

# Truck Accident Involvement With and Without Front-Axle Brakes: Application for Case-Control Methodology

MARTIN E. H. LEE-GOSSELIN, A. J. RICHARDSON, AND GORD TAYLOR

Methods of predicting the change in accident involvement of heavy trucks that would result from the mandatory installation of tractor-trailer front-axle brakes were identified. The choice of method for evaluating the front-axle brake issue became controversial as a result of the finding that a system-wide accident and exposure study would be prohibitively expensive and would take longer to complete than the life expectancy of the regulatory decision. The methodology known as "case-control" seemed an attractive alternative, but its previous use for heavy-truck accidents had been severely challenged. The usefulness of this method may have been obscured during a period of rightful questioning of the interpretation of the results from earlier case-control studies. A detailed analysis was made of the statistical limitations and practical feasibility of the case-control methodology. A computer simulation demonstrated that the method can be used to provide unbiased estimates of the coefficients in a logit-type causal accident model, and that only one control per accident is required. Further, it was recommended that rather than focusing on accident-involvement odds ratios, the model from the case-control methodology should be used in a probabilistic economic analysis to answer the regulatory question. Case-control was found to be a suitable approach for evaluating the front brake issue, but only at a level of threshold economic benefit and not in terms of absolute accident rate (number/veh-km). Moreover, it should be implemented only with certain safeguards, notably the validation of the randomness of control vehicle selection using classified vehicle counts. Estimated costs of implementation, although much below those of system-wide inspection surveys of truck exposure and accidents, were nevertheless substantial.

## THE PROBLEM OF EVALUATING POTENTIAL TRUCK EQUIPMENT REGULATIONS IN CANADA

A research project required the design of methodology to obtain accident and exposure data for comparisons of heavy trucks with certain configurations and equipment. More specifically, data were needed to produce reliable inferences about the benefits of changes to the Canada Motor Vehicle Safety Standard (CMVSS) 121. Standard CMVSS 121 primarily concerns the airbrake systems of heavy trucks.

A number of changes to CMVSS 121 have been under consideration. These changes include the mandatory instal-

lation of front-axle brakes on tractor-trailers, a number of standards applicable to brake performance, and a requirement that front limiting valves, if fitted, be automatic. Although the study did cover a number of other equipment issues, including drive-axle pressure-reducing valves, load-sensing valves, trailer hand-valves, power steering, A- versus B-type trailer hitches, and bobtail configurations, the urgency of the proposed changes to the CMVSS required that most of the methodological effort be devoted to tractor-trailer front-axle brakes.

Any change to the standards must undergo a formal evaluation known as a regulatory impact assessment (RIA). A major part of such studies involves comparing the predicted benefits of a regulation with the costs involved in compliance. In the present context, this assessment implies the use of a fair and statistically competent method of predicting accident-reduction benefits. No such method had been established for an RIA of heavy truck equipment regulations.

Building the statistical case for modifying or leaving intact the CMVSS 121 standard requires the consideration of many technical complexities. It also requires an appreciation of the manner in which amended regulations can reasonably be implemented and enforced, and of the level of proof without which the trucking industry may be reluctant to cooperate. Furthermore, the operation of equipment-specific truck accident and exposure data collection methods would be impossible without the close and sometimes generous cooperation of the trucking industry.

Putting the cost of predicting the benefits of front-brake compliance in perspective, it was estimated that mandatory installation of front brakes on new trucks would only cost about \$1.5 million per year spread over 10,000 to 15,000 new vehicle sales. During the current climate of deregulation and increased competition (which one industry representative described as survival time), it was found that most trucking companies would be willing to cooperate in an evaluation, but that it was unlikely that companies would want to participate in experiments in which equipment was randomly installed or proscribed across a fleet.

Thus, the objective for the study was to develop for Transport Canada a solution that is theoretically sound and relevant to the regulatory decisions, and that respects the reasonable interests of the truck manufacturers and end-users. The findings presented resulted from new methodological research; a 2-day technical workshop for a dozen experts from industry, government, and the research community; and interviews of trucking firms, service companies, public authorities, and others

M. E. H. Lee-Gosselin, Faculty of Administrative Sciences, Laval University, Quebec City, Canada G1K7P4. A. J. Richardson, Department of Transport and Resource Engineering, Royal Melbourne Institute of Technology, GPO Box 2476, Melbourne, Victoria 3001, Australia. G. Taylor, Tandem Engineering and Management Consultants, 190 Bronson, Ottawa, Canada K1R6H4.

who would be involved in carrying out an eventual study of accident involvement.

### Sources of Methodological Difficulty

Fundamental to the adequacy of accident and exposure data collection methods is the difficulty of identifying the status of the truck at the time of measurement. This problem frequently occurs in driver characteristics—age, sex, training, experience, physical and mental state, and so on. But in vehicle characteristics, there are peculiar difficulties because the identification and verification of the devices can only be done in close and direct cooperation with the owner and operators of the equipment. This cooperation is essential because

- No comprehensive data base of trucks is in operation that will identify truck technology at the level of interest;
- Trucks are modified by the dealer, aftermarket, and operator for some of the equipment identified, and there is even less standardization of records at this level;
- Only the owner or operator controls the level of maintenance. A further difficulty may be that the related equipment is sufficiently poorly maintained, resulting in the equipment of interest being verified to be in good condition, but not being able to do its job; and
- Only the owner or operator controls driver quality. The possibility exists that the equipment has been tampered with, such as in the case of disconnected front brakes. Intentional misuse or nonuse is a further possibility, which in the case of front brakes can effectively eliminate their usefulness in as little as 6 months.

Difficulties are also specific to accident data. Canadian provincial accident data records are limited in their description of trucks involved and categories of truck are not always consistently defined across the country. In general, police accident reports do not contain sufficient detail to eliminate from analysis those accidents that are irrelevant to the equipment issues under study. Other sources of accident follow-up, notably insurance companies and fleet management, can sometimes supply the missing detail, but these are not always available for the vehicles selected for study.

In developing exposure data for the vehicles with the equipment under study, verifying that the vehicles selected are not highly atypical would be desirable. Some overall picture of truck use is needed as a basis for comparison. The purpose is not to estimate the absolute safety impact of equipment so much as to ensure that any judgment about the CMVSS made from accident studies cannot be criticized as irrelevant to typical truck use. The best source at the present time is the *Provincial Truck Fleet Study—1986 (1)*. Unfortunately, neither this nor any other source can provide an independent measure of the use patterns of trucks with precisely the equipment-configuration combinations of interest. Therefore, such sources are mostly useful for the design of sampling strategies.

Timing of the regulatory decision also brings some difficulties. Statistical confidence may be satisfied by tracking accidents for reasonable samples of trucks over an extended period. However, the decision cannot be postponed just to satisfy

sampling requirements. After the decision, additional time may be needed before a new or modified standard is implemented to allow the industry to respond to the new requirements. Moreover, the useful lifetime of a standard may be limited because of the evolution of truck technology.

### The Distribution of Trucks with Front-Axle Brake Systems in Canada

The size of truck populations containing, or likely to soon contain, the technologies will place an upper limit on the number of trucks and fleets available to potential experimental designs. If the total truck population using the technology is too small for the collection of data within a reasonable length of time, then it may not be feasible to evaluate comparisons of the type with versus without on the basis of relative accident involvement. The percentage of three-axle trucks with front brakes installed is changing rapidly as a result of a U.S. regulation requiring front brakes on all trucks using federally funded highways in the United States, regardless of origin. Because most large Canadian trucking companies operate across the border to some extent, this regulation has resulted in a large increase in front brake installations on new vehicles and in the retrofit market [on the basis of comments from Bendix, Wabco, Eaton-Yale, and original equipment manufacturers (OEMs)]. Currently, the major OEMs are reporting 80 to 95 percent installation rates.

The average frequency of front brakes in use was measured in a 1986 Transport Canada study (2), which measured a mean installation rate of 54 percent. This rate was found to vary across Canada—40 percent in Alberta and British Columbia, 50 percent in Saskatchewan and Manitoba, 66 percent in Ontario and Québec, and 50 percent in New Brunswick. Truck replacement rates vary from area to area and on the basis of economic cycles. In the last few years, new truck sales have been strong (up to 12 percent replacement of the fleet per year), but an 8 to 10 percent average replacement rate should be assumed for the near future. On the basis of this range of replacement and front brake installation rates for these new vehicles, Table 1 presents low and high forecasts of the average percentage of the total fleet that will be equipped with front brakes out to 1992. This simple forecast assumes that the replaced vehicle is the average of the vehicle pool. In fact, this assumption will underestimate the total penetration as the older vehicles (i.e., the replacement market) are likely to have less than the average front brake installation. One consequence of U.S. regulation is that trucks without front-axle brakes are increasingly a phenomenon of regional or local trucking companies, limiting the potential for studies based on interfleet comparisons.

A further complexity is the installation of limiter valves (called proportioning or automatic limiting valves in the United States), which proportion or eliminate front brake pressure until preset brake pressures are achieved. Thus, under light braking the front brakes are not activated, but under hard braking the front brakes are applied. This device is intended to decrease front wheel lock-up and skidding under light braking action or load. Moreover, these valves are sold in automatic (U.S. regulations) and manual configurations. According to the OEM industry representatives interviewed, between

TABLE 1 LOW-HIGH ESTIMATES OF FRONT BRAKE PENETRATION IN POPULATION

Installation Rate On New Trucks: 75%/90%				
Replacement Rate Of Fleet: 8%/12%				
Region	1986 base%	1988 lo%/hi%	1990 lo%/hi%	1992 lo%/hi%
BC/Alberta	40	46/52	50/61	54/68
Saskat/Manitoba	50	54/60	57/67	60/72
Ontario/Québec	66	67/72	69/76	70/79
Maritimes <sup>1</sup>	50	54/60	57/67	70/72
Wt. Average <sup>2</sup>	55	57/62	59/67	61/71

<sup>1</sup> Assumes N.B. data is indicative of entire Maritime Provinces

<sup>2</sup> Weighted by total truck registrations

62 and 100 percent of the front brake systems are currently specified with pressure limiting valves (based on comments from Freightliner, Navistar, Ford, Volvo GM-White, Mack, and Kenilworth).

### PREVIOUS SOURCES OF DATA ON LARGE TRUCKS

A number of approaches have been adopted to study large truck accidents and exposure, mostly in the United States. Although it is clear that truck operations in Canada differ in some important respects, the U.S. studies offer some insights into the methodological and practical difficulties involved under North American conditions. The most important of the sources reviewed were

- Truck Inventory and Use Survey (TIUS) (3);
- Fleet Accident Evaluation of U.S. FMVSS 121 (4);
- University of Michigan Transportation Research Institute Large-Truck Survey Program, including the Trucks Involved in Fatal Accidents (TIFA) data base, 1980 to 1984, and the National Truck Trip Information Survey (NTTIS), 1985, University of Michigan (5);
- Truck Case-Matching Survey, 1984–1985, Insurance Institute for Highway Safety (IIHS) (6);
- Consolidated Freightways Single and Double Trailer Accident Study, 1978, John C. Glennon Chartered (unpublished data);
- Study of the safety experience of large trucks in Saskatchewan (7);
- Ontario's Commercial Accident Study Program, 1979–1981, Ministry of Transportation, Ontario (MTO) (unpublished data); and
- New York State Economic and Safety Consequences of Increased Truck Weight, 1987, Cornell University (J. Richardson, unpublished data).

Each study involved an ambitious amount of data collection, sometimes at considerable cost. For example, the two recent UMTRI data bases—NTTIS and TIFA—together have cost about \$1 million. Because only the IIHS and MTO studies

involved physical inspection, the level of information about installed equipment is, in general, of insufficient detail for present purposes.

A major additional concern is the size and duration of studies that would be necessary to track differences in accident rate attributable to equipment differences. On the basis of NTTIS and TIFA, an approximation of the level of truck use that might need to be monitored can be obtained to find sufficient accidents to start making such comparisons. For example, the overall accident rate for tractor combinations in the United States is about 240 police-reported accidents per 100 million km. These 240 accidents consist of approximately 155 property damage only, 80 injury, and perhaps four or five fatal involvements. If the trucks monitored averaged 100,000 km/year, one would have to monitor 1,000 trucks to expect 240 accidents. However, although trucks carrying general freight in the United States average close to 160,000 km/year, the overall average for tractors is about half, or 80,000 km, meaning that about 1,250 randomly selected U.S. trucks would have to be monitored to obtain those 240 accidents. If the notion of case and control groups whose rates are to be compared is now introduced, it is easy to demonstrate that only large differences in rates could be detected with samples of a reasonable size over reasonable periods of time.

Important new insights into truck accident rates have been obtained from the studies reviewed. However, the studies also illustrate that the possibilities of developing methodologies to compare absolute accident rates are daunting in the case of rarely found equipment configurations, especially if their contribution to accident reduction turns out to be marginal.

### CHOICE OF METHODOLOGY FOR THE FRONT-AXLE BRAKE ISSUE

#### Consensus on Constraints

Substantial discussion has occurred about whether the collection of system-wide exposure data should be mandated. Much of the debate focuses on the interpretation of the results from the case-control methodology by IIHS, which had been used to avoid a mandate in the state of Washington. In particular, results on relative accident involvement by various factors such as truck configuration seemed to differ markedly from similar comparisons from the UMTRI studies. Although there were no results of direct relevance to the equipment-configuration comparisons of interest in the current study, the same disparity could be expected in those areas as well.

A wide range of practical constraints on data collection in Canada were considered. Of major importance was the emergence of the commercial vehicle safety (CVSA) inspection capability in all provinces. It was generally seen as workable and appropriate for CVSA personnel, without police support, to stop and inspect trucks for control purposes. Serious reservations about fleet-based studies were expressed because of the high likelihood of selection bias. This view was not a criticism of the industry because selection bias can arise from fleet characteristics, such as the purchase of certain types of equipment for use with specialized loads or operations. A related matter was the possible disproportionate importance of small trucking operations with a wide diversity of man-

agement practices and other variations from equipment standards.

In focusing on the development on tractor-trailer front-axle brakes, it was agreed that any methodology must involve physical inspection of trucks in use. It was noted that power steering and limiting valves interact with front-axle brakes and therefore must be taken into account along with other factors, notably brake condition and adjustment, which can be verified only through physical inspection. However, it was recognized that a methodology involving physical inspection randomly distributed across the whole highway system would be likely to cost more than the implementation of the regulation itself.

Thus, the methodological work was constrained to answering the major unresolved issue previously discussed—explaining the discrepancy between the relative accident involvement of differently equipped trucks, as previously estimated from system-wide studies, and those using case-control methods to collect data only at accident sites. In practical terms, this limitation meant developing a recommendation as to which of three candidate solutions was the most efficient at meeting the needs of the RIA on front-axle brakes—a fleet-matching study, a case-control study at accident sites, or a case-control study in which the exposure of controls is validated against data from vehicle inspections carried out randomly across the road system.

### The Rejected Options

#### *Option I: Fleet Matching Studies (Inter or Intra), with Tracking of Accident Rates*

On the basis of further discussions with the trucking industry, fleet following studies did not emerge as a viable alternative. In order to obtain statistically sound estimates of the benefits of mandating front-axle brakes within a time frame appropriate for the regulatory decision, it would be necessary to find a large number of trucks of similar vintage whose front-axle brake status could be established with certainty and guaranteed to be held constant, or at least subject to accurate tracking of changes in status during an extended period of data collection. Because this would involve checks on the installation of front-axle brakes and associated equipment, maintenance, and possible driver readjustment this task was seen as essentially impractical. In addition, many aspects of tracking and record keeping for such vehicles would be likely to interfere to some degree with trucking operations, perhaps even resulting in changes to the duty cycles of case and control vehicles. This option is potentially far more disruptive to the industry than various forms of random inspection on public roads, which would be necessary under the alternatives.

#### *Option II: Case-Control Study with Controls Drawn Network-Wide*

Exploration of this solution consumed a considerable amount of effort but was ultimately rejected, not because it was undesirable, but because it was a less efficient method of answering the regulatory issue than the option selected. This

stage of the methodological work included extensive discussions between statistical and data experts at Transport Canada, the consultant team, and the IIHS about the limitations of case-control methods.

The discrepancy between case-control and system-wide exposure surveys was found to be a question of weighting. In other words, if system-wide estimates of relative risk or accident rates are needed, it would be necessary to augment a case-control study with the random selection of controls across the system. However, in order to achieve these data in sufficient detail to weight all the equipment-configuration comparisons of interest for their system-wide exposure, large-scale sample surveys would be required that would be more elaborate than the multimillion dollar data collection efforts of UMTRI. In this case, the solution would amount to doing both a case-control study at the accident sites and the type of overall system exposure survey for many different cells of the truck population, which was rejected as too ambitious for RIA purposes alone.

System-wide truck accident rates are valuable; however, for the limited purpose of addressing the regulatory decision a less costly solution, which allowed a test of minimum economic benefit, was essential.

#### *Option III: The Recommended Option, a Case-Control Study at Accident Sites*

Therefore there were few alternatives to the case-control approach in which controls are matched to accidents at the accident sites. However, the use proposed for such a case-control approach is that of testing whether a front-axle brake regulation would pass a threshold of economic benefit.

The key conclusions about the case-control method can be stated in terms of limitations on the interpretation of results. Most of Transport Canada's concerns about the method related to the limited nature of the sample obtained when accident-involved trucks and matched controls are selected only at the accident sites, as is proposed here, and as was done in the IIHS-Washington state study. After addressing numerous examples, it was clear that such limited sampling could not be used to obtain truck accident and exposure rates for the road system as a whole, which can be restated as the corollary to the conclusion under Option II with regard to system-wide measurement of controls: The case-control method, confined to accident sites, is not capable of developing population estimates of the prevalence of no front brakes as a cause of accidents over the whole road system, nor can such a method provide accident rates per vehicle-kilometer (with or without front brakes).

Use of the case-control methodology at the accident sites makes possible the calculation of the relative odds that two subclasses of trucks distinguished by the absence or presence of front-axle brakes will be involved in an accident at those sites. Such odds have a particular statistical definition as coefficients in a logit-type regression model. Findings in this form for factors other than brakes have been published in road safety studies. However, as an article (6) on the IIHS-Washington state study points out, relative odds are not comparable to rates that take into account how much each subclass of trucks is used on the road system as a whole. Unfortunately,

results expressed as relative odds have been misinterpreted by some readers as absolute measures of performance in the system.

In the RIA context, the study proposed a limited objective for case-control at the accident sites, such that the interpretation of results is unambiguous. Rather than the publication of relative risk statistics, the proposed use of relative odds is to predict a change in accident frequency for the class of accidents sampled. This prediction would be achieved by using a regression model in which the coefficients have been changed to simulate the effect of universal fitment of front-axle brakes.

The principle is simple—to collect inspection data at accident sites in Canada and to estimate how many fewer or more accidents would likely occur with front-axle brakes installed on all tractors passing those sites. If a reduction in accidents is estimated and if that reduction expressed in dollar savings is sufficiently large to offset the costs of compliance with a universal front-axle brake standard, then the needs of the RIA would be satisfied.

For purposes of evaluating whether a regulation would have sufficient impact, the class of accidents sampled could be a major subset, rather than all heavy-truck accidents. For example, the analysis could apply to all road types, but a subset of accidents defined by road class might be sufficient for RIA purposes, provided that countervailing effects could not logically be expected on road types not sampled.

Enhancements to the case-control procedures, as used previously, are essential if the method is to be used in connection with an RIA on the front-axle brake regulation.

### A Statistical Introduction to Case-Control Methodology

Case-control methodologies have previously been applied in many medical research studies and in a limited number of applications in road safety research, including studies of pedestrian and truck accident causation. The basis of the case-control method is that a case (i.e., an observation of an item with the effect's under investigation being present) is first observed and then another observation is made on a control (i.e., a similar item but without the effect's under investigation being present). The only things allowed to vary between the case and control are those factors being specifically studied for their influence on the presence of the effect.

For example, consider a medical study in which the effect of smoking on the incidence of lung disease is to be studied. The first step would be to identify a sample of persons suffering from lung disease and then to record a number of characteristics for the people in this sample that are thought to be related to lung disease. The next step is to find people without lung disease who match the individuals in the sample with respect to all (or most) characteristics except for their smoking behavior. If successful in matching people on the basis of all the characteristics except smoking behavior, then a simple comparison of the proportion of smokers in the lung disease sample with the proportion of smokers in the non-lung-disease control sample will provide an estimate of the odds ratio for the effect of smoking on lung disease. If smoking does cause lung disease, then the odds ratio should be significantly greater than unity.

The same procedure can be applied to truck accidents and the effect of installing front brakes on the incidence of accidents. In this case, the first step would be to identify a sample of trucks involved in accidents and then to record a number of characteristics for the trucks in this sample that are thought to be related to the incidence of accidents. The next step is to find trucks not involved in accidents that match the trucks in the sample with respect to all (or most) characteristics except for installed front brakes. If successful in matching trucks on the basis of all the characteristics, except installed front brakes, then a simple comparison of the proportion of trucks with front brakes in the accident sample with the proportion of trucks with front brakes in the nonaccident control sample will provide an estimate of the odds ratio for the effect of installation of front brakes on accidents. If installation of front brakes does reduce accidents, then the odds ratio should be significantly less than unity.

Unfortunately, it is rarely possible to have such a closely matched control sample as described because normally more characteristics vary between the case and control sample than just the variable under study. In such situations, it is necessary to control for these other variables by means other than matching. The usual way to do this is to construct a multivariate statistical model of the accident causation process and by statistical inference estimate the likely contributory effects of each of the variables that has not been matched between the case and control samples.

In the truck accident analysis, it is assumed that the cases and controls are matched on the basis of site and time (month of year, day of week, and time of day) by selecting the control observations from the same road as the accident at the same time of day 1 week after the accident. Calculation of the odds ratios is based on the assumption that even though the case and controls are matched with respect to site and time the accident causation process is still a multivariate process. The accident is caused not only by the presence or absence of front brakes, but also by various design features, driver characteristics, and management factors (in addition to various other unspecified factors). Therefore, it is not sufficient to simply calculate naive odds ratios from the raw data but rather it is necessary to estimate a multivariate model of accident causation. The model most often used in this respect is the multivariate logistic regression [or multinomial logit] model. This model is similar in format to that used in many models of transportation demand, such as mode, route, and location choice. The basic format of the model is

$$p\{x\} = 1/\{1 + \exp [-(b_0 + b_1x_1 + \dots + b_nx_n)]\} \quad (1)$$

where

- $p\{x\}$  = probability of an accident's occurring given the set of variables  $\{x\}$ ,
- $\{x\} = \{x_1, x_2, \dots, x_n\}$ ,
- $b_i$  = parameter that estimates the effect of variable  $x_i$  on the probability of an accident, and
- $b_0$  = constant that accounts for the effects of variables that are not specified in the model.

In this simple model, only four types of variables, namely, front brakes ( $x_f$ ), vehicle design features ( $x_v$ ), driver characteristics ( $x_d$ ), and management factors ( $x_m$ ), can be reexpressed in the logit model.

$$p = 1/\{1 + \exp [-(b_0 + b_f x_f + b_v x_v + b_d x_d + b_m x_m)]\} \quad (2)$$

Rearrangement of the logit model results in the following format

$$\ln [p/(1 - p)] = b_0 + b_f x_f + b_v x_v + b_d x_d + b_m x_m \quad (3)$$

This form is similar to a conventional multiple regression model except that the dependent variable is the log of the odds (often termed logits). The transformation of the dependent variable is necessary to bound the probability to lie between zero and one and gives the familiar sigmoid curve for the logit regression model.

The  $b$  parameters in the logit model can be shown (8, p. 233) to be the logarithms of the multivariate odds ratios for each variable (assuming that the variable is dichotomous).

Although the logit model was initially formulated in the context of a cohort study (one in which elements in the population are observed over time to see whether a symptom appears, i.e., to see whether a control becomes a case), Schlesselman (8) observes that the method is just as appropriate in a case-control and a matched case-control study with the only difference being in the magnitude and interpretation of the constant term  $b_0$ . The interpretation of the  $b_i$  coefficients remains the same in each case.

With matched analysis of matched case-control data, each pair of data points is allowed to have an individual value of  $b_0$  (whereas in unmatched case-control the value of  $b_0$  remains the same over all data points). However, the values of  $b_i$  are the same as those obtained from unmatched case-control data.

It is also possible to perform unmatched analysis on matched case-control data (i.e., collect the data in a matched fashion, but perform the analysis as if the data were unmatched). Under these circumstances, two outcomes are possible with respect to the values of  $b_i$ . First, if the cases and controls have been matched on a variable that is associated with the study exposure then the estimates of the  $b_i$  values will be biased towards unity. On the other hand, if the matching is based on a variable that was not associated with the exposure, then the unmatched analysis would not bias the estimates of  $b_i$  and would, in fact, increase the precision of the  $b_i$  estimates. In the current study, it would appear that the matching variables (site and time) are unlikely to be strongly associated with the study exposure variable (the presence or absence of front brakes) and hence an unmatched analysis of the matched data would appear to be justifiable.

The estimation of the  $b_i$  coefficients is performed using maximum likelihood estimation (MLE) methods, with the usual tests of significance associated with MLE (such as likelihood ratio tests) being appropriate. As with all regression models, it is possible to enter transformations and interactions between independent variables into the model by means of specific transformations (e.g., powers of terms) and by multiplication of the independent variables to form a new variable.

In order to provide a secondary means of calculating the sampling error associated with the estimation of the  $b_i$  coefficients, it is possible to use replication methods. The simplest way is to randomly divide the cases into two independent

samples and then perform the logit model estimation independently for each sample. Variance in the parameter estimates obtained from the two samples can be used as confirmation of the parameter variances estimated as part of the MLE estimation procedure.

### The Need To Augment the Case-Control Method Previously Used for Truck Studies

In identifying enhancements to the case-control procedures, the point of reference was again the most relevant previous study, the IIHS-Washington state study. The objective has been to build on IIHS's experience, rather than to criticize their approach. With IIHS's cooperation, it has been possible to identify ways to make the results of a case-control study easier to validate and therefore more credible in an RIA context.

#### *Establish a Population of Accidents Suitable to the Regulatory Context, such that Comparably Defined and Disaggregated Secondary Data are Readily Available from Police Accident Records*

Unlike the IIHS study, which looked at a wide range of accident causation issues, a logic is needed to establish a sampling frame appropriate to the relatively narrow regulatory context. This requirement implies that sampling will not only respect technical requirements, but will also lend credibility to the RIA when it is subjected to political scrutiny.

The objective is to predict a change in accident frequency for a class of heavy-truck accidents that may be sufficient to offset the costs of a front brake regulation. This decision must initially take into account the researcher's judgment about the credibility of a result based, as was previously suggested, on a subset of accidents. There are two important dimensions to this.

First, can a regulation be justified if the cost can be shown to be offset by predicted accident savings on only that part of the system that has been measured? Could a result be defended, for example, only on the basis of a sample of fatal heavy truck accidents on freeways if enough potential savings can be shown there alone to pay for the costs of a regulation? If the answer is yes, it is essential to consider if there is any logic for a substantial reversal of results on any class of heavy truck accidents that have not been covered in the study. For example, will someone argue that front brakes are useful on freeways but a hazard on winding roads not built to freeway standards?

A related matter is the need for the sampling design to take into account the potential for costing the accidents. It might, for example, be efficient to stratify on accident severity to oversample injury accidents, which have higher average costs than property damage accidents.

No matter what the choice of sampling frame may be, it should be recognized that the method of matching controls to accidents is impractical in locations with low truck traffic volumes and some part of the system will inevitably be excluded. A strong case can also be made for excluding accidents on urban roads, other than freeways. Accidents on these roads

tend to be of lower severity. For example, in Ontario urban areas in 1986, 34 percent of accidents involving a heavy truck as the striking vehicle were classified as fatal or injury accidents. This figure compares to 38 percent for freeways and ramps and 46 percent for primary rural undivided roads (Transport Canada, unpublished data). Moreover, the logistics of investigating trucks on urban streets are daunting. Fortunately, it is difficult to imagine a logic that suggests that front-axle brakes are likely to lead to any sort of driving difficulty in urban areas and so the exclusion of urban accidents is unlikely to weaken an RIA that justifies a regulation based on a study of rural roads.

The second dimension involves asking if it is acceptable to extrapolate from a limited data set to a broader but supposedly equivalent situation. The most obvious example would be to justify the regulation by showing that the accident savings on a particular subset of accidents in only one or two provinces is enough to offset the costs of the regulation to the truck fleets in those same provinces and then to assume that the same holds true for the rest of Canada. Case-control methodology cannot be used to describe anything about classes of accidents not included in the sampling frame, but one may wish to assume that the same result would be obtained if the same sampling frame was used in a wider geographical area. Substantial survey cost and operational advantages are obtained by making such an assumption.

The specification of which accidents are to be considered cases must take into account the level of detail in police accident records because the extrapolation will involve calculating a hypothetical change in accident frequency from a baseline that is provided by police records for the total population. The baseline must be available using accident classifications and a level of disaggregation that are comparable to those used for selecting case accidents in the study. For example, it is not possible to extrapolate findings from a case-control study confined to tractor-trailers on limited-access highways if police data lack road class and truck type as accident descriptors. In some types of surveys, it is possible to use sample data to estimate the size of the population when an independent source is unavailable or incomplete. The manner in which the sample is drawn in the present study will not permit such an approach.

*For the Selection of Control Vehicles, Set Up a Method of Obtaining Classified Traffic Count Data at the Control Site, to Ensure that the Selection of Control Vehicles is Truly Random*

Certain additional conditions that were not met are needed for the selection of control trucks in the IIHS-Washington state study. Most important is that a classified traffic count be taken at the time that control trucks are selected and examined. The objective of this count is to ensure that the controls are representative of the total fleet passing the accident sites, at least with respect to observable characteristics. The selection of control trucks by inspectors is always open to the criticism of conscious or unconscious selection bias. For example, it might be that inspectors will tend to do what they are normally required to do—select vehicles in apparently questionable condition. It might be equally true that for

the extra inspections carried out as part of a study some inspectors might select trucks that can be inspected quickly, incurring less objection from the drivers whose trips are interrupted. This selection bias could extend to a tendency to choose or avoid vehicles of certain trucking companies, possibly introducing a selection bias on some nonvisible characteristic that differs by company. If this problem is suspected, and certain companies have sufficient presence in the accident site area, the trucking company could be observed as a count variable.

Even if this count significantly increases the labor cost of the method, it is essential to address perhaps the most serious shortcoming of case-control methodology as it has been applied in the past. Costs may be minimized by using a video camera to record passing traffic. Even if selection bias is shown to have occurred, by comparing the control vehicles to the classified traffic count data obtained at the control sites, the count data could be used to weight the control data so that the composition of the sample of controls matches the composition of the classified counts.

*Use Trace-Back Procedures for Accident-Involved and Control Trucks*

Procedures may be added to trace accident-involved and control trucks back to their operating companies to include fleet management factors in the study. Trace-back procedures relate in particular to the question of maintenance and company policies regarding front-axle brakes.

*Augment the Analysis of Case-Control Data*

Finally, a number of methods are suggested to improve the analysis of data and to estimate the potential impact of the regulation. These methods include performing a secondary estimate of sampling error using replicates and the use of Monte Carlo methods to provide a distribution of accident reductions, which can then be used in a probabilistic economic analysis. These enhancements were subjected to a pilot test in the form of a computer simulation.

**Cost and Feasibility of a Case-Control Study of Front-Axle Brakes in Canada**

Field requirements and costs of a case-control study of front-axle brakes in Canada were verified in a series of interviews with provincial highway departments, police authorities, trucking companies, and the parties involved in the IIHS-Washington state study—IIHS and the Washington State Commercial Vehicle Inspection Division. Richardson and Campbell (9) discuss data collection requirements; experience from previous studies involving roadside truck inspections; the roles of the police, the investigation team, and the provincial vehicle inspectors; and the potential response of contracted mechanic teams.

Such a case-control study would be feasible in Canada, but the costs would be high. A spreadsheet model was developed that could be used to examine, interactively, various decisions

about acceptable statistical error, the expected impact of the regulation, and the number of field teams desired. Two examples of output from the model using the outer limits of reasonable assumptions estimated the likely cost in a range of \$700,000 to \$2,750,000 for studies capable of detecting 40 or 15 percent reductions, respectively, in accidents attributable to the regulation. Almost all of the difference between these estimates resulted from the large increase in cases necessary to detect the smaller reduction—from 383 to 2,314, respectively. Use of provincial vehicle inspectors rather than contracted mechanic teams could reduce these estimates by 6 to 10 percent.

## A COMPUTER SIMULATION TO TEST FOR BIAS IN THE CASE-CONTROL METHOD

### Approach

One of the nagging questions about the case-control methodology is whether the method will provide reliable and unbiased estimates of the odds ratios given that the sample is biased toward accidents and away from the majority of miles traveled without an accident. Put simply, how can one expect to get good estimates of the risk associated with various design features when the data consist of some accidents and a couple of observations that have been matched to each accident?

Although Schlesselman (8) and Manski and Lerman (10) noted that the variable coefficients obtained from a case-control or choice-based sample are reliable and unbiased estimates of those that would be obtained from a full random sample, it was felt desirable to empirically demonstrate the validity of this claim in the context of a case-control accident study. Therefore, the purpose of this analysis was to test the application of case-control methodology to the estimation of reductions in truck accidents following the implementation of a vehicle design feature such as the installation of front brakes.

The analysis was based on a simulation modeling method, wherein a population of accidents was first generated on the basis of an assumed causal model of accident causation. Controls were then selected and the MLE method used to reestimate the underlying (known) causal model. The degree to which the original model coefficients could be reestimated from the simulated data set and the sensitivity of the estimated coefficients to the number of controls selected per accident were primary considerations in assessing the viability of the case-control method for estimating causal accident models.

Because of the nature of simulation, repeated applications of this technique would generate numerically different (but statistically similar) data sets. Therefore, the model coefficients estimated by MLE would not necessarily agree exactly with those of the original model. However, repeated application of the simulation model would generate a distribution of model coefficients and these distributions should not be significantly different from the original model coefficients (if the premise is correct). In addition, if the case-control method yields the same results as from a full random sample, then the results should be independent of the number of controls observed per case (with a full random sample simply being a large number of controls per case).

## Stages in the Simulation

### Stage 1

For the purpose of this simulation, a simple causal model was assumed in which the probability of an accident depended on only two independent variables (*viz.*, the presence or absence of front brakes and the age of the vehicle). One of these variables is discrete whereas the other is continuous. In a more comprehensive analysis, other variables, such as the age of the driver, the number of hours the driver had been driving at the time of the accident, the size of the company operating the truck, the type of road, and the time of day, could be included in the causal model. It was postulated that the probability of an accident's occurring to any particular truck passing a site on the road network is given by the following logit model:

$$p = 1/\{1 + \exp [-(b_0 + b_f X_f + b_v X_v)]\} \quad (4)$$

where

- $p$  = probability of an accident's occurring,
- $b_i$  = coefficient associated with the variables  $X_i$ ,
- $X_f$  = dichotomous variable for the presence of front brakes, and
- $X_v$  = vehicle age in years.

### Stage 2

Coefficients were then selected with plausible signs and with magnitudes such that reasonable estimates of probabilities and changes in probabilities were obtained. A total sample size of approximately 200 accidents was seen as being a feasible and realistic objective. On the basis of these considerations and on the total population of site-time combinations described, the following coefficients were adopted:  $b_0 = -7.2$ ,  $b_f = -0.295$ , and  $b_v = 0.023$ .

### Stage 3

A population of sites was then constructed on an assumed road network with a relatively realistic composition of road type, geometry, and time of day. For example, it was assumed that there were 500 locations, with 200 divided and 300 undivided sites. For divided road sites, it was assumed that 30 percent of these sites were on curves, whereas for undivided road sites it was assumed that 50 percent were on curves. For each site, 12 hr of daytime flows and 12 hr of night time flows were later generated. These assumptions yielded a total of 12,000 site-time combinations.

### Stage 4

At each site-time combination, total hourly truck flows were generated. Time-of-day flow profile was assumed such that between 1 and 8 percent of the day's traffic was observed in each 1-hr period of the day. On divided roads, an average daily flow of 1,000 veh/day was assumed, whereas on undi-

vided roads, an average daily flow of 200 veh/day was assumed. A normal distribution of flows with a coefficient of variation equal to 20 percent of the mean for divided road sites and 30 percent of the mean for undivided road sites was assumed. From these distributions, an expected hourly flowrate was generated for each of the 12,000 site-time combinations.

#### Stage 5

For each hourly flow, average parameters were generated for each of the other variables. Average percentage of trucks with front brakes was set to 60 percent, but the average age of the vehicle was set to be higher on undivided roads (10 years during the day and 11 years at night) than on divided roads (8 years during the day and 7 years at night). It was assumed, for simplicity, that all variables are independently and normally distributed with a coefficient of variation equal to 20 percent of the mean value.

#### Stage 6

For all the 12,000 site-time combinations, a program was written to set up a simulated data matrix for the variables road type, geometry, time, flow, percent with brakes, and vehicle age.

#### Stage 7

For the first site-time combination, the characteristics of each of the vehicles (trucks) passing that site were generated using the normal distribution and coefficient of variation equal to 40 percent of the mean.

#### Stage 8

For each truck passing the site, the probability of an accident was calculated using the causal model and coefficients specified in the initial stage. Then, applying Monte Carlo techniques, a uniform random deviate (between 0 and 1) was generated for each truck. Whether or not an accident occurred was determined by comparing the random deviate with the probability of an accident. If the random deviate was smaller than the probability, then an accident was deemed to have occurred. If an accident occurred at this site-time combination, then the details of this accident vehicle were saved in a separate accident matrix.

#### Stage 9

For each accident included in the sample at this site-time combination, a set of three control vehicles passing that site was generated (as outlined in Stage 7). The population from which these controls were selected corresponds to the flow of vehicles that would have passed the same site 1 week after the accident. The details of these control vehicles were also saved in the accident data matrix. Simulation Stages 7, 8, and 9 were then repeated for all 12,000 site-time combinations.

#### Stage 10

At the end of this procedure, the simulated accident data set now represented the information that would have been obtained had a real case-control survey been conducted. With these data, it was now possible to estimate the accident causation logit model and reestimate the coefficients using MLE.

On completion of Stages 1 to 10, the estimated model was compared with the known causal model developed in Stages 1 and 2. More important, it was also possible to determine whether the estimated model coefficients were affected by the number of controls selected, and hence, by logical extension whether the case-control method itself was able to generate data that provided a means of estimating unbiased coefficients for the accident causation model on the basis of a full random sample of truck travel exposure.

At this point in the analysis, had this been a real study, the next stage would have been to estimate changes in accident probabilities. However, for regulatory purposes the change in the number of accidents would be calculated and the net economic benefits would be estimated. In order to do this, the simulation was extended to demonstrate an approach in three final stages.

#### Stage 11

For each observation (of either a case or a control) in the sample, the front brakes variable was changed to reflect installation of front brakes on the entire fleet by changing all occurrences of a zero for the front brakes variable ( $X_f$ ) to a value of one. The probability of an accident's occurring under these conditions was then calculated by application of the accident causation model estimated in Stages 1 through 10.

#### Stage 12

Given the new probabilities of accident occurrence for each observation, the occurrence of an accident under existing and projected conditions was then simulated using the Monte Carlo technique described in Stage 8. Number of accidents was then summed for existing and projected conditions, and the difference in these summations was an estimate of the reduction in the number of accidents in this sample brought about by the installation of front brakes. The repetition of this stage with a different set of random numbers would result in a different estimate in the number of accidents saved by the installation of front brakes. A full analysis would require the estimation of a distribution in the number of accidents saved, expressed as range of outcomes. (For example, in the imaginary sample of 200 accidents, the average reduction in accidents over 100 runs of the simulation was 11.41, with a standard deviation of 3.52).

#### Stage 13

The change in the number of accidents could then be assigned an economic value (based on standard accident costs) and compared with the cost of retrofitting front brakes to the fleet to determine the economic viability of the retrofit policy.

TABLE 2 COEFFICIENTS DERIVED FROM THE SIMULATION

	Coefficient	S.E.
Average $B_0$	-0.826	0.222
Average $B_1$	-0.372	0.179
Average $B_2$	0.005	0.023

### Results of the Analysis

As a result of the simulation, a data set of approximately 800 observations (200 accidents and 600 controls) was generated. This data set was then read into a data file using the SYSTAT statistical package on an Apple Macintosh personal computer. The nonlinear regression module of SYSTAT was then used to obtain MLE of the coefficients in the logit model of accident causation.

The output from the analysis is a set of coefficients and accompanying standard errors for the underlying logit model. This analysis was repeated four times with independent data sets to obtain an empirical estimate of the variability of the estimated coefficients. The results of the analysis are presented in Table 2.

Average values for the coefficients  $B_1$  and  $B_2$  (for front brakes and vehicle age, respectively) are not in total agreement with the expected values of  $-0.295$  and  $0.023$ . However, their signs are correct and they are within one standard error of the expected value. Therefore, it cannot be rejected that the case-control method and the MLE did in fact succeed in reestimating the coefficients in the underlying causal model. The constant term ( $-0.0826$ ) was not close to the constant term in the underlying model ( $-7.2$ ), but this discrepancy was to be expected because of the higher proportion of accidents in the case-control data set than in the total population of site-time combinations. Naturally, the estimated coefficients could be made more precise by use of a larger data set, particularly one containing more accidents. However, in the current study, the data set was limited by the capabilities of the statistical package in use.

Although the direct comparison of estimated and expected coefficients gives some indication that the case-control meth-

odology gives unbiased estimates of the coefficients, a further test would involve experimenting with the number of controls. In this study, the number of controls was reduced to two and then one by progressively eliminating one or two controls per case from the existing data sets generated in the four runs. For the two-control situation, the third control was eliminated from each case, whereas for the one-control situation both the second and third controls were eliminated. Model coefficients were then estimated with the results presented in Table 3.

Table 3 indicates that as the number of controls is decreased, coefficients  $B_1$  and  $B_2$  fluctuate, but this fluctuation is not systematic. On the other hand, the constant term becomes more positive as the number of controls is decreased.

The fluctuation in  $B_1$  and  $B_2$  may be caused by two sources—either the number of the controls per se, or the composition of the total sample after removal of the controls. Although all the data sets (within one run) are based on the same total set of three controls per accident, the manner in which the controls are removed creates the possibility of creating essentially different data sets when the controls are removed. In order to overcome this possibility, the analysis for the fourth run was redone but with the controls being removed in a systematic fashion. Thus, three two-control data sets were built by removing, in turn, the first, second, and third control in each case. Similarly, three one-control data sets were constructed by including only the first, second, and third controls in each case. In this way, the average of the two-control cases more closely represents the three-control case (because each control is represented a total of two times in the three data sets), whereas in the one-control case each control is represented a total of one time in the three data sets. The results of this analysis are presented in Table 4.

The change in the number of controls has absolutely no effect on the estimation of the coefficients  $B_1$  and  $B_2$ , whereas the constant term  $B_0$  becomes more positive as the number of controls decreases. This finding is consistent with the findings from other areas of transportation research, such as mode choice modeling, in which it has been found that a logit model that is calibrated on a choice-based sample (equivalent to the case-control methodology) will provide unbiased estimators for all coefficients except the alternative-specific constants

TABLE 3 THE EFFECT OF REMOVING ONE, THEN TWO CONTROLS

	<i>Three Controls</i>		<i>Two Controls</i>		<i>One Control</i>	
	Coefficient	S.E.	Coefficient	S.E.	Coefficient	S.E.
Average $B_0$	-0.826	0.222	-0.605	0.233	0.219	0.257
Average $B_1$	-0.372	0.179	-0.309	0.181	-0.405	0.249
Average $B_2$	0.005	0.023	0.008	0.025	0.001	0.027

TABLE 4 AVERAGED COEFFICIENTS AFTER SYSTEMATIC REMOVAL OF CONTROLS

Run #4	<i>Three Controls</i>		<i>Two Controls</i>		<i>One Control</i>	
	Coefficient	S.E.	Coefficient	S.E.	Coefficient	S.E.
Average $B_0$	-1.061	0.266	-0.654	0.256	0.041	0.283
Average $B_1$	-0.219	0.255	-0.219	0.201	-0.219	0.294
Average $B_2$	0.011	0.022	0.011	0.030	0.011	0.011

(10) and that the constant can be corrected by means of a factor relating the market share in the full sample to the market share in the choice-based sample.

### Conclusions from the Simulation

Implications of the simulation study previously described include

1. Case-control methodology can provide data to obtain unbiased estimates of the coefficients in a causal accident model of the logit type, which will be the same as those estimated from a full random sample of travel exposure;

2. The precision of these estimates will be affected by the total sample size. Because of the complexity in the sampling and estimation procedures, it is recommended that replication methods be used to empirically estimate the variances of the estimated coefficients; and

3. A simulation method can be used on the total sample of cases and controls to estimate the reduction in accident numbers resulting from the installation of front brakes to all trucks in the sample. Monte Carlo methods provide a distribution of accident reductions, which can then be used in a probabilistic economic analysis.

### CONCLUSION

This study focused on identifying ways of predicting the change in accident involvement of heavy trucks that would result from the mandatory installation of tractor-trailer front-axle brakes. Initially, it was hoped that any data collection necessary for such predictions would also yield, as a by-product, some broad measures of system safety relevant to trucking operations. Early in the study, it became apparent that such a by-product would be prohibitively expensive and the study was directed toward identifying the most efficient approach to evaluating the regulation without measuring accident rates on the system as a whole.

As a result, the methodology proposed has been developed to answer only this limited question—will the benefits of a proposed mandatory front-axle brake regulation be likely to outweigh the costs entailed? After extensive checking, it was concluded that a modified application of the methodology known as case-control provides the most efficient method of answering that question.

Use of case-control methodology in road safety evaluation has generated considerable controversy in recent years. However, the usefulness of the method has been somewhat obscured during a period of rightful questioning of the interpretations, by some readers, of the results from earlier case-control studies.

This study is not intended to rekindle the controversy. Much of the debate has, unfortunately, hung up on the meaning of the relative involvement ratios that are derived as coefficients in a logistic regression. In the context of equipment regulation, methodology is much more useful if these coefficients are not treated as end results, but rather are used to predict changes in accident frequencies under defined sets of conditions.

That choice-based samples and logit models work in this application is not surprising. The problem is not unlike some

of those in travel demand modeling that have long been using such methods. However, the strength of the previous controversy compels a restatement that this methodology is not a potential source of system-wide heavy-truck accident rates. In particular, it should be noted that the methodology is not designed to yield accident rates per vehicle-kilometer for trucks with different equipment configurations.

### ACKNOWLEDGMENTS

This study was sponsored by Transport Canada. The consultant team included Tony Richardson and Ken Campbell and was led by Martin Lee-Gosselin and Gord Taylor. The thoughtful contributions of Eric Welbourne and Winson Ng, John Lawson, Hans Arora, and Jim White (all of Transport Canada) are gratefully acknowledged, as is the assistance of the Insurance Institute for Highway Safety, especially Ian Jones, in enabling Canada to benefit from the Institute's experience in collecting data on heavy-truck accident involvement. Numerous other experts from the trucking industry, provincial and state police and highway authorities, the Royal Canadian Mounted Police, and several truck repair companies helped the study team assess the feasibility of data collection.

### REFERENCES

1. Statistics Canada. *Provincial Truck Fleet Study—1986*. Ottawa, 1988.
2. *Study of Utilization of Front Brakes and Automatic Slack Adjusters on Heavy-Duty Trucks*. Transport Canada/TES, Quebec, 1986.
3. UMTRI Transportation Data Center. *TIUS Codebooks and Other Documentation Corresponding to the 1982 and Earlier Archived Datasets*. Transportation Research Institute, University of Michigan, Ann Arbor, 1984.
4. K. L. Campbell et al. *Fleet Accident Evaluation of U.S. FMVSS 121*. Final Report UMTRI-79-79, Transportation Research Institute, University of Michigan, Ann Arbor, Aug. 1981.
5. K. L. Campbell, D. F. Blower, G. R. Guy, and A. C. Wolfe. *Analysis of Accident Rates of Heavy-Duty Vehicles*. Final Report UMTRI-88-17, Transportation Research Institute, University of Michigan, Ann Arbor, April 1988.
6. H. S. Stein and I. S. Jones. Crash Involvement of Large Trucks by Configuration: a Case-Control Study. *American Journal of Public Health*, Vol. 78, No. 5, May 1986.
7. G. Sparks et al. *The Safety Experience of Large Trucks in Saskatchewan*. Report to Saskatchewan Highways and Transportation and Transport Canada, 1988.
8. J. J. Schlesselman. *Case-Control Studies: Design, Conduct and Analysis*. Oxford University Press, New York, 1982.
9. A. J. Richardson and K. J. Campbell. *Methodology for Estimating the Accident Involvement of Heavy Duty Trucks After a Change in Equipment Regulations*. Transport Canada, Québec, Feb. 1989.
10. C. F. Manski and S. R. Lerman. The Estimation of Choice Probabilities from Choice-Based Samples. *Econometrica*, Vol. 45, No. 8, 1977–1978.

---

*The views expressed in this paper do not necessarily represent those of the sponsors or of the individuals and organizations who have been kind enough to furnish information.*

*Publication of this paper sponsored by Committee on Traffic Records and Accident Analysis.*