Accident Model of the Traffic Mix: Use of Vehicle Miles to Predict Accidents

THIPATAI CHIRACHAVALA AND JAMES O'DAY

Information on vehicle miles of travel for any two classes of vehicles (e.g., cars versus trucks or vehicles with drunk drivers versus vehicles with nondrunk drivers) can be used together with accident frequency to develop an accident prediction model based on the mix of traffic on roadways. As an illustration, a model was developed to predict the proportion of many possible accident configurations involving cars and trucks (single-car accidents, car-truck accidents, etc.), taking into account the effect of environmental factors such as road class and time of day. Extension of this model to include any number of factors other than road class and time of day is possible. Useful applications of the model include assessment of the relative highway safety of any two vehicle classes that may possess different accident characteristics, assessment of environmental factors that affect the highway safety of these vehicles, and quick input for evaluating policy options concerning of the use of certain types of vehicles.

An accident is considered a single event even when it involves more than one vehicle. Vehicle involvements in accidents, on the other hand, are counts of the number of vehicles involved in accidents. Vehicle involvements in accidents are usually specified by type of vehicle--i.e., car involvements, truck involvements, etc. Therefore, "accidents" and "vehicle involvements" are different concepts and should not be used interchangeably. For example, the number of accidents in 10 collisions, each involving 2 cars, is 10 whereas the number of car involvements in those accidents is 20.

Vehicle miles of travel is a common measure of exposure. The rates (involvements per vehicle mile of travel) provide a more useful comparison of the accident experience of the vehicles than the number of involvements alone. The use of exposure information is not, however, restricted only to the rates comparison. It can be used together with accident information to develop an accident model that permits some inference regarding the nature of accidents. This paper presents a method for achieving such a goal by using an example of cars versus trucks.

Highway accidents can be categorized as single-vehicle, two-vehicle, or more-than-two-vehicle crashes. For traffic consisting of passenger cars and trucks only, the possible accident configurations are (a) single-car (SC), (b) single-truck (ST), (c) car-car (CC), (d) car-truck (CT), (e) truck-truck (TT), and (f) multivehicle (CCC, CCT, TTT, CCTT, CCCC, etc.).

Accidents involving more than two vehicles are relatively rare: They typically account for 5 percent or less of total accidents. They were, therefore, neglected in developing the model.

LITERATURE REVIEW

Scott and O'Day (1) showed that, if involvements in accidents of both cars and trucks were assumed to be proportional to their respective miles of travel, the probability that an accident-involved vehicle (V) was a truck would be equal to the proportion of vehicle miles accumulated by all trucks and the same for cars. Thus, for a population consisting only of cars and trucks,

$$P(V = \text{truck} \mid \text{an accident}) = T \qquad (1)$$

and

$$P(V = \text{car} \mid \text{an accident}) = (1 - T) \qquad (2)$$

where T is the proportion of truck mileage and (1 - T) is the proportion of car mileage.

If the proportion of single-vehicle accidents is represented by S and therefore the proportion of two-vehicle accidents by (1 - S), then the proportions of oneand two-vehicle accident configurations are given by

$$\text{Proportion of SC accidents} = S(1 - T) \qquad (3)$$

$$\text{Proportion of ST accidents} = ST \qquad (4)$$

$$\text{Proportion of CC accidents} = (1 - S)(1 - T)^2 \qquad (5)$$

$$\text{Proportion of CT accidents} = 2(1 - S)T(1 - T) \qquad (6)$$

$$\text{Proportion of TT accidents} = (1 - S)T^2 \qquad (7)$$

These five proportions sum to 1.0 (crashes involving more than two vehicles are neglected).

The limitation of the above model arises from its assumption that the chance of involvement in an accident is the same for a car and a truck. That is, if the car mileage were equal to the truck mileage, the frequencies of their involvements would also be equal. The model does not allow for the fact that trucks and cars might have different potential for being involved in an accident due to different vehicle and/or driver characteristics. In the situation where this assumption is not justified, the model will no longer be valid. This has led to the development of a more general model that does not require such a stringent assumption. This general model will be referred to as the accident model of the traffic mix.

MODEL DEVELOPMENT

Accidents can be viewed as the result of "failures" in a system comprising the vehicles, the drivers, and the environment. To be useful, an accident model ought to reflect such a relation.

The accident model of the traffic mix is a mathematical representation of the probabilities of the occurrence of various possible accident configurations. In considering an example of cars versus trucks, the rationale for such a model is that accident involvements of cars and trucks are some function of their individual characteristics (which collectively reflect the vehicle and/or the driver characteristics), mileage, and environmental factors. That is, P(V = car | an accident) = f(characteristics of cars, car miles, environment); and P(V = truck | an accident) = f(characteristics of trucks, truck miles, environment).

In developing an accident model based on the above assumption, a two-stage modeling procedure was introduced. The first stage derives an accident model of the traffic mix, assuming that the car and truck involvements are some function of their individual characteristics and their respective mileage. The second stage incorporates the environmental factors into the model externally. The two stages are discussed below.

Stage 1

It is assumed that

$$P(V = car|an\ accident) = W_1 (1 - T) \quad (8)$$

and

$$P(V = truck|an\ accident) = W_2 T \quad (9)$$

where W_1 and W_2 are constants representing the involvement propensity of cars and trucks, respectively.

If S represents the proportion of single-vehicle accidents and (1 - S) the proportion of two-vehicle accidents, then the proportions of the five accident configurations involving cars and trucks are given by

$$\text{Proportion of SC accidents} = S(1 - T)/(1 + aT) \quad (10)$$

$$\text{Proportion of ST accidents} = S(1 + a)T/(1 + aT) \quad (11)$$

$$\text{Proportion of CC accidents} = (1 - S)(1 - T)^2/(1 + bT)^2 \quad (12)$$

$$\text{Proportion of CT accidents} = 2(1 - S)(1 + b)T(1 - T)/(1 + bT)^2 \quad (13)$$

$$\text{Proportion of TT accidents} = (1 - S)(1 + b)^2T^2/(1 + bT)^2 \quad (14)$$

The sum of the proportions is 1.0.

The constant a is the difference between the truck involvement rate (per mile of truck travel) and the car involvement rate (per mile of car travel) for single-vehicle accidents, expressed as a percentage of the car involvement rate, or

$$a = (\text{truck involvement rate} - \text{car involvement rate})/\text{car involvement rate} \quad (15)$$

and the constant b is the difference between the truck involvement rate and the car involvement rate for two-vehicle accidents, expressed as a percentage of the car involvement rate, or

$$b = (\text{truck involvement rate} - \text{car involvement rate})/\text{car involvement rate} \quad (16)$$

In this model, the proportionality constant, W_1 in Equation 8, is equal to 1. W_2 in Equation 9 is equal to (1 + a) and (1 + b) for single- and two-vehicle accidents, respectively, which allows the truck involvement rate relative to the car involvement rate to be different for single-vehicle and two-vehicle accident configurations. In the circumstance where a and b are equal, the model as represented by Equations 10-14 still holds. When the accident involvement rates of both cars and trucks are equal, both a and b will be zero and Equations 10-14 simplify to Equations 3-7. Therefore, the model proposed earlier by Scott and O'Day (1) is a special case of the general accident model of the traffic mix.

Stage 2

The model as represented by Equations 10-14 is then applied to each of the cells (or subsets) created by the cross classification of the environmental factors. There may be any number of such variables. For each cell, two independent estimates of a are obtained from Equations 10 and 11 and three independent estimates of b from Equations 12-14. Unique cell estimates of a and b are then determined as a function of the environmental factors. The following model estimation illustrates this point.

MODEL ESTIMATION

The following model estimation is based on a population of vehicles consisting of cars and trucks. The environmental factors considered were road class and time of day. These two factors have been cited in a number of past studies as having an important effect on accident rates. For example, Herd and others (2) reported that on rural roads the overall accident rate (accidents per vehicle mile of travel) was higher at night than during the day. His reported ratio of the night-to-day accident rates was greatest for rural expressways (1.98) and smallest for four-lane roads (1.47). The accident rate at dusk was reported to be higher than that at dawn. The rate for fatal accidents was also reported to be higher at night.

The model estimation involves the following steps:

1. The accident and exposure data are cross-classified by road class and time of day. Let the rows represent the various time-of-day periods and the columns the various categories of road class.
2. The model as represented by Equations 10-14 is applied to the data. For each cell, the estimated values of a can be obtained by solving Equations 10 and 11 and the estimated values of b by Equations 12, 13, and 14. Because all of the factors that affect the occurrences of different accident configurations can never be accounted for in modeling, the two estimated values of a from Equations 10 and 11 may not be exactly identical, though they will be close; the same applies to the three estimated values of b from Equations 12-14. As a result, each combination of road class and time of day will have two independently estimated values of a and three independently estimated values of b.
3. The unique cell estimates of a and b as well as the effects of road class and time of day can be determined as follows. Define

$$a_{+++} = \{\Sigma_i \Sigma_j \Sigma_k a_{ijk}\}/IJK$$
$$a_{i++} = \{\Sigma_j \Sigma_k a_{ijk}\}/JK$$
$$a_{+j+} = \{\Sigma_i \Sigma_k a_{ijk}\}/IK$$

where

I = number of rows,
J = number of columns, and
D = number of a in each cell (i,j), which is 2.

We have

$$\alpha_i^a = a_{i++} - a_{+++} \quad (17)$$

$$\beta_j^a = a_{+j+} - a_{+++} \quad (18)$$

$$\gamma_{ij}^a = a_{ij+} - a_{i++} - a_{+j+} + a_{+++} \quad (19)$$

where a_{ij+} is the average value of a for cell (i,j). Therefore,

$$a_{ij} = a_{+++} + \alpha_i^a + \beta_j^a + \gamma_{ij}^a \quad (20)$$

Similarly, for b with I rows, J columns, and K = 3,

$$\alpha_i^b = b_{i++} - b_{+++} \quad (21)$$

$$\beta_j^b = b_{+j+} - b_{+++} \quad (22)$$

$$\gamma_{ij}^b = b_{ij+} - b_{i++} - b_{+j+} + b_{+++} \quad (23)$$

where b_{ij+} is the average b for cell (i,j). Therefore,

$$b_{ij} = b_{+++} + \alpha_i^b + \beta_j^b + \gamma_{ij}^b \quad (24)$$

The results of the model estimation are represented by Equations 17-24. The values of a_{ij}

and b_{ij} are therefore the unique cell estimates of the model parameters a and b, road class and time of day having been accounted for. α_i^a, β_j^a, γ_{ij}^a, α_i^b, β_j^b, and γ_{ij}^b estimate the main effect of time of day, the main effect of road class, and their interactions on a and b, respectively.

APPLICATION OF THE MODEL

Possible applications of the accident model of the traffic mix include the following:

1. The model can be used to assess the relative involvement rates of any two different classes of vehicles that have different characteristics, taking into account the effect of any number of environmental factors. This will help minimize the undesirable "Simpson's Paradox" caused by confounding factors not otherwise considered.
2. The model can be used to predict the reduction (or the increase) in the number of accidents that results from altering the travel pattern (or the amount of travel) of some vehicles on certain roads and at certain times of day. These results can then be used as input for evaluating various highway-safety policy options concerning the use of certain types of vehicles.

CONCLUSIONS

The accident model of the traffic mix as developed might be expected to predict well when applied to a traffic situation in which the mix of any two different vehicle classes and the overall traffic volume are relatively uniform. The model can potentially be extended to include any number of environmental factors without altering the basic model presented in Equations 10-14. These factors are incorporated into the model in such a way that they partition the accident and exposure data into cells with relatively uniform traffic mix and overall traffic volume. The factors that can be included in the model, of course, depend on the level of detail of the available exposure and accident data. The stability of the estimated model for prediction depends on the ability to search for environmental factors that strongly influence the accident rates of the vehicle classes being investigated. The reliability of the estimated model is a function of the accuracy in measuring exposure. Future research efforts should therefore also be directed to acquiring reliable exposure data with a greater level of detail than is generally available now. Furthermore, there is a need for compatible definitions of the variables in both the accident and exposure data sets.

Computer programs to perform model estimation and prediction are available at the University of Michigan Transportation Research Institute in Ann Arbor.

ACKNOWLEDGMENT

We wish to thank Phyllis A. Gimotty for her constructive editorial comments.

REFERENCES

1. R.E. Scott and J. O'Day. Statistical Analysis of Truck Accident Involvements. NHTSA, FH-11-7614, Dec. 1971.
2. D.R. Herd, K.R. Agent, and R.L. Rizenbergs. Traffic Accidents: Day Versus Night. TRB, Transportation Research Record 753, 1980, pp. 25-30.

Publication of this paper sponsored by Committee on Traffic Records.

Microcomputer-Based Traffic Records System for Small Police Agencies

WILLIAM E. KELSH

In Virginia, there are many small cities, towns, and counties that maintain manual traffic records systems to meet their traffic safety data needs. Of these, the larger localities have a sufficiently high number of motor vehicle crashes and traffic violations to justify the need for automated record-keeping systems. However, the high cost of computer hardware and required technical expertise have discouraged these localities from acquiring the record management capability they need. The advent of the microcomputer has now brought sophisticated record-keeping technology within reach of even the smallest budgets. Still, lacking the staff support and the required applications software, most localities are unable to take advantage of the benefits of the new technology. In an effort to solve the problem, the Virginia Highway and Transportation Research Council, with funding support from the Virginia Department of Transportation Safety, has developed a model user-oriented local traffic records software system for small localities. The system accepts, stores, and recalls data for accidents and traffic offenses rapidly, accurately, and inexpensively. With further development, it will have the capability to be run on most currently marketed microcomputers.

An effective local program for reducing traffic accidents requires the capability to (a) identify traffic safety problems, (b) develop and implement appropriate countermeasures, and (c) evaluate the results of the chosen strategies. To achieve this capability, localities must keep records on the incidence of motor vehicle crashes and violations of traffic ordinances. Further, these records must be organized so they can be easily accessed and analyzed.

Localities also need to keep traffic records for the efficient management and operation of safety programs. Clearly, the key to maximizing the use of limited resources is information about the nature and scope of the traffic safety problem to be addressed. With this information, traffic safety administrators presumably can direct their resources toward the most serious problems or toward those problems that have the highest potential for payoff.

Finally, during these times of economic hardship for local governments, it is important for traffic safety officials to be able to justify traffic