 2.57) 1.97345 (±1.020; 1.94)</td>
</tr>
<tr>
<td>AADT per lane ($10^3$ vehicles per day per lane)</td>
<td>-0.30398 (±0.073; -4.14) -0.38470 (±0.071; -5.39) -0.38376 (±0.071; -5.37)</td>
</tr>
<tr>
<td>Lane width (ft)</td>
<td>---</td>
</tr>
<tr>
<td>Horizontal curvature (degrees/100 ft arc)</td>
<td>---</td>
</tr>
<tr>
<td>Vertical grade (percent)</td>
<td>---</td>
</tr>
<tr>
<td>Sideslope (e.g., 3:1 and 7:1 slopes are recorded as $1/3=0.33$ &amp; $1/7=0.14$, respectively.)</td>
<td>---</td>
</tr>
<tr>
<td>Clear zone width (ft): Paved shoulder width</td>
<td>-0.15338 (±0.035; -4.44)</td>
</tr>
<tr>
<td>Clear zone width (ft): Earth, grass, gravel, or stabilized shoulder width</td>
<td>-0.18211 (±0.027; -6.69)</td>
</tr>
<tr>
<td>Clear zone width (ft): Additional clear zone width beyond shoulders</td>
<td>-0.05198 (±0.015; -3.54)</td>
</tr>
<tr>
<td>Clear zone width (ft): Total available width</td>
<td>--- -0.08101 (±0.014; -5.82) -0.27419 (±0.151; -1.82)</td>
</tr>
<tr>
<td>Clear zone width &gt; 10.5 ft (ft): Clear zone width, if it is greater than 10.5 ft; 0 otherwise</td>
<td>--- 0.10819 (±0.088; 1.23)</td>
</tr>
<tr>
<td>Clear zone width &gt; 20.5 ft (ft): Clear zone width, if it is greater than 20.5 ft; 0 otherwise</td>
<td>--- 0.037183 (±0.035; 1.06)</td>
</tr>
<tr>
<td>Dispersion parameter of the NB model</td>
<td>0.40907 (±0.191; 2.14) 0.60062 (±0.206; 2.91) 0.58807 (±0.204; 2.87)</td>
</tr>
<tr>
<td>Loglikelihood function/n</td>
<td>-0.69 -0.71 -0.70</td>
</tr>
<tr>
<td>Akaike Information Criterion Value/n</td>
<td>1.40 1.42 1.42</td>
</tr>
<tr>
<td>Expected vs. observed total number of accidents</td>
<td>305 vs. 293 309 vs 293 308 vs. 293</td>
</tr>
</tbody>
</table>

Notes: (1) 855 road sections were used in developing these models (i.e., n=855). (2) The average pole spacing of each road section is greater than 0.1 mi. (3) About five years of accident data (1980-1984) were used. (4) Values in parentheses are asymptotic standard deviation and t-statistics of the estimated parameters above; (5) --- indicates "not selected in the final model." (6) The exposure measure, vehicle-miles traveled, in the model is in million vehicle-miles. (7) 1 mi = 1.61 km and 1 ft = 0.3048 m.
Figure 1. An overview of the encroachment-based thinking.

A Vehicle, A Road Segment & A Particular Type of Roadside Objects

**Process 1**

Roadside Encroachment?

**Yes**

A Joint Probability Density of Encroachment Speed and Angle

**Key Determinants of Each Process**

- Mainline Road Design
- Traffic Conditions
- Other Unobservable Driver, Vehicle & Environmental Variables

**Mainline Condition**

**Roadside Condition**

**Process 2**

In Impact Envelope?

**No**

**Process 3**

Collide with the Object?

**No**

**Process 4**

A Reportable Accident?

**Yes**

- Size of Roadside Objects
- Vehicle Swath Width

- Lateral Offsets of Objects
- Surface Type, Slope & Wetness
- Vehicle Braking System & Tire Condition
- Driver’s Response Delay
- Driver’s Braking & Steering Behavior

- Weight/Structural Design of Vehicle
- Vehicle Impact Position
- Driver’s Physical Condition
Figure 2. An example joint probability density function of encroachment speed and angle.

$\phi_{\text{max}}(v)$ \quad \text{Functional Relationship Between Speed and Maximum Possible Angle}

$\phi_{\text{max}}(v_{\text{min}})$

$\phi_{\text{max}}(v_{\text{max}})$

$f(\phi \mid \phi_{\text{max}})$ \quad \text{Angle Density}

$f(\phi_{\text{min}})$

$f(v_{\text{ref}})$
Figure 3. Impact envelope of a roadside object for a given vehicle encroachment angle.

\[ \ell_{\text{enr}} = \ell_{\text{obj}} + W_{\text{veh}} \csc(\phi) + W_{\text{obj}} \cot(\phi) \]
Figure 4. Illustrated conditional probability distributions of impact speeds as a function of the lateral offset of the object when an encroached vehicle is in the hazard envelope.
Figure 5. Probabilities of an encroached vehicle to hit point objects of different sizes at various lateral offsets.

**Point Objects of Different Diameters**

(a) Lateral Offset of the Object (ft)

(b) Diameter of the Object (inches)

P(Hit Object | Encroachment) (%)
Figure 6. Probabilities of an encroached vehicle to hit continuous objects of different lengths at various lateral offsets.
Figure 7. Probabilities of an encroached vehicle to collide with a point and a continuous object at a speed greater than v.

**Point Object**
Diameter: 8 in

**Continuous Object**
Length: 0.25 mi
Width: 1 ft
Figure 8. Comparison of the roadside encroachment frequency estimated from the accident prediction model developed in this study and observed frequencies from earlier studies.
Figure 9. Roadside encroachment frequencies estimated from two utility pole accident models under three different sizes of hazard envelopes.
Report Number (14) ORNL/CP--9436  4
               CONF-980112--

Publ. Date (11) 1997 08
Sponsor Code (18) DOT, XF
UC Category (19) UC-000, DOE/ER