1-1-2003

A database development and analysis of selected side impact collisions in Toronto

John Zaki Bou-Younes
Ryerson University

A DATABASE DEVELOPMENT AND ANALYSIS OF SELECTED SIDE IMPACT COLLISIONS IN TORONTO

By

John Zaki Bou-Younes

Ryerson University, Bachelor of Engineering , Toronto, 2001

A thesis

presented to Ryerson University

in partial fulfillment of the

requirements for the degree of

Master of Applied Science

in the program of

Civil Engineering

Toronto, Ontario, Canada, 2003

UMI Number: EC52878

INFORMATION TO USERS

Ryerson University requires the signatures of all persons using or photocopying this thesis.

Please sign below, and give address and date.

Date	Name	Address	Signature

John Zaki Bou-Younes

A Database Development and Analysis of Selected Side Impact Collisions In Toronto

Ryerson University, Master of Applied Science, 2003

Abstract

This thesis is based on the initial phase of a project that developed an in-depth collision database and performed an analysis of police reported side-impact collisions for City of Toronto intersections between 1998 and 2000. Currently, collision data exists through several different sources in Ontario. The development of a database involving the amalgamation of collision forms, the selection of data fields, and the collection of real collision data from selected, thoroughly investigated side impact collisions involving late model vehicles (1998 and newer), is described. For analysis, Statistical Analysis Software Release 8.02 was used to investigate causation and casual factors of side impact collisions. Statistically significant collision factors determined by fault propensity included apparent driver action, driver age, front seat passenger age, maximum posted speed, approximate vehicle speed, road character, and number of lanes. For intersection collision propensity, statistically significant findings included the system used, presence of flashing signals, intersection legs, roadway volume, and intersection leg road classifications. It is anticipated that the findings from this analysis can provide insight into significant factors in side-impact collisions that will be applied with greater focus to the in-depth collision database, once developed.

Acknowledgements

This study could not have been completed without the generous assistance of my Supervisor and Professor, Dr. Bhagwant Persaud, Ryerson University. Dan Mills, Dr. Chapman and Ian Hale of Ryerson's Vehicle Safety Research Centre are also gratefully acknowledged. The guidance in database development and analysis ideas of Professor Mary Chipman, University of Toronto, and Vartouhi Jazmaji, University of Toronto's Research Services Unit is also graciously appreciated. Remembrance goes to Jim Smith, supervisor for the City of Toronto Traffic Data Centre and Safety Bureau, for his kind assistance and training when using city resources to access numerous MVA reports. I would like to acknowledge Tom Brazier, MTO, Freedom of Information Special Requests, for his assistance in obtaining a portion of data for my research. Thanks to the Network Centres of Excellence and Auto 21, from which I received partial funding. Finally, special thanks to my family, Yoshiko Kaji, Lang'o Odondi and friends for their constant support, understanding and encouragement.

Table of Contents

List of Tables

List of Figures

1 Introduction

1.1 General

This is a thesis based on a portion of an ongoing research project involving a multidisciplinary approach to side-impact collision investigation. The main objective or underlying drive for that research project stems from a statement taken from the U.S. Insurance Institute for Highway Safety (IIHS) [1], that involved the following question and answer:

> **I'm shopping for a new car. Why can't I find driver death rate or insurance loss data for new models?** It takes considerable time to gather and tabulate the real-world data needed to provide statistically significant results for new models. Complete vehicle registration data for each model year typically are released about two years later, and data on fatalities are first available approximately nine months after the end of the calendar year. Similarly, it takes time to amass sufficient insurance claims information to provide meaningful results for a range of vehicles. For vehicles that have not been fundamentally redesigned, previous model year results are good predictors of the current model's experience.

Hence, a case has been made for the need for up-to-date collision data based on new vehicle models, as it comprises an important component for driver death rate and insurance loss data and may be a deciding factor in the purchase of a new or late model vehicle. Furthermore, current late model vehicle safety data are based on mainly two sources: crash vehicle laboratory testing, and databases containing very recent collision records. The pros and cons of laboratory testing are not discussed in this paper. However, limitations exist when using laboratory sanitized conditions to determine factors involved in motor vehicle collisions. Very recent collision data on late model vehicles published by the IIHS has created a perception in the general public that certain models are safer than others, when tables published for various categories of passenger vehicles are compared. Table 1.1 below includes selected excerpts from their online website where particular recent models are tallied with their respective driver death rates.

Model	Driver Death Rates Per Million Registered Vehicle Years
Toyota Camry	37
Volvo 850	39
Mazda MX-6	101
Pontiac Sunfire	206
Ford Explorer	231
Chevrolet Camaro	308

Table 1.1: Driver Death Rates by Model

These rates may make it all too easy for the general public to believe particular models are superior to others. While it may be true that particular vehicle models are dangerously over-powered, have insufficient brakes, terribly high centre of gravities, poor construction, etc., certain figures may cause readers to believe that some models are in fact "dangerous", or perhaps the opposite, giving them a false sense of security. This is particularly the case because of the lack of normalization for exposure and other measures required in making more legitimate and fair comparisons. Normalizing for Driver Death Rates for Million Registered Vehicle Years, as they have, does not necessarily account for the use of these vehicles (i.e., high use vehicles such as taxis, fleet vehicles, as compared to personal vehicles), nor for driver types (i.e. elderly vs young) and a host of other confounding factors. Clearly, there is more to this than just vehicle driver death rates.

It is commonly accepted amongst transportation engineers that there are three main types factors involved in road safety and collision investigation:

- Vehicle factors, as mention above;
- Human (Driver) factors; and
- Environmental factors.

The objective of human factors in road safety is primarily to improve safety through the understanding of the link between humans, vehicles and their environment. Driver workload, fatigue, reaction time are just a few classic examples of studies repeated in search of this link. Humans have capabilities, but they are limited. In fact, some may argue that researchers are most interested in human limitations rather than capabilities. Regardless, human physical, perceptual and cognitive capabilities are significant since all vehicles require a human driver who is in control. One study performed in 1977 by Treat et al determined that human error contributed to 90% of a subset of collisions This supports a commonly accepted notion that "bad drivers" cause collisions.

Environmental factors include external or extrinsic physical conditions that affect and influence the operation of a motor vehicle. The conditions that vehicles are operated in can significantly contribute to the occurrence of vehicle collisions. Several of these factors can lie outside the direct control of a driver, transportation planner, or law-maker (i.e., illumination, road surface, weather conditions, etc.,) while others (i.e., lane widths, signal timing, roadside furniture, etc,.) can lie within this control. Typically, only experts in each field can make contributions that increase efficiency and improve upon the safety aspects of environmental factors.

Managing these three aspects are not typically easily accomplished by a transportation engineer, or for that matter, by any individual professional. Yet, in order to maximize effectiveness, simultaneous consideration of these three factors is required. The solution lies in the combined efforts of multiple professionals, where factors within vehicle, driver and environment can be tackled separately, with appropriate expertise. The concept of a multi-disciplinary approach to collision investigation is only recently emerging, and a portion of this report includes recent efforts made as part of an on-going investigation in side impact collisions. The analysis of these three types of factors depends on collision data. *Recent* collision data are always in demand. Findings based on outdated collision data may be just that - outdated. This report shall illustrate sources, types and methods for analysis of side impact collisions in Toronto, its usefulness as part of a linked data set for analysis, and its analysis to bring out factors associated with these collisions with respect to driver, environmental and vehicle attributes.

Collisions have traditionally been classified in many ways. Side impact collisions are particularly important because they are unique. Unlike other collisions such as frontal and rear, occupants in target vehicles have very little protection. In fact, less than eight (8) centimeters typically separate the occupant on the side that is struck from the striking vehicle. There are essentially no bumpers, engines, etc, to help absorb the energy of the impact. Therefore, these types of impacts can cause severe injury; especially to the occupant on the side the vehicle was struck [2].

In 2001, 39.2% of all Ontario collisions accounted for occurrences of side impact collisions (where side impact collisions are identified by police as: Angle, Sideswipe, and Turning Movement) [3]. However, this number may be conservative, as many collisions that fall outside of these categories may still be considered side impact collisions (i.e., when motor vehicles slide into roadside furniture in a sideways manner).

1.2 Thesis Arrangement

This thesis is divided into six chapters. Following the Introduction, Chapter 2 summarizes the literature reviewed on recent and emerging analyses concerning similar collision data. This chapter also includes analyses from different studies using different data for the purpose of establishing that these techniques indeed have competent applications to the data in this study.

Chapter 3 describes the data used in this study, challenges experienced in managing it and its associations to the research project this report stems from. Chapter 4 provides methodology and analysis selected or developed for this study. Chapter 5 provides results of analyses and interpretations. Chapter 6 provides conclusions and recommendations based on this report.

2 Literature Review

2.1 General

Through the many resources afforded by on line catalogues at Ryerson University and The University of Toronto, a literature review of past, present and emerging studies on the analysis and identification of factors involved in side impact collisions is provided in this chapter.

2.2 Literature on Methods of Collision Assessment

A recent paper by Persaud et al in 2003 reviewed various methods in isolating causes of collisions and assessing effects of treatments, using both routinely available and specifically collected databases. Their review breaks down road safety studies as falling under essentially two categories: Experimental Studies and Observational Studies. Of particular interest is the latter type, where methodology of past and emerging cohort and case-control studies are explored. Cohort studies are described as the identification of two otherwise comparable groups, or cohorts, (alike enough to be compared, at least after controlling for measured confounding variables), that differ by some variable of interest (e.g., exposure to a particular program or a particular engineering feature) and involves following them over time to asses differences in consequences (e.g., their collision record) [4].

The above research paper is based on the case-control study type, used in two separate analyses. This type of study differs slightly from cohort studies in that case-control studies have a known outcome where one group, the case group, is attributed with this known outcome, and the other group, the control, is not attributed with this outcome. Persaud et al (2003) noted that in many transportation safety studies, the estimate of safety effect can be difficult to interpret, because it is usually not possible to statistically control for all confounders such as speeds, sight distance, and geometry [4]. Furthermore, case control and cohort studies are subject to selection biases because sample entities may differ in unacknowledged but nonetheless important ways from the population.

Findings from their paper expose advantages of case-control analysis as a preliminary step in investigating collision safety data. Perhaps the most notable suggestion involves the use of a two part analysis with a case-control preliminary analysis followed by a cohort study that avoids the bias commonly found in a sole case-control analysis. An example of how a case control study could be expected to be constructed following their suggestions can be seen in Figure 2.1.

Example of Applicable Two Step Analysis Recommended

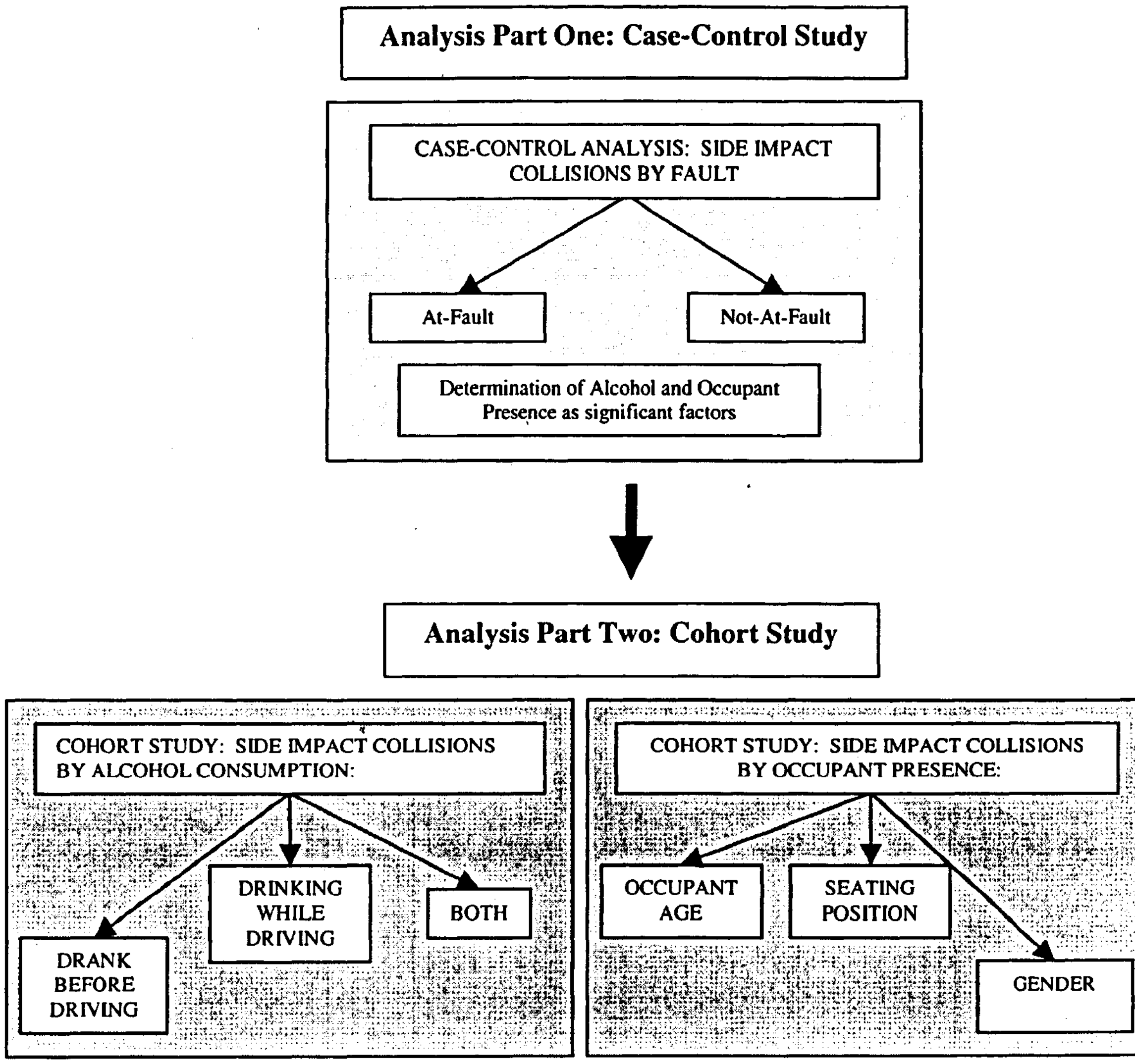


Figure 2.1: Two Part Analysis

The example outlines the use of a case-control and cohort two-part analysis, where the initial case-control analysis is used to identify collision causative factors. The secondary analysis that follows can be used to focus on cohorts targeting specific factors, traditionally, casual factors acting as exposure measures. In the example, alcohol involvement and occupant presence are used and the cohort study would also employ many other factors available in the data, including outcome variables. The reliance on case-control studies alone to determine these relationships expose problems of bias that are much less common in cohort studies, [4] and using both types of analysis, as shown above, overcomes this issue. Furthermore, cohort studies have the advantage of generally being prospective in design- the outcome is not known at the beginning of the study and therefore is less likely to influence the collection of data-and of better collection of data to ensure comparability [5].

2.3 Literature for Side Impact and Left Turning Collision Data Analysis

Past research shows that there is a higher risk involved in making left turning movements at intersections as compared to other intersection movements. The reasons for this include crossing of two directions of traffic and estimating speeds of oncoming traffic [6]. Researchers such as Kirk et al [6] have examined this movement with Kentucky collision data and found it possible to define factors in the driving characteristics of people such as age, gender, occupancy rates, and period of the day. A trend of decreasing crash involvement while the driver ages, was found. Within this trend, a sharp decrease was noted between the ages of 16 and 17, while after the age of 17, this decrease dramatically slowed. It was concluded that crash likelihood decreases rapidly within the first few years of driving. Findings from their study also found no statistical differences:

- between genders for Left Turing Collisions;

- for period of day (day or night); and,
- between occupancy rate.

In addition, the authors note the disagreements between researchers in determining ideal exposure methods for each application, as traditional methods are based on estimating the amount of vehicle-miles traveled (VMT) by simply multiplying the average daily traffic (ADT) with the length of roadway [6]. A quasi-induced exposure is used to avoid limitations existent in estimations of driver exposure from these and other exogenous values such as travel distance, drivers licensed, and vehicles registered. The data itself are used to derive exposure estimates [8]. For instance, it is known that certain driver groups modify their driving pattern towards proportionately more exposure to low-speed and daylight situations and proportionally less exposure to high speed and night situations. Exposure based solely on total distance driven does not account for these changes in driving patterns [7].

Furthermore, these methods have been validated against more conventional techniques [8] and are becoming common practice. Surrogates for vehicle-distance of travel by different classifications of road users are acceptable, granted that an assumption is made bearing a consistent composition of road users.

The format for one such analysis uses relative collision involvement rates (RAIR) [8]. RAIR are calculated by taking the ratio of the percentage of at-fault drivers of one group within a given category, to the percentage of not-at-fault drivers from another category within the same group. It is considered a form of quasi-induced exposure since the distributions of both types of drivers is a representative sample of the same exposures of all drivers in the group.

Of further interest with this method is the necessity to define at-fault and not at-fault drivers. The researchers assigned fault in a crash from information contained in the accident database, where available. This relied on a human factor category that indicated each driver's contribution to the collision. Cases where fault was indeterminate or otherwise when both drivers were simultaneously at fault or not-at-fault were excluded for this study.

Factors selected in their study included age, gender, period of day, and number of occupants. Examination of these factors both individually and in simultaneous combinations were used to focus on particular categories and conditions which were exceedingly unsafe. The analysis used a logistic regression with a statistical significance at the 0.05 level, with a dichotomous independent variable (fault vs not-at-fault), where the probability of the at-fault driver is:

$$P\text{ (driver is at fault)} = 1/(1+e^{-z}) \qquad (2.1)$$

where: $$Z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_N X_N \qquad (2.2)$$

and X_n = driver factors, confounding factors, and interactions [6]. The effect of the independent variables on the crash rates employed a multivariate analysis.

Interestingly, over half (32 out of a total 56) of the published RAIR values in the study were not statistically significant at the 5% level [6].

This methodology and similar types are commonly used by researchers for analysis of categorical collision data. In fact, another study on motor vehicle crashes in Ontario that cross-sectionally examined factors affecting the severity of motor vehicle traffic crashes (MVTCs) by class (i.e., fatal, major, minor, etc.) was performed, where percentage distributions of crashes at each level of severity were examined according to specific factors and tested using bivariate analyses (X^2 test, $P < 0.05$) [9]. The researchers chose to limit the analysis to variables that were both restricted to observations without missing observations and those statistically significant to the 0.05 level (as determined through bivariate analysis). Note that these researchers were fortunate to have a large sample size, as it exceeded 34,000 collisions (after exclusion of observations) with excellent data integrity (less than 2% missing values). Initial findings include determination of 18 factors which achieved statistical significance in the bivariate analysis, listed as follows:

1. Age of Driver
2. Sex of Driver
3. Driver Condition
4. Driver Action
5. Use of Seatbelts
6. Ejection from vehicle
7. Month
8. Day of Week
9. Hour of Day
10. Road Alignment
11. Roadway Configuration
12. Road Surface Condition
13. Speed Limit
14. Weather Conditions
15. Light Conditions
16. Crash Configuration
17. Vehicle Type
18. Vehicle Maneuver

One portion of the analysis consisted of calculating relative risks through crude odds ratios (COR) for a primary comparison, complete with their 95% confidence intervals.

Odds ratios are briefly explained as ratios of probabilities for success/failure, where the probability of success, p and the probability of failure, q = 1-p.

In such a case, the odds of success are defined as

$$\text{Odds(success)} = p/q \qquad (2.3)$$

The odds of failure would be the inverse of the odds of success,

$$\text{Odds(failure)} = q/p \qquad (2.4)$$

The odds of success and failure are clearly reciprocals of each other.

In order to determine the OR ideally, the odds of success for two or more different groups within a category are divided by each other. Variations on this exist and many valid OR scenarios often exist (i.e., where the inverse is true, where the odds of success for two or more different categories for a group are compared, etc.).

For example, given that 9 out of 10 teenage drivers survive side impact collisions while only 6 out of 10 elderly drivers survive in the same conditions, the probabilities for teenage driver survival are:

$$p = 9/10 = 0.9 \qquad (2.5)$$

$$q = 1\text{-}0.9 = 0.1 \qquad (2.6)$$

And for elderly drivers:

$$p = 6/10 = 0.6 \quad (2.7)$$

$$q = 1\text{-}0.6 = 0.4 \quad (2.8)$$

These probabilities are then used to determine the respective OR of success (successful in survival event occurring).

$$\text{Odds(teenage survival)} = p/q = 0.9/0.1 = 9.0 \quad (2.9)$$

$$\text{Odds(elderly survival)} = p/q = 0.6/0.4 = 1.5 \quad (2.10)$$

Finally, we can calculate the OR of teenage vs elderly survival as:

$$\text{Odds Ratio, OR} = 9.0 / 1.5 = 6.0 \quad (2.11)$$

This value is greater than 1.0 which indicates that the odds of teenage vs. elderly survival under the given conditions are 6.0 times greater for teenagers as compared to the elderly, (given identical assumptions). Note that confidence intervals should also be considered and issues arise when confidence intervals straddle a range inclusive of 1.0. In such cases, the propensity and value for a given group to be either a factor that is increasingly or decreasingly causative is indeterminate.

Through screening literature, it was found that selected published research papers included crude odds ratios in their analysis. This is rather troublesome, as crude odds ratios are inferior and do not account for potential correlations between factors. For instance, a portion of the previously mentioned analysis is not ideal, particularly since it did not control for other effects of unaccounted for factors [9]. To account for this, multivariate unconditional logistic regression analyses were employed to estimate each specific factor in relation to crash severity while controlling for these other effects. The importance in controlling for these other effects is very important. Through a multivariate analysis, SAS computed Adjusted Odds Ratios. Using similar criteria to the first portion of their analysis, (the bivariate analyses X^2, $P < 0.05$), SAS computed that the majority of these factors were still significant. However, the magnitude of the adjusted odds rations decreased considerably compared with the crude ORs. On the other hand, the multivariate analyses showed that several factors including the month, hour, road alignment, light conditions, and vehicle type were no longer statistically significant and road surface condition was inversely associated with fatal crashes [9]. The relationships between the findings from the bivariate and multivariate analysis are not clear. Some may argue that it is a worthless comparison.

The trend of selected variables being no longer statistically significant when analyzed in-depth at the multivariate level continues to the final component of the study, where significance similarly at $p < 0.05$ level through the log likelihood test in SAS. This test reveals that only two of the variables remain statistically significant, namely driver age and driver condition. Even at first glance, numerous OR values (and 95% confidence intervals) are seen tottering above and below 1.0. The usefulness of much of the report's published tables on adjusted OR are therefore questionable, but its methodology and purpose are exemplary. In this study, quasi-induced exposure is recommended due to the difficulty and present unavailability of travel estimates.

Interestingly, another similar study where experimentation with logistic regression models were used to explain similar dependent and independent variables was performed [10]. Several regression analyses of driving record variables over a six-year time period (1986 –1991) were performed to compare the results obtained using various techniques and found that this did not yield different results. The techniques compared were ordinary least squares, weighted least squares, Poisson, negative binomial, linear probability and logistic regression models.

Through the process of literature review, an appreciation for normalized data and meaningful analysis became apparent. In similar studies, many weaknesses exist. Even one recently (2001) published paper by Pulugurtha et al [11] is easily criticized as tables and results are limited to frequencies and percentages by category, with no mention of missing neither data nor consideration for level of significance.

Another superior study used logistic regression to analyze collision data to identify driver, highway, and environmental factors that differentiate collisions with a dichotomous independent outcome variable [12]. Logistic regression determines β coefficients that make the observed outcome most "likely," using the maximum-likelihood method. Since the model is nonlinear, an iterative algorithm is used to estimate coefficients [13]. The validity of each variable's presence in the regression model is checked by examining the statistical level of significance (LOS) for its β coefficient. A typical acceptance level for LOS is 0.05, which indicates that if the β parameter were equal to zero, the probability of seeing a value of β as extreme or more extreme than the one observed is less than 0.05 (5%) [12].

Independent (explanatory) variables included in their research paper, falling under three major categories, are:

Driver:

- age, gender and alcohol involvement;

Highway:

- location, horizontal alignment, vertical alignment, setting, speed limit, and type of highway; and,

Environmental:

- light conditions, pavement conditions.

Interestingly, where categorical (zero-one) variables were needed to represent different conditions within a given category, the number of zero-one columns is always one less than the number of conditions (i.e., three categorical zero-one variables were needed to represent four different light conditions, and two variables were used for three pavement conditions) [12].

Zero-one variables were set with hypothesized collision contributing conditions as 1. The "base" condition (all categorical variables equal to zero) represented an intersection on a flat section of straight roadway in an urban area during daylight when the pavement was dry and the driver had not been drinking. From this base condition, the increase in collision odds attributed to each independent variable can be computed from the logistic regression coefficients [12]. Of particular interest is the method used to split the categories. For instance, age groupings were determined by varying the range of ages in a model until the age variable was no longer statistically significant within the age group [12]. They were fortunate to have over 400,000 observations and achieve low levels of statistical significance (0.00005) for several of their models.

A similar breakdown of independent explanatory variables by driver, vehicle and environmental variables are found in a report titled "An Evaluation of Severity and Outcome of Injury by Type of Object Struck (First Object Struck Only) for Motor Vehicle Crashes in Connecticut" [14]. Variables were derived from several sources of data, all falling under the umbrella of the State of Connecticut Department of Transportation (DOT). Record Types from Crash Summary Records, Traffic Unit Information Records, and Involved Person Records, were linked together. This created information pertinent to the:

- crash as a whole;
- identification of each vehicle or pedestrian involved in a crash; and,
- information about vehicle operators, struck pedestrians, passengers, and witnesses.

This linking process allows an in-depth analysis of severity and outcome of injury, providing a more complete picture of the effects of crashes with any particular collision. Their study used variables similar to the other reviewed papers above, with variables classified in seven different categories (as opposed to three), namely:

1. demographic factors including age and gender;
2. geographic factors including location of the crash and location of the fixed object struck;
3. subjective factors including speeding, following too closely, violating traffic controls, unsafe use of highway, etc;

4. objective factors including driver illness, vehicle involved in emergency;
5. road and weather/season condition including construction and road surface;
6. police judgment/investigation including whether or not the driver had been drinking; and.
7. clinical variables.

Interestingly, the categories within each of these variables differ from report to report. In their study, categorization was as follows:

- drivers' age having five subgroups: age less than 25 years, 25 to 44, 45 to 64, 65 to 74, and greater than 74 years.);
- length of (hospital) stay categorized into three groups: Emergency Department (ED) treated and released, inpatient with length of stay equal to 1 day, and inpatient with length of stay greater than 1 day; and,
- mortality categorized as died at the crash site, ED death (died in hospital with zero length of stay), died as inpatient (died in hospital with length of stay equal to or greater than 1 day), and died after discharge.

The methodology used was almost identical to the above mentioned report (McGuinnis et al, 1998) where all categorical variables were converted into binary variables, as required for the analysis. Stepwise logistic regression models were constructed with Level of Significance (LOS) of 0.01 and 0.05 for entrance and exit, respectively, to identify parsimonious sets of independent variables.

A Review of the Evidence for Factors Affecting Incidence and Severity [15] reveal several key components summarized from a wealth of related and recent collision studies. Driver, vehicle and environmental characteristics are determined to play relevant parts in these crashes, as the review authors stated that,

> "Older drivers, the variability in the size and weight of the vehicles sharing the roads, the characteristics of intersections and the mechanisms used to influence traffic flow all combine to affect the rates of occurrence of side impact crashes. Older vehicle occupants, presumably because of increased physical frailty, are more likely to be killed or injured, especially in the target vehicle."

The paper also reveals "evidence that wearing the seatbelts will reduce the risk and severity of injury, although this clearly does not affect the risk of collision in the first place" [15].

But the argument exists that seatbelts *do indeed* "affect the risk of collision in the first place", as seatbelts are often overlooked for their added benefit of keeping the driver restrained and in control of the vehicle. The Canadian Health Network states that "Besides protecting people in a collision, seat belts also keep the driver in place so they can stay in control of the car" [16]. Should the vehicle require a sharp turn or hard deceleration, seatbelts may make the difference between a driver remaining in an ideal seating position versus moving loosely around, losing control and potentially having a collision. In fact, this paper [16] reveals,

> "Seat belts reduce your chances of dying in an accident by 45 to 55%. They reduce your chances of serious injury by about 50%. Child passenger restraints (infant car seats) reduce the risk of death and serious injuries by about 70%. In 1997 in Alberta, people who used restraints were much less likely to be injured in a crash (14.7%) than those who didn't use them (35.7%). Transport Canada estimates that since 1989, the increased use of seatbelts has avoided 66,000 injuries."

Without further criticism of one specific comment from an otherwise outstanding paper, other findings include that evidence exists to better engineer intersections "whether as roundabouts, or by adding appropriate traffic controls, or improving sightlines for approaching vehicles" [15]. Several studies reviewed in the paper identify numerous vehicle design factors such as "the disparity in the size and weight of vehicles sharing the road... a particular challenge to vehicle designers". Furthermore, the authors noted,

> "A number of design features - in brakes, to make it easier for the bullet vehicle to stop to avoid a crash, in the strength and energy-absorbing capacity of the target vehicle, to protect its occupants from intrusion, in the design of interior surfaces of the target vehicle, in the visibility of either vehicle so that drivers have more warning to avoid a crash..."

In addition to vehicle design features, intersection features are also of great interest in determining causes of side impact collisions. Of particular importance is signal phasing, especially since many side impact collisions occur at intersections, where left-turning vehicles narrowly cross other vehicle's paths with synchrony. A recently published report by the National Cooperative Highway Research Program (NCHRP) comprehensively evaluates safety and effectiveness of signal displays, with emphasis on Protective Permissive Left-Turn Control (PPLT). A key concern with PPLT control is the "yellow trap," which occurs during the change from permitted left turns in both directions to a lagging protected left turn in one direction [17]. In layman's terms, this happens when a vehicle enters an intersection on a priority (or protected) phase attempting a left-turn maneuver when suddenly, the priority phase turns amber, and the opposing thru traffic commences, thereby leaving the driver vulnerable to the oncoming traffic. This occurs for a number of reasons, and the report describes this issue as it pertains to many of the phasing systems (shown in Figure 2.2) used throughout the United States.

As evidenced by the multiple phasing and display options identified above, "accommodating left-turning vehicles at signalized intersections has been an ongoing concern for transportation engineers as they seek a balance between intersection capacity and safety through signal phasing techniques" [17]. Through surveys, field studies and controlled lab experiments, measurements on the comprehension of these signal systems and phasing schedules were determined. Confusion based on the sampled driver population's lack of understanding of these signal systems, particularly priority phases and flashing vs. steady-green lights, shows issues such as the "yellow trap" amongst others, can be overcome through a more understandable display and a less complicated phasing plan.

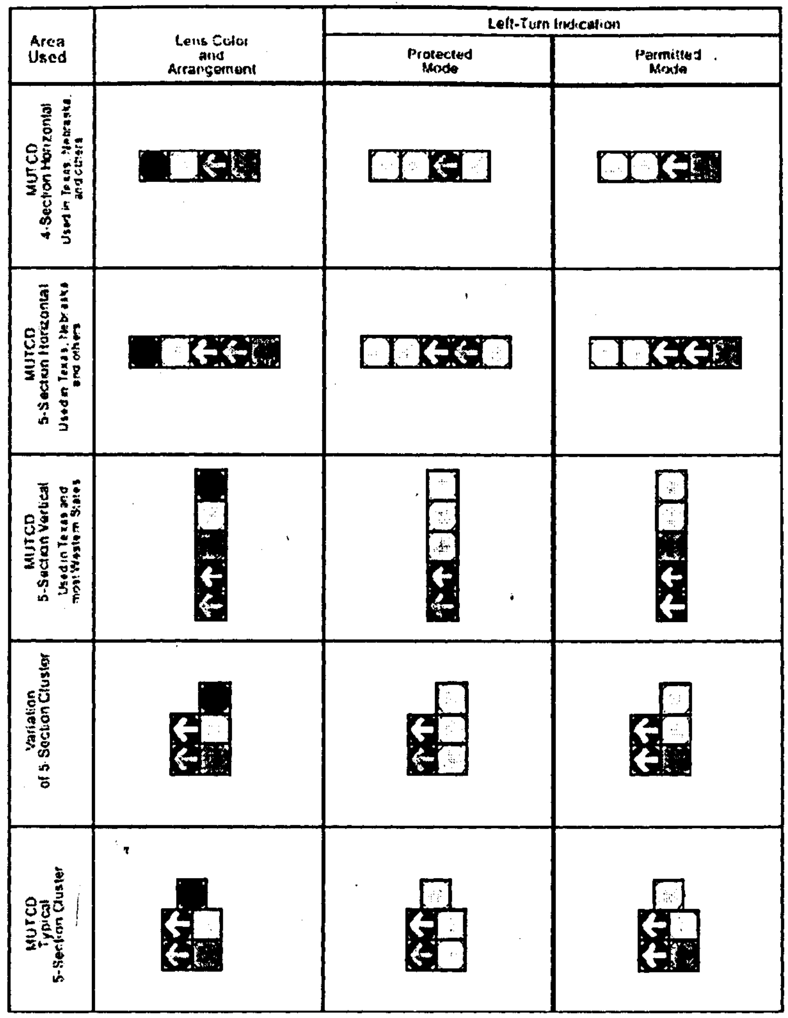


Figure 2.2: Variations in PPLT Displays

The "Dallas Display" signal system (as shown in Figure 2.3) has been known to overcome these issues through its simple and effective design.

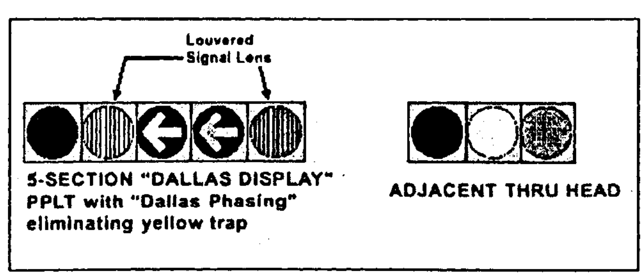


Figure 2.3: Illustration of Dallas Display

A portion of their report focus on measures of average response time for various signal systems and phases. In general, it was determined that response time was generally lower for the flashing permissive indications. Furthermore, "a trend in average response time by age was very evident as drivers over the age of 65 took between 2 and 4 sec of additional time to respond when compared with drivers under the age of 24" [17]. This is likely attributed to the effects of dementia associated with aging.

Their research also includes crash data analysis left-turn crash rates associated with field traffic conflicts for PPLT displays. Observed traffic events were categorized by four potential left-turning conflicts types, (see diagrams in Figure 2.4), namely:

- Type 1—opposing left-turn conflicts;
- Type 2—left-turn/same direction conflicts;
- Type 3—left-turn/lane change conflicts; and
- Type 4—secondary conflicts, such as those involving a pedestrian or bicyclist or resulting from a lane overflow.

A total of 11 hours of data were collected at each of the 24 study intersections for a total of 264 hours of observation time. During the observation period, the research team observed approximately 2,000 vehicles; of which 5,000 were left-turn vehicles; and 17,000 were through vehicles [17]. A summary of the observed quantities for each type are shown in Figure 2.5.

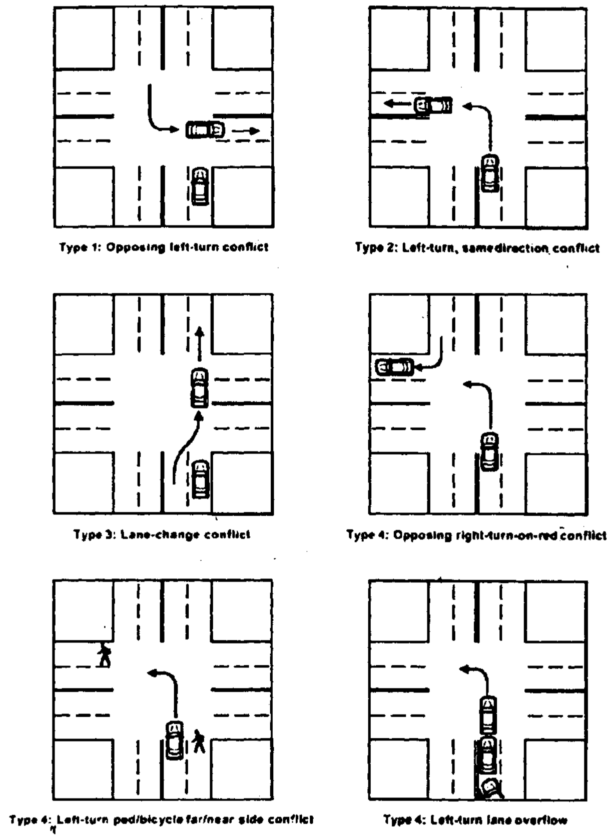


Figure 2.4: Illustration of Conflict Types

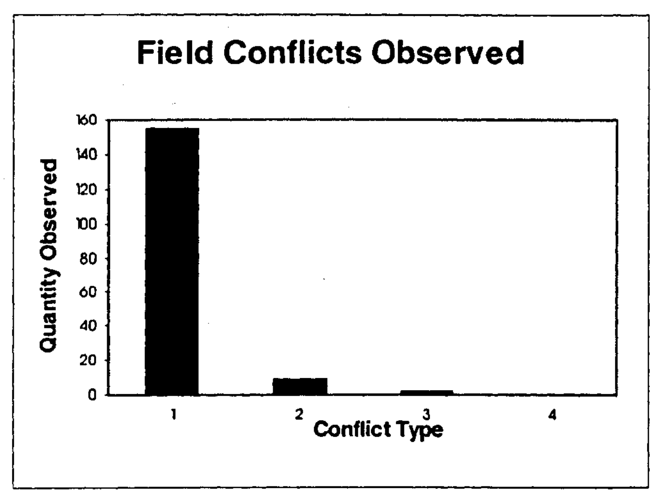


Figure 2.5: PPLT Field Conflicts

From this study, there were several important findings, namely:

- Most left-turn conflicts were related to hesitation at the onset of the green indication;
- Aggressive driving appeared to be the major cause of Type 1 conflicts. Drivers continued to make left-turn maneuvers during the yellow and all-red phase following the protected left-turn phase and were in conflict with the opposing through traffic.
- Type 2 conflicts were primarily the result of a driver's hesitating to turn left on the left-turn permissive indication. The sudden hesitation would cause a conflict with following vehicles. There appeared to be a relationship between the driver's understanding of the permissive circular green indication and the observed Type 2 conflicts;
- The few Type 3 conflicts were a result of driver error and not the lack of understanding of the PPLT signal display; and
- There were no Type 4 conflicts observed.

In addition to field conflicts, classifications for "Traffic Events" paralleling the four conflict events were:

- Type 1: Driver hesitating on the left-turn protected indication;
- Type 2: Driver hesitating on the left-turn permissive indication;
- Type 3: Driver going through the circular red indication; and
- Type 4: Driver backing a vehicle out of the intersection, back into the left-turn lane.

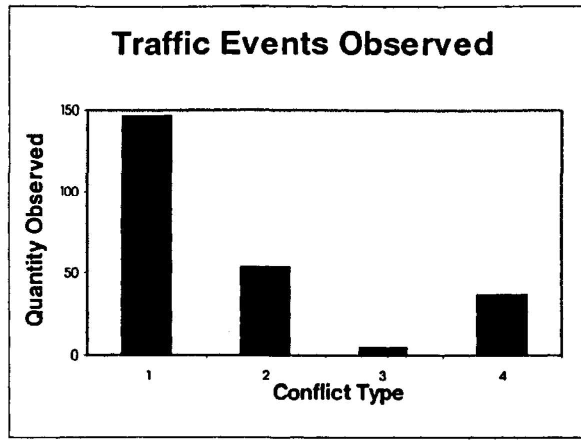


Figure 2.6: PPLT Field Study Traffic Events

- The largest occurrence of Type 1 traffic events involved a five-section horizontal PPLT signal display arrangement. The simultaneous illumination of the green arrow and the circular red indications appeared to increase the workload of the driver, resulting in an increase in driver uncertainty.
- Type 2 traffic events were observed at each of the study intersections. The occurrence did not appear to be related to the PPLT signal arrangement or phasing or indication.
- Numerous drivers were observed proceeding through the all-red indication (i.e., red light runners). Therefore, Type 3 traffic events were recorded only when the action was clearly a function of driver misunderstanding. For this reason, only five Type 3 events were recorded, and the occurrences showed no pattern to suggest an influence of the PPLT signal display, indication, or phasing.

- 33 Type 4 events were associated with a flashing permissive indication. The driver would enter the intersection during the permissive phase and not have the opportunity to make the left-turn maneuver. The driver would then choose to back up.

Another portion of the above mentioned study involved a "Crash Data Analysis", where selected components of a crash database created in 1988 as part of a FHWA study were examined through comparisons of left-turn crash rates associated with various PPLT signal displays. This initiative was sparked by the past "inadequate documentation of causes of a crash related to traffic signal display and operation" [17]. Traffic volumes, signal display information, and 3 years' of crash data for the study intersections (the same intersections studied in the operational study and the conflict study) were evaluated in the crash analysis. General findings on the use of PPLT signals for intersections with left-turning vehicles include:

- leading protected left-turn phasing has the lowest crash rate (as compared with lagging protection;
- vehicle delay decreases;
- fuel usage decreases;
- vehicle progression is improved; and
- vehicle crashes increase.

As the number of left-turning vehicles increase, average delay and accident potential for both through and left-turning vehicles increase [17].

The report shows a relationship exists between driver's understanding of signal systems and its effect on safety. A focus on the weaknesses associated with left-turns at intersections is made with "conflicts" and "traffic events" as key indicators. There are a number of limitations associated with their methodology. The narrow classifications used for the four "conflicts" and four "traffic events" do not account for other situations, nor for other confounding factors. For instance, the duration of each of the traffic events varies, as the amount of available time for each of the classes (i.e., hesitating on the left-turn permissive indication; vs. going through the circular red indication) differs. Also, traffic way volumes are not explicitly accounted for either. This may bear particular significance as individual jurisdictions may have their own specifications (i.e., minimum left-turning volumes, etc.) that call for particular signal systems to be used. Furthermore, there are a handful of issues associated with using crash rates as a measure of safety, such as:

- Regression-to-mean or bias-by-selection effect must be accounted for, otherwise over-estimation of the effectiveness of any condition or treatment may result. In fact, some situations with high crash rates may have fewer crashes in subsequent periods, even if no treatment is carried out; and
- The use of the crash rates to compare the safety 'before ' and 'after' or 'with' and 'without' a treatment, leads to a paradox; by comparing crash rates one

would be led to the incorrect conclusion about which selected site is safer, or whether a treatment helps or impairs safety. This paradoxical result stems from the fact that the crash rates fails to separate the effect of traffic flow on safety from the effect of a treatment on safety. In fact, when safety is improved from a given treatment, the expected crash rate may actually increase. This kind of error will be present always when the relationship between crash frequency and traffic flow is not a straight line through the origin [18]. To remedy this, crash frequency should be used instead of crash rates. Using crash frequency avoids this kind of error through modifying the crash frequency.

An interesting limitation of their study found throughout the report was their failure to measure influences of the intersection geometry. They reasoned that this is "could not be measured because the study simulated only exclusive left-turn lane configurations" [17].

Another weakness shows "in the aggregate, the crash analysis findings ... did not perform consistently within a selection of four crash statistics. The ranking of one crash statistic [observed conflicts] did not match that of another crash statistic [traffic events]" [17]. Also, the results of the crash rate analysis and the conflict study had no correlation in rank ordering. This poor synchronicity certainly reduces confidence in their findings.

Findings and oversights from the literature reviewed above have been incorporated and considered where possible in this report.

2.4 Summary of Literature Review

In general, findings from the literature reviewed point to a number of methods for the analysis of collision data, including side-impact collisions. The use of logistic regression and statistical programs is superior to simple summary statistics and crude odds ratios (COR). In addition, odds ratios (OR) are most typically deemed statistically significant and publishable when criteria for $P < 0.05$.

A number of significant factors in collision involvement have been shown to appear in past research as human, vehicle and environmental factors emerged. Also, the range of variables available and used in the reviewed studies differed from study to study, but commonalities exist where driver, vehicle and environmental factors are essential. The fact that certain factors are absent in particular studies is disheartening, and it must be stressed that great importance lies in the availability and inclusion of all relevant factors for a worthwhile analysis. It has been revealed that great advantages lie in the linking of separate data sets, as the linking process allows an in-depth analysis of severity and outcome of injury, providing a wider picture of the effects of crashes with any particular collision. Chapter 3.0 expands on this idea of the data linking process used to construct a database.

3 Development of a Collision Database for In-Depth study of Side Impact Collisions

3.1 General

In order to fully benefit and accommodate an analysis of collision data with experts in multiple disciplines, all available Toronto data-sets were linked together. Catering to each specialized expert is accomplished through the accumulation and linkage of data-sets containing vehicle, driver and environmental information. Talents for each expert are harnessed in this way. This chapter focuses on the Toronto data sources available and those used to develop a database for further multi-disciplinary research efforts.

Within The City of Toronto, there are a number of official documents that are used to record collision information, as follows:

- Ontario Ministry of Transportation Motor Vehicle Accident Reports (MVAs) (Figure 3.1) and accompanying investigating officer field notes
- Ministry of Transportation Self Reporting Collision Reports (Figure 3.2 and Figure 3.3)
- Transport Canada Vehicle Safety Research Team Collision Investigation Reports (Booklet, Appendix A)

In addition to these documents, Transport Canada's Vehicle Safety Research Teams have access to VIN (vehicle identification number) code reading software, which can be used in combination with vehicle VIN's to obtain important characteristics such as vehicle mass, weight distributions, and other dimensions.

Also, it was decided that provisions exist for "control vehicles" as part of the linked-data set. This is discussed in the following sections.

In order to develop a superior and unique database, it is proposed that all of the above available data sources are merged together, as shown in Figure 3.4.

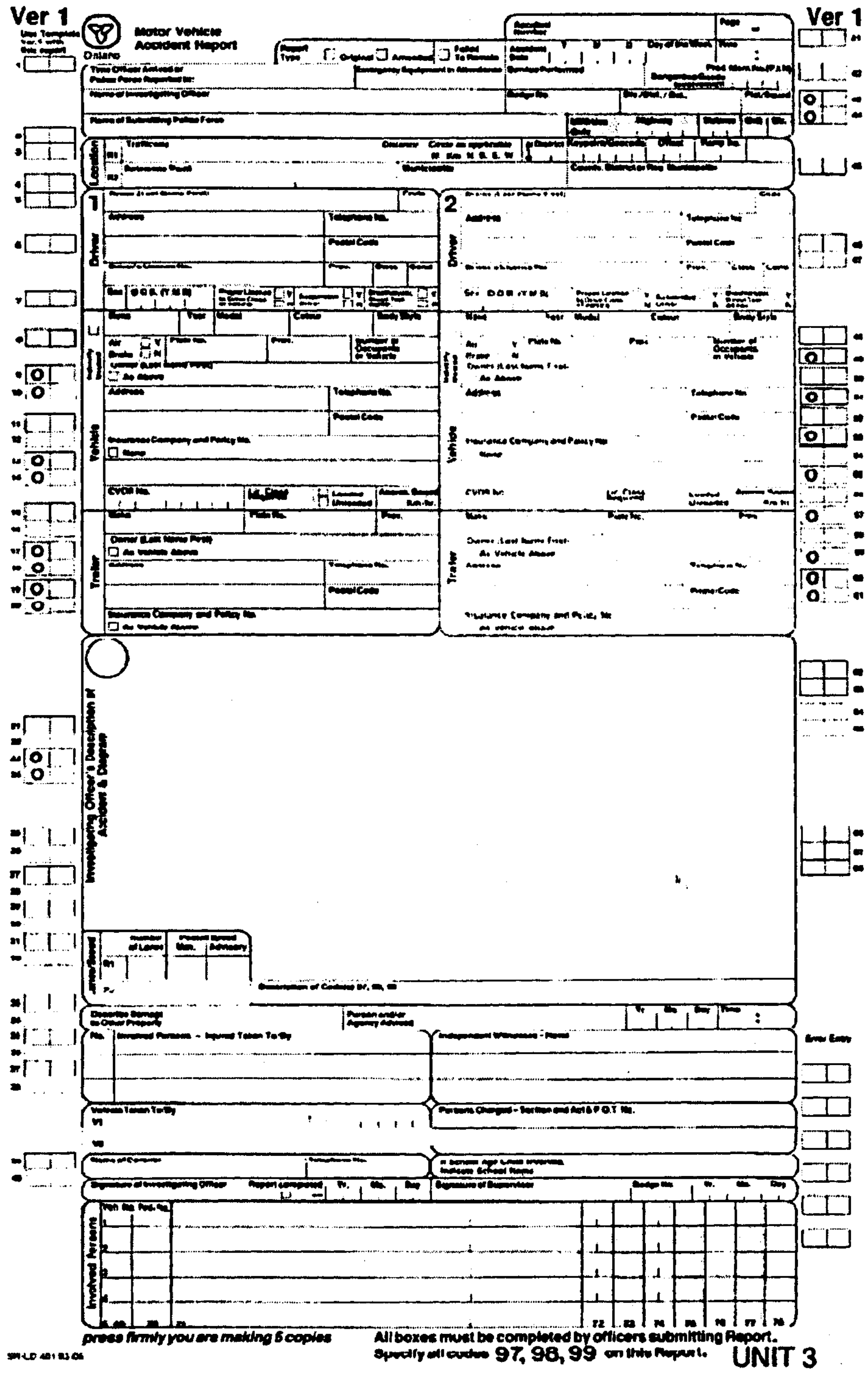


Figure 3.1: Ministry of Ontario Motor Vehicle Accident Report [19]

Ministry of Transportation
Ontario

Self Reporting Collision Report

Collision Report No. 265963 | Page 1 of 2

1. For MTO and Police Use Only
Reporting Location: EAST C. R. C.
Collision Date (Y/M/D): 98 12 24 | Day: THUR | Time (Hrs): 1630

Police Service of Jurisdiction: Toronto Police Service
Number of Vehicles Involved: 2
☐ Property Damage ☐ Failed to Remain | ☒ Supplementary Form Attached | ☒ Original ☐ Amendment

☒ at ☐ near | R 1 - Number and Name of Street, Intersection, Highway, etc.: W/B STEELES RD
Road Jurisdiction: 01 Munic 02 Prov Hwy 03 Twp 04 City/Dist 05 Reg Munic 06 Prv Prop 07 Fed 99 Other | 01

R 2 - Name of Shopping Mall, Plaza or Other Commercial or Institutional Property, etc.: VICTORIA PARK AVENUE [redacted]
County, District or Reg. Municipality: York

Highway	Distance	Unit	Dir.	District	Keypoint/Geocode	Offset	Ramp No.

Vehicle Damage: 01 none 02 light 03 moderate 04 severe 05 demolished | 02

Damage Viewed: ☒ yes ☐ no | Viewed by (Name & Badge No.): [redacted] | Date (Y/M/D): 98 12 24 | Time (Hrs): 2040 | Damage Est.: $300

Describe Damage to Vehicle or Additional Remarks for Investigator: R/F wheel / R/F door / R/R 1/4 panel / R/R door / R/F bumper

Describe Damage to Other Property:

Initial Impact: 02 | Secondary Impact:

Rept Rec'd by (Name, Badge No., Pol. Serv): [redacted] | Time (Hrs): 2045 | Person and/or Agency Advised: [redacted] | Date (Y/M/D) | Time (Hrs)

Officer Assigned | Date Assigned (Y/M/D) | Date Cleared (Y/M/D) | Charges Laid (Item No.)

2. Your Information - Driver | Last Name, First Name: [redacted] | MTO Code
Address - Street No. & Name, City, Town, Province, Postal Code | Telephone Nos. Home [redacted] Work [redacted]

Driver's Licence No	Prov	Class	Cond	Sex	Date of Birth (Y/M/D)	Age	Proper Licence for Class of Vehicle	Suspended Driver
[redacted]	ON	G2	—	M	1982/03/05	16	☒ yes ☐ no	☐ yes ☒ no

3. Your Vehicle [redacted]

Prov.	Year	Make	Model	Colour	Body Style	Number of Occupants in Vehicle
ON	91	TOYOTA	COROLLA	WHITE	4-DOOR	1

Air Brake ☐ yes ☒ no | Lic. Class Req'd: G | ☐ Loaded ☒ Unloaded | Insurance Company's Name & Policy No. / Expiry Date (Y/M/D): [redacted]

Vehicle Taken / Towed To | Vehicle Taken / Towed By

Owner ☐ Same as Driver | Last Name, First Name: [redacted]
Address - Street No. & Name, City, Town, Province, Postal Code: [redacted] | Telephone Nos. Home [redacted] Work [redacted]

4. Your Trailer | Plate No. | Prov. | Make | Insurance Company's Name | ☐ Same as Vehicle | Policy No. | Expiry Date (Y/M/D)
Owner ☐ Same as Driver | Last Name, First Name | Address - Street No. & Name, City, Town, Province, Postal Code: [redacted]

5. Other Driver | Last Name, First Name: [redacted] | Driver's Licence No: [redacted]
Address - Street No. & Name, City, Town, Province, Postal Code: [redacted] | Telephone Nos. Home [redacted] Work [redacted]

6. Other Vehicle | Plate No.: [redacted] | Prov | Year | Make | Model | Colour: Turquoise (?) | Body Style | Number of Occupants in Vehicle: 1
Other Owner ☐ Same as Driver | Last Name, First Name | Insurance Company's Name & Policy No.: [redacted] | Expiry Date (Y/M/D)

7. Witness | Last Name, First Name | Address - Street No. & Name, City, Town, Prov, Postal Code: [redacted] | Telephone Nos. Home [redacted] Work [redacted]

8. Collision Details - See instructions sheet. Fill in the boxes below with the numbers that best describe the collision. If "other (specify)" is selected, please give details.

01 Collision Location 03
On Highway: 01 non intersection 02 intersection related 03 at intersection 04 at / near private drive 05 at railway crossing 06 underpass / tunnel 07 overpass / bridge 98 other (specify)
Off Highway: 08 trail 09 frozen lake / river 10 parking lot 99 other (specify)

02 Impact Location 02
01 within intersection 02 thru lane 03 left turn lane 04 right turn lane 05 right turn channel 06 two way left turn lane 07 passing lane 08 left shoulder 09 right shoulder 10 not on roadway - left side 11 not on roadway - right side 12 off highway 99 other (specify)

03 Weather (multiple choice) 01
01 clear 02 rain 03 snow 04 freezing rain 05 drifting snow 06 strong wind 07 fog, mist, smoke, dust 99 other (specify)

04 Light 01
01 daylight 02 daylight - artificial 03 dawn 04 dawn - artificial 05 dusk 06 dusk - artificial 07 dark 08 dark - artificial 99 other (specify)

05 Traffic Control 01
01 traffic signal 02 stop sign 03 yield sign 04 pedestrian crossover 05 police control 06 school guard 07 school bus 08 traffic gate 09 traffic controller 10 no control 99 other (specify)

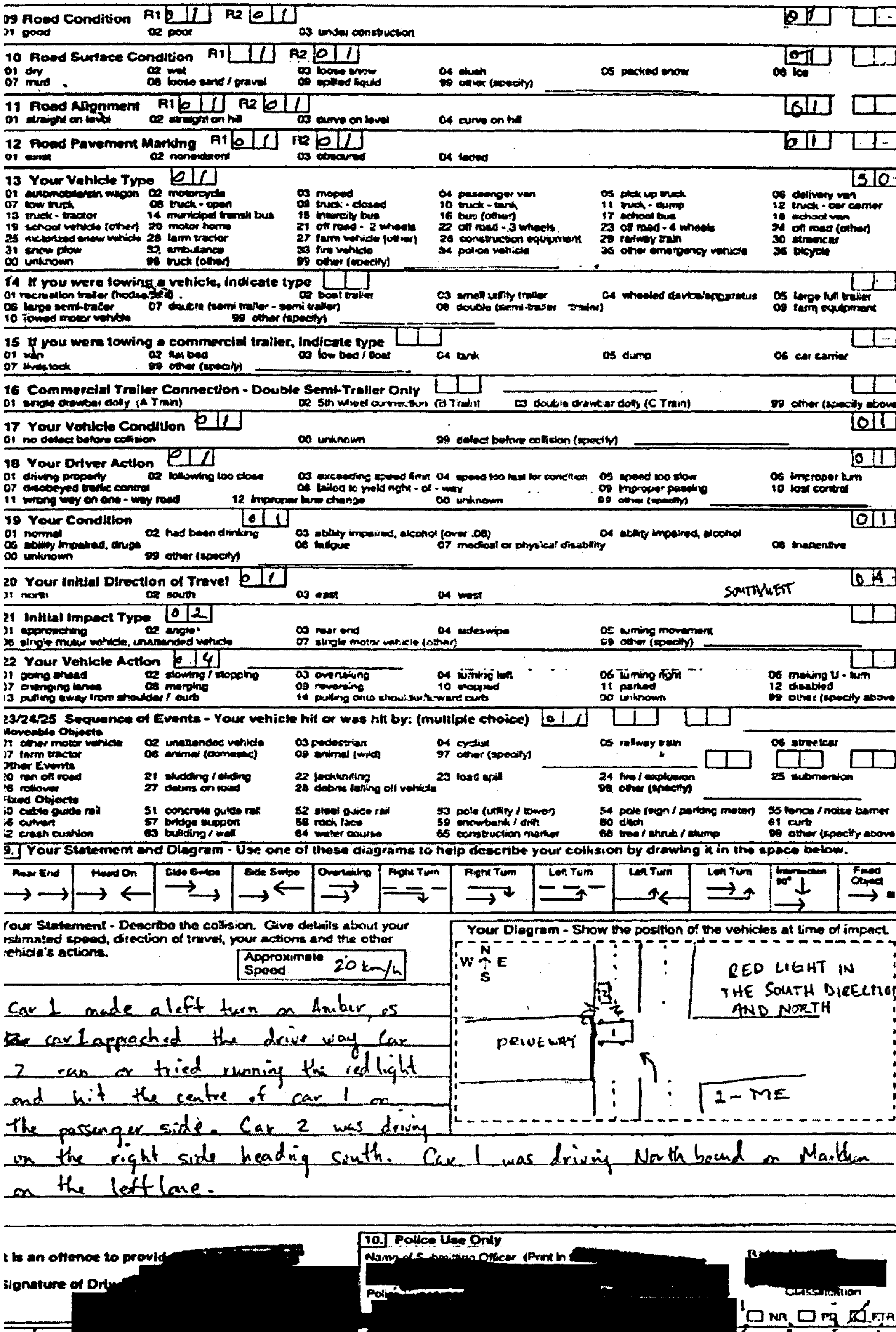
06 Traffic Control Condition 01
01 functioning 02 not functioning 03 obscured 04 missing / damaged

07 Road Character R1 02 R2 02
01 undivided - one way 02 undivided - two way 03 divided with barrier 04 divided - no barrier 05 ramp 06 collector lane 07 express lane 08 transfer lane

08 Road Surface R1 01 R2 01
01 asphalt 02 oil treated gravel 03 gravel/crushed stone 04 concrete 05 earth 06 wood 07 steel 08 brick / interlocking stone 99 other (specify)

The Self Reporting Collision Report is a public record available for a fee

Figure 3.2: Ontario Self-Reporting Collision Report, Page 1



09 Road Condition R1 01 R2 01 — 01
01 good 02 poor 03 under construction

10 Road Surface Condition R1 01 R2 01 — 01
01 dry 02 wet 03 loose snow 04 slush 05 packed snow 06 ice
07 mud 08 loose sand / gravel 09 spilled liquid 99 other (specify)

11 Road Alignment R1 01 R2 01 — 01
01 straight on level 02 straight on hill 03 curve on level 04 curve on hill

12 Road Pavement Marking R1 01 R2 01 — 01
01 exist 02 nonexistent 03 obscured 04 faded

13 Your Vehicle Type 01 — 50
01 automobile/stn wagon 02 motorcycle 03 moped 04 passenger van 05 pick up truck 06 delivery van
07 tow truck 08 truck - open 09 truck - closed 10 truck - tank 11 truck - dump 12 truck - car carrier
13 truck - tractor 14 municipal transit bus 15 intercity bus 16 bus (other) 17 school bus 18 school van
19 school vehicle (other) 20 motor home 21 off road - 2 wheels 22 off road - 3 wheels 23 off road - 4 wheels 24 off road (other)
25 motorized snow vehicle 26 farm tractor 27 farm vehicle (other) 28 construction equipment 29 railway train 30 streetcar
31 snow plow 32 ambulance 33 fire vehicle 34 police vehicle 35 other emergency vehicle 36 bicycle
00 unknown 98 truck (other) 99 other (specify)

14 If you were towing a vehicle, indicate type
01 recreation trailer (hodse/tent) 02 boat trailer 03 small utility trailer 04 wheeled device/apparatus 05 large full trailer
06 large semi-trailer 07 double (semi trailer - semi trailer) 08 double (semi-trailer trailer) 09 farm equipment
10 towed motor vehicle 99 other (specify)

15 If you were towing a commercial trailer, indicate type
01 van 02 flat bed 03 low bed / float 04 tank 05 dump 06 car carrier
07 livestock 99 other (specify)

16 Commercial Trailer Connection - Double Semi-Trailer Only
01 single drawbar dolly (A Train) 02 5th wheel connection (B Train) 03 double drawbar dolly (C Train) 99 other (specify above)

17 Your Vehicle Condition 01 — 01
01 no defect before collision 00 unknown 99 defect before collision (specify)

18 Your Driver Action 01 — 01
01 driving properly 02 following too close 03 exceeding speed limit 04 speed too fast for condition 05 speed too slow 06 improper turn
07 disobeyed traffic control 08 failed to yield right - of - way 09 improper passing 10 lost control
11 wrong way on one - way road 12 improper lane change 00 unknown 99 other (specify)

19 Your Condition 01 — 01
01 normal 02 had been drinking 03 ability impaired, alcohol (over .08) 04 ability impaired, alcohol
05 ability impaired, drugs 06 fatigue 07 medical or physical disability 08 inattentive
00 unknown 99 other (specify)

20 Your Initial Direction of Travel 01 — 04
01 north 02 south 03 east 04 west SOUTHWEST

21 Initial Impact Type 02
01 approaching 02 angle 03 rear end 04 sideswipe 05 turning movement
06 single motor vehicle, unattended vehicle 07 single motor vehicle (other) 99 other (specify)

22 Your Vehicle Action 04
01 going ahead 02 slowing / stopping 03 overtaking 04 turning left 05 turning right 06 making U - turn
07 changing lanes 08 merging 09 reversing 10 stopped 11 parked 12 disabled
13 pulling away from shoulder / curb 14 pulling onto shoulder/toward curb 00 unknown 99 other (specify above)

23/24/25 Sequence of Events - Your vehicle hit or was hit by: (multiple choice) 01
Moveable Objects
01 other motor vehicle 02 unattended vehicle 03 pedestrian 04 cyclist 05 railway train 06 streetcar
07 farm tractor 08 animal (domestic) 09 animal (wild) 97 other (specify)
Other Events
20 ran off road 21 skidding / sliding 22 jackknifing 23 load spill 24 fire / explosion 25 submersion
26 rollover 27 debris on road 28 debris falling off vehicle 98 other (specify)
Fixed Objects
50 cable guide rail 51 concrete guide rail 52 steel guide rail 53 pole (utility / tower) 54 pole (sign / parking meter) 55 fence / noise barrier
56 culvert 57 bridge support 58 rock face 59 snowbank / drift 60 ditch 61 curb
62 crash cushion 63 building / wall 64 water course 65 construction marker 66 tree / shrub / stump 99 other (specify above)

9. Your Statement and Diagram - Use one of these diagrams to help describe your collision by drawing it in the space below.

Rear End | Head On | Side Swipe | Side Swipe | Overtaking | Right Turn | Right Turn | Left Turn | Left Turn | Left Turn | Intersection 90° | Fixed Object

Your Statement - Describe the collision. Give details about your estimated speed, direction of travel, your actions and the other vehicle's actions.

Approximate Speed 20 km/h

Car 1 made a left turn on Amber, as car 1 approached the drive way Car 2 ran or tried running the red light and hit the centre of car 1 on the passenger side. Car 2 was driving on the right side heading south. Car 1 was driving North bound on Markham on the left lane.

Your Diagram - Show the position of the vehicles at time of impact.
N W E S
RED LIGHT IN THE SOUTH DIRECTION AND NORTH
DRIVEWAY
1 – ME

It is an offence to provid[redacted]
Signature of Dri[redacted]

10. Police Use Only
Name of Submitting Officer (Print in [redacted]
Poli[redacted]
Classification
☐ NR ☐ PQ ☒ FTA

Figure 3.3: Ontario Self-Reporting Collision Report, Page 2

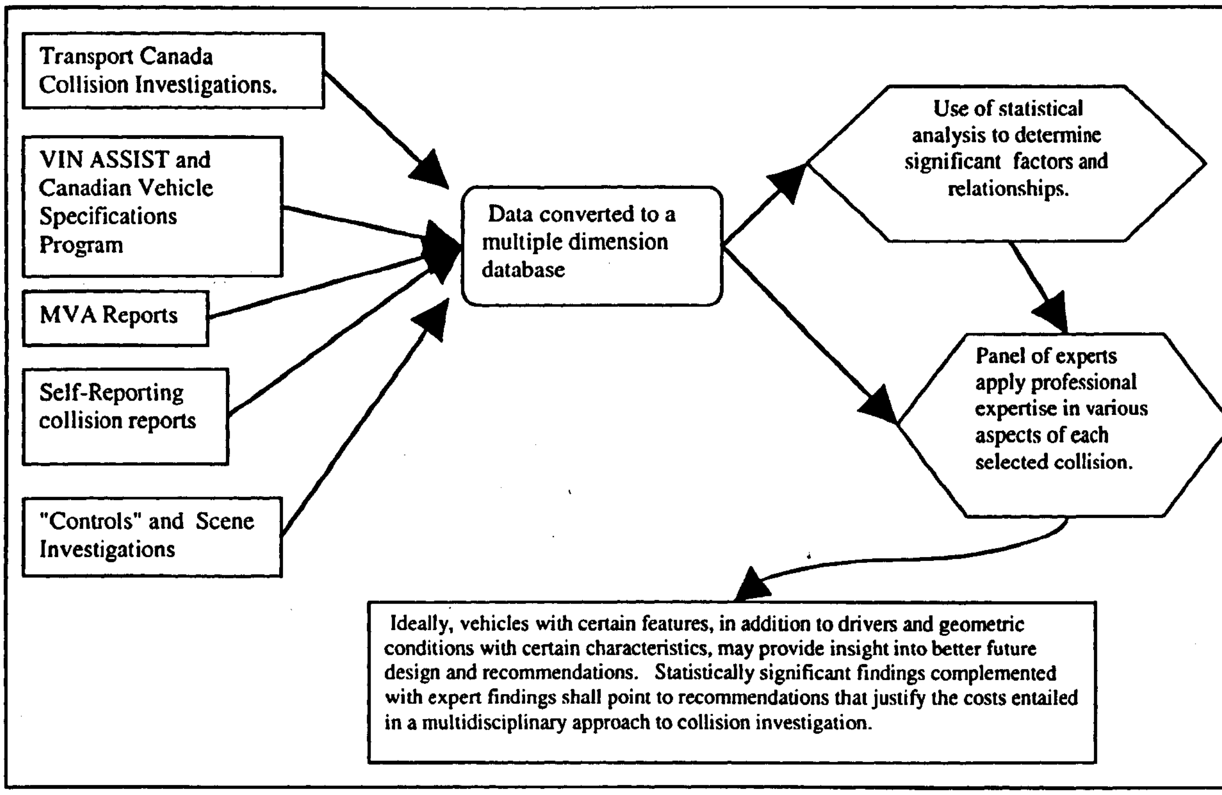


Figure 3.4: Proposed Database Multiple Data Sources Flowchart

3.2 Ministry of Transportation Ontario Motor Vehicle Accident (MVA) Reports

The Ministry of Transportation of Ontario (MTO) maintains a database on every reportable motor vehicle collision that occurs in Ontario. MVA reports are official forms completed by investigating officers typically at a given collision scene. They are important records that consist of three parts: collision information, driver/vehicle information, and involved person information. Some details of each include:

- Collision Information: Date, time, street name, reference location, collision severity, weather conditions, etc.

- Driver/Vehicle Information: License class, administration of breathalyzer, safety equipment use, vehicle maneuver, etc.

- Involved person Information: Age, gender, seating position, injury, etc.

Titles for these MVA report form variables and the classes they contain are shown on Figure 3.5, and additional fields are also shown in Figure 3.6.

A comprehensive list of the variables found in this report is listed in Appendix B. The definitions for the terms found in this report are listed in the Glossary. The data handwritten in these reports are essentially permanent, aside from occasional updates within 30 days that involve changes in involved person's status of health, (e.g., in the event of hospitalization followed by death).

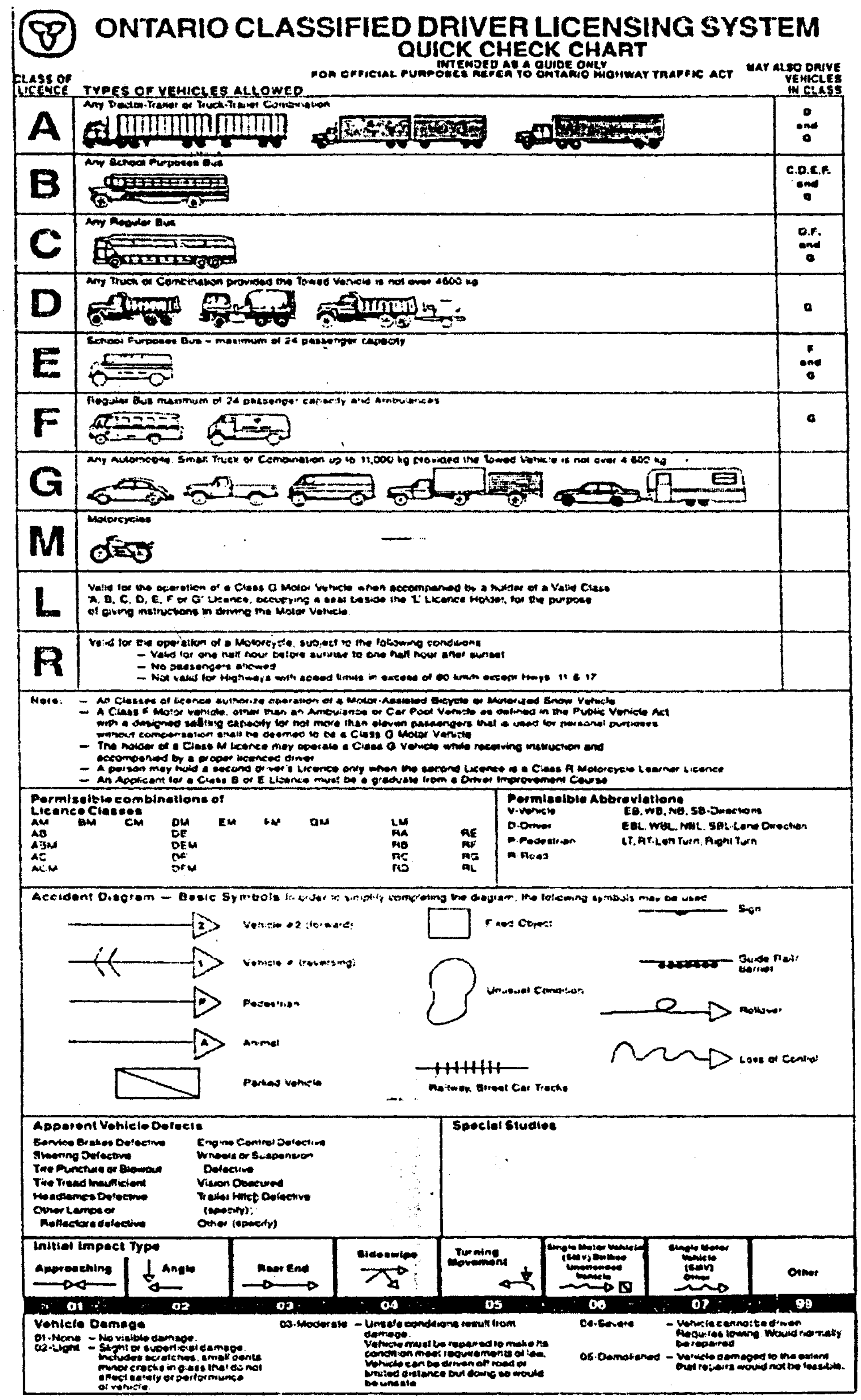


Figure 3.6: Additional MVA Information

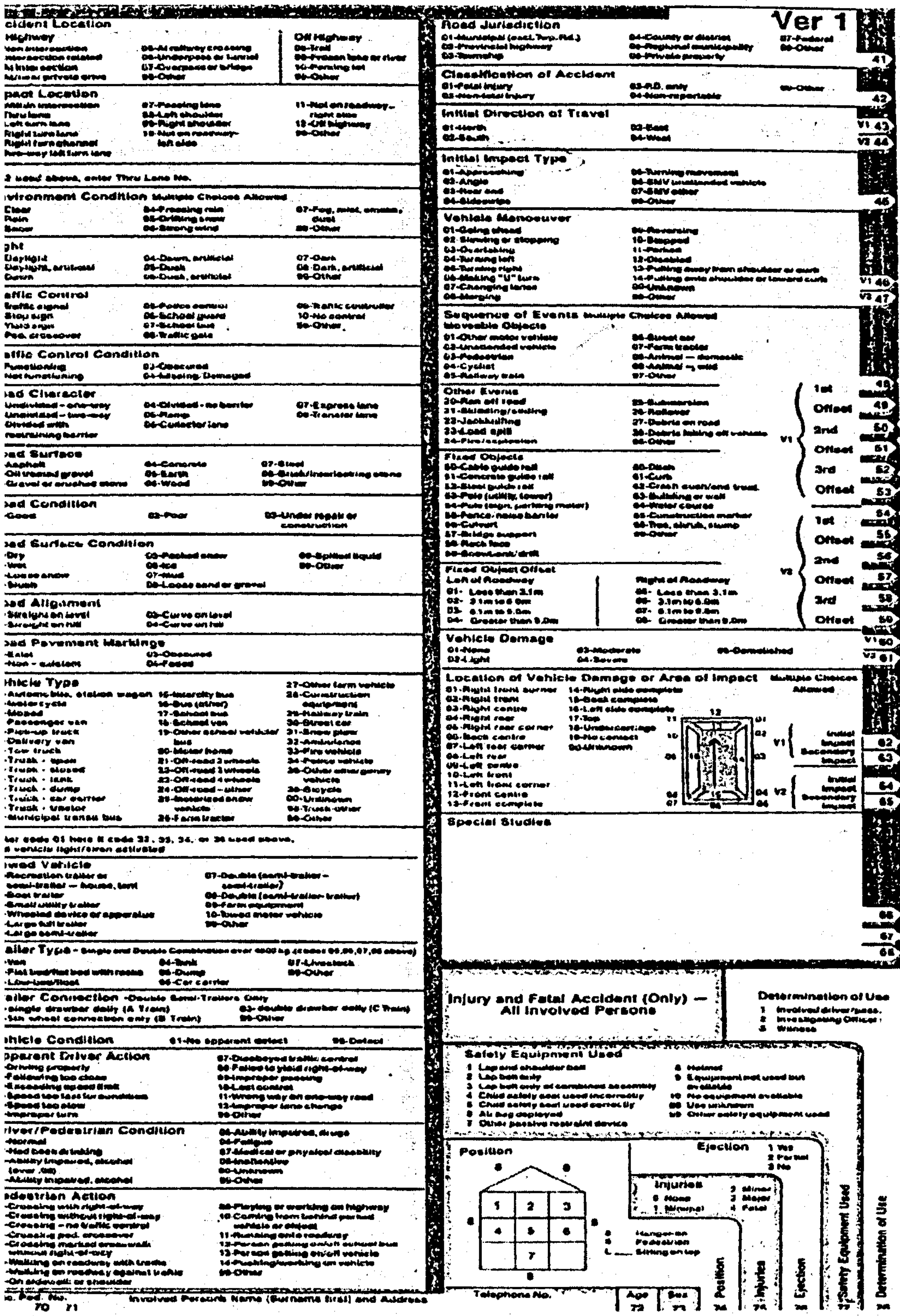
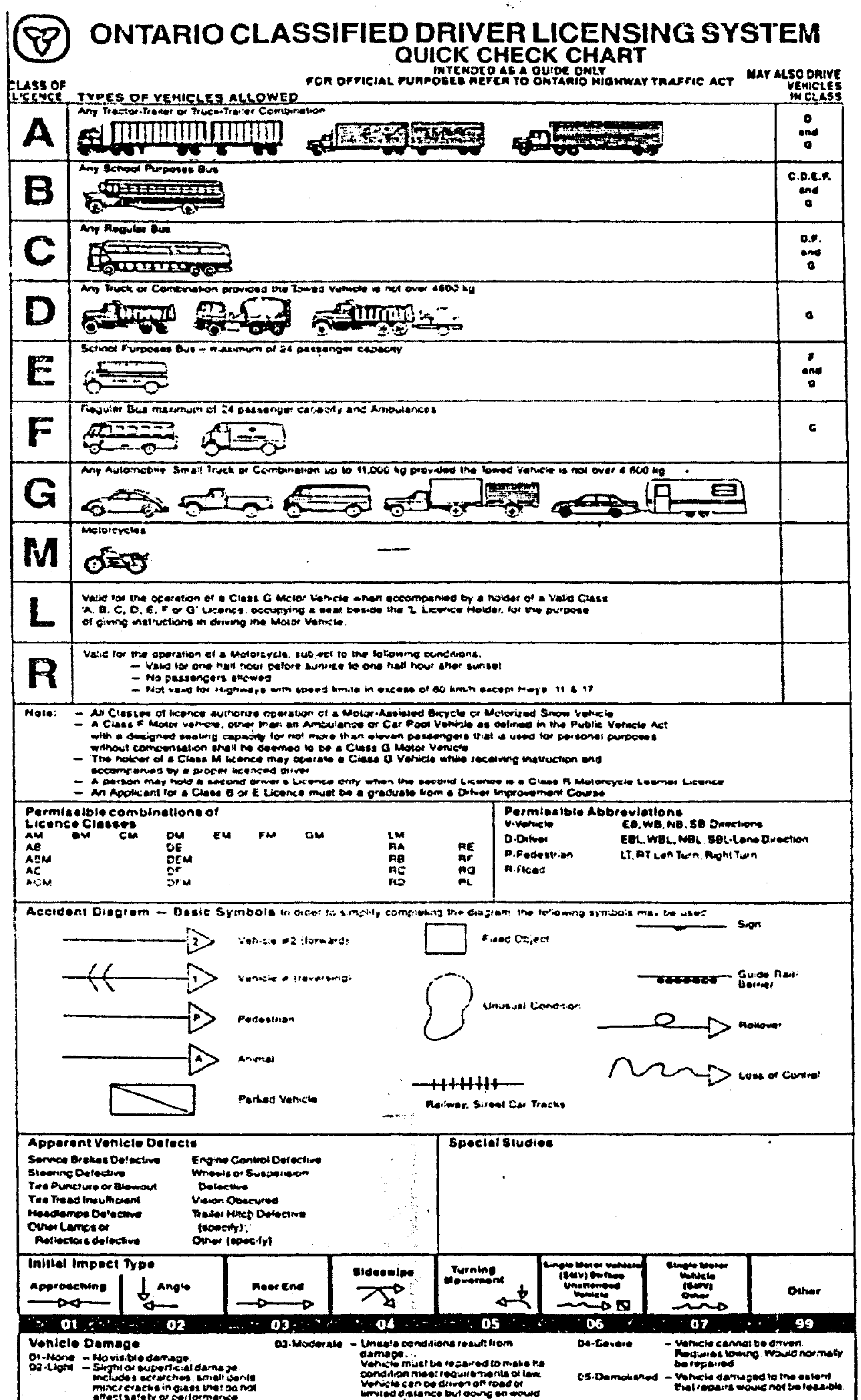


Figure 3.5: MVA Report Variables and Classes



ONTARIO CLASSIFIED DRIVER LICENSING SYSTEM
QUICK CHECK CHART

INTENDED AS A GUIDE ONLY
FOR OFFICIAL PURPOSES REFER TO ONTARIO HIGHWAY TRAFFIC ACT

CLASS OF LICENCE	TYPES OF VEHICLES ALLOWED	MAY ALSO DRIVE VEHICLES IN CLASS
A	Any Tractor-Trailer or Truck-Trailer Combination	D and G
B	Any School Purposes Bus	C, D, E, F, and G
C	Any Regular Bus	D, F, and G
D	Any Truck or Combination provided the Towed Vehicle is not over 4600 kg	G
E	School Purposes Bus – maximum of 24 passenger capacity	F and G
F	Regular Bus maximum of 24 passenger capacity and Ambulances	G
G	Any Automobile, Small Truck or Combination up to 11,000 kg provided the Towed Vehicle is not over 4 600 kg	
M	Motorcycles	
L	Valid for the operation of a Class G Motor Vehicle when accompanied by a holder of a Valid Class 'A, B, C, D, E, F or G' Licence, occupying a seat beside the 'L' Licence Holder, for the purpose of giving instructions in driving the Motor Vehicle.	
R	Valid for the operation of a Motorcycle, subject to the following conditions: – Valid for one half hour before sunrise to one half hour after sunset – No passengers allowed – Not valid for Highways with speed limits in excess of 80 km/h except Hwys. 11 & 17	

Note:
- All Classes of licence authorize operation of a Motor-Assisted Bicycle or Motorized Snow Vehicle
- A Class F Motor vehicle, other than an Ambulance or Car Pool Vehicle as defined in the Public Vehicle Act with a designed seating capacity for not more than eleven passengers that is used for personal purposes without compensation shall be deemed to be a Class G Motor Vehicle
- The holder of a Class M licence may operate a Class G Vehicle while receiving instruction and accompanied by a proper licenced driver
- A person may hold a second driver's Licence only when the second Licence is a Class R Motorcycle Learner Licence
- An Applicant for a Class B or E Licence must be a graduate from a Driver Improvement Course

Permissible combinations of Licence Classes

AM	BM	CM	DM	EM	FM	GM	LM	
AB			DE				RA	RE
ABM			DEM				RB	RF
AC			DF				RC	RG
ACM			DFM				RD	RL

Permissible Abbreviations

V-Vehicle — EB, WB, NB, SB Directions
D-Driver — EBL, WBL, NBL, SBL-Lane Direction
P-Pedestrian — LT, RT Left Turn, Right Turn
R-Road

Accident Diagram — Basic Symbols In order to simplify completing the diagram, the following symbols may be used

Apparent Vehicle Defects

Service Brakes Defective
Steering Defective
Tire Puncture or Blowout
Tire Tread Insufficient
Headlamps Defective
Other Lamps or Reflectors defective
Engine Control Defective
Wheels or Suspension Defective
Vision Obscured
Trailer Hitch Defective (specify);
Other (specify)

Special Studies

Initial Impact Type

Approaching	Angle	Rear End	Sideswipe	Turning Movement	Single Motor Vehicle (SMV) Unattended Vehicle	Single Motor Vehicle (SMV) Other	Other
01	02	03	04	05	06	07	99

Vehicle Damage

01-None – No visible damage.
02-Light – Slight or superficial damage includes scratches, small dents minor cracks in glass that do not affect safety or performance of vehicle.
03-Moderate – Unsafe conditions result from damage. Vehicle must be repaired to make its condition meet requirements of law. Vehicle can be driven off road or limited distance but doing so would be unsafe
04-Severe – Vehicle cannot be driven. Requires towing. Would normally be repaired
05-Demolished – Vehicle damaged to the extent that repairs would not be feasible.

Figure 3.6: Additional MVA Information

In many instances, symptoms of these injuries will not occur immediately. Instead they may take up to 72 hours to manifest themselves. Some of the common symptoms for the injuries include headache, blurred vision, loss of taste, smell, or hearing, blood in urine or stool, swelling, loss of motion to the area/stiffness/tightness and more [2].

As a result, deaths and injury severity may still be underestimated for a variety of reasons. These are discussed in the Limitations section of this report.

3.3 Ministry of Transportation Self Reporting Collision Reports

Due to the large frequency of motor vehicle collisions combined with the limited resources of investigating Police Forces, officers are limited to the investigation of collisions at the scene when there is at least some injury component defined in four categories as:

1. Minor: person did not go to hospital when leaving the scene of the accident. Includes minor abrasions, bruises and complaint of pain.
2. Minimal: person went to hospital and was treated in the emergency room but not admitted.
3. Major: person admitted to hospital. Includes person admitted for observation.

4. Fatal: person killed immediately or within 30 days of the motor vehicle crash. [20]

To account for the limited police resources, partnerships between government (police), Insurance Providers and private enterprise have developed a solution; Collision Self-Reporting Centres. The concept of Self Reporting of collisions is the main function of the Collision Reporting Centre (CRC). Drivers involved in property damage collisions report within 24 hours to the CRC where a police officer inspects the vehicle damage. Here, drivers themselves complete a simplified version of a collision form that is checked by a police officer [21].

Under the Ontario Highway Traffic Act [22], when one is in a collision in which there is only property damage (no injury or death, and, among other conditions, no criminal activities such as impaired driving) the involved person(s) may report the collision immediately by proceeding with one's vehicle to a Collision Reporting Centre. Self-reporting of a collision was introduced on January 1, 1997 [23]. (There are other requirements, such as a $1000 minimum reportable level for property damage only collision, but these bear no significance to this report and are not discussed.) Immediately, digital pictures are taken and a CRC sticker is permanently attached to indicate that the damage has been reported. A great benefit of this process is its quick resolving nature and the convenience it affords to the typical driver through the 24 hour grace period. This service is free of charge to motorists as it is funded in part through insurance companies is open 24 hours a day, 7 days a week.

Further improvements in CRCs are well underway, particularly through electronic claims reporting Web-based software. For example, Allianz Canada successfully completed the first ever pilot project on February 19th, 2003. A motor vehicle accident report has been electronically submitted from North York Accident Support Services Ltd., one of Toronto's three CRCs, directly to Allianz Canada's Toronto claims office. Allianz believes that having the driver record and submit the data themselves can save duplicate work typical of this type, in addition to doing the work of what would have otherwise been an employee.

3.4 Transport Canada Vehicle Safety Research Team Collision Investigation Reports

Through the Government of Canada, the Road Safety and Motor Vehicle Regulation of Transport Canada has a mandate to contribute to a reduction in deaths, injuries, and property damage resulting from motor vehicle use, through improved safety of the motor vehicles.

Transport Canada satisfies this through five areas of activity, including:

1. Directed studies Investigations
2. Defect Investigations
3. Special Investigations
4. Community Involvement and Education

5. Professional Development

Of particular importance are the Directorate's established Directed Studies Investigations and Special Investigations. These investigations consist of in-depth and impartial analysis of collisions where multi-disciplinary research teams are established typically working out of Universities. The term "Multi-disciplinary" indicates that the members of the teams are either full-time staff or consultants hailing from a diverse group of professions. Scientists, Engineers, Physicians, Coroners, Psychologists, Police Officers and other specialists scientifically analyze motor vehicle collision data to determine vehicle crashworthiness and injury causation. Subsequently, the teams recommend to the Directorate improvements to the Canadian Motor Vehicle Safety Standards (CMVSS) or the need for new standards. Included in their in depth investigations are many fields relating to crush, injuries, estimated speeds, injury sources and more. These and other forms from their investigations are attached in Appendix A. These forms are completed through vehicle, occupant and scene inspections conducted by their team of experts, often followed with occupant and investigating police officer interviews.

Vehicle interiors are investigated to determine the dynamics of occupant and vehicle contact. Interior vehicle environment and particularly safety features are also thoroughly examined and recorded. Features such as air bags, seat belts, seat back structure and roof strength are examined to determine their usage and effectiveness in a collision. Through confidential communication with the occupants and other confidential sources (i.e., medical treatment records, coroner, police, witness, etc.,) further information is determined. After harvesting this information, the sensitive components are purged from their files upon completion of the case in order to ensure anonymity of those involved. Also performed are collision scene inspections "measuring skid marks, fluid spills and gouges".

Aside from the large amount of detailed collision information this source provides, the team members that conduct the investigation are typically more specialized in collision investigation, as compared to police officers and the self-reporting public. In fact, members of these teams have offered seminars and other collision investigation training to police departments. Included in their reports is output from two programs, namely, VIN ASSIST, and Canadian Vehicle Specifications. These programs decipher VIN numbers of involved vehicles and output characteristics such as vehicle mass, weight distributions, height, width, and more. A notably important field is the seventeen character Vehicle Identification Number (VIN). This is a unique code that is found on all motor vehicles. It has been recognized by Transport Canada as an important field. Identifying a vehicle for detailed investigation is best achieved through its VIN. Ideally the VIN of each vehicle involved would be part of a police report. This will be potentially reliable when electronic readers are practical and widely available.

Electronic readers are becoming widely used, as many emission testing facilities in Toronto are scanning VIN codes located on vehicle pillars for easy tracking.

A sample of VIN ASSIST and CANADIAN VEHICLE SPECS output appears in Figure 3.7 and Figure 3.8.

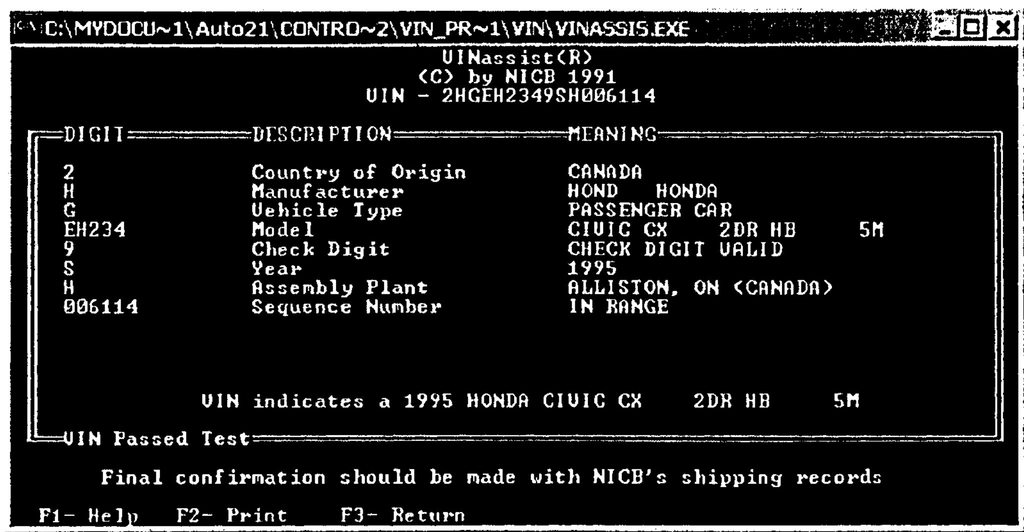


Figure 3.7: VIN ASSIST Software Output Sample

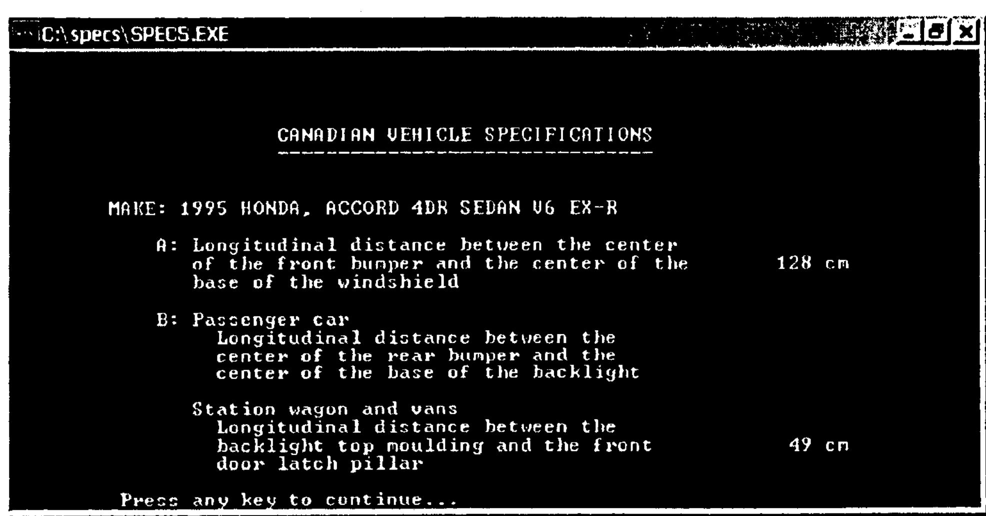


Figure 3.8: Canadian Vehicle Specifications VIN Software Output Sample

3.5 Control Investigations and Form Creation

Experts involved in this research project insisted on obtaining additional information for the database from "control vehicles". (This is useful for comparisons and contrasts in order to determine what factors differ form collision and non-collision experiencing vehicles, drivers and environments.) To date, Transport Canada has not performed this before. It was decided that collecting additional information by returning to collision scenes at the same hour, seven days after the collision, to note traffic and weather conditions and to select control vehicles traveling at the crash scene would generate useful information about non-crashing vehicles.

It was proposed that the license plates of control vehicles be used to determine the VIN numbers and driver's names and addresses (assuming registered owner's are driver's). Through government resources, this information is available, although it is guarded. For a marginal fee, vehicle history searches can be completed online through the Ministry of Transportation Online Services (http://www.mto.gov.on.ca/english/dandv/catalogue.htm). In addition, Ontario Private Driver and Vehicle License Issuing Offices also offer this service.

Road safety statistics experts would argue that this report's focus is in fact that of a comparison group and not a control group. Hauer supports the notion that "when the assignment to treatment is 'at random', it is legitimate to speak of a 'statistical experiment' which involves a 'control group'" [24].

Despite efforts to select control vehicles at random, controlling for day of week, time of day, and day of year was possible, but not other fields such as weather. Thus, despite any efforts, this study never had a true control group, since regardless of the size, as "even if both entities are very large, they will differ systematically with respect to some casual factors" [24]. It should be noted that vehicles were not selected entirely at random. In fact, control vehicles are selected with a ratio of 4:1, four control vehicles per case vehicle, respectively, (with the intention of sending drivers surveys at a later date). Furthermore, since the study's focus is on late model vehicles, efforts have been made at control scenes to collect only relatively late model vehicles (defined as 1998 and newer) traveling in target and bullet vehicle directions.

"In contrast, when the assignment of entities to the treatment group is not made at random, then...they will differ systematically with respect to some causal factors" [24]. In part, a goal of this report is to identify and quantify these systematic differences between case and control vehicles, with consideration of driver fault. Finally, for the reasons explained above, the term "control" shall be replaced with "comparison" throughout the remainder of this report.

The construction of the control investigation forms followed the format provided by Transport Canada's Vehicle Safety Research Teams Collision Investigation Fields. When composing these forms, field selection was carefully matched with as similar as possible a manner. Having comparable fields ensures easily comparable results.

Furthermore, additional fields were recorded for use in analysis other than the comparison group type. Provisions to accommodate signal timing, sight distance, approximate vehicle speeds, volume counts, and more were created and included. The completed forms are attached to the Vehicle Safety Research Teams Collision Investigation Booklet in Appendix C.

3.6 Database Creation for Data on All Available Forms

Ideally, all possible fields recorded about a particular collision would be of greatest benefit to researchers and analysts. However, this is often unpractical as it would seem natural for collisions of particularly minor severities to be given less attention and detail than those of a more serious nature.

Regardless, using the commonalities that exist between all motor vehicle collisions, and even between the three different collision report forms found in Ontario, the link of them into one all encompassing and all inclusive data set increases its potential for analysis. Also, great potential exists in making them compatible with one another, such that fields for the common data (i.e., weather conditions, match piecemeal with variable numbers). Difficulties in this process arose when conflicts between identical fields (such as weather conditions) had differing options. For instance, Transport Canada's field for "Weather

Condition", contains class "3 - Raining", while the MVA report's corresponding "Environment Condition" class "3 - Snow" differs, yet other classes are matched exactly. Although many parallels exist with the available variables, this was a tedious and time-consuming process.

For this report, a database has been constructed that meets this requirement, and is essentially all encompassing and far more complete than any one of the existing three collision data collection schemes. The similar nature of many of the data fields found in the explained three report types required decisions to be made as to the prioritizing of data.

For instance, the police reported collision severity may differ from the results of the interview performed with the occupant several days later. Often injuries may be discovered after the initial excitement of a collision. At times, the body tries to stop these signals by creating chemicals that help block pain signals. These chemicals, called endorphins, are morphine-like painkilling substances that decrease the pain sensation. [25] Afterwards, soreness may appear days later, where occupants then seek medical treatment. One example of the triggering mechanism behind endorphin release is one's own thoughts and emotions. For example, a father who is driving his children is hurt in a car accident. He is so worried about his children that he doesn't feel the pain of his own broken arm. The concern for his children has caused the natural release of endorphins, which block the pain signal and prevent him from noticing the pain.

As explained earlier in this report, different formats for similar data from the multiple sources of data existed. In such cases, the format from the most reliable source was adopted. More specifically, reliability was based on a combination of considerations. Typically, investigating officers are first to arrive immediate at scene, so fields such as road surface environmental condition, weather and illumination are taken from their reports. However, other non-time sensitive information such as pre- and post-collision vehicle positions, target and bullet vehicle actions, etc, are recorded from the Vehicle Safety Research Team investigations. The synopsis composed by summation of hours of work by the Research Teams for each collision held the highest regard as an information source for each collision. As a result, fields from MVA reports, self reports and Vehicle Safety Research Teams were all incorporated. The size of the generated linked spreadsheet initially contained over 3000 observations. From this, carefully selected factors have been selected from the data sets and are discussed and used for regression modeling in the following chapters. The shall provide researchers insight on the selected collisions, as it exists as a preliminary investigation based on currently available linked-data using methods derived from the literature review in Chapter 2.

4 Preliminary Investigation of Toronto Police Reported Side Impact Collisions - Research Approach

4.1 General

This chapter describes the data and methodology used for the preliminary analyses 1 and 2, with data available from MVA reports and electronic intersection data files, respectively. Using electronically available data and through several linked data sets, 1835 signalized intersections for years 1998-2000 inclusive within the City of Toronto were analyzed to identify factors affecting intersections with collisions vs. those without. In addition, 1718 motor vehicle accident reports representing 1466 motor vehicle collisions (as there are frequently multiple MVA reports per collision) were selected and further scrutinized manually to identify only the side impact collisions. These collisions were linked with the intersection data and analyzed too. The results of this report's research efforts illustrate the usefulness of using linked data sets to perform analysis, as mentioned in the literature review section. Individually, each data set could not provide the degree of insight and depth that has been afforded only through this linking process.

As mentioned earlier, there are multiple sources of data. These are merged into only two databases and are used for two analyses (see Figure 4.1: Data for the Two Separate Analysis).

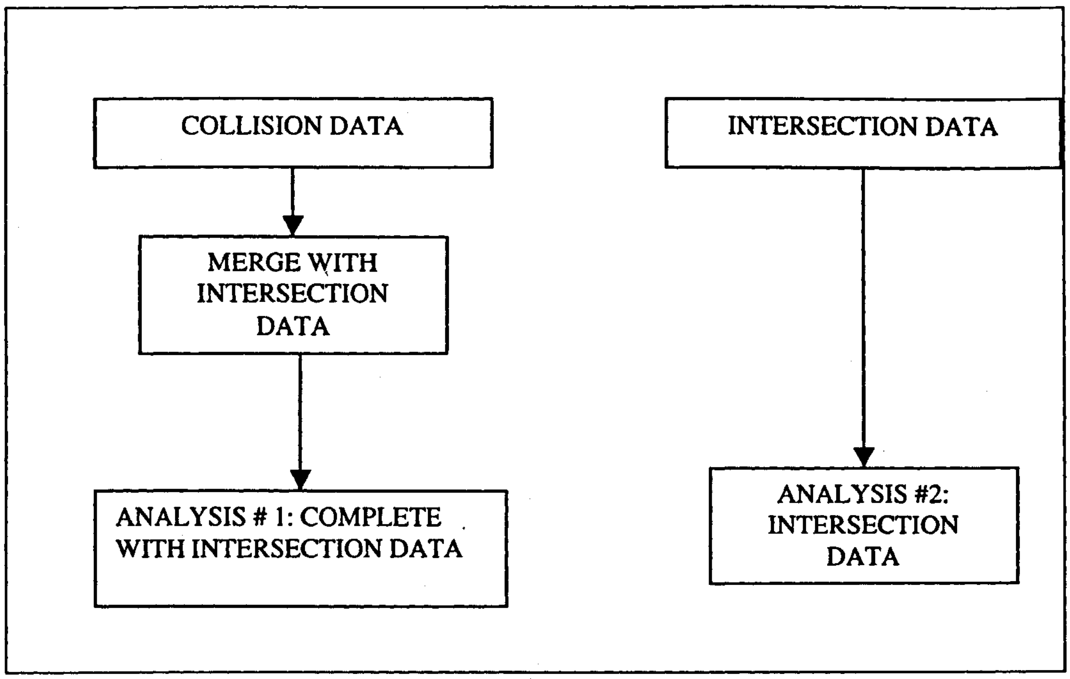


Figure 4.1: Data for the Two Separate Analysis

Researchers typically compose a null hypothesis in hope that it can be discredited. Both analyses performed in this study follow this idea. Briefly explained, the null hypothesis,

$$H_0: \quad y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots \beta_n X_n = 0 \tag{4.1}$$

And when summarized takes the following form:

$$y = \alpha + \sum_{i=1}^{n} \beta_n x_n = 0 \tag{4.2}$$

where,

H_0: Null Hypothesis

y:	Outcome variable (in this report it is dichotomous, zero-one)
α :	y intercept, (constant)
β_n:	Explanatory predictor coefficients
X_n:	Predictor variable names

When there is no evidence of a significant relationship between the response variable and the predictors, the null hypothesis is accepted. Contrasting this is the alternate hypothesis,

$$H_A: \quad y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots \beta_n X_n \neq 0 \tag{4.3}$$

And when summarized takes the following form:

$$y = \alpha + \sum_{i=1}^{n} \beta_n x_n \neq 0 \tag{4.4}$$

where,

H_A:	Alternate Hypothesis
β_n: $\neq 0$	Explanatory predictor parameters with at least one variable β_n

When there is evidence of a significant relationship between the response variable and predictors, the alternate hypothesis is accepted.

Throughout the hypothesis testing process, error has traditionally been quantified through classifications of Type 1 and 2. Type 1 errors are those that arise from rejecting what is true. Existing as a measure of probability, Type I error is designated by the Greek letter alpha (α) and is called the Type I error rate. The smallest error probability this report has allowed itself to accept is 5%, thereby fixing α at 0.05. With this said, rejecting what is true is henceforth fixed at a maximum of 0.05. "Statistical significance", P, is defined by this concept and its value throughout this report. P can be defined as the probability of a more extreme absolute value than the observed value if the true value was zero or null.

This differs from a Type 2 error, where accepting what is false occurs. Designated by the Greek letter beta (β), this error rate is only an error in the sense that an opportunity to reject the null hypothesis correctly was lost. It is not an error in the sense that an incorrect conclusion was drawn since no conclusion is drawn when the null hypothesis is not rejected.

In summary, this paper strives to determine the salient factors that cause a defined response, (either fault or collision occurrence, as in analysis 1 and 2, respectively). The null hypothesis assumes that there is not even a single significant predictor. The alternate hypothesis assumes that there is at least one, possibly even many.

Linear regression is a common way of studying relationships between a dependent variable and independent factor(s). Since the independent variables in this study are not linear, a more appropriate method, namely logistic regression, is used.

4.2 Logits and Odds Ratios

There are statistical computing learning resources provided by the University of California, Los Angelos, (UCLA) [26] that provide a great explanation of the intimate relationship between the coefficients produced by logit and the odds ratios produced by logistic. In this explanation, a logit defined as the log base e (log) of the odds,

$$[1] \quad \text{logit}(p) = \log(\text{odds}) = \log(p/q) \tag{4.5}$$

Logistic regression is explained as simply ordinary regression using a logit response or outcome variable as the response variable,

$$[2] \quad \text{logit}(p) = a + bX \tag{4.6}$$

or

$$[3] \quad \log(p/q) = a + bX \tag{4.7}$$

This means that the coefficients in logistic regression are in terms of the log odds, that is, the coefficient 1.69 implies that a one unit change in gender results in a 1.69 unit change in the log of the odds.

Equation [3] can be expressed in odds by eliminating the log. This is done by taking "e" to the power for both sides of the equation.

[4] $p/q = e^{a + bX}$ (4.8)

The end result of all the mathematical manipulations is that the odds ratio can be computed by raising e to the power of the logistic coefficient,

[5] $OR = e^{b} = e^{1.69} = 5.44$ (4.9)

When reading odds, it is important to recognize that an odds value on its own has little meaning. For instance, if the eldest graduate student cohort's odds of passing a thesis defense was 3, this shows promise that the odds are favourable for success for members of this cohort. However, whether the odds are 3, or 4 or for that matter 10, one can only draw similarly vague "favourably successful" conclusions.

In order to produce more meaningful and comparable results, comparisons to other cohorts for passing a thesis defense could be made. Odds determined from the other younger cohorts could be compared using odds ratios calculated against a given base group. For comparisons against the eldest cohort, the base group would be selected as this eldest cohort and similarly for other comparisons. Throughout the results Chapter of this report, odds ratios are determined with base group rows highlighted. These can also be easily identified as having an odds ratio value (against themselves) of 1.

4.3 Logistic Analysis

Logistic regression was carried out in this analysis through SAS's GENMOD procedure. Although the typical logistic regression in SAS would use the LOGISTIC procedure, this was not ideal for this study since; the data in this report contained categorical (classification) and character explanatory variables. In fact, using the LOGISTIC procedure requires explanatory variables to be numeric, and cannot easily accept categorical variables as they are. It is possible that LOGISITC could have been used. However a significant amount additional coding to convert and essentially construct indicator variables in advance would be required. Furthermore, this process often results in too few observations per parameter for the necessary sample size and is not appropriate for these statistics. These limitations are overcome through the GENMOD procedure, and in particular through its provision of a class statement, as explained later, for specifying categorical (classification) variables. The GENMOD procedure can fit logistic regression models for response data using maximum-likelihood estimation.

The GENMOD procedure fits generalized linear models including, not only classical linear models but also logistic and probit models for binary data, loglinear models for multinomial data, and Poisson regression models for Poisson data [27].

Note that a generalized linear model has three components:

- a response variable [y_i] with some probability distribution, $i = 1, 2, \ldots, n$
- a set of explanatory variables $\mathbf{x}_i$ and parameter vector β
- a monotonic link function g that describes how the expected value of y_i,θ_i, is related to $\mathbf{x}_i'\beta$:

$$g(\theta_i) = \mathbf{x}_i'\beta \qquad (4.10)$$

Consider the relationship of a dichotomous outcome variable to a set of explanatory variables. Such situations can arise from clinical trials where the explanatory variables are treatments, stratification variables, and background covariables; another common source of such analyses are observational studies where the explanatory variables represent factors for evaluation and background variables [28].

The model for θ, the probability of an event, can be specified as follows [28]:

$$\theta = \frac{\exp\left(\alpha + \sum_{k=1}^{t} \beta_k x_k\right)}{1 + \exp\left(\alpha + \sum_{k=1}^{t} \beta_k x_k\right)} \qquad (4.11)$$

Recall that the exp refers to the exponential or raising to the power "e" of a given value or estimate. It is the natural logarithm base, to a power.

It follows that the odds are written:

$$\frac{\theta}{1-\theta}=\exp\left(\alpha+\sum_{k=1}^{t}\beta_k x_k\right)=\exp\text{ (regression model's estimates)}$$

(4.12)

This paper uses a generalized linear model by choosing an appropriate link function and response distribution. In the classical linear model, the probability distribution is the normal and the link function is the identity: $g(\theta) = \theta$. For logistic regression, the distribution is the binomial and the link function is the logit which is linear [27]:

$$g(\theta_i)=\log\left(\frac{\theta}{1-\theta}\right)=\alpha+\sum_{\kappa=1}^{t}\beta_\kappa x_\kappa \alpha+\beta_1 x_1+\beta_2 x_2+\ldots+\beta_i x_i \qquad (4.13)$$

The exp (β_k) are the odds ratios for unit changes in x_k, that is, the amount by which $\theta/(1-\theta)$ is multiplied per unit change in x_k. Recall from the previous section that a close relative of the Odds is the log-odds or logit. One can easily come back from the logit $X = \text{logit}[\theta]$ to the probability, where:

$$\theta=\frac{\exp(X)}{1+\exp(X)} \qquad (4.14)$$

An advantage of using a logit to link the predictors to the outcome is that it transforms an unrestricted interval to a restrctied interval [0,1] (the probability). This is the essence of why GENMOD uses this link function - to link the binary outcome variable to a probability based on the effects of a set of essentially unrestricted independent variables (β_n).

GENMOD uses Newton-Raphson algorithms. These algorithms are commonly accepted by statisticians as computing standard errors that are more conservative than those yielded with other procedures. When using the class statement in PROC GENMOD, the qualitative explanatory variables listed generate a model matrix. Parameterizations of these variables throughout this analysis are incrementally estimated for all levels. By default, SAS utilizes the last level within a class as a reference. Using features such as CONTRAST allows user-defined customized estimates. Throughout this paper, the CONTRAST statement was used to make more relevant comparisons, provided that a given parameter was statistically significant. Details on this and more are covered in the following analysis-focusing sections, complimented with actual programming used for this thesis.

4.3.1 Exclusion of Intercepts

In almost all types of modeling, an intercept is a typical part of the model and is almost always significantly different from zero. For every model constructed, SAS by default tests whether this parameter is equal to zero. If the intercept were zero (equivalent to having no intercept in the model), the resulting model implies that the response function must be exactly zero when all the predictors are set to zero. For an ordinary regression model this means that the mean of the response variable is zero. For a logistic model it means that the logit (or log odds) is zero, which implies that the event probability is 0.5. This is a very strong assumption that is sometimes reasonable, but more often is not. So, a highly significant intercept in a model is generally not a problem. By the same token, if the intercept is not significant, it is not a good idea to remove it from the model because this could create a model that says that the response function must be zero when the predictors are all zero.

Since the nature of this report's models are based on outcomes of fault and intersection collision propensity, the intercept need not be zero. In this case, modeling is such that removal of the intercept is not required, but note that there is essentially no meaning in this intercept. For the intersection modeling, the intercept is a misnomer, because it gives the odds of intersections having collisions which is not as useful since continuous covariates are included. For instance, since some of the covariates used were presence of FAG, presence of SCOOT, and roadway volume, then the predicted odds of the intersection collision propensity would be for those intersections without FAG, without SCOOT, and with a roadway volume of 0. Should it be decided that the constants be removed, SAS has several automatic procedures for this.

The intercept can be useful though, provided that the predictor levels were constructed in a different manner. Any category may be chosen as the reference category, but the results might be easier to interpret if you choose the one that has the lowest proportion on the outcome variable. In such cases of logistic modeling, the intercept would have little practical value in itself, but could still be useful for reconstruction of probabilities of injury for specific groups.

Values of intercepts determined from the models developed in this report can be found in Appendix E: Model Output.

4.4 Analysis 1: MVA Reported Collisions

The data used were three years (1998-2000) of accident data from the City of Toronto Traffic Data Centre and Safety Bureau.

Through descriptive fields selected from Ontario Motor Vehicle Accident Report Forms, collisions meeting particular criteria were selected. Inclusion criteria included:

- accident location at intersections or at least reported intersection related;
- a traffic signal exists at the collision scene;
- the class of accident involves a minimum of some injury component reported at the scene;
- at least one vehicle involved was turning left; and,
- the collision resulted in a side impact with two passenger vehicles.

The focus on this report is on passenger vehicles and their interactions. Therefore, collisions involving trucks, busses, streetcars, motorcycles, pedestrians, cyclists, horses and other road users were excluded.

Other excluded collisions are:

- those "not investigated at scene";
- not side impact;
- those with poles; and,
- hit and runs due to lack of information.

From these data, collision reports were printed and manually further scrutinized to ensure only collisions resulting in a side impact existed. Furthermore, when no clear identification differentiating target and bullet vehicle could be made, the collision was excluded from the study altogether.

One could argue that this is an entire population, since it includes *all* collisions meeting the above criteria for the specified time interval. However, under a stricter sense, the statistical community would consider this claim as having external validity issues. This research paper shall not explore external validity issues in-depth but shall instead carefully state limitations of data and analysis with respect to results. Without being sidetracked with comprehensive discussions on statistical issues such as external validity, one should recognize that in this case, missing elements and entire observations may be a result of other confounding factors which give rise to these issues (of external validity). Examples in this research effort's data could consist of poor recording/absence of detailed scene and collision information at large, complicated intersections, when weather conditions are adverse. Please refer to the section on Limitations in this paper, where the existence of these and other issues are recognized.

For this analysis, the MVA source collision data was used in a different capacity than Analysis I. Since the MVA report's data is based on the existence of a collision and not aggregated by intersection collision frequencies ,this data can be considered disaggregate. As a result, merging data into the linked spreadsheet (as discussed in the above section covering Analysis I) invoked the use of a similar key with "Duplicates" permitted. For each collision, the city identifying “PX” label was manually entered and set as “key” for linking spreadsheets electronically. This enabled intersection data to merge with MVA report data.

The spreadsheet formed from the data available on the selected MVA reports was entered manually and is explained in the two following sections.

4.4.1 Dependent Variables

Separated into two analyses for target and bullet vehicles, the analysis employs driver Fault and Not-At-Fault as single, dichotomous, binary, variables. It is a label that is granted to both target and bullet vehicles, where sufficient evidence warrants. Collisions exist where both drivers are at-fault, neither drivers are at-fault, and combinations within. The modeling methodology used in the analysis takes this into account, and does not necessarily rule out collisions falling outside of the classic ideal of a single driver at-fault coupled with another driver not-at-fault.

Assessing fault is not a simple task. Ontario's Statutes and Regulations as written in Regulation 668 made under the Insurance Act [29] provide a comprehensive document on Fault Determination Rules. The regulation states that "an insurer shall determine the degree of fault of its insured for loss or damage arising from the use or operation of an automobile in accordance with these rules". Interestingly, the degree of fault (of an insured) is determined without regard to:

- The circumstances in which the incident occurs, including weather conditions, road conditions, visibility or the actions of pedestrians; or
- The location on the insured's automobile of the point of contact with any other automobile involved in the incident [29];
- Independently of charges laid by a police officer. A charge under the Highway Traffic Act does not necessarily mean that the person charged was "at fault". In the same way, a lack of charges does not mean that the person was "not at fault" [30].

The regulation illustrates some twenty types of accidents occurring on public highways, and sets out rules dealing with intersection accidents, accidents in parking lots and other matters. In each case the rule specifies the degree of fault to be assessed against each driver. The following excerpt specific to this report and its signalized intersection focus was taken from this Regulation 668, 276/90, Section 15 [29]:

- This section applies with respect to an incident that occurs at an intersection with traffic signals.
- If the driver of automobile "B" fails to obey a traffic signal, the driver of automobile "A" is not at fault and the driver of automobile "B" is 100 per cent at fault for the incident.
- If it cannot be established whether the driver of either automobile failed to obey a traffic signal, the driver of each automobile shall be deemed to be 50 per cent at fault for the incident.
- If the traffic signals at the intersection are inoperative, the degree of fault of the drivers shall be determined as if the intersection were an all-way stop intersection.

Furthermore, Rules for Automobiles Traveling in Opposite Directions also bear a direct relevance to this side impact study, and are summarized from the Regulations as:

When automobile "A" collides with automobile "B", and the automobiles are traveling in opposite directions and in adjacent lanes:

- If neither automobile "A" nor automobile "B" changes lanes and both automobiles are on or over the centre lane when the incident (a "sideswipe") occurs, the driver of each automobile is 50 per cent at fault for the incident. (See Figure 4.2).

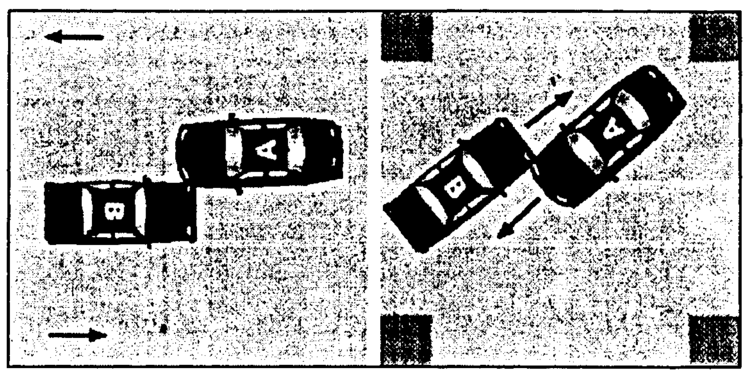


Figure 4.2: Sideswipe Collisions

- If the location on the road of automobiles "A" and "B" when the incident (a "sideswipe") occurs (see Figure 4.2) cannot be determined, the driver of each automobile is 50 per cent at fault for the incident.

- If automobile "B" is over the centre line of the road (Figure 4.3) when the incident occurs, the driver of automobile "A" is not at fault and the driver of automobile "B" is 100 per cent at fault for the incident.

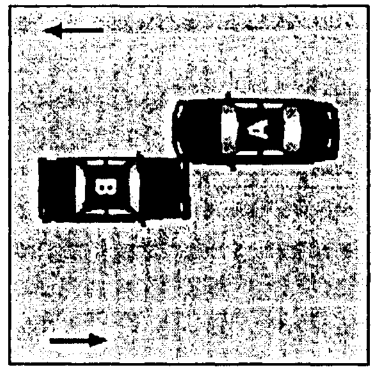


Figure 4.3: Vehicle Over Centre Line

- If automobile "B" turns left into the path of automobile "A", (Figure 4.4) the driver of automobile "A" is not at fault and the driver of automobile "B" is 100 per cent at fault for the incident.

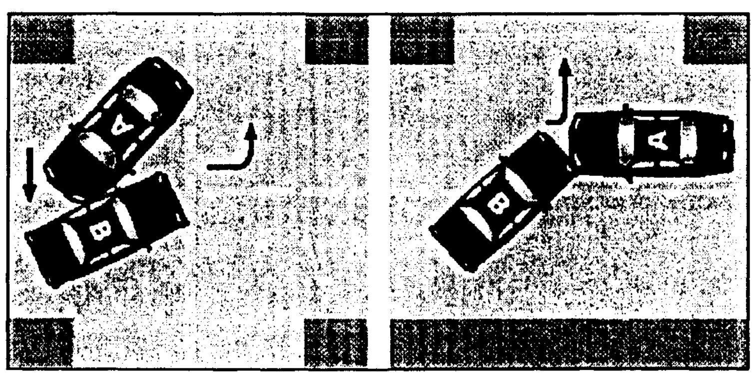


Figure 4.4: Vehicle Turning in Another Vehicle's Path

- If automobile "B" is leaving a parking place or is entering the road from a private road or driveway, (Figure 4.5) and if automobile "A" is overtaking to pass another automobile when the incident occurs, the driver of automobile "A" is not at fault and the driver of automobile "B" is 100 per cent at fault for the incident.

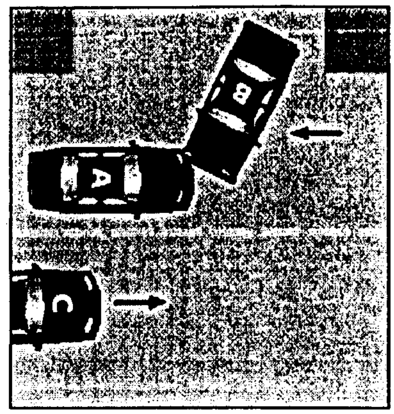


Figure 4.5: Vehicle Collision From Entrance

Other examples include "Rules for automobiles traveling in the same direction in adjacent lanes". This section applies when automobile "A" collides with automobile "B", and both automobiles are traveling in the same direction in adjacent lanes.

If the incident occurs when automobile "A" is turning left at an intersection and automobile "B" is overtaking automobile "A" to pass it, the driver of automobile "A" is 25% at fault and the driver of automobile "B" is 75% at fault for the incident.

The Fault Determination Rules are comprehensive and lengthy, but are not fully discussed in this report. As a whole, the Fault Determination Rules referred to above were followed for this study. Regardless, in every collision, it came down to a human decision, which is subject to error and bias. Following the fault determination rules helped to provide some degree of consistency and is recommended for future use.

Through screening collisions, driver statements, investigating officer opinions, directions of travel, vehicle maneuvers, driver actions, driver conditions and witness accounts (where available) as appearing on motor vehicle accident reports, collision reports, and investigating officer's field notes, were reviewed and faults assessed. These same sources and an identical procedure were employed to differentiate between the target and bullet vehicle. When no clear indication of either being at fault or not at fault existed, the fault observation was left blank. Two variables created, namely D1F and D2F, served as flags for fault for target and bullet vehicle, respectively. There are included in the input statement as labeled in Figure 4.6. For consistency, the designation 1 and 2 has been assigned to variables where identification of target and bullet vehicle, respectively, is required. For instance, the notation "2" found on the fault outcome variable "D2F" indicates bullet, and an identical model for target "1" was also formulated.

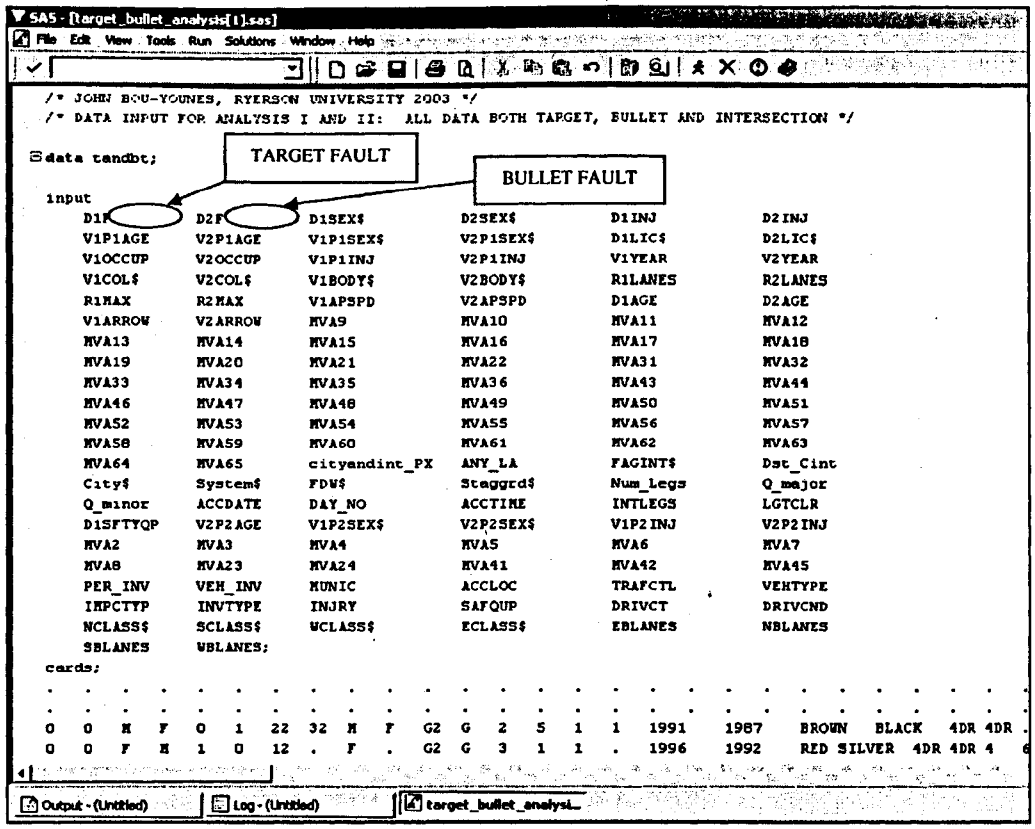


Figure 4.6: SAS Program Editor Window

4.4.2 Independent Variables

Independent variables are the individual factors that can be defined as circumstances contributing to a result (the dependent variable).

As discussed earlier, the prime source of data in this analysis was derived from MVA reports. These reports contained a wide range of variables available existing in many forms; Discrete; continuous; integer; ordinal; nominal; and non-integer data included. For a comprehensive listing of the variables found in the MVA report see Appendix B.

From the Toronto Police MVA data available at the time, a limited number of variables were recorded electronically. Their identity and observed ranges are:

- Number of persons involved (1-7);
- Vehicle inventory: Number of occupants, including driver (1-12);
- Traffic control presence: "Control" being the device type for intersection signaling (1-Traffic Signal, 3-Yield Sign);
- Vehicle type: Note that the VAN vehicle type includes minivan and "work vans" and all passenger vans. Similarly, SUV includes jeeps and 4 door trucks (1-Automobile, 2-Motorcycle, 4-Passenger Van, 5-Pick-up Truck, 6-Delivery Van, 7-Tow Truck, 32-Ambulance, 33-Fire Vehicle, 34-Police Vehicle, 98-Other Truck);
- Vehicle number (1,2);
- Safety equipment use and availability (1-10);
- Driver action (1-14);

  and

- Driver condition (1-8);

In order to increase the scope of this research, the reports were scoured to gather more information while recognizing target and bullet vehicles distinctly, and the following three actions were taken to supplement the existing electronic MVA data with independent variables:

1. Recording information available on MVA reports that are coded but had not been included in the electronic spreadsheet provided, including:

- Environment condition (1-7);
- Light condition (1-8);
- Traffic control (Scoot, MTSS, other);
- Traffic control condition (1-4);
- Road character (1-4);
- Road surface (1-8);
- Road condition (1-3);
- Road surface condition (1-9);
- Road alignment (1-4);
- Road pavement markings (1-4);
- Classification of accident (1-3);
- Initial direction of travel (1-4);
- Initial impact type (1-7);
- Vehicle color;
- Vehicle year (1938-1999);
- Vehicle damage (1-99);
- Safety equipment used (1-10);
- Gender (M or F);
- Age (16-96 years old);
- Seating Position (1-7);

and

- Ejection (1-3).

2. Coding and including other recorded data on MVA's that are explicitly recorded but previously uncoded, including:

    - Day of week (1-7);
    - Time (2400 hrs);
    - City (Etobicoke, East York, North York, Scarborough, Toronto, York)
    - Failed to remain at scene (0,1);
    - Driver's license class (G1, G2, G, GM, Other);
    - Proper license to drive class of vehicle (0, 1);
    - Driver license conditions/suspension (0, 1);
    - Administration of breath/blood test (0, 1);
    - Approximate speed (0-85);
    - Maximum posted speed (15-100);
    - Number of lanes (1-12);
    - Charges laid (0,1);

      and
    - Vehicle towed (0,1).

3. Creating additional fields for data that were not explicitly recorded for coding but became apparent when hand drawn scene diagrams were viewed and investigating officer or self reporting notes/comments/statements were reviewed. These variables include:

    - Number of intersection legs (3,4);

- Existence of entry immediately adjacent to intersection (0,1);
- Intersection perpendicularity (0,1);

  and

- Intersection stagger (0,1);

Note: for the above three actions, distinctions were made for target and bullet in all cases, where applicable.

Clearly, having this data electronically included from the start would be ideal, particularly since it is already recorded; however, police data entry resources are limited. This process of building on existing data is similar to the method used [31] where a review of each report was undertaken to create essentially more data fields that indicate characteristics about a collision that were previously uncoded yet apparent from scanning scene diagrams.

The analysis of Target and Bullet vehicles was performed separately, so that comparisons could ultimately be made between the two. For each of these analysis, at fault vs. not at-fault was the dichotomous dependent outcome variable. The tendency of certain predictor factors being attributed to any of these groups provides insight into differences and commonalities between them but must properly be interpreted.

A mix of different data types exist in the above fields, including:

- Binary: Two possible values, the higher value indicates "true" or "present", the lower value signifying "no" or "absent". Throughout this report 1 and 0 are used, respectively.
- Nominal: Integer codes with no logical sequence. These codes could be used to identify safety equipment used, Driver Action, etc.
- Ordinal: Integer codes having a logical sequence. One such example is injury severity, ranging from 0, none to 4, fatal.
- Continuous: Continuous scale (i.e., Driver's age).

It was intended to compare the binary fault outcome for both target and bullet for each collision. Ideally, one could expect to find statistically proven differences between target and bullet drivers by fault. Throughout the analysis, contrasts are made between classes of binary, nominal and ordinal predictors.

Using hardcopy printouts from the collision reports, in addition to data obtained from the other sources mentioned, the data was entered in MS Excel 2002 as a collection of records. A Pentium 260,088 KB RAM 1.80 GHz CPU computer was used throughout the analysis. Each collision record occupied one row, with variables occupying individual columns. Formatting the data for transfer to Statistical Analysis Software (SAS) Release 8.02 TS Level 02M0 required converting missing observations to periods (".") in addition to maintaining consistent capitalization, particularly between class variables.

For compatibility with SAS, the Excel spreadsheet was converted to the text tab delimited format (as shown in Figure 4.7) and inserted directly into the SAS program editor.

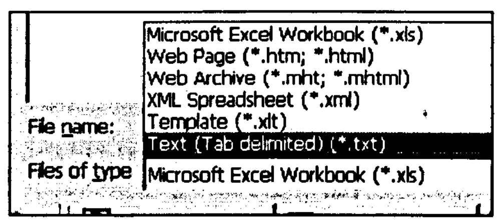


Figure 4.7: Data Saved as Tab Delimited

Programming for SAS necessary to input the data followed the format:

```
data tandbt;
input
                D1F         D2F         D1SEX$ D2SEX$
                D1AGE                   D2AGE
                                          ...(and more variables)
cards;
```

Briefly explained, the data step provides a filename for the storage of the data. In this analysis, the filename chosen was "tandbt.dat". The *.dat extension is automatically added to the file by SAS.

The input statement instructs SAS to identify, in order of appearance, the observations included in the data file. A semi-colon follows this statement. When programming in SAS, a semicolon must appear at the end of any command line, including the final line of the tab delimited data.

After the cards statement, the text tab delimited data can be cut and paste into the SAS program editor window as shown in Figure 4.6: SAS Program Editor Window.

At this point, a "run" statement follows.

After careful examination of the data, it became evident that distinctions could be made between driver, vehicle and environmental factors. Since comparisons by dichotomous fault outcome for Target and Bullet vehicles were desired, SAS required a framework for this logistic type modeling.

For modeling purposes, a meaningful model does not simply include all available variables together and hope for statistically significant results to appear. The benefits of using multiple models for collision data analysis as opposed to a single model are generally accepted. A recent paper found that additional models were advantageous due to it being not possible to account in the model for all the factors that cause differences in accident potential (e.g. weather, geometrics) [32]. With this in mind, careful selection of variables based on findings from past research was used as a basis for initial consideration of variables within model selection.

It was decided that this analysis would separated the driver, environmental and vehicle factors into two: human and extra-human, while maintaining fault as the outcome variable. Model building began with classic hypothesis testing, as reviewed in section 4.0. The null and alternate hypotheses are constructed for human and extra-human modeling as follows.

4.4.3 Human Factors

Given that,

$$H_0: y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots \beta_n X_n = 0 \quad (4.15)$$

Which, when summarized takes the following form:

$$y = \alpha + \sum_{i=1}^{n} \beta_n x_n = 0 \quad (4.16)$$

where,

H_0: Null Hypothesis

y: Outcome variable (in this report it is dichotomous, i.e., zero-one, not at fault, at fault, respectively)

α: Y intercept

β_n: Human explanatory predictor estimated coefficients for (gender, age, etc.)

X_n: Variable names

The null hypothesis is accepted when the all the human factors show no evidence of a significant relationship between the fault indicating response variable, y, where: $y = \beta_1 = \beta_2 = \beta_3 = \ldots = \beta_n = 0$.

Contrasting this is the alternate hypothesis, where the same notations as above apply and

H_A: $$y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots \beta_n X_n \neq 0 \quad (4.17)$$

And when summarized takes the following form:

$$y = \alpha + \sum_{i=1}^{n} \beta_n x_n \neq 0 \quad (4.18)$$

where,

H_A: Alternate Hypothesis

β_n: Human explanatory predictor parameters such as (gender, age, etc.) with at least one variable $\beta_n \neq 0$

When there is evidence of a significant relationship between the response variable (y) and the human factor predictor variables, the alternate hypothesis is accepted.

Recall that PROC GENMOD is a procedure for fitting a somewhat wider range of Generalized Linear Models (GLM) and Generalized Additive Models (GAM), and differs in that the sampling distribution of the dependent variable may be specified as binomial (amongst selected others); and the links between the expected value or probability of Y and the linear additive function of X may be: logistic (amongst selected others). For binary data, these models operate to predict a function of the mean probability of occurrence for all cases having the same scores on X. The outcome for each case is assumed to be from an independent binomial "experiment" (each line of the SAS input used for this study reads in this manner). The expected value of the distribution of the dependent variable (which is binary) is the probability of one of the two outcomes, where the models are predicting the probability of a given (summed up) score in any one trial, conditional on the predictor variables. In this case, the X variables are treated as a linear additive function (i.e. $y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots \beta_n X_n$). Like other linear models, it is possible to transform or link the X's in any way (so long as it remains linear in the b parameters) and include interactions. Unlike classical linear models, the link function may be any of the exponential family distributions (such as binomial) and the link function may be any monotonic differentiable function, (such as logit) [33]. That is, the shape of the curve that relates the predicted score on X (that is, the sum or $\beta_i X_i$) to the expected value of outcome, y.

For the purposes of this research, human factors depend on the individual attributes of driver's and are very limited in number. Fields such as seatbelt use, proper license to drive class of vehicle, driver's license class, age, vehicle speed, passengers etc., are generally accepted to be of interest in such a study. Of these variables, only a select few were included in the initial model as shown in Figure 4.8. For instance, driver injury was excluded, since it is an outcome of a collision and an after-effect not clearly foreseeable prior to the collision.

```
/* First Model Built for Bullet HUMAN Model*/
proc genmod data=tandbt desc;

class    MVA34   MVA36    D2SEX    V2P1SEX    D2LIC;
model    D2F=    MVA34    MVA36 D2SEX    D2AGE    V2P1AGE    V2OCCUP
         V2P1SEX    D2LIC/D=bin link=logit type3;
/*MVA34               7     1 3 4 6 7 8 10 */
contrast '3 Vs 1' MVA34 -1 1 0 0 0 0 0;
contrast '4 Vs 1' MVA34 -1 0 1 0 0 0 0;
contrast '6 Vs 1' MVA34 -1 0 0 1 0 0 0;
contrast '7 Vs 1' MVA34 -1 0 0 0 1 0 0;
contrast '8 Vs 1' MVA34 -1 0 0 0 0 1 0;
contrast '10 Vs 1' MVA34 -1 0 0 0 0 0 1;
/*MVA36               5     0 1 2 6 8 */
contrast '0 Vs 1' MVA36 -1 1 0 0 0;
contrast '2 Vs 1' MVA36 0 -1 1 0 0;
contrast '6 Vs 1' MVA36 0 -1 0 1 0;
contrast '8 Vs 1' MVA36 0 -1 0 0 1;
                      /*1 3 4 6 7 8 10*/
run;
```

Figure 4.8: First Model Built for Human Factors

The "proc genmod" procedure is followed with a descending option statement, "desc". This ensures that the first ordered value is 1. This results in the model being based on probability, P{bullet driver at fault}. The "class" statement in the programming above serves to identify only the class variables. Represented by the abbreviated variable names MVA34, MVA36, D2SEX, D2AGE, V2P1AGE, V2OCCUP, V2P1SEX and D2LIC, the predictors included in this model were driver action, driver condition, driver gender, driver age, front passenger age, number of occupants, front passenger gender, and driver license class, respectively. (These variables were selected as predictors for the model testing based on literature reviewed on findings in other reports, as outlined in this reports chapter 2-literature review, subject to availability within the used data.) The "class" statement can be combined with the "type3" setting to automatically provide likelihood ratio test statistics for the effect of each term in the model. When a Type 3 analysis is requested, PROC GENMOD produces a table that contains the likelihood ratio statistics, degrees of freedom, and *p*-values based on the limiting chi-square distributions for each effect in the model statement [34]. The test will be based on (n-1) degrees of freedom for categorical variables with n levels. The default analysis is to compute likelihood ratio statistics for the contrasts or score statistics for GEEs. Wald statistics are computed if the WALD option is also specified. By default, GENMOD uses a corner point parameterization for categorical variables where the last category of each variable is used as the reference category [35]. One method for specifying a reference category is to define a format for the variable using a space as the first character of the formatted value for all categories except the reference category and specifying the order=formatted option in PROC GENMOD. Since a space is sorted before all other characters, GENMOD will use the desired category as the reference. This parameterization is particularly useful ultimately when constructing desired comparisons for odds ratios.

The "D=bin" text indicates the binomial response probability distribution used. The "link=logit" specifies a logit link function as the means for SAS to link the predictors with the outcome variable, as explained in section 4.3. In addition, the "type3" function was specified in order to take advantage of the likelihood ratio (LR) statistics for the effect of each term in the model through the table shown below in Figure 4.9. Briefly, a Type 3 estimatible function (contrast) for an effect is a linear function of the model parameters that involves the parameters of the effect and any interactions with that effect. A test of hypothesis that the Type 3 contrast for a main effect is equal to 0 is intended to test the significance of the main effect in the presence of interactions [28]. Note that a Type 3 analysis does not depend on the order in which the terms for the model are specified.

LR Statistics For Type 3 Analysis

Source	DF	Chi-Square	Pr > ChiSq
MVA34	5	30.93	<.0001
MVA36	2	9.69	0.0079
D2SEX	1	0.00	0.9649
D2AGE	1	2.84	0.0918
V2P1AGE	1	4.52	0.0334
V2OCCUP	1	0.01	0.9278
V2P1SEX	1	0.43	0.5115
D2LIC	4	4.12	0.3894

Figure 4.9: Type 3 Analysis Statistics

Figure 4.9 can be viewed as serving a similar role to that of an ANOVA table. It includes likelihood ratio tests for each of the parameters effects [27].

Also included within this feature are various default measures for assessing data fit, as shown in the following Figure 4.10. This is important since once the model is applied, an assessment of how well it fits the data are required. This output includes several tests of overall model adequacy which test the global null hypothesis that none of the independent variables in the model are related to changes in probability of even occurrence.

Two traditional goodness-of-fit tests are the Pearson chi-square Q_P, and the likelihood ratio chi-square, Q_L, also known as the deviance. If the model fits, both Q_P and Q_L are approximately distributed as chi-square. In this report, when goodness-of-fit values/Degrees of Freedom (DF) falls below 1.0, they indicate a model with good fit. The deviance is the log likelihood statistic for the difference between this main effects model and the saturated model [27].

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	123	117.1191	0.9522
Scaled Deviance	123	117.1191	0.9522
Pearson Chi-Square	123	133.6206	1.0863
Scaled Pearson X2	123	133.6206	1.0863
Log Likelihood		-58.5596	

Algorithm converged.

Figure 4.10: Goodness of Fit

Tests of the statistical significance of each independent variable are also provided where the criteria displayed are approximate chi-square statistics.

One study discussed in the literature review performed additional chi-square tests or non-parametric tests for each bivariate analysis [14]. In dealing with problematic areas of fit, partial residual plots were used as the evaluation tool. Goodness-of-fit was evaluated by comparing fitted probabilities with observed value of dependent variables within deciles of probability, and calculating the corresponding observed chi-square statistic. In addition, an area under the receiver operator curve for logistic models was calculated to evaluate the models predictive power.

In this report, these additional tests were not explicitly nor manually conducted. However, the goodness of fit statistics described previously and the TYPE 3 function mentioned on the previous page utilizes the Wald test for the interaction terms as a goodness-of-fit test for the main effects model. Odds ratio estimates are all considered alongside in their analysis through 95% Wald confidence limits. So long as the 95% Wald confidence limits do not contain the value 1, these values are automatically considered significant (see Appendix E).

Figure 4.11 shows contrast results that were custom programmed for levels within driver actions as comparisons to a base level "1", namely, driving normally. This was done to avoid potential limitations occurring with SAS default contrasts, where the where the last category of each variable is used as the reference category, which may or may not be relevant for comparison purposes. Manually programmed contrasts such as this are used throughout the analysis and are vital to obtaining meaningful results, complete with their statistical significance (p values).

Contrast Results

Contrast	DF	Chi-Square	Pr > ChiSq	Type
3 Vs 1	1	0.31	0.5782	LR
4 Vs 1	0	.	.	LR
6 Vs 1	1	16.65	<.0001	LR
7 Vs 1	1	.	.	LR
8 Vs 1	1	7.21	0.0072	LR
10 Vs 1	1	3.42	0.0643	LR
0 Vs 1	0	.	.	LR
2 Vs 1	1	0.11	0.7411	LR
6 Vs 1	1	9.27	0.0023	LR
8 Vs 1	0	.	.	LR

Figure 4.11: Contrasts Manually Selected

Parameter estimates can be conceptualized as how much mathematical impact a unit changes in the value of the independent variable has on increasing or decreasing the probability that the dependent variable will achieve the value of one in the population from which the data are assumed to have been randomly sampled.

Note that the likelihood ratio test for a contrast is twice the difference between the log likelihood of the current fitted model and the log likelihood of the model fitted under the constraint that the linear function of the parameters defined by the contrast is equal to zero [27].

Simply looking at this test, successively higher log likelihoods with advancing model iterations can be a measure of success. Of course, other concepts, (such as estimate magnitudes and polarity) must be simultaneously considered.

In Appendix E, SAS output for all of the models developed in this report are comprehensively shown. A selection of all converging and final models used in this are presented and discussed in the Results Chapter.

To demonstrate the actual regression model evolution process, only the analysis titled "First Model Built for Human Factors: Bullet Vehicle", is explained in full detail. The steps used in this analysis are essentially duplicated for all other models built, which serve as exemplary representation of all logistic modeling using GENMOD in this report.

In this first model, with programming shown in Figure 4.8, there were 624 missing values and 140 observations used.

Within 0.75 seconds, SAS calculated the following results (as shown in Figure 4.9):

- MVA34 and MVA36, representing bullet vehicle apparent driver action and driver condition were found to be (highly) statistically significant at $p < 0.0001$ and $p < 0.008$, respectively;
- D2SEX, representing driver gender was not statistically significant ($p=0.965$);
- D2AGE, representing driver age, was found to be statistically significant ($p=0.0918$);
- V2P1AGE, representing bullet vehicle front seat passenger age was found to be statistically significant ($p < 0.033$);
- V2OCCUP, representing number of occupants in the vehicle was not found to be statistically significant ($p=0.928$);
- V2P1SEX, representing front passenger gender, was not found to be statistically significant ($p=0.512$)
- D2LIC, representing driver license class was not sufficiently statistically significant ($p = 0.389$)

Computed estimates and magnitudes are for these results and more are shown in Chapter 5. Selection for an evolving second iteration (see Figure 4.12) followed a backwards-stepwise method, were variables were selected based on the results of the first analysis and specifically their calculated p values. "Backwards" stepwise regression begins with the model including all of the potential independent variables, and successively eliminates those which cost the least in terms of reduction of the coefficient of determination [36].

```
proc genmod data=tandbt desc;

class   MVA34  MVA36   D2SEX   D2LIC;
model   D2F=   MVA34   MVA36 D2SEX   D2AGE   V2P1AGE   V2OCCUP
        D2LIC/D=bin link=logit type3; run;
```

Figure 4.12: Second Step of Backwards Stepwise Regression for Bullet "Human" Predictor's Model

The evolution of this model essentially used criteria similar to the automatic version of the LOGISTIC procedure's backwards stepwise selection process. Using this method, elimination for variables selected in the model occurred when the $p < 0.05$ condition was not met. In this iteration, the least costing predictor in this iteration was V2P1SEX, since it holds the highest p value with $p = 0.857$. In three subsequent iterations, D2SEX ($p = 0.965$), V2OCCUP ($p = 0.930$) and V2P1SEX ($p = 0.516$) (as shown in Figure 4.13: 4th iteration) were removed from the model statement. A 5th and 6th iteration (see Figure 4.14: 6th iteration) were also performed. In the 5th iteration, although driver condition met the 0.05 condition, it was removed in this step from the model on account of over-parameterization. The basis for this stems from the eight classes within this predictor, of which only four levels are found in the observable data. Having many predictors and so few observations intuitively outweighs the benefit of including this predictor, as the concern for over-parameterization arises. This predictor was removed in order to avoid this issue.

```
/* First Model Built for Bullet HUMAN Model*/
proc genmod data=tandbt desc;

class   MVA34  MVA36   D2SEX   V2P1SEX   D2LIC;
model   D2F=   MVA34   MVA36 D2SEX   D2AGE   V2P1AGE   V2OCCUP
        V2P1SEX   D2LIC/D=bin link=logit type3;
/*MVA34             7    1 3 4 6 7 8 10 */
contrast '3 Vs 1' MVA34 -1 1 0 0 0 0 0;
contrast '6 Vs 1' MVA34 -1 0 0 1 0 0 0;
contrast '8 Vs 1' MVA34 -1 0 0 0 0 1 0;
contrast '10 Vs 1' MVA34 -1 0 0 0 0 0 1;
/*MVA36             5    0 1 2 6 8 */
contrast '0 Vs 1' MVA36 -1 1 0 0 0;
contrast '2 Vs 1' MVA36 0 -1 1 0 0;
contrast '6 Vs 1' MVA36 0 -1 0 1 0;
contrast '8 Vs 1' MVA36 0 -1 0 0 1;
                    /*1 3 4 6 7 8 10*/
run;
```

Figure 4.13: 4th iteration

These predictors were therefore backward stepwise eliminated to consist of the remaining apparent driver action (MVA34), front seat passenger age (V2P1AGE) and driver age (D2AGE), as found the 6th in iteration below.

```
proc genmod data=tandbt desc;

class   MVA34;
model   D2F=   MVA34   D2AGE   V2P1AGE
        /D=bin link=logit type3;
/*MVA34             7    1 3 4 6 7 8 10 */

contrast '3 Vs 1' MVA34 -1 1 0 0 0 0 0;
contrast '6 Vs 1' MVA34 -1 0 0 1 0 0 0;
contrast '8 Vs 1' MVA34 -1 0 0 0 0 1 0;
contrast '10 Vs 1' MVA34 -1 0 0 0 0 0 1;

run;
```

Figure 4.14: 6th iteration

Having fewer predictors in the model allows for increasingly more available models since the criteria requiring all model predictors to exist is reduced from seven variables to two. As a result, 500 observations were used with only 264 missing values.

Realistically, there is no perfect answer just as there is no monopoly of knowledge. The 4th, 5th and 6th iterations produce models that are stable and converge under SAS's rather stringent type 3 analysis criteria. Perhaps the 6th iteration can be considered superior since it is more composed of predictors with classes having met the $P < 0.05$ criteria. When further iterations were continued, SAS issued warnings indicating that the model no longer converged. Stripping predictors in this backwards regression can increase propensity to fail algorithm convergence and collapse models, particularly when so few predictors remain and so many other confounding factors are left out. Attempting to use extremely few predictors to model any complex situation is poor practice.

Results from this model and other final converging models are provided in Chapter 5.

4.4.4 Extra-Human Factors

H_0: $$y = \alpha + \square 1X1 + \beta_2 X_2 + \beta_3 X_3 + \ldots \beta_n X_n = 0 \quad (4.19)$$

When summarized takes the following form:

$$y = \alpha + \sum_{i=1}^{n} \beta_n x_n = 0 \quad (4.20)$$

where,

H_0: Null Hypothesis

y: Outcome variable (in this report it is dichotomous, zero-one, not at fault, at fault, respectively)

α: Y intercept

β_n: Extra Human explanatory predictor parameters coefficients (for road surface condition, etc.)

X_n: Predictor variable names

The null hypothesis is accepted when the all the extra-human factors show no evidence of a significant relationship between the fault indicating response variable, y, where:

$$y = \beta_1 = \beta_2 = \beta_3 = \ldots = \beta_n = 0. \qquad (4.21)$$

Contrasting this is the alternate hypothesis, where the same notations as above apply and

H_A: $$y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots \beta_n X_n \neq 0 \qquad (4.22)$$

And when summarized takes the following form:

$$y = \alpha + \sum_{i=1}^{n} \beta_n x_n \neq 0 \qquad (4.23)$$

where,

H_A: Alternate Hypothesis

β_n: Extra-Human explanatory predictor parameters such as (road surface condition, age, etc.) with at least one factor $\beta_n \neq 0$

When there is evidence of a significant relationship between the response variable (y) and the Extra Human factor predictor variables, the alternate hypothesis is accepted.

The variables selected having little or no predictive association with "human interaction" and which were used as predictors for the initial "extra-human" regression modeling were represented by the abbreviated variable names RLANES, RMAX, VAPSPD, VBODY, VYEAR, VCOL, MVA10, MVA14, MVA16, MVA18, MVA 20 MVA32 and MVA47, representing number of lanes, maximum posted speed, approximate vehicle speed, body style, vehicle manufacture year, vehicle colour, road character, road surface, road condition, road surface condition, road alignment, road pavement markings and vehicle maneuver, respectively. These variables were selected as predictors for the model testing based on literature reviewed on findings in other reports, as outlined in this reports chapter General literature review, subject to availability within the used data.

Since the outcome variable for this analysis was fault, and the predictors selected here were extra-human, the results of this modeling aimed at finding factors associated with fault that are just teetering beyond the absolute control of drivers. For instance, the vehicle colour may have been chosen prior to the collision which can be considered weakly associated with the absolute control of the driver, at the *time of collision.* The purpose of this predictor being modeled in this way would be to determine various colour involvements and propensity in side impact collisions through target and bullet comparisons made.

4.5 Analysis 2: Intersections With and Without Collisions

For the second analysis using intersection data, the Queries tool in Microsoft Access was used to link intersection data in spreadsheet format. Synchronization of the data fields was maintained through the "key" feature, where in any given spreadsheet at least one variable (and its entire column of records) must be selected as "key". Since the first part of the analysis proposed uses a single dependent variable based on occurrences of collisions at each given intersection, the intersection "PX" identification numbers (as exist in the data sets) were set as key. These ID numbers are unique for intersections within Toronto, and to avoid repetition and other potential errors within the various data sets, all key ID numbers' attributes were set to "indexed" and "No Duplicates". Note that this portion of data is considered aggregate since they are aggregated by intersection number, and accident counts/frequencies.

For the intersection analysis, the dependant outcome variable in this analysis was chosen as the intersections having no collisions of any kind within the three years of available data for left-turn and angle classifications. This included collisions that were property damage only PDO, injury, and fatality.

4.5.1 Dependent Variable

The dependent outcome variable followed a typical binary coding as 0 and 1 for (1) collision or (0) non-collision experiencing intersection for years 1998, 1999 and 2000.

In this case, the odds of a collision occurring are defined as:

$$\frac{\text{Pr (collision at intersection)}}{\text{Pr (No collision at intersection)}} = e^{\alpha + B_1X_1 + B_2X_2 + \ldots + B_nX_n} = e^{\alpha} e^{B_1X_1} \ldots e^{B_nX_n} \qquad (4.24)$$

There were also two other fields provided in the data with flags for collisions with:

- Injuries and fatalities combined; and,
- injuries exclusively.

These have potential to be used in combination with the data to yield differences in intersection features that vary with intersections experiencing different level of collision severity, namely, PDO, injury, fatality, and the selected combinations mentioned. Differences found between these combinations may help identify areas of concern for intersection safety. However, these differences are not specifically explored due to the exceedingly small number of fatalities found in the data. As a result, only modeling comparisons are made for intersections with and without injury collisions of any kind.

4.5.2 Independent Variables

Intersection data consists of the following potential factors:

- Intersection PX reference number

- `System$`: Intersection control system, essentially a flag for the presence of SCOOT (Split Cycle Offset Optimization Technique) or MTSS (City's Main Traffic Signal System)
- `FAGINT$`: Presence of Flashing Advance green (FAG)
- `FDW$`: Presence of FDW (Flashing Don't Walk)
- `Num_Legs`: Number of legs at intersection
- `Staggered$`: Staggered intersection alignment
- `City$`: municipality or area of Toronto
- `Q_major Q_minor`: Major and minor roadway 24 hour peak volumes, respectively
- `DstInt`: Distance to closest adjacent intersection
- `Any_LA$`: Presence of Left turn Arrow (LA)
- `N$, S$, W$, E$`: Road classification by each direction

The backwards stepwise regression procedure used in the target and bullet fault Analysis 1 of this chapter was used in an identical manner for this analysis on intersections with and without collisions. Similarly, results from this model and other final converging models are provided in the following Chapter 5.0

5 Preliminary Investigation - Results

5.1 General

The results are presented in two separate sub-sections where simple summary statistics and calculated odds ratios are featured and are followed with the more technical findings from the logistic regressions

5.2 Target and Bullet Fault Comparison

The categories along with their frequencies and percentages are shown in Table 5.1. Any missing frequencies are not included in the percentage calculations. However, the frequency of missing data should be observed so as to grasp a sense of which fields tend to experience highly missing observations. The effect of missing data is described in the limitations section of this report.

	Frequency						
Category	Target		Bullet		Crude Odds:	Crude Odds:	Crude Odds Ratios:
	Frequency	% of Total	Frequency	% of Total	Target	Bullet	Target VS Bullet
Fault							
At-Fault reference	**370**	**69.5**	**171**	**32.0**	**2.28**	**0.47**	**1.00**
Not-At Fault	162	30.5	363	68.0	0.44	2.12	0.21
Frequency missing	232		230				

Table 5.1: Fault Comparisons for Target and Bullet Drivers

Crude odds (CO) reveal that target drivers are associated with being at fault, while bullet drivers are not at fault. The convention of this report uses the highlighted rows, in this case, "At-Fault reference", indicative of the base odds category for use in constructing crude odds ratios (COR). The base category will always have an OR = 1.0, since this is a product of the cancellation from dividing the numerator and base denominator. The odds ratio immediately above reveal that there is a low tendency shown (COR = 0.21) for target vs. bullet being not-at fault compared with at fault. This shows that target vehicles are inclined to be at fault (with CO = 2.28) as compared with bullet vehicles (with CO = 0.47), all other things assumed equal. Recall that these OR are essentially only crude summary statistics. More meaningful results are possible through the regression modeling as explained in the previous chapter, and the results of these regressions are explicitly shown in the following pages.

5.3 Human Factors

Beginning with Table 5.2, with predictors considered, each of the variables used in the regression modeling are shown. From a simplistic summary perspective, driver gender revealed that males are more inclined to be drivers for both target and bullet vehicles involved in the collisions under study, with OR greater than 1.0 calculated as 1.28 and 1.96 respectively. When target vs bullet vehicles are compared by driver gender, female drivers vs. male drivers yield an OR of 1.53. Simply stated, females in the data sampled appear to have a higher involvement in these collisions as Target drivers. However, under careful consideration of other predictors from both the target and bullet regression modeling process, this predictor was deemed not statistically feasible for model inclusion ($Pr < 0.05$ criteria not met) and was removed in the 4th and 2nd iterations, respectively (see Appendix D for modeling text and regression output).

Category	Frequency: Target Frequency	Frequency: Target % of Total	Frequency: Bullet Frequency	Frequency: Bullet % of Total	Crude Odds: Target	Crude Odds: Bullet	Crude Odds Ratios: Target VS Bullet
Driver Gender							
Male	374	56.1	438	66.2	1.28	1.96	1.00
Female	293	43.9	224	33.8	0.78	0.51	1.53
Frequency missing	97		102				
Driver Age Group							
15-24	126	19.1	167	25.3	0.24	0.34	0.51
25-34	177	26.8	157	23.8	0.37	0.31	0.76
35-44	144	21.8	139	21.1	0.28	0.27	0.70
45-54	88	13.3	104	15.8	0.15	0.19	0.57
55-64	58	8.8	46	7.0	0.10	0.08	0.85
65+	68	10.3	46	7.0	0.11	0.08	1.00
Frequency missing	103		105				
Number of Occupants							
1	438	65.2	426	63.7	1.87	1.76	1.00
2	155	23.1	151	22.6	0.30	0.29	1.00
3	49	7.3	60	9.0	0.08	0.10	0.79
4	23	3.4	23	3.4	0.04	0.04	0.97
5+	7	1.0	8	1.2	0.01	0.01	0.85
Frequency missing	92		97				
Driver License Class							
G1	11	1.7	12	1.8	0.02	0.02	0.90
G2	130	19.6	127	19.3	0.24	0.24	1.00
Gm_any	8	1.2	11	1.7	0.01	0.02	0.71
G	490	73.8	481	73.0	2.82	2.70	1.00
other	25	3.8	28	4.2	0.04	0.04	0.88
Frequency missing	100		105				
1st Passenger's Gender							
Male	72	33.3	83	39.5	0.50	0.65	1.00
Female	144	66.7	127	60.5	2.00	1.53	1.31
Frequency missing	548		554				
1st Passenger's Age Group							
15-24	57	31.3	58	32.4	0.46	0.48	1.00
25-34	40	22.0	42	23.5	0.28	0.31	0.97
35-44	31	17.0	28	15.6	0.21	0.19	1.13
45-54	20	11.0	19	10.6	0.12	0.12	1.07
55-64	13	7.1	16	8.9	0.08	0.10	0.83
65+	21	11.5	16	8.9	0.13	0.10	1.34
Frequency missing	582		585				
Apparent Driver Action							
1-Driving properly	213	33.5	398	61.9	0.50	1.62	1.00
3-Exceeding speed limit	6	0.9	11	1.7	0.01	0.02	1.02
~~4-Speed to fast for condition~~	~~2~~	~~0.3~~	~~6~~	~~0.9~~	~~0.00~~	~~0.01~~	~~0.62~~
6-Improper turn	135	21.3	48	7.5	0.27	0.08	5.26
~~7-Disobeyed traffic control~~	~~41~~	~~6.5~~	~~126~~	~~19.6~~	~~0.07~~	~~0.24~~	~~0.61~~
~~8-Failed to yield right-of-way~~	~~238~~	~~37.5~~	~~54~~	~~8.4~~	~~0.60~~	~~0.09~~	~~6.24~~
Frequency missing	124		115				
Driver Condition							
1-Normal	575	87.7	604	91.0	7.10	10.07	1.00
2-Had been drinking	9	1.4	10	1.5	0.01	0.02	0.95
3-Ability impaired, alcohol > 0.8	4	0.6	0	0.0	0.01	0.00	-
4-Ability impaired, alcohol	0	0.0	1	0.2	0.00	0.00	0.00
6-Fatigue	2	0.3	2	0.3	0.00	0.00	1.05
8-Inattentive	66	10.1	47	7.1	0.11	0.08	1.48
Frequency missing	101		97				

Table 5.2: Human Predictors Considered

Driver age was readily available in the collision data used. Ages were grouped into cohorts as shown in Figure 5.1 and the cohorts were chosen following similar formats of ORSAM [23] and the "An Evaluation of Severity And Outcome Of Injury By Type Of Object Struck (First Object Struck Only) for

Motor Vehicle Crashes in Connecticut" [14], which used similar groupings. This was purposely done so that comparisons could be made between these reports. For target vehicles, the odds seem most favourable for collision involvement in the 25-34 cohort. Bullet vehicles appear to have higher odds in the younger 15-24 cohort, which supports the commonly accepted notion of young, inexperienced drivers having high collision involvements. With growing age, speculation on increased maturity levels and driver experience seem to correspond with lower odds values. A slight increase in the odds occurs in the 65+ cohort. One might speculate that this is a phenomena associated directly with decreasing mental and physical abilities that are inevitable with aging. The OR set with the 65+ base cohort indicates notably lower values for the 15-24 and 45-54 age categories, as compared to all other cohorts. One might speculate that this phenomena could be explained by the most immature and inexperienced nature of the youngest cohort, while this wouldn't apply to the 45-54 cohort. Perhaps experienced drivers feel "too comfortable" and perceived risk phenomena occur. Furthermore, recall that measures for exposure and other limitations must be considered particularly since certain cohorts may experience significantly different exposures, which are not accounted for completely in the quasi-induced exposure method used in this report.

However, under careful consideration of other predictors from both the target and bullet regression modeling process, driver age was not statistically feasible for model inclusion (as Pr < 0.05 criteria not met) and was eliminated during the backwards regression process. However, driver age for the bullet vehicle is considered marginally significant since Pr = 0.074, and through inspection of the measures afforded by the stringent type 3 analysis in SAS, its continued convergence of the model and improvement in model chi-square statistics, it was decidedly included. The estimate for this predictor as a whole equals 1.03, (equivalent to odds of exp(1.03) = 2.80) which is a result that is interpreted differently from an OR. This estimate is for the predictor as a whole, and not for any class comparison within. As such, only a very general observation can be made on its positive magnitude, which suggests that increasing age is associated positively with being at fault for bullet vehicles. No additional comments can be legitimately made since statistical significance was not achieved for selected class comparisons within this predictor, but potential for further study of this predictor exists through the construction of different class contrasts, set with different base classes for comparison. It is entirely possible that statistical significance can be discovered within different contrasts for this marginally statistically significant predictor.

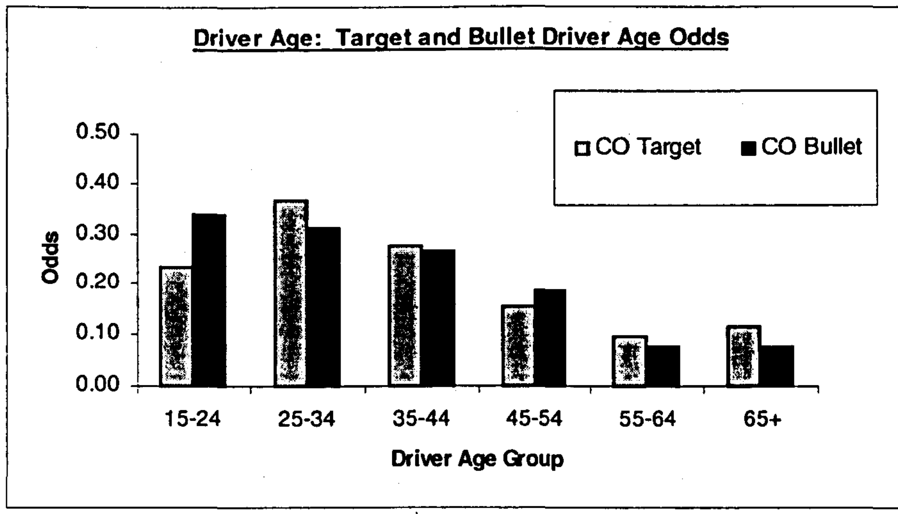


Figure 5.1: Target and Bullet by Age of Driver Odds

Findings based on the number of occupants for target and bullet vehicles revealed both target and bullet vehicles shared similar occupancy rates and showed a similar decreasing trend in Odds with increasing occupancy (Figure 5.2). The gross majority of the vehicles (over 60%) had only a single occupant at the time of collision. One possible conclusion drawn from these results would suggest that number of occupants has a beneficial effect since fewer incidences occur with increasing occupancy. However, exposure issues must be considered since situations where number of occupants exceeds a single occupant are likely far less frequent and are not accounted for in this data.

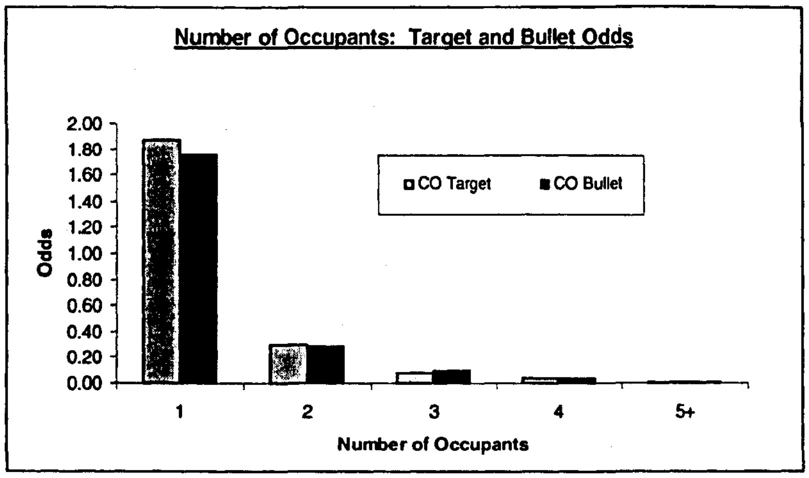


Figure 5.2: Target and Bullet by Number of Occupants Odds

When occupancy was considered in the regression model combined along with all other "human" predictors, the number of occupant's predictor was determined to be statistically significant (Pr < 0.039) in the bullet model but not statistically significant for the target model. This predictor is, therefore, also included in the bullet model. The estimate for this predictor in the bullet model is 1.95, which suggests an increased propensity for fault exists with an increasing number of occupants.

Driver's license class summary statistics indicate an overwhelming majority of fully (G) licensed driver's involvement in the selected collisions. Crude odds (CO) for target and bullet drivers amongst the various license classes (G1, G2, any GM combination and full G licensed drivers) are shown in Figure 5.3. The CO for G licensed drivers is significantly greater than any of the other classes. This may be attributed as a direct result of exposure, in addition to the sheer number of G licensed drivers as compared to all other classes.

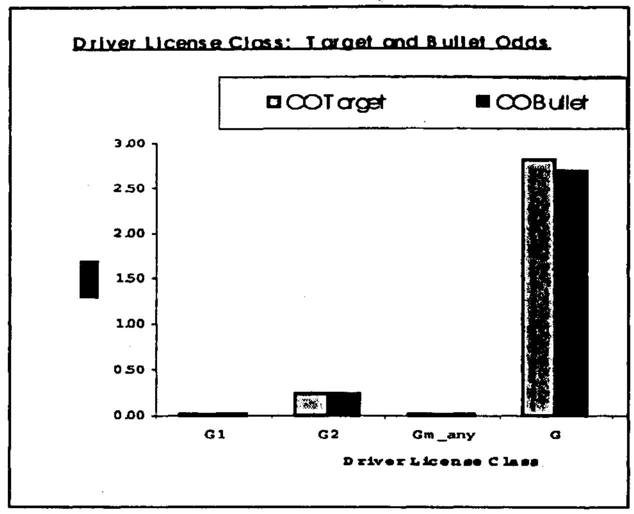


Figure 5.3: Target and Bullet by Driver License Class Odds

When included in the "human" regression model with the other predictors, driver's license class was determined to be not statistically significant and was eliminated from the model.

The gender of the front seated passenger was also included in the study. Examining the CO in Figure 5.4 individually for male vs. female show a three to four-fold increase across both target and bullet vehicles. For female vs. male passenger presence in target vs. bullet vehicles, a COR of 1.31 was determined. This finding suggests female presence has an increased incidence of collision. Perhaps the predominantly male driving population is more easily distracted by font seat passengers who are members of the opposite sex. However, when this predictor was included in the regression model along with other "human" variables, it was determined to be not statistically significant.

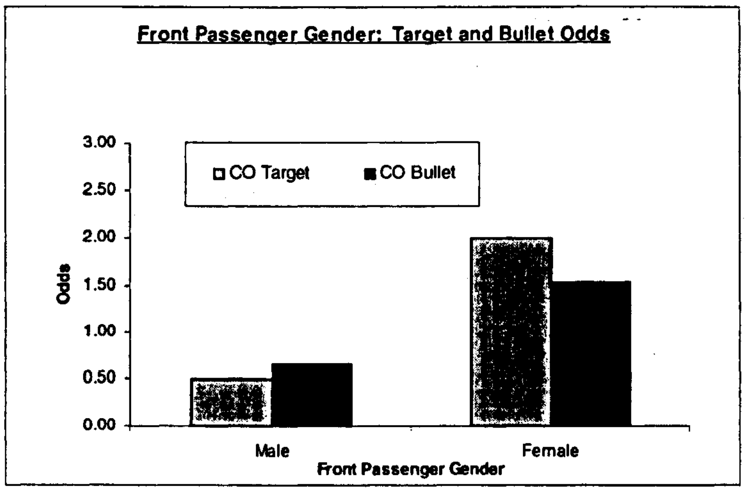


Figure 5.4: Target and Bullet by Front Passenger Gender Odds

The age of passengers seated in the front seat was also considered in this study as part of the human variables. The CO determined for all cohorts were significantly less than 1.0 (Figure 5.5) and this is indicative of a low propensity for there to be front seat passengers of any age in target and bullet vehicles.

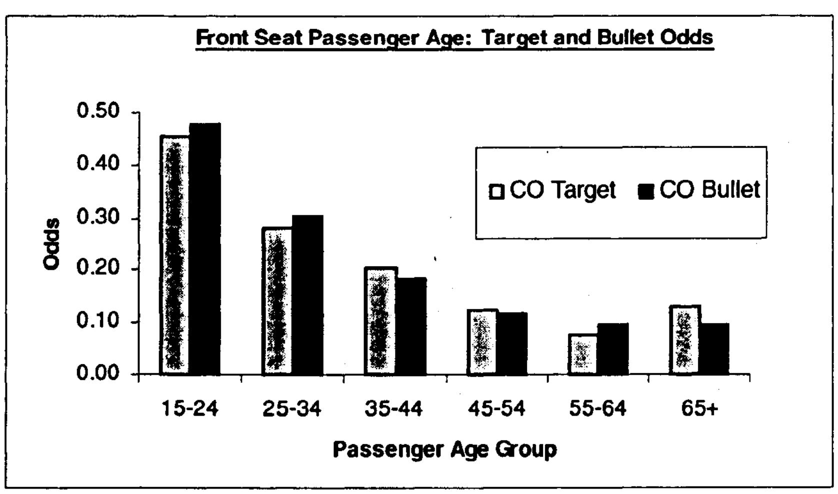


Figure 5.5: Target and Bullet by Age of Front Seat Passenger Odds

When considered in the human regression model, this predictor was marginally statistically significant ($Pr < 0.087$) for bullet vehicles and was included in the model. The regression revealed that all contrasts constructed within this predictor did not yield any statistically significant values. Therefore, no OR values for corresponding contrasts were available. However, the predictor as a whole yielded an estimate of -0.027 (odds = exp(-0.027) = 0.97), which indicates in (negative) polarity and (small) magnitude that increasing age is slightly associated with decreasing fault. This compliments the findings developed from the CO shown above.

For the target vehicle model, this predictor was determined to be not statistically significant (Pr < 0.379) as was eliminated from the model.

Apparent driver actions were also examined. From the values in Table 5.2, it is all to easy to draw immediate conclusions without diligence. For instance, the OR value of 1.02 indicates no appreciable difference between exceeding the speed limit and the base condition of driving normally. However, due to the extremely low frequencies (6 and 11 for target and bullet, respectively) the values of this class explain too little and are of little meaning in the scope of things. Another class within this predictor, titled "Speeding to fast for condition" suffers from this same issue. In fact, a handful of other classes available in the data were removed altogether from Table 5.2 due to extremely low and zero frequencies. Also, several of the levels within this predictor were deemed to be too highly associated with the outcome variable. For instance, "Disobeying a Traffic Control" is directly associated with being at fault. These variables are shown in Table 5.2 with a strikethrough across them. When these classes were included in the regression model for "human" factors, both target and bullet models eliminated this class as Pr = 0.998 and were even deemed "non-estimatible" by SAS through a lack of convergence.

Improper turns as compared with driving properly for target vs bullet yielded a COR=5.26 through summary statistics, with odds from the summary statistics of 0.27 and 0.08. From the summary statistics, this class seems problematic for target vehicles. From manually screening and entering each collision, officer's field notes repeatedly expressed a similar sentiment. This notion is further evidenced by the higher propensity for target drivers to be at fault, as apparent in Table 5.1. The regression modeling target and bullet "human" model odds = 43.41 and 18.16 with Pr = 0.001 and Pr = 0.001, respectively. From these fault regression model values, both target and bullet, suffer from high odds. Furthermore, through taking the OR of these two odds, it can be shown that targets suffer from this fault class comparison more than double (43.4:18.2 = 2.4) the odds that bullets do.

When another class within driver actions was considered, namely disobeying traffic control, the opposite result was shown. This class as compared with driving normally for target vs. bullet yield a COR=0.61. Thus, a lower propensity for target driver's to disobey traffic controls exist in the data. It is easy to conceptualize that, for a large part, disobeying traffic controls is common in the bullet vehicle. For instance, imagine a classic side impact collision where a bullet vehicle runs a red light and strikes a target vehicle. Since the above OR indicates that there is a lesser propensity for these to occur on target drivers a greater propensity exists for these to occur on bullet drivers. The regression results for this class of the predictor compared with the base class of "Driving Properly" are highly statistically significant as P = 0.024 and P = 0.001 and for target and bullet respectively. The corresponding odds values determined from the regression modeling are 6.49 and 11.84 respectively. The OR from these (6.5:11.8 = 0.54) also indicates a lower propensity for targets to be at fault under these conditions. Most likely this is attributed to the intimate relationship that may exist with fault and driver actions that are "improper", "disobeying", "failed" etc. The connotations of some of the classes within this predictor potentially serve as partial indicators of fault outcome, which may overwhelmingly serve as a powerful individual class factor. However, this could also be attributed to other confounding factors which may cloud the values of the salient factors. In addition, this may very well be a rare case statistical significance achieved by chance alone. Regardless, it is a class within this predictor that is debatable.

When the "Failure to yield right of way" class was examined against driving normally for target vs. bullet a COR=8.24 was determined. There appears to be a concern with target drivers failing to yield right of way which creates an inclination to be caught in the path of other (bullet) vehicles. From this target drivers seem to frequently be in *the wrong place at the wrong time*. Combining this finding with the bullet driver's tendency indicated above to apparently disobey traffic controls spells a recipe for disaster. From the regression results, an OR = 6.94 was determined for target vs. bullet for "failure to yield right of way" vs. "driving normally". This indicates a lower propensity for target vs. bullet in this contrast,. Similarly to the previous two class levels within this predictor, it is believed that this class serves strongly as an indicator of fault, which may overwhelm the fault outcome used in the regression. As a result, this contrast is also considered debatable.

The final predictor considered in the "human" regression modeling was Driver Condition. From the summary statistics of Table 5.2, essentially all of the collisions (87.7% and 91.0% for target and bullet, respectively) fell within the "Normal" base class. Only a handful of more inattentive target drivers as compared with bullet drivers were found, and this resulted in a OR = 1.48. When this predictor was considered in the logistic regression modeling for both target and bullet, it did not meet the statistical criteria where $Pr < 0.05$ and was therefore backwards stepwise eliminated.

Table 5.2 shows the Human Variables simple frequency summaries and crude OR with all predictors considered in the "Human" logistic regression modeling.

From the above results and discussion, there appears to be certain faults and human predictors which are more attributed to target, while others are more attributed to bullet. Further identification and focus on these can direct efforts for future researchers to help increase safety from both the target and bullet driver's perspective.

5.4 Extra-Human Factors

Summary frequencies and calculated OR for variables having little or no predictive association with human interaction and which were used as predictors for the initial "extra-human" regression modeling are shown in Table 5.4.

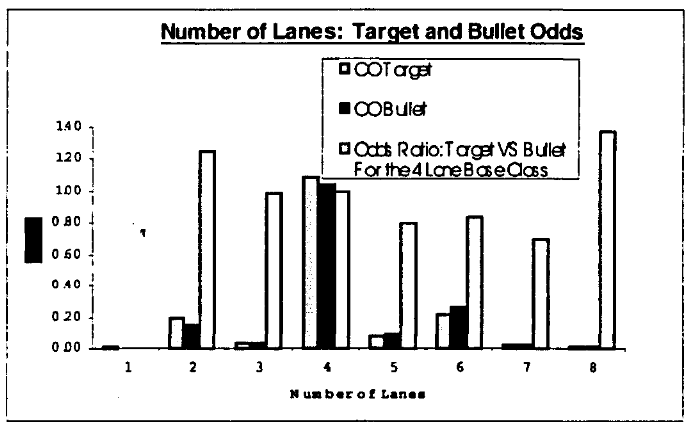


Figure 5.6: Target and Bullet by Number of Lanes Odds

In the selected collisions, Both Table 5.4 and Figure 5.6 show that target and bullet have virtually identical distributions. Unlike other figures used to convey odds up to this point in the report, COR have also been included on the figure above for convenience. When the number of lanes was included in the regression analysis as a class variable along with other extra-human variables, it was determined to be a highly statistically significant predictor (Pr = 0.001) for the target model only, with odds = 1.29 . However, when additional contrasts were programmed in the regression modeling to measure the individual effects of particular classes, none of the classes were individually estimateable.

Vehicle body style tended to be predominantly 4 door for both target and bullet vehicle.

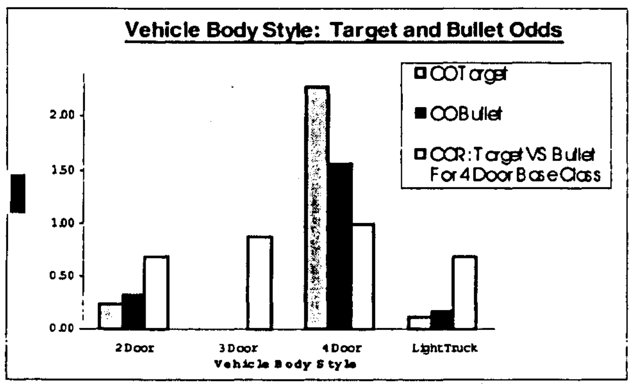


Figure 5.7: Target and Bullet by Vehicle Body Style Odds

COR values are all less than 1.0 for target vs. bullet vehicle body styles when compared with the base 4 door class. Propensity for 4 door vehicles to act as both target and bullet vehicles in these collisions appears to rank highest.

Of the nine classes of vehicle colours selected, silver was most common throughout and was selected as a base class for comparison. Black yielded rather similar (nearest COR to 1.0) to the base silver (with COR = 0.96), while other colours showed higher propensities for collision involvement. Vehicle colours in order of ascending COR are brown (COR = 1.19), blue (COR = 1.35), maroon (COR = 1.48), white (COR = 1.48) red (COR = 1.51) and green (COR = 1.52).

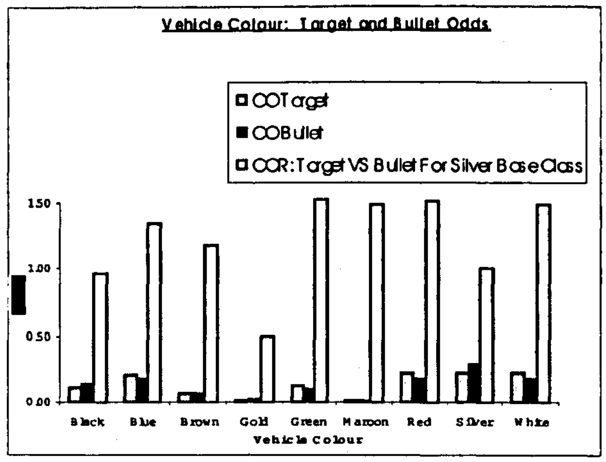


Figure 5.8: Target and Bullet by Vehicle Colour Odds

When vehicle colour was included in the extra-human regression modeling for both target and bullet models, it failed to allow the model to converge under SAS's stringent type 3 analyses, and was therefore eliminated from the models. As expected, measures of goodness of fit increased after this variable was eliminated from all models. Perhaps this can be attributed to over-parameterization issues associated with the (rather large number of) nine separate classes. Having many descriptors and a limited number of observations creates great concern as models struggle to converge. This area of concern is discussed in the limitations section of the report.

Road character and corresponding COR determined from the frequency summary is shown in Figure 5.9. Undivided one-way roads in addition to roads without a qualifying barrier have COR marginally greater than 1.

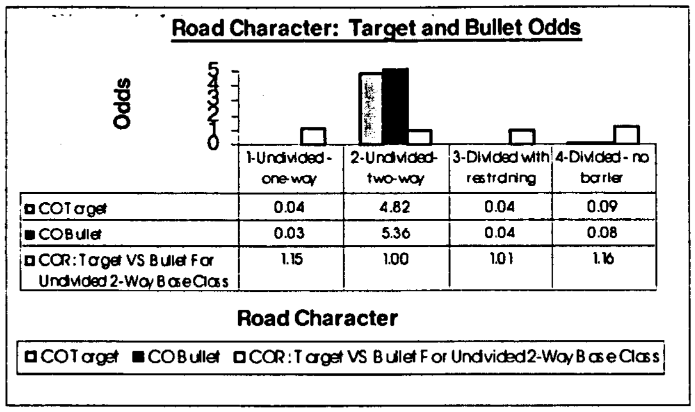


Figure 5.9: Target and Bullet by Road Character Odds

When road character was included in the regression models, both target and bullet models found this predictor to be marginally statistically significant, with Pr = 0.053 and Pr = 0.064 as highlighted in Table 5.3, respectively.

	BULLET					TARGET				
PREDICTOR	DF	Chi-Square	Pr > ChiSq	estimated value	EXP of est = ODDS	DF	Chi-Square	Pr > ChiSq	estimated value	EXP of est = ODDS
Road Character	4	9.33	0.0534	-	-	4	8.90	0.0637	-	-
1-Undivided -one-way	-	-	-	-	-	-	-	-	-	-
2-Undivided-two-way	1	0.88	0.3486	0.7689	2.16	1	0.02	0.8950	0.0768	1.08
3-Divided with restraining barrier	1	0.28	0.6336	-0.4934	0.61	1	0.85	0.8456	0.1548	1.17
4-Divided - no barrier	1	2.08	0.1517	1.2846	3.61	1	0.49	0.4928	-0.4651	0.63

Table 5.3: Extra-Human Regression Model Results for Road Character on Target and Bullet

Through highlighted rows, as similar to the format consistent throughout this report, Table 5.3 shows the undivided-two-way class has been chosen as the base class. The individual classes had contrasts (essentially comparisons) programmed and their resulting estimates between all classes vs. the base class are shown. However, the statistics associated with these are not statistically significant, as Pr < 0.05 criteria is not met. In order to clarify, the predictor as a whole is statistically significant, but the contrasts programmed within its levels all yield OR with poor statistical significance.

Frequencies and COR for Road Alignment shown in Table 5.4 indicate that the straight-on-level has been set as base class and exists for over 88% of the selected collisions. COR = 1.14 for the straight-on-hill class contrasted with the base class suggests that only a slightly higher propensity for target vs. bullet fault vehicles exists. Other classes for this predictor are obscenely too infrequent and the analysis of these variables would likely cause more harm than good. As discussed in the limitations section, Simpson's Paradox and other issues associated with rare occurrences tends to discredit any results when using OR as a measurement tool. This is simply one of the limitations within the realm of Odds and OR. A significantly larger database may help overcome this issue.

Road alignment was included as a predictor in the extra-human target and bullet models. Like the majority of other predictors in this report, it too was determined to be not statistically significant and was backwards stepwise eliminated in the 5^{th} and 9^{th} iterations for target and bullet models, respectively.

Vehicle Maneuver fell almost entirely into two classes, namely Going ahead and Turning left. This is a direct result of the selection criteria for these collisions, where left-turning side impact collisions were sought (as discussed in

Chapter *Analysis 1: MVA Reported Collisions*). The sum of the left turning percentage distributions of 78.9% and 19.5% for target and bullet (totaling 98.4%, slightly less than 100% due to missing and incomplete observations,) confirm this. From Table 5.3, target vs. bullet vehicles turning left are compared with the base condition of going ahead. A calculated COR = 15.97 indicates a high propensity for this condition to exist. This is likely attributed to the consequences associated and conflicts arising from making left turns through: crossing the paths of other oncoming vehicles; judging available gaps; watching out for pedestrians; and monitoring signal phase. However, under the regression modeling for target and bullet, this predictor was determined to be not statistically significant by the 6^{th} and 7^{th} iterations, respectively, and was therefore eliminated from any further modeling.

From Table 5.4, Road Surface Condition is predominantly dry, followed by wet. More importantly, there is no appreciable difference in the distributions within the various classes. This seems logical, since the conditions on the target road are likely to be identical to the conditions on the bullet road, with exceptions on one side or the other due to rare circumstances such as improper drainage, debris, spills, partial salting/plowing, etc. When this predictor was included in the target and bullet regression modeling, it was determined to be not statistically significant at the 4^{th} and 6^{th} iterations, respectively, and was therefore eliminated.

An asphalt Road surface exists in almost entirely all of the collisions studied. Even so, it was included in the initial modeling. This predictor was eventually backwards stepwise eliminated as it became apparent that it was not statistically significant in the regression modeling. It is not explored any further in this report. Other remaining predictors including road condition, road pavement markings and vehicle condition all followed similar conditions, where extremely few observations were noted outside of the basic and normal condition. These predictors were also eliminated in the regression process due to poor statistical significance and a lack of convergence of the models.

Note that all predictors discussed up to this point were categorical, having multiple levels within. Rather unique to these other variables, RMAX and VMAX representing maximum posted speed and maximum pre-collision vehicle speed, respectively, are treated as continuous and included in the regression analysis for target and bullet models. Because of the nature of continuous variables, it is not possible to construct or compare contrasts and OR for these two predictors. However, through regression it was determined for the bullet model that RMAX was highly statistically significant, P = 0.003, with an estimate of 0.05. The positive polarity of this value indicates a marginally increasing propensity for target vehicle to be at fault with higher posted maximum speeds. Perhaps this positive association is a result of the lesser amount of time associated with smaller gaps that creates fault causing conditions. For the corresponding bullet model, this predictor failed to meet statistical significance criteria and was eliminated. The regression modeling yielded both target and bullet models Maximum Pre-Collision Vehicle Speed to be highly statistically significant, as P = 0.0001 and P = 0.0001 with estimates of , respectively. The negative association of these estimates suggests that both target and bullet are inversely fault associated with higher pre-collision speeds. While this may seem counter-intuitive, it may be a direct result of the nature and circumstance in which this data field is recorded. The pre-collision vehicle speed is typically determined through simply questioning the involved persons, and involved parties may intentionally understate their actual traveled speed in fearing consequences of admitting high speeds to investigating officers. If this is widespread enough, a bias may exist within this predictor.

Table 5.5 lists remaining variables available which were excluded from any modeling in this analysis. These were excluded from modeling through a combination of logical grounds and a lack of past research showing any fault outcome associations.

Category	Frequency: Target: Frequency	Frequency: Target: % of Total	Frequency: Bullet: Frequency	Frequency: Bullet: % of Total	Crude Odds: Target	Crude Odds: Bullet	Crude Odds Ratios: Target VS Bullet
Number of Lanes							
1	4	0.6	0	0.0	0.01	0.00	-
2	104	16.3	82	12.8	0.19	0.15	1.25
3	18	2.8	18	2.8	0.03	0.03	0.98
4	333	52.1	328	51.1	1.08	1.04	1.00
5	45	7.0	55	8.6	0.08	0.09	0.81
6	115	18.0	135	21.0	0.22	0.27	0.84
7	12	1.9	17	2.6	0.02	0.03	0.70
8	7	1.1	5	0.8	0.01	0.01	1.38
other	1	0.2	2	0.3	0.00	0.00	0.49
Frequency missing	125		122				
Vehicle Body Style							
2 Door	78	18.9	99	24.2	0.23	0.32	0.68
3 Door	1	0.2	1	0.2	0.00	0.00	0.87
4 Door	287	69.5	249	60.9	2.28	1.56	1.00
Light Truck	47	11.4	60	14.7	0.13	0.17	0.68
Frequency missing	351		355				
Vehicle Colour							
Black	42	9.5	52	12.3	0.10	0.14	0.96
Blue	75	16.9	66	15.6	0.20	0.19	1.35
Brown	27	6.1	27	6.4	0.06	0.07	1.19
Gold	5	1.1	12	2.8	0.01	0.03	0.49
Green	50	11.3	39	9.2	0.13	0.10	1.52
Maroon	5	1.1	4	0.9	0.01	0.01	1.48
Red	80	18.0	63	14.9	0.22	0.18	1.51
Silver	80	18.0	95	22.5	0.22	0.29	1.00
White	80	18.0	64	15.2	0.22	0.18	1.48
Frequency missing	340		342				
Road Character							
1-Undivided -one-way	25	3.8	22	3.4	0.04	0.03	1.15
2-Undivided-two-way	545	82.6	552	84.3	4.82	5.36	1.00
3-Divided with restrain	27	4.1	27	4.1	0.04	0.04	1.01
4-Divided - no barrier	55	8.4	48	7.3	0.09	•0.08	1.16
5-Ramp	5	0.8	6	0.9	0.01	0.01	0.84
6-Collector lane	0	0.0	0	0.0	0.00	0.00	-
7-Express lane	0	0.0	0	0.0	0.00	0.00	-
8-Transfer lane	1	0.2	0	0.0	0.00	0.00	-
Frequency missing	106		109				
Road Surface							
1-Ashphalt	657	99.2	656	99.8	131.40	655.00	1.00
2-Oil treated gravel	4	0.6	0	0.0	0.01	0.00	-
3-Gravel or crushed st	0	0.0	0	0.0	0.00	0.00	-
4-Concrete	1	0.2	1	0.2	0.00	0.00	1.00
5-Earth	0	0.0	0	0.0	0.00	0.00	-
6-Wood	0	0.0	0	0.0	0.00	0.00	-
7-Steel	0	0.0	0	0.0	0.00	0.00	-
8-Brick/interlocking stc	0	0.0	0	0.0	0.00	0.00	-
Frequency missing	102		108				
Road Condition							
1-Good	653	98.5	650	98.8	65.30	81.25	1.00
2-Poor	6	0.9	4	0.6	0.01	0.01	1.49
3-Under repair or cons	4	0.6	4	0.6	0.01	0.01	1.00
Frequency missing	101		106				
Road Surface Condition							
1-Dry	628	82.4	624	82.0	4.69	4.55	1.00
2-Wet	127	16.7	130	17.1	0.20	0.21	0.97
3-Loose snow	2	0.3	2	0.3	0.00	0.00	0.99
4-Slush	2	0.3	2	0.3	0.00	0.00	0.99
5-Packed snow	2	0.3	1	0.1	0.00	0.00	1.99
6-Ice	1	0.1	1	0.1	0.00	0.00	0.99
7-Mud	0	0.0	0	0.0	0.00	0.00	-
8-Loose sand or grave	0	0.0	1	0.1	0.00	0.00	0.00
9-Spilled liquid	0	0.0	0	0.0	0.00	0.00	-
Frequency missing	2		3				
Road Alignment							
1-Straight on level	581	88.3	592	90.1	7.56	9.11	1.00
2-Straight on hill	65	9.9	58	8.8	0.11	0.10	1.14
3-Curve on level	10	1.5	3	0.5	0.02	0.00	3.40
4-Curve on hill	2	0.3	4	0.6	0.00	0.01	0.51
Frequency missing	106		107				
Road Pavement Markings							
1-Exist	641	97.4	642	98.0	37.71	49.38	1.00
2-Non-existant	13	2.0	9	1.4	0.02	0.01	1.45
3-Obscured	3	0.5	3	0.5	0.00	0.00	1.00
4-Faded	1	0.2	1	0.2	0.00	0.00	1.00
Frequency missing	106		109				
Vehicle Condition							
1-No apparent defect	658	99.8	661	99.8	658.00	661.00	1.00
99-Defect	1	0.2	1	0.2	0.00	0.00	1.00
Vehicle Manoeuver							
1-Going ahead	136	20.2	537	79.9	0.25	3.98	1.00
2-Slowing or stopping	1	0.1	2	0.3	0.00	0.00	1.97
3-Overtaking	2	0.3	0	0.0	0.00	0.00	-
4-Turing left	530	78.9	131	19.5	3.73	0.24	15.97
5-Turing right	0	0.0	1	0.1	0.00	0.00	0.00
6-Making "U" turn	1	0.1	0	0.0	0.00	0.00	-
7-Changing lanes	0	0.0	1	0.1	0.00	0.00	0.00
8-Merging	0	0.0	0	0.0	0.00	0.00	-
9-Reversing	0	0.0	0	0.0	0.00	0.00	-
10-Stopped	2	0.3	0	0.0	0.00	0.00	-
11-Parked	0	0.0	0	0.0	0.00	0.00	-
12-Disabled	0	0.0	0	0.0	0.00	0.00	-
13-Pulling away from s	0	0.0	0	0.0	0.00	0.00	-
14-Pulling onto should	0	0.0	0	0.0	0.00	0.00	-
Frequency missing	91		92				

Table 5.4: Extra-Human Variables Considered

Category	Target Frequency	Target % of Total	Bullet Frequency	Bullet % of Total	Crude Odds: Target	Crude Odds: Bullet	Crude Odds Ratios: Target VS Bullet
Driver Injuries							
0 None	99	14.8	273	41.6	0.17	0.71	1.00
1 Minimal	319	47.8	232	35.3	0.92	0.55	3.79
2 Minor	234	35.1	146	22.2	0.54	0.29	4.42
3 Major	15	2.2	6	0.9	0.02	0.01	6.89
4 Fatal	0	0.0	0	0.0	0.00	0.00	-
Frequency missing	97		107				
Driver's Safety Equipment Use							
1-Lap and shoulder belt	618	94.6	638	92.1	17.17	11.60	1.00
2-Lap belt only	6	0.9	8	1.2	0.01	0.01	0.77
3-Lap belt only of combined assembly	3	0.5	6	0.9	0.00	0.01	0.52
6-Air bag deployed	19	2.9	31	4.6	0.03	0.05	0.63
9-Equipment not used but available	8	1.2	10	1.4	0.01	0.01	0.83
Other safety equipment used	9		8				
1st Passenger's Injuries							
0 None	60	27.8	83	39.7	0.38	0.66	1.00
1 Minimal	75	34.7	68	32.5	0.53	0.48	1.53
2 Minor	75	34.7	53	25.4	0.53	0.34	1.96
3 Major	6	2.8	5	2.4	0.03	0.02	1.66
Frequency missing	548		555				
Vehicle Type							
1-Automobile, station wagon	584	87.6	552	83.3	7.04	4.97	1.00
4-Passenger Van	46	6.9	55	8.3	0.07	0.09	0.79
5-Pick-up truck	16	2.4	25	3.8	0.02	0.04	0.60
6-Delivery van	8	1.2	17	2.6	0.01	0.03	0.44
Other	13	1.9	14	2.1	0.02	0.02	0.88
Frequency missing	95		99				
Initial Direction of Travel							
1-North	158	23.6	177	26.5	0.31	0.36	0.82
2-South	175	26.1	159	23.8	0.35	0.31	1.02
3-East	180	26.9	166	24.9	0.37	0.33	1.00
4-West	157	23.4	165	24.7	0.31	0.33	0.88
Frequency missing	94		95				
Location of Vehicle Damage or Area of Impact: Initial Impact							
1-Right front corner	90	13.5	106	16.1	0.16	0.19	2.63
2-Right front	78	11.7	14	2.1	0.13	0.02	17.24
3-Right centre	82	12.3	11	1.7	0.14	0.02	23.07
4-Right rear	63	9.4	8	1.2	0.10	0.01	24.38
5-Right rear corner	5	0.7	2	0.3	0.01	0.00	7.74
6-Back centre	2	0.3	1	0.2	0.00	0.00	6.19
7-Left rear corner	0	0.0	1	0.2	0.00	0.00	0.00
8-Left rear	20	3.0	2	0.3	0.03	0.00	30.95
9-Left centre	37	6.5	8	1.2	0.06	0.01	14.32
10-Left front	36	5.4	12	1.8	0.06	0.02	9.29
11-Left front corner	59	8.8	112	17.0	0.10	0.21	1.63
12-Front centre	63	9.4	195	29.6	0.10	0.42	1.00
13-Front complete	64	9.6	173	26.3	0.11	0.36	1.15
14-Right side complete	50	7.5	5	0.8	0.08	0.01	30.95
15-Back complete	3	0.4	2	0.3	0.00	0.00	4.64
16-Left side complete	15	2.2	6	0.9	0.02	0.01	7.74
Frequency missing	96		104	15.8			
Location of Vehicle Damage or Area of Impact: Secondary Impact							
1-Right front corner	17	11.1	5	5.4	0.13	0.06	3.86
2-Right front	6	3.9	8	8.6	0.04	0.09	0.85
3-Right centre	13	8.5	2	2.2	0.09	0.02	7.39
4-Right rear	12	7.8	1	1.1	0.09	0.01	13.64
5-Right rear corner	9	5.9	1	1.1	0.06	0.01	10.23
6-Back centre	5	3.3	1	1.1	0.03	0.01	5.68
7-Left rear corner	6	3.9	7	7.5	0.04	0.08	0.97
8-Left rear	15	9.8	9	9.7	0.11	0.11	1.89
9-Left centre	10	6.5	2	2.2	0.07	0.02	5.68
10-Left front	1	0.7	4	4.3	0.01	0.04	0.28
11-Left front corner	6	3.9	4	4.3	0.04	0.04	1.70
12-Front centre	22	14.4	25	26.9	0.17	0.37	1.00
13-Front complete	9	5.9	19	20.4	0.06	0.26	0.54
14-Right side complete	2	1.3	2	2.2	0.01	0.02	1.14
15-Back complete	1	0.7	1	1.1	0.01	0.01	1.14
16-Left side complete	6	3.9	1	1.1	0.04	0.01	6.82
17-Top	7	4.6	0	0.0	0.05	0.00	-
18-Undercarriage	6	3.9	1	1.1	0.04	0.01	6.82
Frequency missing	611		671				

Table 5.5: Additional Variables Available Not Used in Analysis

5.5 Summary of Results for Human and Extra-Human Factors

A summary from Analysis I: MVA Report Fault Comparisons of the four different models obtained using the backwards stepwise regressions and Pr < 0.05 criteria are shown below. Predictors eliminated in the backwards stepwise regression have been excluded according to the mentioned criteria, and it is the opinion of this paper that these are acceptable but not perfect models.

Bullet:

- Human (6 Iterations):

  D2F= MVA34 D2AGE V2P1AGE

- Extra-human (10 Iterations):

  D2F= R2MAX V2APSPD MVA10

Target:

- Human (6 Iterations):

  D1F= MVA33 V1OCCUP

- Extra-human (9 Iterations):

  D1F = R1LANES V1APSPD MVA9

5.6 Intersection Collision Factors

Table 5.6 shows all of the intersection variables and their frequencies in the data used for this analysis. Data containing missing observations is not included below.

Category	Frequency		Crude Odds
	Frequency	% of Total	
Traffic Control Condition			
1-Functioning	651	98.5	65.10
2-Not functioning	8	1.2	0.01
3-Obscured	1	0.2	0.00
4-Missing/damaged	1	0.2	0.00
	103		
Environment Condition			
1-Clear	666	87.4	6.94
2-Rain	85	11.2	0.13
3-Snow	7	0.9	0.01
4-Freezing rain	1	0.1	0.00
5-Drifting snow	2	0.3	0.00
6-Strong wind	1	0.1	0.00
7-Fog, mist, smoke, dust	0	0.0	0.00
Light			
1-Daylight	337	68.2	2.15
2-Daylight, artificial	2	0.4	0.00
3-Dawn	2	0.4	0.00
4-Dawn, artificial	2	0.4	0.00
5-Dusk	14	2.8	0.03
6-Dusk, artificial	4	0.8	0.01
7-Dark	42	8.5	0.09
8-Dark, artificial	91	18.4	0.23
Frequency missing	270		
	764		
Presence of Left-Turn Arrow			
Present	565	74.0	2.84
Not Present	199	26.0	0.35
Presence of Flashing Adv. Green			
Present	400	52.4	1.10
Not Present	364	47.6	0.91
City/Ward Collision Occurred			
Etobicoke	154	20.2	0.25
East York	14	1.8	0.02
North York	239	31.3	0.46
Scarborough	222	29.1	0.41
Toronto	119	15.6	0.18
York	16	2.1	0.02
Signalized System Used			
MTSS	565	74.0	2.84
SCOOT	199	26.0	
Presence of Flashing Don't Walk			
Present	559	73.2	2.73
Not Present	205	26.8	0.37
Intersection Legs			
3	110	14.4	0.17
4	654	85.6	5.95
Day of Week			
Sunday	115	15.1	0.18
Monday	112	14.7	0.17
Tuesday	112	14.7	0.17
Wednesday	93	12.2	0.14
Thursday	137	18.0	0.22
Friday	119	15.6	0.19
Saturday	73	9.6	0.11

Table 5.6: General Intersection Data

Figure 5.10 shows that the majority of intersections featured LTA in the data under analysis.

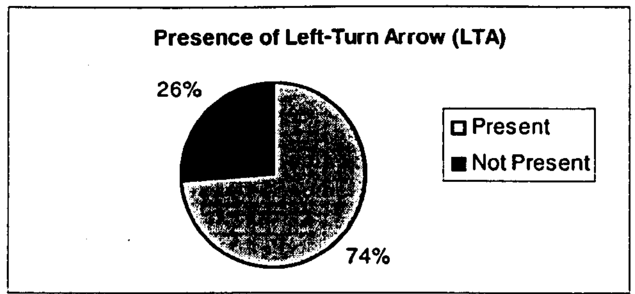


Figure 5.10: LTA Presence

Figure 5.11 shows that slightly over half of the intersections featured FAG.

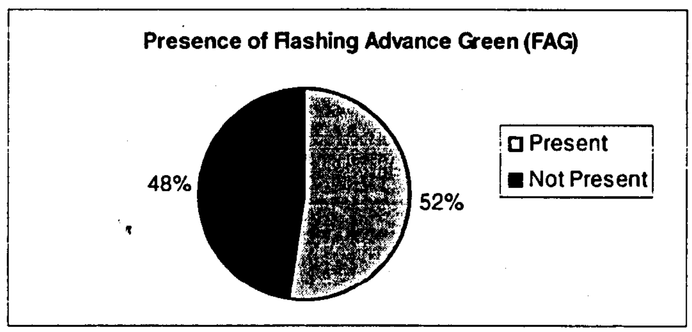


Figure 5.11: FAG Presence at Intersection

In Toronto, many intersections are under the fixed control of MTSS, while others are activated with the magnetic loop detector SCOOT system. In the intersections studied, Figure 5.12 shows that the majority are controlled by MTSS.

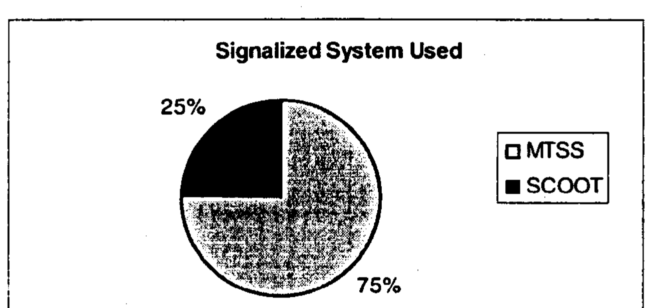


Figure 5.12: Signal System

The flashing hand and pedestrian system is shown in Figure 5.13 exists in 73% of the intersections under study.

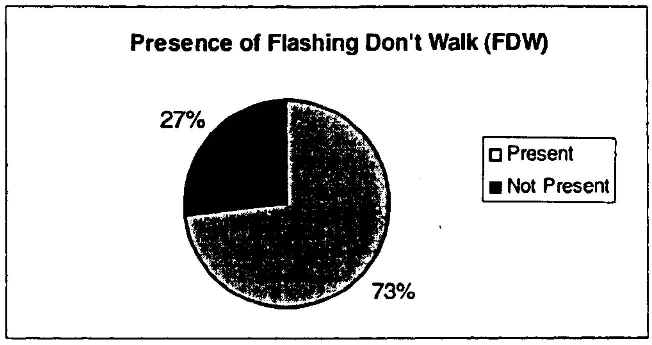


Figure 5.13: FDW Presence

The intersections under study were entirely either 3 or 4 legged. The gross majority are shown in the following Figure 5.14 to be of the 4 legged configuration.

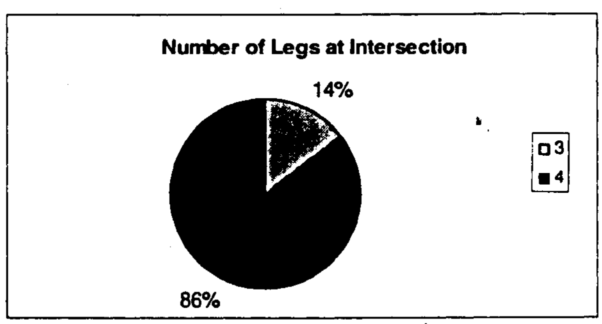


Figure 5.14: Intersection Legs

When collisions were examined by day of week, Figure 5.15 shows that Thursday and Friday lead in highest frequency.

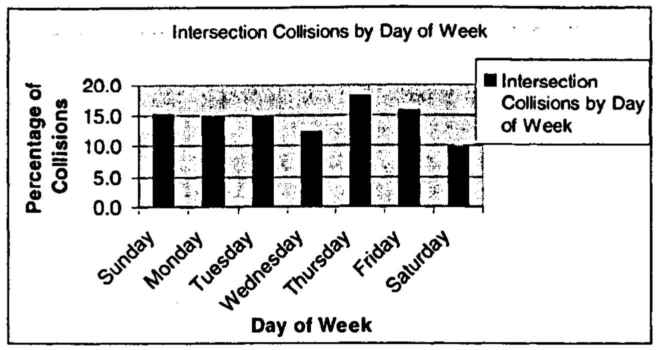


Figure 5.15: Collisions by day of week

Figure 5.16 shows a breakdown of how weekday collisions dominated in frequency weekend collisions. Of course, this could not be considered an entirely fair comparison since more days of the week exist in the weekdays vs. the weekend, but differences in travel patterns and travel demand potentially confound this.

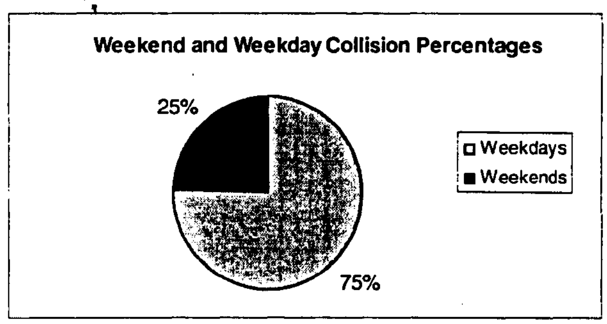


Figure 5.16: Weekend vs. Weekday Comparison

The above are all merely summary findings associated with the intersections selected which are modeled for collision existence.

The initial modeling (and first iteration) for this intersection analysis consisted of the following 14 variables: System, FAGINT, FDW, Num_Legs, Staggered, City, Q_major, Q_minor, DstInt, Any_LA, N, S, W, and E. Recall that these variables have been described previously. From the first iteration, SAS revealed that the model converges. Due to insufficient statistical significance (where $P > 0.05$), the following summary of backwards stepwise eliminated variables is shown in Table 5.7 below.

Iteration that Variable is Stepwise Eliminated	Variable Name
2	DstInt
3	Any_LA
4	Q_minor
5	City
6	E
7	FDW
8	Staggered

Table 5.7: Intersection Regression Variable Elimination

The remaining variables all had sufficient statistical significance to remain in the model. These factors for intersections with and without injury, property damage only and fatality collisions, yielded after 8 iterations are:

```
Inj_fat_PDO = System FAGINT Num_Legs Q_major N S W
```

A summary of P statistics associated with these factors are shown in Table 5.8 below.

Parameter	Pr > ChiSq
System	0.0366
FAGINT	0.0001
Num_Legs	0.0159
Q_major	0.0003
N	0.0289
S	0.0282
W	0.0001

Table 5.8: Parameter Statistics

Although all of the parameters within Table 5.8 are statistically significant, only select class contrasts within levels of several of these predictors were also statistically significant. These contrasts are shown in Table 5.9 below.

Parameter	Estimate	Odds = exp (Estimate)
FAG: No vs. Yes	-1.12	0.32
Num Legs 3 vs 4 lanes	-0.90	0.41
Q_major	0.00	1.00
Northbound: Local vs Major Arterial	-1.27	0.28
Southbound: Local vs Major Arterial	-1.34	0.26
Southbound: All other road classes vs Major Arterial	-2.85	0.06
Westbound: Collector vs Major Arterial	-1.25	0.29
Westbound: Local vs Major Arterial	-2.67	0.07
Westbound: All other road classes vs Major Arterial	-3.89	0.02

Table 5.9: Statistically Significant Class Contrasts for Intersection Analysis

Odds derived from the exponentiated estimates of the regression model results for intersections with vs. without collisions yielded the following statistically significant results:

- An odds of 0.32 shows a lower propensity exists for No vs. Yes for the presence of appearance of FAG Intersection Systems at intersections with collisions.
- 3 legged vs. 4 legged intersections had an odds of 0.41, which shows a lower propensity for selected collisions to occur at 3 vs. 4 legged intersections exists.
- Q_major statistically shows no effect with increasing or decreasing odds. This suggests that the data sampled shows no statistical difference for intersections experiencing collisions vs. those without based on volume on the major roadway.

- Major Arterial appeared to outweigh road classes in northbound, southbound and westbound directions.
- Through additional examination of this intersection data, outside of the regression modeling, a manual approach to explore the FAG factor was conducted. Manually calculated, separated and plotted odds show (Figure 5.17) differences in FAG and FAG-free intersections. An apparent trend occurs near the middle of the graph, which is likely attributed to a high proportion of vehicles on both roadways (especially those with crossing paths) resulting in a high number of potential conflicts. More importantly, it can be seen that the FAG curve lies above the without FAG curve throughout the graph, which indicates a higher odds of a given intersection having a collision. However, it should be noted that minor roadway volumes are not statistically significant when considered in the regression with all other predictors, but that is an analysis which exists in isolation of this.

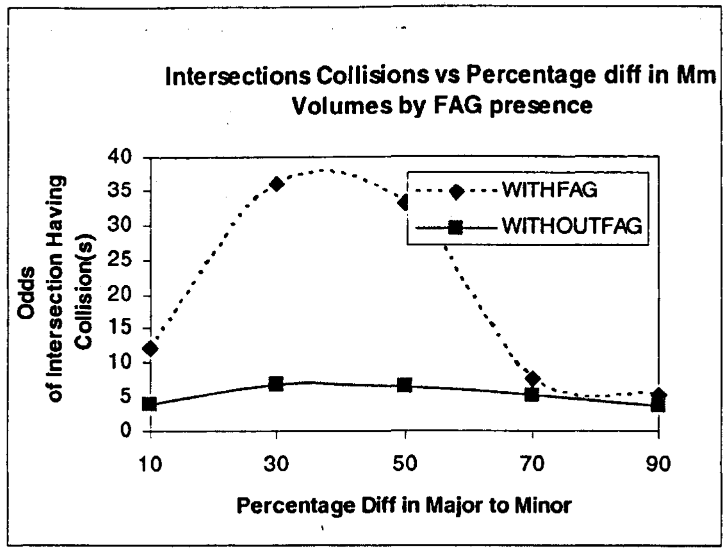


Figure 5.17: FAG Presence and Major Minor Flows

SAS output including entire statistics associated with both predictors and contrasts custom programmed appear in the printed model output in Appendix E.

5.7 Summary of Results for Intersection Collision Factors

For intersections, system type (scoot or mtss), presence of FAG, the number of legs, the major trafficway flow, and the road classifications of the North, South and West legs were found to be statistically significant factors.

Even with the parameter contrasts described above, it is possible to gain a general understanding but difficult extract an exact meaning behind several of the results. Statistically, it has been determined that these findings do exist. Pessimistically speaking, reasons for their existence could be attributed to chance since no current literature supporting this exists. Optimistically, one could speculate that this is a valid new discovery, perhaps even a function of Toronto's road structure, where there are significantly different attributes of city intersections by their approaches - attributes which have bearing on collision propensity at intersections. Specific examples of this could be the commuter nature of the city, where more people enter than leave during work-trips in the morning/afternoon hours, or perhaps even commuters traveling with sunlight in their eyes both ways, from their dwellings in the east end, etc. both of which may "clash" when crossing paths at intersections. Certainly, other such confounding factors have the potential to cloud salient factors from emerging. Regardless, inferences have been made from the data and never have any beliefs been imposed - the like of which would sabotage a fair analysis. Rather, including variables that were believed to be of relevance quantifies a belief.

5.8 Limitations

Although it is difficult to quantify, it is the opinion of this author that the greatest limitation of this report results from the inability to capture and include all relevant factors for analysis. As discussed throughout the report, other confounding factors likely exist. Literature reviewed and discussed pointedly shows this. It is better to do less exhaustively than to do more haphazardly.

The second most significant limitation is underestimation, which stems from numerous sources. According to the data available for the year 1998, there are notably on a sliver of reported collisions that occur in "Metro Toronto", and a massive disproportion of locations listed as "Other", as compared to later years where many more collisions are listed within "Metro Toronto". This is like attributed to the City of Toronto's amalgamation on January 1st, 1998, when many collision reporters (police officers) were in a sort of transition phase where they were partial to label Metro Toronto's "new" territory correctly.

Also, it is likely that the number of collisions going unreported increased first in 1997 with the introduction of self-reporting, and again in 1998 (in combination with the mentioned effect in paragraph above), when the property damage minimum for reporting increased to $1000. It is also likely that the numbers of collisions involving minimal injuries are underreported since injuries such as whiplash or soft tissue injuries are not apparent at the scene of the collision. The reverse holds true, albeit weaker, when retribution for losses is required.

As mentioned earlier in the report, injury severity and death is updated 30 days after the collision and any change of status of injury severity (and possible death) will be left out of the data. This can lead to underestimation, which may not be captured on the file. This also occurs when injury victims seek medical attention after the collision is reported. Perhaps electronically linked medical and collision records could overcome this.

Another source of underestimation exists in reported "Driver Condition", which is indicated as a separate field within the Ontario Motor vehicle Accident Report's Field (see Appendix B for complete list of fields). Since investigating officers must make judgment calls on driver's ability and condition as related to alcohol, drugs, medical condition, fatigue etc, and these judgments are subject to error. Also, since investigating officers arrive after the collision, driver conditions may change with time - provided that the driver is still even in attendance of the scene. Often times, injured drivers have left to seek medical treatment.

However, there are issues with the data used in this study. Similar to other studies is the issue of completeness of police reports on motor vehicle collisions. There may be considerate under reporting of minimal and minor injury classified collisions. Another concern is the reliability of injury coding, as misclassification between levels is possible. Although police should follow-up through hospital records for 30 days in order to identify fatal-injuries, patients who die following this period may not be reported appropriately [9]. Furthermore, it is expected that there is an underreporting of alcohol involvement, but this paper shall not focus on approximating this objectively.

It is generally accepted that many motorists wish to avoid insurance coverage increases associated with reporting collisions and fault assessments. Examining the Fault Determination Rules reveals that if no degree of fault is assessed against a driver, his renewal premium should not be affected. However, if a degree of fault is assessed against the driver, it is likely that there will be some effect on the renewal premium [40].

This concept is increasingly likely to be true with decreasing severity of collisions.

Furthermore, "[i]f an insured feels that fault has been improperly assessed and cannot resolve the matter with the insurance company, his or her remedy would be to commence an action in court and have the degree of fault established by the court [40]. The hassle and intimidation of going to court may further disway driver's from reporting collisions.

As indicated in a study by Applied Research and Evaluation Services, another significant problem with these data is that for the most part, police in Canada are not trained in recognizing driver intoxication by either alcohol or drugs (especially considering the effects of alcohol impairment can resemble those caused by a head trauma), so their estimations probably under-represent actual frequencies of impaired driving in the crashes that they attend and report upon [37]. This report also argues that the extent of drug-impaired crashes is underestimated due to a variety of reasons, primarily due to a lack of testing and testing sensitivity. They point to the results of a classic study by in 1982 by

Terhune that found that driver BACs had to reach 0.20 or over before 58% of the police recognized impairment level [38].

However, even the collisions which are investigated are subject to many other types of limitations. For instance, not all police officers are vehicle collision experts. As a result, data obtained from their reports may be erroneous. Furthermore, their reports must be manually entered from handwritten forms to electronic spreadsheets. This process is currently still the responsibility of Police, and since it is a manual human process, it introduces potential errors. Contributing to this error source is the fact that Toronto's current Police chief is renowned for "getting officers out from behind desks, and back onto the streets" - Unknown source.

Currently in Toronto, linking self reporting collisions from one to another is becoming a daunting task. CRCs are often abused as driver's claim collisions occurred at differing intersections and under different conditions that the truth. Furthermore, drivers are trying to test the system by reporting to different CRCs throughout the city with differing false statements. Perhaps the greatest weakness of this process can be attributed to drivers reporting at separate CRC's with falsified intersection names and explanations. This makes it increasingly difficult to link collisions to a specific intersection. Fortunately, the system discussed earlier in this report developed by Allianz, predicts that widespread use of their electronic self-reporting system will speed up and simplify the process of matching reports from the various parties involved in the same collision.

As mentioned in a summary by the Chronic Disease and Injuries - Injury Prevention and Substance Abuse Prevention [20], there are limitations associated with the Ontario collision data location fields. In particular, collision data is limited to location of collision entered geographically by place of occurrence. Their report expresses hope for future work on behalf of the Ministry of Transportation to include additional information by residence of driver in future publications since analysis based on place of occurrence can be misleading. Collisions may occur to people who do not live in that area, particularly in areas frequented by tourists and commuters. This has particular importance when considering underreporting.

In addition, there are issues with stepwise variable selection process used in the analysis's that should be stated. Here are *some* of the problems with stepwise variable selection.

1. R-squared values are biased high, yielding confidence intervals for effects and predicted values that are falsely narrow (See Altman and Anderson, *Statistics in Medicine*).

2. It yields P-values that do not have the proper meaning and the proper correction for them is apparently a very complex issue.
3. It has problems in the presence of collinearity.
4. Increasing the sample size doesn't help very much (see Derksen and Keselman).
5. It allows us to not think about the problem.
6. It uses a lot of paper.

[39]

In addition to analysis limitations, this report is affected by issues associated with exposure. As with the quasi-induced exposure techniques used commonly in similar studies, one must remember that OR used in this report are a measure rather different than one that can account for driver miles, licensed drivers, and a host of other exposures.

Furthermore, the interpretation of odds and OR are surrounded by limitations, perhaps most profoundly by that of Simpson's Paradox. The fact that a marginal table may exhibit an association completely different from the partial tables is known as Simpson's Paradox (Simpson 1951, Yule 1903) [28]. Missing values have significant implications when dealing with OR's as they can affect the denominator of the ratio and yield exaggerated results. These become even more apparent when over-parameterization issues arise, as many predictors in the variables used contain several classes. The large number of parameters and relatively limited quantity of data available create over-parameterization issues

The fault assessment used to create flags indicating target and bullet driver faults was not a simple task. Ontario's Statutes and Regulations as written in the Insurance Act [40] provide comprehensive Fault Determination Rules, which were used in this study. One weakness inherent in using the fault determination rules stems from the its use of percentage attributions of fault in specific scenarios. These percentage attributions, particularly partial fault assessments (i.e., of 25% and 75% at fault) do not directly translate to a binary fault outcome. Under the given conditions, this was decided to be the best method. Note that when no clear indication of either being at fault or not at fault existed, the fault observation was left blank. It is also possible for both vehicles to simultaneously be either at fault or not-at-fault. In light of this, support in following the fault determination rules is achieved through its provision of a consistent means of assessment, and is recommended for future use.

As in any report, it is essential that these limitations be recognized throughout all of the presented results. Further to this is the potential multiplicative negative effect of all these mentioned limitations, some of which may act in conjunction with others to cloud the salient factors from emerging.

6 Conclusions and Recommendations

6.1 General

At a minimal cost, a preliminary analysis the identification of factors involved in side impact left turning collisions was performed. Crude odds ratios in addition to regression model computed and exponentiated estimates for odds ratios were examined. Differences between these were primarily attributed to the fact that other confounding variables exist and lie outside the scope of this research. Limitations aside, statistically significant values for factors associated with these collisions were discussed and certain factors showed clear associations with the fault outcome.

It is not surprising to see different factors appearing in separate models. It seems possible that bullet vehicles are sensitive to maximum posted speeds. After all, it is almost expected that the characteristics that define target and bullet vehicles be different, particularly because they are two unique and independent items.

Missing values for each variable was rather problematic throughout the data used, especially in comparison to the large sample sizes in other studies discussed in the literature review.

6.2 Conclusions

The factors that were statistically significant and included in the bullet models differed slightly from the factors found in the target models. Both target and bullet human factor models shared Apparent Driver Action and Front Seated Passenger Age as significant factors. However, only the bullet model included Driver Age and only the target model included Number of Vehicle Occupants. Similarly, for the extra-human factors, both Road Character and Approximate Vehicle Speed appeared in target and bullet models. However, the bullet model contained Maximum Posted Speed while the target model contained Number of Lanes.

From the intersection analysis, intersections were studied to determine causes and factors associated with intersections experiencing collisions and those without with classifications of left-turning and angle, over a period of three years. Certain features and combinations of features showed statistically significant propensities to be associated with intersections with collisions as compared to those without. These factors were intersection system (SCOOT or MTSS), presence of FAG, Number of Legs, Flow on the major roadway, and classifications of North, South and West roads. While these predictors as a whole were statistically significant, only selected contrasts within yielded statistically significant odds.

Crude odds (CO) and crude odds ratios (COR) were used to exhibit findings from summary statistics for the combination of *both* target and bullet vehicles. Odds and odds ratios (OR) were used throughout the report as measures between predictors and predictor contrasts for statistically significant results ($P < 0.05$) of target vs. bullet comparisons. The organization of the data through classifications and sub-classifications (i.e., Human and Extra-Human) make distinctions between variables for separate models more apparent. This creates ease in model building. The logistic modeling with the commonly accepted $P < 0.05$ criteria as shown for other reports reviewed is easy to interpret.

Furthermore, it was intended to compare the known case target and bullet vehicle driver's characteristics. Ideally, one could expect to find statistically proven differences between target and bullet drivers involved in collisions, particularly matched with fault outcomes. Unfortunately, due to confidentiality issues, this information on comparison vehicles drivers was never made available. Perhaps with continued persistent efforts, policy makers and in particular, Metropolitan Police decision makers will provide better access to confidential data for legitimate studies such as this.

Salient factors associated with selected side impact collisions and intersection collisions in general have been reported. These provide insight into the analysis of data that shall be applied in further research in a more focused analysis of the in-depth collision database that has been developed.

6.3 Recommendations

Throughout the composition of this report, a number of recommendations became apparent. These have been separated into two categories, namely, recommendations for data, and future work.

Recommendations with respect to Data:

- Investigating officers should have a provision on the MVA reports to record info on the object struck, specifically what happened to it (i.e., breakaway pole, did it indeed breakaway? Did it fall over, deform, displace the roadside barrier, etc.).

- Include other environmental weather conditions, linked electronically, such as temperature, humidity, etc.

- The MVA report author's should create a field within Vehicle Type for "Sport Utility Vehicle", and "Mini-Van" and "Full sized Passenger Van", as opposed to the existing "Passenger Van" field which is too limited.

- MVA reports have a limited vehicle occupant "position" (or seating) fields that do not account for the rear rows of minivans seats. This field should be modernized to account for this.

- MVA reports should be perfectly linked with Self-reporting forms, using a similar numeric style and sequence for codes depicting collision details. This would increase the ease with which these documents could be merged. Transport Canada should also consider adopting a similar format, in addition to the extra observations they make. This would allow further compatibility and ease for linked data set efforts.

Recommendations with respect to future work:

- Following the fault determination rules helped to provide some degree of consistency and is recommended for future use. Perhaps a fault field directly on an MVA would be of some added benefit.

- MVA reports should have a field indicating more than just two different types of LT collisions. There are many more combinations of LT collision, and addressing this concern would certainly enrich the data.

- MVA reports should indicate whether the second generation air bags deployed, or at least indicate which phase and type of deployment occurred. Insurance agencies would have a particular interest in this for many reasons.

- Through manual data entry, it was repeatedly observed that investigating police officer's labeled roadways for vehicle one and two as R1 and R2. Unfortunately, they did this at their discretion without any apparent nor consistent procedure.

- Police recorded roadway designations often were not respective of vehicles one and two, V1 and V2. This creates confusion in differentiating V1 and V2 between the R2 label, "reference point" which differs from the R1 label "trafficway". Officers may tend to simply label R1 the more major of the two streets. This has negative consequences when entering and organizing data for analysis and comparison of target and bullet vehicles.

- A good next step would be to take the intersection analysis to another level and focus on identifying side impact collisions. From this, comparisons between intersections with collisions and with side impact collisions could be made. Comparisons with intersections experience varying levels of collision could also be made.
- Combinations of contrasts built on existing contrasts would be an interesting way to analyze areas of known concern further. With additional data, it would be interesting to exploring combinations of features such as FDW and FAG, which wouldn't be possible in the current data due to its size limitations.
- In future work, it may be better suited to model fault outcomes as continuous percentages, rather than binary dichotomous outcomes. This is particularly true since the Ontario Fault Determination Rules [29] produce fault in terms of percentages, and not simply as binary outcomes.

Appendix A: Transport Canada Vehicle Safety Research Investigation Forms

CASE: ________-________ VEHICLE #: ____ OCCUPANT: _____

Transport Canada Transports Canada

FRONT AIR BAG FORM

ab1 AIR BAG DEPLOYMENT

Yes No

ab2 SEAT OCCUPIED

Yes No

ab3 HORIZONTAL DIMENSION

________cm

ab4 VERTICAL DIMENSION

________cm

ab5 NUMBER OF VENTS

____

ab6 EVIDENCE OF OCCUPANT CONTACT

Yes No

ab7 AIR BAG DAMAGE DUE TO DEPLOYMENT

Yes No

ab8 ABRASIONS

Yes No Unknown

ab9 THERMAL BURNS

Yes No Unknown

ab10 CHEMICAL BURNS

Yes No Unknown

ab11 IRRITATION FROM RESIDUE

Yes No Unknown

ab12 OTHER AIR BAG RELATED INJURIES

Yes No Unknown

ab13 DUAL DEPLOYMENT THRESHOLD

Yes No

ab14 SEAT OCCUPANCY DETECTION SENSOR

Yes No

ab15 OCCUPANT PROXIMITY SENSOR

Yes No

SKETCH AIR BAG COVER DESIGN:

FRTAIRB.DOC 21/05/97

CASE: ________-________

Transport Canada / Transports Canada **CASE FORM**

ca1 PROVINCIAL COLLISION CASE NUMBER

____

POLICE FORCE: ____________________

ca2 PROVINCE

____

ca3 COLLISION DATE

___/___/___
yy mm dd

ca4 COLLISION TIME (00 - 23)

____

ca5 SAMPLING DATE

___/___/___
yy mm dd

ca6 POLICE REPORTED COLLISION SEVERITY

1 Property damage only
2 Non-fatal injury
3 Fatal

ca7 NUMBER OF VEHICLE FORMS

____

VEHICLE INSPECTION DATES

VEH: ______________ DATE: ___/___/___

VEH: ______________ DATE: ___/___/___

VEH: ______________ DATE: ___/___/___
yy mm dd

SCENE INSPECTION DATE

___/___/___
yy mm dd

LIGHT CONDITION

1 Dawn - one hour before sunrise
2 Daylight - between sunrise and sunset
3 Dusk - one hour after sunset
4 Dark - between dusk and dawn
5 Artificial illumination

WEATHER CONDITION

1 Clear
2 Drizzle
3 Raining
4 Freezing rain
5 Snowing
6 Fog, smog
7 Cloudy
Q Other

ROAD SURFACE ENVIRONMENTAL CONDITION

1 Dry
2 Wet
3 Snow covered
4 Snow patches
5 Ice covered
6 Ice patches
7 Slush / wet snow
8 Muddy
9 Sand / dirt / oil

CASE.DOC 21/05/97

CASE: _________-_________

Transport Canada Transports Canada **CASE FORM**

ROAD SURFACE MATERIAL	*LEGAL SPEED*
1 Concrete	____________km/h
2 Asphalt	
3 Earth	*ADVISORY SPEED*
4 Gravel	
Q Other	____________km/h

CASE.DOC 21/05/97

CASE: ________-________ **VEHICLE #:** ____

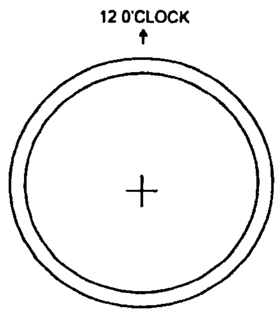
Transport Canada / Transports Canada

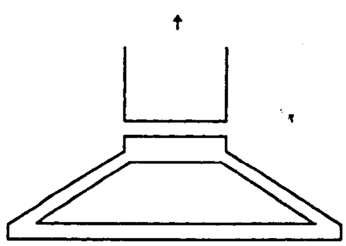
DAMAGE FORM

da8 STEERING WHEEL DAMAGE DUE TO OCCUPANT LOADING

Yes   No   Unknown

ANNOTATE DAMAGE ON DIAGRAMS:





TOP VIEW

da9 COLUMN COMPRESSION DUE TO OCCUPANT LOADING

________cm

LEFT SHEAR CAPSULE: ________cm

RIGHT SHEAR CAPSULE: ________cm

OR

ORIGINAL LENGTH: ________cm

COMPRESSED LENGTH: ________cm

da10 STEERING COLUMN REARWARD DISPLACEMENT

________cm

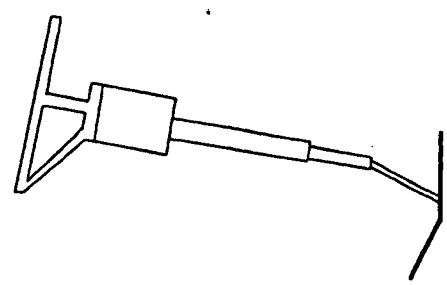
ORIGINAL DIMENSION:________cm

DAMAGED DIMENSION:________cm

FROM: ____________________

TO: ____________________

DESCRIBE COLUMN DESIGN:



da11 OCCUPANT COMPARTMENT INTRUSION LOCATION

F Front
B Back
L Left
R Right
T Top
U Underside
N None

da12 MAXIMUM EXTENT OF INTRUSION

________cm

CASE: ________-________ **VEHICLE #:** ____

Transport Canada / Transports Canada **DAMAGE FORM**

LEFT HOOD HINGE

1 Functional
2 Released during crash
3 Failed during crash (separated)
4 Jammed
5 Deformed

RIGHT HOOD HINGE

1 Functional
2 Released during crash
3 Failed during crash (separated)
4 Jammed
5 Deformed

DESCRIBE DAMAGE:

da3 FUEL SYSTEM TYPE

1 Gasoline
2 Diesel
3 LPG
4 CNG
Q Other

da4 FUEL SYSTEM INTEGRITY LOSS

Yes No

da5 VEHICLE FIRE

Yes No

da6 PRE-CRASH TIRE FAILURE

Yes No Unknown

da7 DOOR LATCHES AND DOOR HINGES

1	Functional	LF	______
2	Released during crash	LR	______
3	Failed during crash	RF	______
4	Jammed	RR	______
N	Not applicable	CARGO	______
U	Unknown		

DESCRIBE DAMAGE:

DAMAGE.DOC 21/05/97

CASE: ________-________ VEHICLE #: ____

Transport Canada | Transports Canada

DAMAGE FORM

da13 Lf: ________cm

da14 Lc: ________cm

da15 Cmax: ________cm

da16 C1: ________cm

C2: ________cm

C3: ________cm

C4: ________cm

C5: ________cm

C6: ________cm

da17 OBJECT CONTACTED: ______________

PDOF: _______

CDC: ______________

da18 OBJECT CONTACTED: ______________

PDOF: _______

CDC: ______________

da19 OBJECT CONTACTED: ______________

PDOF: _______

CDC: ______________

da20 EBS ________km/h

da21 DELTA-V ________km/h

DELTA-V LATERAL________km/h

DELTA-V LONGITUDINAL________km/h

da22 PRE-IMPACT SPEED OF BULLET VEHICLE

________km/h

(DO NOT CODE FOR ACRS)

da23 VEHICLE WRITTEN OFF

Yes No

da24 TOTAL REPAIR COST OR WRITE-OFF VALUE

______________$

da25 PORTION OF REPAIR COST DUE TO AIRBAG DEPLOYMENT

______________$

PERCENTAGE OF REPAIR COST DUE TO AIRBAG DEPLOYMENT

______________%

da26 SOURCE OF REPAIR COST OR WRITE-OFF VALUE

1 Body shop
2 Insurance
3 Red Book
4 Dealer

DAMAGE.DOC 21/05/97

CASE: ________-________ VEHICLE #: ____

Transport Canada Transports Canada **NON-CASE OCCUPANT FORM**

OCCUPANT: _____

nc1 GENDER Male Female

nc2 AGE ________

nc3 INJURY SEVERITY

- **1** No injury
- **2** Minimal
- **3** Minor
- **4** Major
- **5** Fatal
- **6** Death or injury due to natural causes
- **7** Injured, extent unknown
- **U** Unknown

nc4 SEAT BELT USED

Yes No Unknown

OCCUPANT: _____

nc1 GENDER Male Female

nc2 AGE ________

nc3 INJURY SEVERITY

- **1** No injury
- **2** Minimal
- **3** Minor
- **4** Major
- **5** Fatal
- **6** Death or injury due to natural causes
- **7** Injured, extent unknown
- **U** Unknown

nc4 SEAT BELT USED

Yes No Unknown

OCCUPANT: _____

nc1 GENDER Male Female

nc2 AGE ________

nc3 INJURY SEVERITY

- **1** No injury
- **2** Minimal
- **3** Minor
- **4** Major
- **5** Fatal
- **6** Death or injury due to natural causes
- **7** Injured, extent unknown
- **U** Unknown

nc4 SEAT BELT USED

Yes No Unknown

OCCUPANT: _____

nc1 GENDER Male Female

nc2 AGE ________

nc3 INJURY SEVERITY

- **1** No injury
- **2** Minimal
- **3** Minor
- **4** Major
- **5** Fatal
- **6** Death or injury due to natural causes
- **7** Injured, extent unknown
- **U** Unknown

nc4 SEAT BELT USED

Yes No Unknown

NCOCC.DOC 21/05/97

CASE: ________-________ VEHICLE #: ____ OCCUPANT: _____

Transport Canada Transports Canada **OCCUPANT FORM**

oc1 GENDER	oc3 HEIGHT
Male Female	_________ft, in _________cm *(inches x 2.54 = cm)*
oc2 AGE	**oc4 MASS**
_________	_________lb _________kg *(lb x 0.4536 = kg)*

DESCRIBE ALL INJURIES

Do you remember when the air bag deployed?	*Did you notice any smoke from the air bag?*

OCCUPANT.DOC 21/05/97

CASE: ________-________ VEHICLE #: ____ OCCUPANT: ______

Transport Canada / Transports Canada

OCCUPANT FORM

Do you have any special medical condition?

oc5 SPECIAL MEDICAL CONDITION

- **1** Musculoskeletal
- **2** Cardiovascular
- **3** Diabetes
- **4** Pregnancy
- **5** Respiratory
- **N** None
- **Q** Other
- **U** Unknown

Did you take any medication in the last 24hrs?

oc6 MEDICATION / DRUG USAGE (prior 24hrs)

- **1** ASA
- **2** Sleeping pills
- **3** Tranquilizers
- **4** Stimulants
- **5** Insulin
- **6** Heart drugs
- **N** None
- **Q** Other
- **U** Unknown

Did you consume any alcohol?

oc7 BAC ________(mg%)

Where do you normally position the seat?

oc8 OCCUPANT REPORTED SEAT POSITION

- **1** Fully forward
- **2** Forward of middle
- **3** Middle
- **4** Rearward of middle
- **5** Fully rearward
- **N** Not applicable
- **U** Unknown

How were you sitting just prior to the collision?

Do you recall the location of your hands on the steering wheel?

Did you brace yourself with your hands and feet?

OCCUPANT.DOC 21/05/97

CASE: ________-________ VEHICLE #: ____ OCCUPANT: ______

Transport Canada / Transports Canada

OCCUPANT FORM

oc9 PRE-CRASH POSITION

1 Normal seating position
2 Braced
3 Leaning forward
4 Slumped forward
5 Slouched in seat
6 Reclined in seat
7 Feet on dash / windows
8 Rotated left
9 Rotated right
10 Lying down
11 On lap
12 Standing on seat
13 Standing on floor
Q Other
U Unknown

oc10 SEAT BELT USED

Yes No Unknown

Was the lap belt over the stomach or low over the hips?

Was the shoulder belt over the shoulder, under the arm or behind the back?

Was there any slack in the lap or shoulder belt?

Do you adjust the belt prior to driving?

oc11 SEAT BELT USE MODE

1 Used correctly
2 Belt extended
3 Lap belt slack
4 Torso belt slack
5 Lap and torso slack
6 Lap belt not used
7 Torso belt not used
8 Lap belt on abdomen
9 Torso belt off shoulder
10 Torso belt under arm
11 Torso belt behind back
12 Multiple occupants
13 Improper CRS installation
N Not applicable
Q Other
U Unknown

oc12 CHILD RESTRAINT USED

Yes No

Were you wearing glasses or contact lenses?

Were they damaged?

oc13 WEARING EYE WEAR

Yes No Unknown

OCCUPANT.DOC 21/05/97

CASE: ________-________ VEHICLE #: ____ OCCUPANT: _____

Transport Canada / Transports Canada **OCCUPANT FORM**

Can you describe what clothing were you wearing?

Was your clothing damaged?

(If female) were you wearing any makeup?

Which hospital were you transported to?

How long were you at the hospital?

oc14 INJURY SEVERITY

- **0** No injury
- **1** No codeable injury
- **2** Injured
- **3** Fatal

oc15 MEDICAL TREATMENT

- **1** Injured, sought no treatment
- **2** First aid at scene
- **3** Treated by general practitioner
- **4** Examined and released from hospital
- **5** Admitted to hospital
- **6** Fatal
- **N** No Injury
- **U** Unknown

oc16 INJURY INCREASED BY SEATING POSITION

- **1** Definite
- **2** Probable
- **3** Possibly
- **4** Definitely not
- **U** Unknown

oc17 INJURY INCREASED BY REAR LOADING TO SEATBACK

- **1** Definite
- **2** Probable
- **3** Possibly
- **4** Definitely not
- **U** Unknown

oc18 INJURY INCREASED BY INTRUSION

- **1** Definite
- **2** Probable
- **3** Possibly
- **4** Definitely not
- **U** Unknown

oc19 INJURY INCREASED BY LOOSE OBJECTS

- **1** Definite
- **2** Probable
- **3** Possibly
- **4** Definitely not
- **U** Unknown

oc20 INJURY INCREASED BY OCCUPANT TO OCCUPANT INTERACTION

- **1** Definite
- **2** Probable
- **3** Possibly
- **4** Definitely not
- **U** Unknown

oc21 INTERVIEW TYPE

- **1** Personal interview
- **2** Telephone interview
- **3** Questionnaire
- **4** Interview with other occupant or relative
- **N** No interview

OCCUPANT INTERVIEW DATE ____/____/____ (yy mm dd)

OCCUPANT.DOC 21/05/97

CASE: ________-________ VEHICLE #: ____ OCCUPANT: ______

Transport Canada Transports Canada **RESTRAINT FORM**

re1 SEAT BELT TYPE

1 Lap belt only
2 Body mounted lap and torso
3 Door mounted lap and torso
4 Motorized
Q Other

re2 TONGUE TYPE

1 Fixed
2 Locking
3 Modified locking sliding
4 Sliding

re3 ADJUSTABLE UPPER SEAT BELT ANCHOR

Yes No

D-RING TRAVEL LENGTH ________cm

POSITION OF D-RING FROM TOP________cm

re4 TENSION RELIEVING DEVICE

Yes No

re5 LOADING EVIDENCE

Yes No

DESCRIBE LOADING EVIDENCE

D-RING:

TONGUE:

WEBBING:

OTHER:

RSTRNT.DOC 21/05/97

CASE: ________-________ VEHICLE #: ____ OCCUPANT: ______

Transport Canada / Transports Canada **SEAT FORM**

se15 SEAT POSITION

1 Fully forward
2 Forward of middle
3 Middle
4 Rearward of middle
5 Fully rearward
N Not applicable
U Unknown

IF SEAT CANNOT BE MOVED, DETERMINE SEAT TRACK POSITION:

OCCUPANT REPORTED SEAT POSITION:

se16 SEAT ANCHORAGE DAMAGE

1 Deformed
2 Failed
N No damage

DESCRIBE DAMAGE:

se17 SEAT CUSHION ADJUSTER DAMAGE

1 Deformed
2 Failed
N No damage, not applicable
U Unknown

DESCRIBE DAMAGE:

se18 SEATBACK DAMAGE

1 Deformed forwards
2 Deformed rearwards
3 Failed forwards
4 Failed rearwards
5 Deformed laterally
N No damage

DESCRIBE DAMAGE:

se19 REAR LOADING TO SEATBACK

1 Loaded by vehicle occupant
2 Loaded by object
N None

SEAT.DOC 21/05/97

CASE: ________-________ VEHICLE #: ____

Transport Canada / Transports Canada

VEHICLE FORM

ve1 CASE VEHICLE

Yes No

ve2 YEAR

____

ve3 MAKE

____

ve4 MODEL

____

ve5 BODY TYPE

____

ve7 GROSS MASS

__________ Vehicle

__________ Occupants

__________ Cargo

__________ TOTAL kg

ve8 WHEELBASE

__________ cm

ve9 VEHICLE TOWING A TRAILER

Yes No

ve10 VEHICLE TOWED FROM SCENE

Yes No

ve11 NUMBER OF OCCUPANT FORMS

______

ve6 V.I.N.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17

D.O.M: _____/_____ (mm / yy)

ODOMETER: ______________ km

ENGINE DISPLACEMENT: ______________

TRANSMISSION TYPE:

Auto Manual

TRANSMISSION SHIFT LEVER:

Column Floor

DRIVE TYPE:

FWD RWD AWD 4x4

BRAKE TYPE:

FRONT: disc drum

REAR: disc drum

VEHICLE.DOC 21/05/97

Appendix B: MVA Variable List

Collision File

- Collision Date
- Collision Day
- Collision Location
  - -County/Region
  - -Highway
  - -Highway Suffix
  - -Municipality
- Collision Number
- Collision Severity
- Collision Time
- Collision Scene
  - -Emergency Equipment
  - -Traffic Control Condition
  - -Traffic Control Type
- Dangerous Goods
  - -Product ID
  - -Product Summary
- Failure to Remain
- Number of Fatalities
- Impact Location
- Initial Impact
- Lighting
- Number of Vehicles Involved
- Police Force
- Road Characteristics
  - -Alignment
  - -General Condition
  - -Lane Type
  - -Posted Speed Limit
  - -Speed Limit
  - -Surface Condition
  - -Surface Type
- Road Location Characteristics
- Road Jurisdiction
- Weather Conditions

Driver-Vehicle File

- Collision Number

Driver Information
- -Actions
- -Age
- -Age Category - Census
- -Age Category - MTO
- -Birth date
- -Charges
- -Condition
- -License Class
- -Province
- -Sex
- -Suspension Status

- Driver-Vehicle Involvement
- Hit/Run Apprehension
- Vehicle Information
  - -Condition
  - -Damage Control
  - -Direction of Travel
  - -Maneuver
  - -Number of Occupants
  - -Province
  - -Speed
  - -Speed Category
  - -Type
- Vehicle Number

Involved Person File

- Collision Number
- Involved Persons Information
  - -Age
  - -Ejection
  - -Injury Level
  - -Pedestrian Action
  - -Pedestrian Condition
  - -Position in Vehicle
  - -Safety Device Used
  - -Sex
- Person Number
- Vehicle Number

Appendix C: Control Investigation Forms

Sheet 1

x

CONTROL COLLISION INVESTIGATION DATA FORMS

LAST UPDATED April 15th 2003

(CIRCLE) COLLISION DATA SOURCE: TORONTO LONDON MONTREAL

PROVINCIAL COLLISION CASE NUMBER: ______________________

OTHER (T.C.) COLLISION REFERENCE NUMBER OR INTERSECTION NAME: ______________________

DATE OF CONTROL STUDY: ______________
MM/DD/YY

DATE OF ORIGINIAL COLLISION: ______________
MM/DD/YY

TIME OF CONTROL STUDY: ______________
0-2400 HRS

TIME OF ORIGINIAL COLLISION: ______________
0-2400 HRS

CONTROL DAY OF WEEK:

Sun Mon Tues Wed Thurs Fri Sat

ORIG. COLLISION DAY OF WEEK:

Sun Mon Tues Wed Thurs Fri Sat

TARGET R1 STREETNAME AND VEHICLE OF INTEREST'S DIRECTION OF TRAVEL: ______________________

BULLET R2 STREETNAME AND VEHICLE OF INTEREST'S DIRECTION OF TRAVEL: ______________________

INVESTIGATOR'S (YOUR) NAME: ______________________

X Sheet 2 Data Collection Sheet

CONTROL CASE FORM

PARKING ARRANGEMENT WITHIN 30 meters of intersection by approach

DRIVEWAY ACTIVITY/ACCESS POINTS WITHIN 30 METRES OF INTERSECTION (Identify by approach)

UNUSUAL TRAFFIC CHARACTERISTICS (e.g., with respect to vehicle types)

UNUSUAL DISTRACTIONS (e.g. Video Screens)

CONTROL ROAD CONDITION

01 Good
02 Poor
03 Underconstruction

ROAD CHARACTER

R1 :________ R2:________

01 Undivided - one way *
02 Undivided - two way
03 Divided with barrier
04 Divided - no barrier
05 Ramp
06 Collector lane
07 Express lane
08 Transfer lane

CONTROL ROAD SURFACE MATERIAL

1 Concrete
2 Asphalt
3 Earth
4 Gravel
5 Other

CONTROL LIGHT CONDITION

1 Dawn - one hour before sunrise
2 Daylight - between sunrise and sunset
3 Dusk - one hour after sunset
4 Dark - between dusk and dawn
5 Artificial Illumination

CONTROL WEATHER CONDITION

1 Clear
2 Drizzle
3 Raining
4 Freezing rain
5 Snowing
6 Fog, smog, smoke
7 Cloudy
O Other

CONTROL ROAD SURFACE ENVIRONMENTAL CONDITION (PICK WORST OF R1 and R2)

1 Dry
2 Wet
3 Snow covered
4 Snow patches
5 Ice covered
6 Ice patches
7 Slush / wet snow
8 Muddy
9 Sand / dirt / oil

LEGAL SPEEDS

ON R1: ____________ **km/h**

ON R2: ____________ **km/h**

CONTROL AVERAGE SPEED

ON R1: ____________ **km/h**

ON R2: ____________ **km/h**

A divided roadway is a section with a median type of curbed or positive barrier or median width greater than or equal to one metre. An undivided roadway is a section without a qualifying median.

GEOMETRIC FORM i

ge1 TOTAL L LANES IN A ____.	
ge2 TOTAL LT LANES IN A ____.	
ge3 TOTAL LTR LANES IN A ____.	
ge4 TOTAL T LANES IN A ____.	
ge5 TOTAL TR LANES IN A ____.	
ge6 TOTAL R LANES IN A ____.	
ge7 TOTAL LR LANES IN A ____.	
ge8 TOTAL L LANES IN B ____.	
ge9 TOTAL LT LANES IN B ____.	
ge10 TOTAL LTR LANES IN B ____.	
ge11 TOTAL T LANES IN B ____.	
ge12 TOTAL TR LANES IN B ____.	
ge13 TOTAL R LANES IN B ____.	
ge14 TOTAL LR LANES IN B ____.	
ge15 TOTAL L LANES IN C ____.	
ge16 TOTAL LT LANES IN C ____.	
ge17 TOTAL LTR LANES IN C ____.	
ge18 TOTAL T LANES IN C ____.	
ge19 TOTAL TR LANES IN C ____.	
ge20 TOTAL R LANES IN C ____.	
ge21 TOTAL LR LANES IN C ____.	
ge22 TOTAL L LANES IN D ____.	
ge23 TOTAL LT LANES IN D ____.	
ge24 TOTAL LTR LANES IN D ____.	
ge25 TOTAL T LANES IN D ____.	
ge26 TOTAL TR LANES IN D ____.	
ge27 TOTAL R LANES IN D ____.	
ge28 TOTAL LR LANES IN D ____.	

ge29 IS THE INTERSECTION ANGLE APPROX. PERPENDICULAR?

0 - NO 1 - YES

ge30 MAXIMUM VERT. ALIGNMENT (SLOPE)

0 = VISUALLY FLAT 1 = GRADE

(i.e. VISUAL UP/DOWNHILL)

ge31 MAXIMUM EMBANKMENT

0 = VISUALLY FLAT 1 = EMBANKED
(i.e. RACE CAR TRACKS ARE EMBANKED)

ge32 HORIZONTAL ALIGNMENT

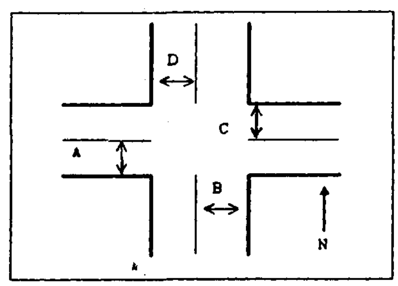
0 - STRAIGHT 1 - CURVED

ge33 APPROX. MINIMUM SIGHT DISTANCE
(IN EITHER PERTAINING DIRECTION)
____ (METERS)

ROAD CLASSIFICATION

ge34 R1		ge35 R2
1	PROVINCIAL EXPRESSWAY	1
2	CITY EXPRESSWAY	2
3	MAJOR ARTERIAL	3
4	MINOR ARTERIAL	4
5	COLLECTOR	5

DRAW R1 AND R2 ON DIAGRAM BELOW



15 MINUTE VEHICLE COUNT

ge36 R1 =

ge37 R2 =

ge38 NUMBER OF INTERSECTION LEGS

Sheet 4

Cycle 1	Phase	FAG or LTGA?		Yellow plus all red	Permitted Movements by Phase											
					Eastbound			Northbound			Westbound			Southbound		
					R	T	L	R	T	L	R	T	L	R	T	L
1	1															
	2															
	3															
	4															
	5															
	6															
2	1															
	2															
	3															
	4															
	5															
	6															
3	1															
	2															
	3															
	4															
	5															
	6															
4	1															
	2															
	3															
	4															
	5															
	6															
5	1															
	2															
	3															
	4															
	5															
	6															
Average (Compute in office)	1															
	2															
	3															
	4															
	5															
	6															

Instructions:

Go through 5 cycles when signal timing is not fixed. Otherwise after two identical cycles we can assume not fixed

Record movements that proceed in each phase, the green time for the phase and the yellow plus all red (if any) at the end of the phase. Treat advanced green as a separate phase.

X Sheet 5 Data Collection Sheet

LICENSE PLATES OF CONTROL VEHICLES

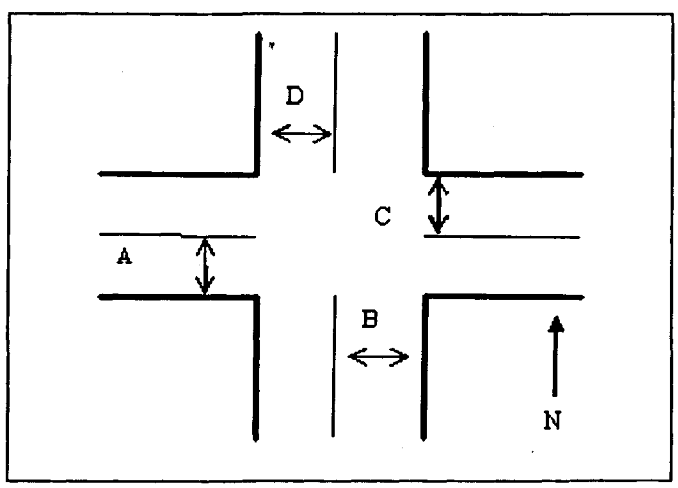
DIRECTION 1: NORTH / SOUTH / EAST / WEST VEHICLE INFO: LICENSE PLATE ____________________	DIRECTION 2: NORTH / SOUTH / EAST / WEST VEHICLE INFO: LICENSE PLATE ____________________
DIRECTION 1: NORTH / SOUTH / EAST / WEST VEHICLE INFO: LICENSE PLATE ____________________	DIRECTION 2: NORTH / SOUTH / EAST / WEST VEHICLE INFO: LICENSE PLATE ____________________
DIRECTION 1: NORTH / SOUTH / EAST / WEST VEHICLE INFO: LICENSE PLATE ____________________	DIRECTION 2: NORTH / SOUTH / EAST / WEST VEHICLE INFO: LICENSE PLATE ____________________
DIRECTION 1: NORTH / SOUTH / EAST / WEST VEHICLE INFO: LICENSE PLATE ____________________	DIRECTION 2: NORTH / SOUTH / EAST / WEST VEHICLE INFO: LICENSE PLATE ____________________

X Sheet 6 Data Collection Sheet

FORMv

A SUMMARY/DIAGRAM OF THE COLLISION SCENE & EVENTS

Note: This page is primarily for the control person's reference, to refresh when at scene.



X Sheet 7 Data Collection Sheet

GEOMETRIC FORM ii: Only complete for TARGET and BULLET VEHICLE LANE(S).

gee1 AVG L LANES WIDTH IN A ___.

gee2 AVG LT LANES WIDTH IN A ___.

gee3 AVG LTR LANES WIDTH IN A ___.

gee4 AVG T LANES WIDTH IN A ___.

gee5 AVG TR LANES WIDTH IN A ___.

gee6 AVG R LANES WIDTH IN A ___.

gee7 AVG LR LANES WIDTH IN A ___.

gee8 AVG L LANES WIDTH IN B ___.

gee9 AVG LT LANES WIDTH IN B ___.

gee10 AVG LTR LANES WIDTH IN B ___.

gee11 AVG T LANES WIDTH IN B ___.

gee12 AVG TR LANES WIDTH IN B ___.

gee13 AVG R LANES WIDTH IN B ___.

gee14 AVG LR LANES WIDTH IN B ___.

gee15 AVG L LANES WIDTH IN C ___.

gee16 AVG LT LANES WIDTH IN C ___.

gee17 AVG LTR LANES WIDTH IN C ___.

gee18 AVG T LANES WIDTH IN C ___.

gee19 AVG TR LANES WIDTH IN C ___.

gee20 AVG R LANES WIDTH IN C ___.

gee21 AVG LR LANES WIDTH IN C ___.

gee22 AVG L LANES WIDTH IN D ___.

gee23 AVG LT LANES WIDTH IN D ___.

gee24 AVG LTR LANES WIDTH IN D ___.

gee25 AVG T LANES WIDTH IN D ___.

gee26 AVG TR LANES WIDTH IN D ___.

gee27 AVG R LANES WIDTH IN D ___.

gee28 AVG LR LANES WIDTH IN D ___.

gee29 ROAD PAVEMENT MARKINGS?

R1		R2
1	Exist	1
2	NON-EXISTENT	2
3	OBSCURED	3
4	FADED	4

(i.e. PROTECTION = ISLAND, CURB, BARRIER...)

PROTECTION OTHER THAN PAINTED LINES IN:

gee30	A	0 NO	1 YES
gee31	B	0 NO	1 YES
gee32	C	0 NO	1 YES
gee33	D	0 NO	1 YES

(i.e. COMPLETE PAINTED LINES FOR LANE)

gee34 PAINTED LINES IN A?	0 NO	1 YES
gee35 PAINTED LINES IN B?	0 NO	1 YES
gee36 PAINTED LINES IN C?	0 NO	1 YES
gee37 PAINTED LINES IN D?	0 NO	1 YES

Note "Complete" painted lines meaning when multi-lane, markings must exist for centre line and lanes. When single lane, just centreline or more is acceptable

Appendix D: Regression Models Built

```
                         programmingforthesisprint
/* ALL DATA BOTH TARGET, BULLET AND INTERSECTION */

data tandbt;

input
        D1F             D2F                 D1SEX$  D2SEX$  D1INJ   D2INJ   V1P1AGE
V2P1AGE
        V1P1SEX$                 V2P1SEX$                 D1LIC$  D2LIC$  V1OCCUP
V2OCCUP
        V1P1INJ V2P1INJ V1YEAR  V2YEAR  V1COL$  V2COL$  V1BODY$ V2BODY$
        R1LANES R2LANES R1MAX   R2MAX   V1APSPD V2APSPD D1AGE   D2AGE
        V1ARROW V2ARROW MVA9    MVA10   MVA11   MVA12   MVA13   MVA14
        MVA15   MVA16   MVA17   MVA18   MVA19   MVA20   MVA21   MVA22
        MVA31   MVA32   MVA33   MVA34   MVA35   MVA36   MVA43   MVA44
        MVA46   MVA47   MVA48   MVA49   MVA50   MVA51   MVA52   MVA53
        MVA54   MVA55   MVA56   MVA57   MVA58   MVA59   MVA60   MVA61
        MVA62   MVA63   MVA64   MVA65   cityandint_PX   ANY_LA  FAGINT$
        Dst_Cint        City$   System$ FDW$    Staggrd$        Num_Legs
        Q_major Q_minor ACCDATE DAY_NO  ACCTIME INTLEGS LGTCLR  D1SFTYQP
        V2P2AGE V1P2SEX$                V2P2SEX$                V1P2INJ V2P2INJ MVA2
        MVA3    MVA4    MVA5    MVA6    MVA7    MVA8    MVA23   MVA24
        MVA41   MVA42   MVA45   PER_INV VEH_INV MUNIC   ACCLOC  TRAFCTL
        VEHTYPE IMPCTYP INVTYPE INJRY   SAFQUP  DRIVCT  DRIVCND NCLASS$
        SCLASS$ WCLASS$ ECLASS$ EBLANES NBLANES SBLANES WBLANES;
cards;

d
a
t
a
;
run;

proc print data=tandbt1;run;
proc freq data=tandbt;run;


/*  ****************Human Factors************************** */

/* Diagnostics - multicollinearity for factors of human based model, for Bullet*/

data tandbt;
set tandbt1;
if D2SEX='M' then SEX=1;
if D2SEX='F' then SEX=0;run;
proc print noobs data=tandbt;run;
proc corr data=tandbt1;
var      SEX /* D2SEX */   D2AGE    V2P1AGE    V2OCCUP    /*V2P1SEX*/    /*D2LIC*/
MVA34   MVA36;
run;

/* First Model Built for Bullet HUMAN Model*/
proc genmod data=tandbt desc;

class   MVA34  MVA36   D2SEX   V2P1SEX   D2LIC;
model   D2F=   MVA34   MVA36 D2SEX   D2AGE   V2P1AGE   V2OCCUP
        V2P1SEX   D2LIC/D=bin link=logit type3;
/*MVA34               7     1 3 4 6 7 8 10 */
contrast '3 vs 1' MVA34 -1 1 0 0 0 0 0;
contrast '4 vs 1' MVA34 -1 0 1 0 0 0 0;
contrast '6 vs 1' MVA34 -1 0 0 1 0 0 0;
contrast '7 vs 1' MVA34 -1 0 0 0 1 0 0;
                                              Page 1
```

```
                                  programmingforthesisprint
contrast '8 vs 1' MVA34 -1 0 0 0 0 1 0;
contrast '10 vs 1' MVA34 -1 0 0 0 0 0 1;
/*MVA36               5     0 1 2 6 8 */
contrast '0 vs 1' MVA36 -1 1 0 0 0;
contrast '2 vs 1' MVA36 0 -1 1 0 0;
contrast '6 vs 1' MVA36 0 -1 0 1 0;
contrast '8 vs 1' MVA36 0 -1 0 0 1;
                                 /*1 3 4 6 7 8 10*/
run;
/* Second Model Built for Bullet HUMAN Model*/
/* Changes from previous one include:  removal of D2SEX  since it holds
highest P value with P = 0.9649   */
proc genmod data=tandbt desc;

class   MVA34  MVA36   V2P1SEX   D2LIC;
model   D2F=   MVA34   MVA36   D2AGE   V2P1AGE   V2OCCUP   V2P1SEX
        D2LIC/D=bin link=logit type3; run;
/*MVA34               7     1 3 4 6 7 8 10 */
contrast '3 vs 1' MVA34 -1 1 0 0 0 0 0;
contrast '4 vs 1' MVA34 -1 0 1 0 0 0 0;
contrast '6 vs 1' MVA34 -1 0 0 1 0 0 0;
contrast '7 vs 1' MVA34 -1 0 0 0 1 0 0;
contrast '8 vs 1' MVA34 -1 0 0 0 0 1 0;
contrast '10 vs 1' MVA34 -1 0 0 0 0 0 1;
/*MVA36               5     0 1 2 6 8 */
contrast '0 vs 1' MVA36 -1 1 0 0 0;
contrast '2 vs 1' MVA36 0 -1 1 0 0;
contrast '6 vs 1' MVA36 0 -1 0 1 0;
contrast '8 vs 1' MVA36 0 -1 0 0 1;
                                 /*1 3 4 6 7 8 10*/
run;
/* Third Model Built for Bullet HUMAN Model*/
/* Changes from previous one include:  removal of V2OCCUP since it holds
highest P = 0.9303 */
proc genmod data=tandbt desc;

class   MVA34  MVA36   V2P1SEX   D2LIC;
model   D2F=   MVA34   MVA36   D2AGE   V2P1AGE   V2P1SEX
        D2LIC/D=bin link=logit type3; run;
/*MVA34               7     1 3 4 6 7 8 10 */
contrast '3 vs 1' MVA34 -1 1 0 0 0 0 0;
contrast '4 vs 1' MVA34 -1 0 1 0 0 0 0;
contrast '6 vs 1' MVA34 -1 0 0 1 0 0 0;
contrast '7 vs 1' MVA34 -1 0 0 0 1 0 0;
contrast '8 vs 1' MVA34 -1 0 0 0 0 1 0;
contrast '10 vs 1' MVA34 -1 0 0 0 0 0 1;
/*MVA36               5     0 1 2 6 8 */
contrast '0 vs 1' MVA36 -1 1 0 0 0;
contrast '2 vs 1' MVA36 0 -1 1 0 0;
contrast '6 vs 1' MVA36 0 -1 0 1 0;
contrast '8 vs 1' MVA36 0 -1 0 0 1;
                                 /*1 3 4 6 7 8 10*/
run;

/* Fourth Model Built for Bullet HUMAN Model*/
/* Changes from previous one include:  removal of V2P1SEX since it holds
highest P = 0.5161. */
proc genmod data=tandbt desc;

class   MVA34  MVA36   D2LIC;
model   D2F=   MVA34   MVA36   D2AGE   V2P1AGE
        D2LIC/D=bin link=logit type3;
/*MVA34               7     1 3 4 6 7 8 10 */
                                              Page 2
```

```
/*contrast '3 vs 1' MVA34 -1 1 0 0 0 0 0;
contrast '4 vs 1' MVA34 -1 0 1 0 0 0 0; */
contrast '6 vs 1' MVA34 -1 0 0 1 0 0 0;
contrast '7 vs 1' MVA34 -1 0 0 0 1 0 0;
contrast '8 vs 1' MVA34 -1 0 0 0 0 1 0;
contrast '10 vs 1' MVA34 -1 0 0 0 0 0 1;

/ MVA36              5    0 1 2 6 8 */
contrast '0 vs 1' MVA36 -1 1 0 0 0;
contrast '2 vs 1' MVA36 0 -1 1 0 0;
contrast '6 vs 1' MVA36 0 -1 0 1 0;
contrast '8 vs 1' MVA36 0 -1 0 0 1;

run; /* /* converges


/* Fifth Model Built for Bullet HUMAN Model*/
/* Changes From previous one include:  Removed D2LIC since it had
highest unacceptable P value =  0.4004 */
proc genmod data=tandbt desc;
class    MVA34  MVA36;
model    D2F=   MVA34   MVA36   D2AGE   V2P1AGE
         /D=bin link=logit type3;run;
/*MVA34               7    1 3 4 6 7 8 10 */
contrast '3 vs 1' MVA34 -1 1 0 0 0 0 0;
contrast '4 vs 1' MVA34 -1 0 1 0 0 0 0;
contrast '6 vs 1' MVA34 -1 0 0 1 0 0 0;
contrast '7 vs 1' MVA34 -1 0 0 0 1 0 0;
contrast '8 vs 1' MVA34 -1 0 0 0 0 1 0;
contrast '10 vs 1' MVA34 -1 0 0 0 0 0 1;
           /*     5    0 1 2 6 8 */
contrast '0 vs 1' MVA36 -1 1 0 0 0;
contrast '2 vs 1' MVA36 0 -1 1 0 0;
contrast '6 vs 1' MVA36 0 -1 0 1 0;
contrast '8 vs 1' MVA36 0 -1 0 0 1;

run;  /* no convergence... CONTINUED WITH a sixth analysis, as shown below
*/

/* Sixth Model Built for Bullet HUMAN Model*/
/* Changes from previous one include:  decision was made between choosing MVA36 and
V2P1AGE, and MVA36 was choosen to be removed.  Having many parameters (5) for
a rather small number of observations creates over-parameterization issues.
Removing this predictor seems prudent, but not "perfect".*/

proc genmod data=tandbt desc;

class    MVA34;
model    D2F=   MVA34   D2AGE   V2P1AGE
         /D=bin link=logit type3;
/*MVA34               7    1 3 4 6 7 8 10 */

contrast '3 vs 1' MVA34 -1 1 0 0 0 0 0;
contrast '4 vs 1' MVA34 -1 0 1 0 0 0 0;
contrast '6 vs 1' MVA34 -1 0 0 1 0 0 0;
contrast '7 vs 1' MVA34 -1 0 0 0 1 0 0;
contrast '8 vs 1' MVA34 -1 0 0 0 0 1 0;
contrast '10 vs 1' MVA34 -1 0 0 0 0 0 1;

estimate '3 vs 1' MVA34 -1 1 0 0 0 0 0;
estimate '4 vs 1' MVA34 -1 0 1 0 0 0 0;
estimate '6 vs 1' MVA34 -1 0 0 1 0 0 0;
estimate '7 vs 1' MVA34 -1 0 0 0 1 0 0;
```



```
estimate '8 vs 1' MVA34 -1 0 0 0 0 1 0;
estimate '10 vs 1' MVA34 -1 0 0 0 0 0 1;

run;
/* At this point the model converged, and this analysis has stopped at this point.
However, there are now less significant predictors! Therefore model 4 is considered
a good finishing point. */


/*****************************************************************************
****/
        /*              Bullet Analysis Continued:
                        Extra-Human factors:  Environmental and vehicle predictors
*/

/* Diagnostics - multicollinearity for factors of Environmental and vehicle
predictors based model, for Bullet*/

proc corr data=tandbt;
var     R2LANES R2MAX   V2APSPD  /* V2BODY */ V2YEAR    /* V2COL */
MVA10   MVA14   MVA16   MVA18   MVA20   MVA32
MVA47
/* MVA54         MVA55   MVA56   MVA57   MVA58   MVA59 MVA61     MVA64   MVA65  all
not good predictors and therefore removed */
;run;

/* First Model Built for Bullet Extra-Human Model*/
proc genmod data=tandbt desc;
class   V2BODY   V2COL   MVA10  MVA14   MVA16   MVA18   MVA20   MVA32  MVA47;
/*      MVA22 we have better variable for this (vehicle type) included already */
/* MVA44 fault and direction no way, not in this context. but ideal for further work
 */

model D2F= R2LANES   R2MAX   V2APSPD   V2BODY   V2YEAR   V2COL   MVA10
                MVA14     MVA16      MVA18     MVA20    MVA32  MVA47
/D=bin link=logit type3;

contrast '3DR VS 2DR' V2BODY -1 1 0 0;
contrast '4DR VS 2DR' V2BODY -1 0 1 0;
contrast 'LTTRUCK VS 2DR' V2BODY -1 0 0 1;

contrast '2 VS 1' MVA10 -1 1 0 0 0;
contrast '3 VS 1' MVA10 -1 0 1 0 0;
contrast '4 VS 1' MVA10 -1 0 0 1 0;
contrast '5 VS 1' MVA10 -1 0 0 0 1;

contrast '2 VS 1' MVA14 -1 1 0;
contrast '3 VS 1' MVA14 -1 0 1;

contrast '2 VS 1' MVA16 -1 1 0 0 0 0;
contrast '3 VS 1' MVA16 -1 0 1 0 0 0;
contrast '4 VS 1' MVA16 -1 0 0 1 0 0;
contrast '6 VS 1' MVA16 -1 0 0 0 1 0;
contrast '8 VS 1' MVA16 -1 0 0 0 1 0;

contrast '2 VS 1' MVA18 -1 1 0 0;
contrast '3 VS 1' MVA18 -1 0 1 0;
contrast '4 VS 1' MVA18 -1 0 0 1;

contrast '2 VS 1' MVA20 -1 1 0 0;
contrast '3 VS 1' MVA20 -1 0 1 0;
contrast '4 VS 1' MVA20 -1 0 0 1;
```

```
contrast '2 VS 1' MVA32 -1 1;

contrast '2 VS 1' MVA47 -1 1 0 0;
contrast '4 VS 1' MVA47 -1 0 1 0;
contrast '7 VS 1' MVA47 -1 0 0 1;
run;

/* Second Model Built for Bullet Extra-Human Model */
/* MVA20 and MVA32 removed */
proc genmod data=tandbt desc;
class   V2BODY   V2COL   MVA10  MVA14   MVA16   MVA18   MVA47;

model D2F= R2LANES   R2MAX   V2APSPD   V2BODY   V2YEAR   V2COL   MVA10
                 MVA14       MVA16        MVA18      MVA47
/D=bin link=logit type3;

contrast '3DR VS 2DR' V2BODY -1 1 0 0;
contrast '4DR VS 2DR' V2BODY -1 0 1 0;
contrast 'LTTRUCK VS 2DR' V2BODY -1 0 0 1;

contrast '2 VS 1' MVA10 -1 1 0 0 0;
contrast '3 VS 1' MVA10 -1 0 1 0 0;
contrast '4 VS 1' MVA10 -1 0 0 1 0;
contrast '5 VS 1' MVA10 -1 0 0 0 1;

contrast '2 VS 1' MVA14 -1 1 0;
contrast '3 VS 1' MVA14 -1 0 1;

contrast '2 VS 1' MVA16 -1 1 0 0 0 0;
contrast '3 VS 1' MVA16 -1 0 1 0 0 0;
contrast '4 VS 1' MVA16 -1 0 0 1 0 0;
contrast '6 VS 1' MVA16 -1 0 0 0 1 0;
contrast '8 VS 1' MVA16 -1 0 0 0 1 0;

contrast '2 VS 1' MVA18 -1 1 0 0;
contrast '3 VS 1' MVA18 -1 0 1 0;
contrast '4 VS 1' MVA18 -1 0 0 1;

contrast '2 VS 1' MVA47 -1 1 0 0;
contrast '4 VS 1' MVA47 -1 0 1 0;
contrast '7 VS 1' MVA47 -1 0 0 1;
run;
/* Third Model Built for Bullet Extra-Human Model */
/* MVA14 removed */

proc genmod data=tandbt desc;
class   V2BODY   V2COL   MVA10  MVA16   MVA18   MVA47;

model D2F= R2LANES   R2MAX   V2APSPD   V2BODY   V2YEAR   V2COL   MVA10
                 MVA16           MVA18      MVA47
/D=bin link=logit type3;

contrast '3DR VS 2DR' V2BODY -1 1 0 0;
contrast '4DR VS 2DR' V2BODY -1 0 1 0;
contrast 'LTTRUCK VS 2DR' V2BODY -1 0 0 1;

contrast '2 VS 1' MVA10 -1 1 0 0 0;
contrast '3 VS 1' MVA10 -1 0 1 0 0;
contrast '4 VS 1' MVA10 -1 0 0 1 0;
contrast '5 VS 1' MVA10 -1 0 0 0 1;

contrast '2 VS 1' MVA16 -1 1 0 0 0 0;
```



```
contrast '3 VS 1' MVA16 -1 0 1 0 0 0;
contrast '4 VS 1' MVA16 -1 0 0 1 0 0;
contrast '6 VS 1' MVA16 -1 0 0 0 1 0;
contrast '8 VS 1' MVA16 -1 0 0 0 1 0;

contrast '2 VS 1' MVA18 -1 1 0 0;
contrast '3 VS 1' MVA18 -1 0 1 0;
contrast '4 VS 1' MVA18 -1 0 0 1;

contrast '2 VS 1' MVA47 -1 1 0 0;
contrast '4 VS 1' MVA47 -1 0 1 0;
contrast '7 VS 1' MVA47 -1 0 0 1;
run;

/* Fourth Model Built for Bullet Extra-Human Model*/
/* V2COL removed */

proc genmod data=tandbt desc;
class   V2BODY   MVA10  MVA16   MVA18   MVA47;

model D2F= R2LANES   R2MAX   V2APSPD   V2BODY   V2YEAR   MVA10
                 MVA16          MVA18      MVA47
/D=bin link=logit type3;

contrast '3DR VS 2DR' V2BODY -1 1 0 0;
contrast '4DR VS 2DR' V2BODY -1 0 1 0;
contrast 'LTTRUCK VS 2DR' V2BODY -1 0 0 1;

contrast '2 VS 1' MVA10 -1 1 0 0 0;
contrast '3 VS 1' MVA10 -1 0 1 0 0;
contrast '4 VS 1' MVA10 -1 0 0 1 0;
contrast '5 VS 1' MVA10 -1 0 0 0 1;

contrast '2 VS 1' MVA16 -1 1 0 0 0 0;
contrast '3 VS 1' MVA16 -1 0 1 0 0 0;
contrast '4 VS 1' MVA16 -1 0 0 1 0 0;
contrast '6 VS 1' MVA16 -1 0 0 0 1 0;
contrast '8 VS 1' MVA16 -1 0 0 0 1 0;

contrast '2 VS 1' MVA18 -1 1 0 0;
contrast '3 VS 1' MVA18 -1 0 1 0;
contrast '4 VS 1' MVA18 -1 0 0 1;

contrast '2 VS 1' MVA47 -1 1 0 0;
contrast '4 VS 1' MVA47 -1 0 1 0;
contrast '7 VS 1' MVA47 -1 0 0 1;
run;
/* Fifth Model Built for Bullet Extra-Human Model*/
/* V2BODY removed */

proc genmod data=tandbt desc;
class   MVA10   MVA16   MVA18   MVA47;

model D2F= R2LANES   R2MAX   V2APSPD   V2YEAR   MVA10
                 MVA16          MVA18      MVA47
/D=bin link=logit type3;

contrast '2 VS 1' MVA10 -1 1 0 0 0;
contrast '3 VS 1' MVA10 -1 0 1 0 0;
contrast '4 VS 1' MVA10 -1 0 0 1 0;
contrast '5 VS 1' MVA10 -1 0 0 0 1;

contrast '2 VS 1' MVA16 -1 1 0 0 0 0;
```

```
contrast '3 VS 1' MVA16 -1 0 1 0 0 0;
contrast '4 VS 1' MVA16 -1 0 0 1 0 0;
contrast '6 VS 1' MVA16 -1 0 0 0 1 0;
contrast '8 VS 1' MVA16 -1 0 0 0 1 0;

contrast '2 VS 1' MVA18 -1 1 0 0;
contrast '3 VS 1' MVA18 -1 0 1 0;
contrast '4 VS 1' MVA18 -1 0 0 1;

contrast '2 VS 1' MVA47 -1 1 0 0;
contrast '4 VS 1' MVA47 -1 0 1 0;
contrast '7 VS 1' MVA47 -1 0 0 1;
run;

/* Sixth Model Built for Bullet Extra-Human Model*/
/* MVA16 removed */

proc genmod data=tandbt desc;
class   MVA10   MVA18   MVA47;

model D2F= R2LANES   R2MAX   V2APSPD   V2YEAR   MVA10
                  MVA18               MVA47
/D=bin link=logit type3;

contrast '2 VS 1' MVA10 -1 1 0 0 0;
contrast '3 VS 1' MVA10 -1 0 1 0 0;
contrast '4 VS 1' MVA10 -1 0 0 1 0;
contrast '5 VS 1' MVA10 -1 0 0 0 1;

contrast '2 VS 1' MVA18 -1 1 0 0;
contrast '3 VS 1' MVA18 -1 0 1 0;
contrast '4 VS 1' MVA18 -1 0 0 1;

contrast '2 VS 1' MVA47 -1 1 0 0;
contrast '4 VS 1' MVA47 -1 0 1 0;
contrast '7 VS 1' MVA47 -1 0 0 1;
run;

/* Seventh Model Built for Bullet Extra-Human Model*/
/* MVA47 removed */

proc genmod data=tandbt desc;
class   MVA10   MVA18;

model D2F= R2LANES   R2MAX   V2APSPD   V2YEAR   MVA10
                  MVA18
/D=bin link=logit type3;

contrast '2 VS 1' MVA10 -1 1 0 0 0;
contrast '3 VS 1' MVA10 -1 0 1 0 0;
contrast '4 VS 1' MVA10 -1 0 0 1 0;
contrast '5 VS 1' MVA10 -1 0 0 0 1;

contrast '2 VS 1' MVA18 -1 1 0 0;
contrast '3 VS 1' MVA18 -1 0 1 0;
contrast '4 VS 1' MVA18 -1 0 0 1;
run;
/* Eigth Model Built for Bullet Extra-Human Model*/
/* V2YEAR removed */

proc genmod data=tandbt desc;
class   MVA10   MVA18;
```



```
model D2F= R2LANES   R2MAX   V2APSPD  MVA10
                  MVA18
/D=bin link=logit type3;

contrast '2 VS 1' MVA10 -1 1 0 0 0;
contrast '3 VS 1' MVA10 -1 0 1 0 0;
contrast '4 VS 1' MVA10 -1 0 0 1 0;
contrast '5 VS 1' MVA10 -1 0 0 0 1;

contrast '2 VS 1' MVA18 -1 1 0 0;
contrast '3 VS 1' MVA18 -1 0 1 0;
contrast '4 VS 1' MVA18 -1 0 0 1;
run;
/* Ninth Model Built for Bullet Extra-Human Model*/
/* MVA18 removed */

proc genmod data=tandbt desc;
class   MVA10;

model D2F= R2LANES   R2MAX   V2APSPD  MVA10

/D=bin link=logit type3;

contrast '2 VS 1' MVA10 -1 1 0 0 0;
contrast '3 VS 1' MVA10 -1 0 1 0 0;
contrast '4 VS 1' MVA10 -1 0 0 1 0;
contrast '5 VS 1' MVA10 -1 0 0 0 1;

run;
/* Tenth and final Model Built for Bullet Extra-Human Model*/
/* R2LANES removed */
proc genmod data=tandbt desc;
class   MVA10;

model D2F=  R2MAX   V2APSPD  MVA10
/D=bin link=logit type3;

contrast '2 VS 1' MVA10 -1 1 0 0 0;
contrast '3 VS 1' MVA10 -1 0 1 0 0;
contrast '4 VS 1' MVA10 -1 0 0 1 0;
contrast '5 VS 1' MVA10 -1 0 0 0 1;

estimate '2 VS 1' MVA10 -1 1 0 0 0;
estimate '3 VS 1' MVA10 -1 0 1 0 0;
estimate '4 VS 1' MVA10 -1 0 0 1 0;
estimate '5 VS 1' MVA10 -1 0 0 0 1;

/* All significant and converging */
run;


/*******************************************************************************
*******************************************************************************/

/* Now, here begins the Target Fault and not at fault analysis July 1st 2003
john bouyounes */

/* Diagnostics - multicollinearity */

proc corr data=tandbt;
var
/* D1SEX */      D1AGE   V1P1AGE  V1OCCUP        /* V1P1SEX */,    /* D1LIC */
```

```
MVA33   MVA35;
run;
/* First Model Built for Target HUMAN Model */
proc genmod data=tandbt desc;
class   MVA33   MVA35  D1SEX   V1P1SEX   D1LIC;
model  D1F=   MVA33   MVA35   D1SEX   D1AGE        V1P1AGE   V1OCCUP      V1P1SEX
D1LIC/D=bin link=logit type3;

contrast '3 vs 1' MVA33 -1 1 0 0 0 0 0 0;
contrast '4 vs 1' MVA33 -1 0 1 0 0 0 0 0;
contrast '5 vs 1' MVA33 -1 0 0 1 0 0 0 0;
contrast '6 vs 1' MVA33 -1 0 0 0 1 0 0 0;
contrast '7 vs 1' MVA33 -1 0 0 0 0 1 0 0;
contrast '8 vs 1' MVA33 -1 0 0 0 0 0 1 0;
contrast '12 Vs 1' MVA33 -1 0 0 0 0 0 0 1;
/*MVA35             5    1 2 3 6 8 */
contrast '2 Vs 1' MVA35 -1 1 0 0 0;
contrast '3 vs 1' MVA35 -1 0 1 0 0;
contrast '6 vs 1' MVA35 -1 0 0 1 0;
contrast '8 Vs 1' MVA35 -1 0 0 0 1;
run;

/* Second Model Built for Target HUMAN Model*/
/* Changes from previous one include:  removal of MVA35 SINCE IT SHOWS POOR P
VALUES AND HAS MANY PAREMTERS THUS HAVING POTENTIAL FOR OVER-PARAMETERIZATION ISSUES
*/

proc genmod data=tandbt desc;
class   MVA33   D1SEX   V1P1SEX   D1LIC;
model  D1F=   MVA33   D1SEX   D1AGE        V1P1AGE   V1OCCUP      V1P1SEX
D1LIC/D=bin link=logit type3;

contrast '3 vs 1' MVA33 -1 1 0 0 0 0 0 0;
contrast '4 vs 1' MVA33 -1 0 1 0 0 0 0 0;
contrast '5 vs 1' MVA33 -1 0 0 1 0 0 0 0;
contrast '6 vs 1' MVA33 -1 0 0 0 1 0 0 0;
contrast '7 vs 1' MVA33 -1 0 0 0 0 1 0 0;
contrast '8 vs 1' MVA33 -1 0 0 0 0 0 1 0;
contrast '12 vs 1' MVA33 -1 0 0 0 0 0 0 1;

run;

/* 3 Model Built for Target HUMAN Model*/
/* Changes from previous one include:  removal of D1LIC AS P = 0.8577  */

proc genmod data=tandbt desc;
class   MVA33   D1SEX   V1P1SEX;
model  D1F=   MVA33   D1SEX   D1AGE        V1P1AGE   V1OCCUP      V1P1SEX
/D=bin link=logit type3;

contrast '3 Vs 1' MVA33 -1 1 0 0 0 0 0 0;
contrast '4 vs 1' MVA33 -1 0 1 0 0 0 0 0;
contrast '5 vs 1' MVA33 -1 0 0 1 0 0 0 0;
contrast '6 vs 1' MVA33 -1 0 0 0 1 0 0 0;
contrast '7 vs 1' MVA33 -1 0 0 0 0 1 0 0;
contrast '8 vs 1' MVA33 -1 0 0 0 0 0 1 0;
contrast '12 vs 1' MVA33 -1 0 0 0 0 0 0 1;

run;

/* 4 Model Built for Target HUMAN Model*/
/* Changes from previous one include:  removal of D1SEX AS P = 0.8819  */
```



```
proc genmod data=tandbt desc;
class   MVA33   V1P1SEX;
model  D1F=   MVA33   D1AGE        V1P1AGE   V1OCCUP     V1P1SEX
/D=bin link=logit type3;

contrast '3 Vs 1' MVA33 -1 1 0 0 0 0 0 0;
contrast '4 Vs 1' MVA33 -1 0 1 0 0 0 0 0;
contrast '5 Vs 1' MVA33 -1 0 0 1 0 0 0 0;
contrast '6 vs 1' MVA33 -1 0 0 0 1 0 0 0;
contrast '7 vs 1' MVA33 -1 0 0 0 0 1 0 0;
contrast '8 Vs 1' MVA33 -1 0 0 0 0 0 1 0;
contrast '12 Vs 1' MVA33 -1 0 0 0 0 0 0 1;

run;

/* 5 Model Built for Target HUMAN Model */
/* Changes from previous one include:  removal of D1AGE AS P = 0.8460  */

proc genmod data=tandbt desc;
class   MVA33   V1P1SEX;
model  D1F=   MVA33   V1P1AGE   V1OCCUP        V1P1SEX
/D=bin link=logit type3;

contrast '3 Vs 1' MVA33 -1 1 0 0 0 0 0 0;
contrast '4 Vs 1' MVA33 -1 0 1 0 0 0 0 0;
contrast '5 Vs 1' MVA33 -1 0 0 1 0 0 0 0;
contrast '6 vs 1' MVA33 -1 0 0 0 1 0 0 0;
contrast '7 Vs 1' MVA33 -1 0 0 0 0 1 0 0;
contrast '8 vs 1' MVA33 -1 0 0 0 0 0 1 0;
contrast '12 Vs 1' MVA33 -1 0 0 0 0 0 0 1;

run;

/* 6 Model Built for Target HUMAN Model*/
/* Changes from previous one include:  removal of V1P1SEX P = 0.3727  */

proc genmod data=tandbt desc;
class   MVA33;
model  D1F=   MVA33   V1P1AGE   V1OCCUP
/D=bin link=logit type3;

contrast '3 vs 1' MVA33 -1 1 0 0 0 0 0 0;
contrast '4 vs 1' MVA33 -1 0 1 0 0 0 0 0;
contrast '5 vs 1' MVA33 -1 0 0 1 0 0 0 0;
contrast '6 vs 1' MVA33 -1 0 0 0 1 0 0 0;
contrast '7 vs 1' MVA33 -1 0 0 0 0 1 0 0;
contrast '8 vs 1' MVA33 -1 0 0 0 0 0 1 0;
contrast '12 vs 1' MVA33 -1 0 0 0 0 0 0 1;

estimate '3 vs 1' MVA33 -1 1 0 0 0 0 0 0;
estimate '4 vs 1' MVA33 -1 0 1 0 0 0 0 0;
estimate '5 vs 1' MVA33 -1 0 0 1 0 0 0 0;
estimate '6 vs 1' MVA33 -1 0 0 0 1 0 0 0;
estimate '7 vs 1' MVA33 -1 0 0 0 0 1 0 0;
estimate '8 vs 1' MVA33 -1 0 0 0 0 0 1 0;
estimate '12 vs 1' MVA33 -1 0 0 0 0 0 0 1;

run; /*  Great.  */

/*************************************************************************************
****/
                    /*               TARGET Analysis Continued:
```

```
*/

/* First Model Built for Target Extra-Human Model /

proc genmod data=tandbt desc;
class   V1BODY   V1COL   MVA9   MVA11   MVA13   MVA15   MVA17   MVA19   MVA31 MVA46;
model   D1F = R1LANES   R1MAX   V1APSPD   V1BODY   V1YEAR   V1COL   MVA9   MVA11
MVA13   MVA15   MVA17   MVA19   MVA31  MVA46
/D=bin link=logit type3;
/*
contrast '3DR VS 2DR' V2BODY -1 1 0 0;
contrast '4DR VS 2DR' V2BODY -1 0 1 0;
contrast 'LTTRUCK VS 2DR' V2BODY -1 0 0 1;

contrast '2 VS 1' MVA10 -1 1 0 0 0;
contrast '3 VS 1' MVA10 -1 0 1 0 0;
contrast '4 VS 1' MVA10 -1 0 0 1 0;
contrast '5 VS 1' MVA10 -1 0 0 0 1;

contrast '2 VS 1' MVA14 -1 1 0;
contrast '3 VS 1' MVA14 -1 0 1;

contrast '2 VS 1' MVA16 -1 1 0 0 0 0;
contrast '3 VS 1' MVA16 -1 0 1 0 0 0;
contrast '4 VS 1' MVA16 -1 0 0 1 0 0;
contrast '6 VS 1' MVA16 -1 0 0 0 1 0;
contrast '8 VS 1' MVA16 -1 0 0 0 1 0;

contrast '2 VS 1' MVA18 -1 1 0 0;
contrast '3 VS 1' MVA18 -1 0 1 0;
contrast '4 VS 1' MVA18 -1 0 0 1;

contrast '2 VS 1' MVA20 -1 1 0 0;
contrast '3 VS 1' MVA20 -1 0 1 0;
contrast '4 VS 1' MVA20 -1 0 0 1;

contrast '2 VS 1' MVA32 -1 1;

contrast '2 VS 1' MVA47 -1 1 0 0;
contrast '4 VS 1' MVA47 -1 0 1 0;
contrast '7 VS 1' MVA47 -1 0 0 1;

THIS IS NOT NECESSARY AND HAS BEEN COMMENTED OUT, INCASE ANY OF THESE ARE ACTUALL
REQUIRED */
run;
/* Second Model Built for Target Extra-Human Model*/
/*  Removed MVA11, MVA13 and MVA19.   Poor data distribtion (all are in the first
level) */
proc genmod data=tandbt desc;
class   V1BODY   V1COL   MVA9   MVA15   MVA17   MVA31 MVA46;
model   D1F = R1LANES   R1MAX   V1APSPD   V1BODY   V1YEAR   V1COL   MVA9
MVA15   MVA17   MVA31  MVA46
/D=bin link=logit type3;run;

/* Third Model Built for Target Extra-Human Model */
/*  Removed MVA31.  Not at all significant. */
proc genmod data=tandbt desc;
class   V1BODY   V1COL   MVA9   MVA15   MVA17   MVA46;
model   D1F = R1LANES   R1MAX   V1APSPD   V1BODY   V1YEAR   V1COL   MVA9
MVA15   MVA17  MVA46
/D=bin link=logit type3;run;
```



```
/* Fourth Model Built for Target Extra-Human Model*/
/*  Removed MVA15.  Not at all significant. */
proc genmod data=tandbt desc;
class   V1BODY   V1COL   MVA9   MVA17   MVA46;
model   D1F = R1LANES   R1MAX   V1APSPD   V1BODY   V1YEAR   V1COL   MVA9
MVA17  MVA46
/D=bin link=logit type3;run;

/* Fifth Model Built for Target Extra-Human Model */
/*  Removed MVA17.  Not at all significant. */
proc genmod data=tandbt desc;
class   V1BODY   V1COL   MVA9   MVA46;
model   D1F = R1LANES   R1MAX   V1APSPD   V1BODY   V1YEAR   V1COL   MVA9
MVA46
/D=bin link=logit type3;run;

/* Sixth Model Built for Target Extra-Human Model*/
/*  Removed MVA46.  Low significance and over-parameterization issues. */
proc genmod data=tandbt desc;
class   V1BODY   V1COL   MVA9;
model   D1F = R1LANES   R1MAX   V1APSPD   V1BODY   V1YEAR   V1COL   MVA9
/D=bin link=logit type3;run;

/* Seventh Model Built for Target Extra-Human Model*/
/*  Removed V1BODY.  Low significance and over-parameterization issues. */
proc genmod data=tandbt desc;
class   V1COL   MVA9;
model   D1F = R1LANES   R1MAX   V1APSPD   V1YEAR   V1COL   MVA9
/D=bin link=logit type3;run;
/* Eighth Model Built for Target Extra-Human Model*/
/*  Removed V1year.  P  */
proc genmod data=tandbt desc;
class   V1COL   MVA9;
model   D1F = R1LANES   R1MAX   V1APSPD     V1COL   MVA9
/D=bin link=logit type3;run;

/* Ninth Model Built for Target Extra-Human Model*/
/*  Removed V1COL.  P */
proc genmod data=tandbt desc;
class   MVA9;
model   D1F = R1LANES   R1MAX   V1APSPD  MVA9
/D=bin link=logit type3;run;
/* FINAL AND Tenth Model Built for Target Extra-Human Model*/
/*  Removed R1MAX.  P */
proc genmod data=tandbt desc;
class   MVA9;
model   D1F = R1LANES   V1APSPD  MVA9
/D=bin link=logit type3;

contrast '2 VS 1' MVA9 -1 1 0 0 0;
contrast '3 VS 1' MVA9 -1 0 1 0 0;
contrast '4 VS 1' MVA9 -1 0 0 1 0;
contrast '5 VS 1' MVA9 -1 0 0 0 1;

estimate '2 VS 1' MVA9 -1 1 0 0 0;
estimate '3 VS 1' MVA9 -1 0 1 0 0;
estimate '4 VS 1' MVA9 -1 0 0 1 0;
estimate '5 VS 1' MVA9 -1 0 0 0 1;

run;
/* FINAL, ALL REMAINING VARIABLES ARE SIGNIFICANT */
```

```
run;

/**** INTERSECTION ANALYSIS ON ALL INTERSECTIONS*************************/

data tandbt;
/* USEFUL FOR OTHER ANALYSIS ON THE 363 R1LANES R2LANES R1MAX    R2MAX
MVA9    MVA10   MVA11   MVA12   MVA13   MVA14   MVA15   MVA16
MVA17   MVA18   MVA19   MVA20   MVA43   MVA44   MVA7    MVA41

ANY_LA$  Dst_Cint      City$  System$        FDW$    StaggrdS        Num_Legs

NCLASS$ SCLASS$ WCLASS$ ECLASS$
EBLANES NBLANES SBLANES WBLANES
*/

data intersctn;
input
PX      Inj     Inj_fat Inj_fat_PDO     System$ FAGINT$ FDW$     Num_Legs
Staggered$      City$
Q_major Q_minor DstInt   Any_LA$        N$      S$      W$       E$;

cards;
d
a
t
a
;
run;
;
proc freq;
run;
/* These are the other TWO models one could eventually build in later time: Inj
Inj_fat,
and ultimately compare against the all coll models namely, Inj_fat_PDO. Actually,
since there are only a few fatalities, this is not a good idea with the limited data
uesd. */

proc genmod data=intersctn desc;
class   System   FAGINT   FDW   Num_Legs   Staggered   City   Any_LA
N   S   W   E;
model Inj_fat_PDO       =
System   FAGINT   FDW   Num_Legs   Staggered   City   Q_major
Q_minor   DstInt   Any_LA   N   S   W   E/D=bin link=logit type3;
run;
/* wow, even from the first run, convergence occurs */

 /* 2nd run, with DstInt removed */
proc genmod data=intersctn desc;
class   System   FAGINT   FDW   Num_Legs   Staggered   City   Any_LA
N   S   W   E;
model Inj_fat_PDO       =
System   FAGINT   FDW   Num_Legs   Staggered   City   Q_major
Q_minor      Any_LA   N   S   W   E/D=bin link=logit type3;
run;

 /* 3rd run, with Any_LA removed */
proc genmod data=intersctn desc;
class   System   FAGINT   FDW   Num_Legs   Staggered   City
N   S   W   E;
model Inj_fat_PDO       =
System   FAGINT   FDW   Num_Legs   Staggered   City   Q_major
```



```
Q_minor         N   S   W   E/D=bin link=logit type3;
run.

 /* 4thd run, with Q_minor removed */
proc genmod data=intersctn desc;
class   System   FAGINT   FDW   Num_Legs   Staggered   City
N   S   W   E;
model Inj_fat_PDO       =
System   FAGINT   FDW   Num_Legs   Staggered   City   Q_major
      N   S   W   E/D=bin link=logit type3;
run;

 /* 5th run, with  City removed */
proc genmod data=intersctn desc;
class   System   FAGINT   FDW   Num_Legs   Staggered
N   S   W   E;
        model Inj_fat_PDO       =
        System   FAGINT   FDW   Num_Legs   Staggered      Q_major
           N   S   W   E/D=bin link=logit type3;

contrast '3 vs 4 lanes' Num_Legs  1 -1 0;
contrast 'other vs 4 lanes' Num_Legs 0 -1 1;
estimate '3 vs 4 lanes' Num_Legs 1 -1 0;
estimate 'other vs 4 lanes' Num_Legs 0 -1 1;

contrast 'Northbound Collector vs Major Arterial' N 1 0 -1 0 0;
contrast 'Northbound Local vs Major Arterial' N 0 1 -1 0 0;
contrast 'Northbound Minor vs Major Arterial' N 0 0 -1 1 0;
contrast 'Northbound All other road classes vs Major Arterial' N 0 0 -1 0 1;
estimate 'Northbound Collector vs Major Arterial' N 1 0 -1 0 0;
estimate 'Northbound Local vs Major Arterial' N 0 1 -1 0 0;
estimate 'Northbound Minor vs Major Arterial' N 0 0 -1 1 0;
estimate 'Northbound All other road classes vs Major Arterial' N 0 0 -1 0 1;

contrast 'Southbound Collector vs Major Arterial' S 1 0 -1 0 0;
contrast 'Southbound Local vs Major Arterial' S 0 1 -1 0 0;
contrast 'Southbound Minor vs Major Arterial' S 0 0 -1 1 0;
contrast 'Southbound All other road classes vs Major Arterial' S 0 0 -1 0 1;
estimate 'Southbound Collector vs Major Arterial' S 1 0 -1 0 0;
estimate 'Southbound Local vs Major Arterial' S 0 1 -1 0 0;
estimate 'Southbound Minor vs Major Arterial' S 0 0 -1 1 0;
estimate 'Southbound All other road classes vs Major Arterial' S 0 0 -1 0 1;

contrast 'Westbound Collector vs Major Arterial' W 1 0 -1 0 0;
contrast 'Westbound Local vs Major Arterial' W 0 1 -1 0 0;
contrast 'Westbound Minor vs Major Arterial' W 0 0 -1 1 0;
contrast 'Westbound All other road classes vs Major Arterial' W 0 0 -1 0 1;
estimate 'Westbound Collector vs Major Arterial' W 1 0 -1 0 0;
estimate 'Westbound Local vs Major Arterial' W 0 1 -1 0 0;
estimate 'Westbound Minor vs Major Arterial' W 0 0 -1 1 0;
estimate 'Westbound All other road classes vs Major Arterial' W 0 0 -1 0 1;

contrast 'Eastbound Collector vs Major Arterial' E 1 0 -1 0 0;
contrast 'Eastbound Local vs Major Arterial' E 0 1 -1 0 0;
contrast 'Eastbound Minor vs Major Arterial' E 0 0 -1 1 0;
contrast 'Eastbound All other road classes vs Major Arterial' E 0 0 -1 0 1;
estimate 'Eastbound Collector vs Major Arterial' E 1 0 -1 0 0;
estimate 'Eastbound Local vs Major Arterial' E 0 1 -1 0 0;
estimate 'Eastbound Minor vs Major Arterial' E 0 0 -1 1 0;
estimate 'Eastbound All other road classes vs Major Arterial' E 0 0 -1 0 1;
run;

 /* 6th run, with  E removed */
```

```
proc genmod data=intersctn desc;
class   System   FAGINT   FDw   Num_Legs   Staggered
N   S   w   ;
        model Inj_fat_PDO        =
        System   FAGINT   FDw   Num_Legs   Staggered      Q_major
           N   S   w   /D=bin link=logit type3;

contrast '3 vs 4 lanes' Num_Legs  1 -1 0;
contrast 'other vs 4 lanes' Num_Legs 0 -1 1;
estimate '3 vs 4 lanes' Num_Legs 1 -1 0;
estimate 'other vs 4 lanes' Num_Legs 0 -1 1;

contrast 'Northbound Collector vs Major Arterial' N 1 0 -1 0 0;
contrast 'Northbound Local vs Major Arterial' N 0 1 -1 0 0;
contrast 'Northbound Minor vs Major Arterial' N 0 0 -1 1 0;
contrast 'Northbound All other road classes vs Major Arterial' N 0 0 -1 0 1;
estimate 'Northbound Collector vs Major Arterial' N 1 0 -1 0 0;
estimate 'Northbound Local vs Major Arterial' N 0 1 -1 0 0;
estimate 'Northbound Minor vs Major Arterial' N 0 0 -1 1 0;
estimate 'Northbound All other road classes vs Major Arterial' N 0 0 -1 0 1;

contrast 'Southbound Collector vs Major Arterial' S 1 0 -1 0 0;
contrast 'Southbound Local vs Major Arterial' S 0 1 -1 0 0;
contrast 'Southbound Minor vs Major Arterial' S 0 0 -1 1 0;
contrast 'Southbound All other road classes vs Major Arterial' S 0 0 -1 0 1;
estimate 'Southbound Collector vs Major Arterial' S 1 0 -1 0 0;
estimate 'Southbound Local vs Major Arterial' S 0 1 -1 0 0;
estimate 'Southbound Minor vs Major Arterial' S 0 0 -1 1 0;
estimate 'Southbound All other road classes vs Major Arterial' S 0 0 -1 0 1;

contrast 'westbound Collector vs Major Arterial' w 1 0 -1 0 0;
contrast 'westbound Local vs Major Arterial' w 0 1 -1 0 0;
contrast 'westbound Minor vs Major Arterial' w 0 0 -1 1 0;
contrast 'westbound All other road classes vs Major Arterial' w 0 0 -1 0 1;
estimate 'westbound Collector vs Major Arterial' w 1 0 -1 0 0;
estimate 'westbound Local vs Major Arterial' w 0 1 -1 0 0;
estimate 'westbound Minor vs Major Arterial' w 0 0 -1 1 0;
estimate 'westbound All other road classes vs Major Arterial' w 0 0 -1 0 1;

run;

 /* 7th run. with  FDw removed */
proc genmod data=intersctn desc;
class   System   FAGINT   Num_Legs   Staggered
N   S   w   ;
        model Inj_fat_PDO        =
        System   FAGINT   Num_Legs   Staggered      Q_major
           N   S   w   /D=bin link=logit type3;

contrast '3 vs 4 lanes' Num_Legs  1 -1 0;
contrast 'other vs 4 lanes' Num_Legs 0 -1 1;
estimate '3 vs 4 lanes' Num_Legs 1 -1 0;
estimate 'other vs 4 lanes' Num_Legs 0 -1 1;

contrast 'Northbound Collector vs Major Arterial' N 1 0 -1 0 0;
contrast 'Northbound Local vs Major Arterial' N 0 1 -1 0 0;
contrast 'Northbound Minor vs Major Arterial' N 0 0 -1 1 0;
contrast 'Northbound All other road classes vs Major Arterial' N 0 0 -1 0 1;
estimate 'Northbound Collector vs Major Arterial' N 1 0 -1 0 0;
estimate 'Northbound Local vs Major Arterial' N 0 1 -1 0 0;
estimate 'Northbound Minor vs Major Arterial' N 0 0 -1 1 0;
estimate 'Northbound All other road classes vs Major Arterial' N 0 0 -1 0 1;
```



```
contrast 'Southbound Collector vs Major Arterial' S 1 0 -1 0 0;
contrast 'Southbound Local vs Major Arterial' S 0 1 -1 0 0;
contrast 'Southbound Minor vs Major Arterial' S 0 0 -1 1 0;
contrast 'Southbound All other road classes vs Major Arterial' S 0 0 -1 0 1;
estimate 'Southbound Collector vs Major Arterial' S 1 0 -1 0 0;
estimate 'Southbound Local vs Major Arterial' S 0 1 -1 0 0;
estimate 'Southbound Minor vs Major Arterial' S 0 0 -1 1 0;
estimate 'Southbound All other road classes vs Major Arterial' S 0 0 -1 0 1;

contrast 'westbound Collector vs Major Arterial' w 1 0 -1 0 0;
contrast 'westbound Local vs Major Arterial' w 0 1 -1 0 0;
contrast 'westbound Minor vs Major Arterial' w 0 0 -1 1 0;
contrast 'westbound All other road classes vs Major Arterial' w 0 0 -1 0 1;
estimate 'westbound Collector vs Major Arterial' w 1 0 -1 0 0;
estimate 'westbound Local vs Major Arterial' w 0 1 -1 0 0;
estimate 'westbound Minor vs Major Arterial' w 0 0 -1 1 0;
estimate 'westbound All other road classes vs Major Arterial' w 0 0 -1 0 1;

run;
/* 8th run. with  Staggered removed */
proc genmod data=intersctn desc;
class   System   FAGINT   Num_Legs
N   S   w   ;
        model Inj_fat_PDO        =
        System   FAGINT   Num_Legs   Q_major
           N   S   w   /D=bin link=logit type3;

contrast '3 vs 4 lanes' Num_Legs  1 -1 0;
contrast 'other vs 4 lanes' Num_Legs 0 -1 1;
estimate '3 vs 4 lanes' Num_Legs 1 -1 0;
estimate 'other vs 4 lanes' Num_Legs 0 -1 1;

contrast 'Northbound Collector vs Major Arterial' N 1 0 -1 0 0;
contrast 'Northbound Local vs Major Arterial' N 0 1 -1 0 0;
contrast 'Northbound Minor vs Major Arterial' N 0 0 -1 1 0;
contrast 'Northbound All other road classes vs Major Arterial' N 0 0 -1 0 1;
estimate 'Northbound Collector vs Major Arterial' N 1 0 -1 0 0;
estimate 'Northbound Local vs Major Arterial' N 0 1 -1 0 0;
estimate 'Northbound Minor vs Major Arterial' N 0 0 -1 1 0;
estimate 'Northbound All other road classes vs Major Arterial' N 0 0 -1 0 1;

contrast 'Southbound Collector vs Major Arterial' S 1 0 -1 0 0;
contrast 'Southbound Local vs Major Arterial' S 0 1 -1 0 0;
contrast 'Southbound Minor vs Major Arterial' S 0 0 -1 1 0;
contrast 'Southbound All other road classes vs Major Arterial' S 0 0 -1 0 1;
estimate 'Southbound Collector vs Major Arterial' S 1 0 -1 0 0;
estimate 'Southbound Local vs Major Arterial' S 0 1 -1 0 0;
estimate 'Southbound Minor vs Major Arterial' S 0 0 -1 1 0;
estimate 'Southbound All other road classes vs Major Arterial' S 0 0 -1 0 1;

contrast 'westbound Collector vs Major Arterial' w 1 0 -1 0 0;
contrast 'westbound Local vs Major Arterial' w 0 1 -1 0 0;
contrast 'westbound Minor vs Major Arterial' w 0 0 -1 1 0;
contrast 'westbound All other road classes vs Major Arterial' w 0 0 -1 0 1;
estimate 'westbound Collector vs Major Arterial' w 1 0 -1 0 0;
estimate 'westbound Local vs Major Arterial' w 0 1 -1 0 0;
estimate 'westbound Minor vs Major Arterial' w 0 0 -1 1 0;
estimate 'westbound All other road classes vs Major Arterial' w 0 0 -1 0 1;

run;
/* this is good */
```

Appendix E: Model Output

firstbullethuman

First Model for Bullet Human Factors

The SAS System 00:43 Thursday, August 28, 2003 69

The GENMOD Procedure

Model Information

Data Set	WORK.TANDBT
Distribution	Binomial
Link Function	Logit
Dependent Variable	D2F
Observations Used	140
Missing Values	624

Class Level Information

Class	Levels	Values
MVA34	7	1 3 4 6 7 8 10
MVA36	4	0 1 2 8
D2SEX	2	F M
V2P1SEX	2	F M
D2LIC	5	G G1 G2 GM_any other

Response Profile

Ordered Value	D2F	Total Frequency
1	1	45
2	0	95

PROC GENMOD is modeling the probability that D2F='1'.

Parameter Information

Parameter	Effect	MVA34	MVA36	D2SEX	V2P1SEX	D2LIC
Prm1	Intercept					
Prm2	MVA34	1				
Prm3	MVA34	3				
Prm4	MVA34	4				
Prm5	MVA34	6				
Prm6	MVA34	7				
Prm7	MVA34	8				
Prm8	MVA34	10				
Prm9	MVA36		0			
Prm10	MVA36		1			
Prm11	MVA36		2			
Prm12	MVA36		8			
Prm13	D2SEX			F		
Prm14	D2SEX			M		

The SAS System 00:43 Thursday, August 28, 2003 70

The GENMOD Procedure

Parameter Information

Page 1

firstbullethuman

Parameter	Effect	MVA34	MVA36	D2SEX	V2P1SEX	D2LIC
Prm15	D2AGE					
Prm16	V2P1AGE					
Prm17	V2OCCUP					
Prm18	V2P1SEX				F	
Prm19	V2P1SEX				M	
Prm20	D2LIC					G
Prm21	D2LIC					G1
Prm22	D2LIC					G2
Prm23	D2LIC					GM_any
Prm24	D2LIC					other

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	123	117.1191	0.9522
Scaled Deviance	123	117.1191	0.9522
Pearson Chi-Square	123	133.6206	1.0863
Scaled Pearson X2	123	133.6206	1.0863
Log Likelihood		-58.5596	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept		1	27.4704	2.0335	23.4848	31.4559	182.49	<.0001
MVA34	1	1	-26.6916	0.9840	-28.6201	-24.7630	735.86	<.0001
MVA34	3	1	-48.5448	135477.4	-265579	265482.3	0.00	0.9997
MVA34	4	0	0.0000	0.0000	0.0000	0.0000	.	.
MVA34	6	1	-23.5108	1.1650	-25.7941	-21.2275	407.29	<.0001
MVA34	7	1	-24.4278	0.9731	-26.3349	-22.5206	630.22	<.0001
MVA34	8	0	-24.0953	0.0000	-24.0953	-24.0953	.	.
MVA34	10	0	0.0000	0.0000	0.0000	0.0000	.	.
MVA36	0	0	0.0000	0.0000	0.0000	0.0000	.	.
MVA36	1	1	-3.0928	1.2471	-5.5371	-0.6485	6.15	0.0131
MVA36	2	1	-3.6034	1.9336	-7.3931	0.1864	3.47	0.0624
MVA36	8	0	0.0000	0.0000	0.0000	0.0000		.
D2SEX	F	1	-0.0239	0.5432	-1.0885	1.0407	0.00	0.9649
D2SEX	M	0	0.0000	0.0000	0.0000	0.0000	.	.

Page 2

firstbullethuman

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
D2AGE		1	0 0333	0.0199	-0 0057	0 0722	2 80	0.0945

28, 2003 71 The SAS System 00:43 Thursday, August

The GENMOD Procedure

Analysis Of Parameter Estimates

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
V2P1AGE		1	-0.0358	0.0172	-0.0694	-0.0021	4.34	0.0371
V2OCCUP		1	-0.0289	0.3186	-0.6533	0.5956	0.01	0.9279
V2P1SEX	F	1	0.3389	0.5206	-0.6814	1.3591	0.42	0.5151
V2P1SEX	M	0	0.0000	0.0000	0.0000	0.0000	.	.
D2LIC	G	1	0 0699	0.9447	-1.7817	1.9216	0.01	0.9410
D2LIC	G1	1	-0.3920	1.6828	-3.6901	2.9062	0.05	0.8158
D2LIC	G2	1	-0.9131	1.1402	-3.1479	1.3216	0.64	0.4232
D2LIC	GM_any	1	2.4514	1.8646	-1.2031	6.1059	1.73	0.1886
D2LIC	other	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale		0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

LR Statistics For Type 3 Analysis

Source	DF	Chi-Square	Pr > ChiSq
MVA34	5	30.93	<.0001
MVA36	2	9.69	0.0079
D2SEX	1	0.00	0.9649
D2AGE	1	2.84	0.0918
V2P1AGE	1	4.52	0.0334
V2OCCUP	1	0.01	0.9278
V2P1SEX	1	0.43	0.5115
D2LIC	4	4.12	0.3894

Non-Estimable Rows

Contrast	Row
4 vs 1	1

Non-Estimable Rows

firstbullethuman

Contrast	Row
0 vs 1	1

28, 2003 72 The SAS System 00:43 Thursday, August

The GENMOD Procedure

Non-Estimable Rows

Contrast	Row
8 vs 1	1

Contrast Results

Contrast	DF	Chi-Square	Pr > ChiSq	Type
3 vs 1	1	0.31	0.5782	LR
4 vs 1	0	.	.	LR
6 vs 1	1	16.65	<.0001	LR
7 vs 1	1	.	.	LR
8 vs 1	1	7.21	0.0072	LR
10 vs 1	1	3.42	0.0643	LR
0 vs 1	0	.	.	LR
2 vs 1	1	0.11	0.7411	LR
6 vs 1	1	9.27	0.0023	LR
8 vs 1	0	.	.	LR

02finalbullethuman.txt
The SAS System

The GENMOD Procedure

Model Information

Data Set	WORK.TANDBT
Distribution	Binomial
Link Function	Logit
Dependent Variable	D2F
Observations Used	147
Missing Values	617

Class Level Information

Class	Levels	Values
MVA34	7	1 3 4 6 7 8 10

Response Profile

Ordered Value	D2F	Total Frequency
1	1	48
2	0	99

PROC GENMOD is modeling the probability that D2F='1'.

Parameter Information

Parameter	Effect	MVA34
Prm1	Intercept	
Prm2	MVA34	1
Prm3	MVA34	3
Prm4	MVA34	4
Prm5	MVA34	6
Prm6	MVA34	7
Prm7	MVA34	8
Prm8	MVA34	10
Prm9	D2AGE	
Prm10	V2P1AGE	

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	139	139.5178	1.0037
Scaled Deviance	139	139.5178	1.0037

The SAS System 2003

The GENMOD Procedure

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF

Page 1

02finalbullethuman.txt

Pearson Chi-Square	139	146.0317	1.0506
Scaled Pearson X2	139	146.0317	1.0506
Log Likelihood		-69.7589	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept		1	23.9308	0.9189	22.1298	25.7317	678.25	<.0001
MVA34	1	1	-26.0695	0.8440	-27.7237	-24.4153	954.11	<.0001
MVA34	3	1	-48.3788	112260.3	-220075	219977.8	0.00	0.9997
MVA34	4	0	0.0000	0.0000	0.0000	0.0000	.	
MVA34	6	1	-23.1702	1.0509	-25.2300	-21.1104	486.09	<.0001
MVA34	7	1	-23.5979	0.8692	-25.3015	-21.8942	737.00	<.0001
MVA34	8	0	-24.1026	0.0000	-24.1026	-24.1026	.	
MVA34	10	0	0.0000	0.0000	0.0000	0.0000	.	
D2AGE		1	0.0308	0.0175	-0.0035	0.0650	3.10	0.0783
V2P1AGE		1	-0.0275	0.0162	-0.0593	0.0044	2.86	0.0909
Scale		0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

LR Statistics For Type 3 Analysis

Source	DF	Chi-Square	Pr > ChiSq
MVA34	5	44.46	<.0001
D2AGE	1	3.18	0.0745
V2P1AGE	1	2.93	0.0870

Contrast Estimate Results

Label	Estimate	Standard Error	Alpha	Confidence Limits		Chi-Square	Pr > ChiSq
3 vs 1	-22.3093	112260.3	0.05	-220048	220003.9	0.00	0.9998
4 vs 1	Non-est	.	.	.	.	.	
6 vs 1	2.8993	0.7457	0.05	1.4377	4.3609	15.12	0 0001

The SAS System 13:29 Tuesday, September
Page 2

The GENMOD Procedure

Contrast Estimate Results

Label	Estimate	Standard Error	Alpha	Confidence Limits		Chi-Square	Pr > ChiSq
7 vs 1	2.4717	0.4885	0.05	1.5143	3.4291	25.60	<.0001
8 vs 1	1.9669	0.8440	0.05	0.3128	3.6211	5.43	0.0198
10 vs 1	26.0695	0.8440	0.05	24.4153	27.7237	954.11	<.0001

Non-Estimable Rows

Contrast	Row
4 vs 1	1

Contrast Results

Contrast	DF	Chi-Square	Pr > ChiSq	Type
3 vs 1	1	0.70	0.4012	LR
4 vs 1	0	.	.	LR
6 vs 1	1	.	.	LR
7 vs 1	1	.	.	LR
8 vs 1	1	5.31	0.0212	LR
10 vs 1	1	3.58	0.0584	LR

firstbulletxhuman

First Model for Bullet Extra-Human Factors

The SAS System 00:43 Thursday, August 28, 2003 76

The GENMOD Procedure

Model Information

Data Set	WORK.TANDBT
Distribution	Binomial
Link Function	Logit
Dependent Variable	D2F
Observations Used	253
Missing Values	511

Class Level Information

Class	Levels	Values
V2BODY	4	2DR 3DR 4DR LTTRUCK
V2COL	9	BLACK BLUE BROWN GOLD GREEN MAROON RED SILVER WHITE
MVA10	5	1 2 3 4 5
MVA14	3	1 2 3
MVA16	6	1 2 3 4 6 8
MVA18	4	1 2 3 4
MVA20	4	1 2 3 4
MVA32	2	1 2
MVA47	4	1 2 4 7

Response Profile

Ordered Value	D2F	Total Frequency
1	1	103
2	0	150

PROC GENMOD is modeling the probability that D2F='1'.

Parameter Information

Parameter	Effect	V2BODY	V2COL	MVA10	MVA14	MVA16	MVA18	MVA20	MVA32	MVA47
Prm1	Intercept									
Prm2	R2LANES									
Prm3	R2MAX									
Prm4	V2APSPD									
Prm5	V2BODY	2DR								
Prm6	V2BODY	3DR								
Prm7	V2BODY	4DR								
Prm8	V2BODY	LTTRUCK								
Prm9	V2YEAR									

The SAS System 00:43 Thursday, August 28, 2003 77

The GENMOD Procedure

firstbulletxhuman

Parameter Information

Parameter	Effect	V2BODY	V2COL	MVA10	MVA14	MVA16	MVA18	MVA20	MVA32	MVA47
Prm10	V2COL		BLACK							
Prm11	V2COL		BLUE							
Prm12	V2COL		BROWN							
Prm13	V2COL		GOLD							
Prm14	V2COL		GREEN							
Prm15	V2COL		MAROON							
Prm16	V2COL		RED							
Prm17	V2COL		SILVER							
Prm18	V2COL		WHITE							
Prm19	MVA10			1						
Prm20	MVA10			2						
Prm21	MVA10			3						
Prm22	MVA10			4						
Prm23	MVA10			5						
Prm24	MVA14				1					
Prm25	MVA14				2					
Prm26	MVA14				3					
Prm27	MVA16					1				
Prm28	MVA16					2				
Prm29	MVA16					3				
Prm30	MVA16					4				
Prm31	MVA16					6				
Prm32	MVA16					8				
Prm33	MVA18						1			
Prm34	MVA18						2			
Prm35	MVA18						3			
Prm36	MVA18						4			
Prm37	MVA20							1		
Prm38	MVA20							2		
Prm39	MVA20							3		
Prm40	MVA20							4		
Prm41	MVA32								1	
Prm42	MVA32								2	
Prm43	MVA47									1
Prm44	MVA47									2
Prm45	MVA47									4
Prm46	MVA47									7

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	219	244.6514	1.1171
Scaled Deviance	219	244.6514	1.1171

The SAS System 00:43 Thursday, August 28, 2003 78

The GENMOD Procedure

Criteria For Assessing Goodness Of Fit

firstbulletxhuman

Criterion	DF	Value	Value/DF
Pearson Chi-Square	219	220.2222	1.0056
Scaled Pearson X2	219	220.2222	1.0056
Log Likelihood		-122.3257	

WARNING: Negative of Hessian not positive definite.

Analysis Of Parameter Estimates

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept		1	-98.5949	78.0822	-251.633	54.4434	1.59	0.2067
R2LANES		1	0.3864	0.1406	0.1107	0.6620	7.55	0.0060
R2MAX		1	0.0287	0.0288	-0.0278	0.0852	0.99	0.3190
V2APSPD		1	-0.0278	0.0111	-0.0496	-0.0060	6.24	0.0125
V2BODY	2DR	1	-0.0607	0.5333	-1.1059	0.9845	0.01	0.9094
V2BODY	3DR	1	24.7134	118499.2	-232229	232278.8	0.00	0.9998
V2BODY	4DR	1	-0.2040	0.4585	-1.1027	0.6946	0.20	0.6563
V2BODY	LTTRUCK	0	0.0000	0.0000	0.0000	0.0000	.	
V2YEAR		1	0.0588	0.0393	-0.0182	0.1358	2.24	0.1344
V2COL	BLACK	1	-1.1579	0.7052	-2.5401	0.2243	2.70	0.1006
V2COL	BLUE	1	-0.5039	0.5871	-1.6546	0.6468	0.74	0.3907
V2COL	BROWN	1	-1.0895	0.7489	-2.5573	0.3784	2.12	0.1457
V2COL	GOLD	1	-0.2959	0.9902	-2.2365	1.6448	0.09	0.7651
V2COL	GREEN	1	-1.8870	0.7613	-3.3792	-0.3948	6.14	0.0132
V2COL	MAROON	1	-0.7911	1.9344	-4.5825	3.0003	0.17	0.6826
V2COL	RED	1	-0.2289	0.5604	-1.3272	0.8694	0.17	0.6829
V2COL	SILVER	1	-0.4911	0.5302	-1.5304	0.5481	0.86	0.3543
V2COL	WHITE	0	0.0000	0.0000	0.0000	0.0000	.	
MVA10	1	1	-67.6158	51471.18	-100949	100814.0	0.00	0.9990
MVA10	2	1	-44.6509	0.5992	-45.8252	-43.4766	5553.65	<.0001
MVA10	3	1	-46.0630	1.3352	-48.6800	-43.4461	1190.14	<.0001
MVA10	4	0	-43.8494	0.0000	-43.8494	-43.8494	.	
MVA10	5	0	0.0000	0.0000	0.0000	0.0000	.	
MVA14	1	1	1.5099	1.7683	-1.9559	4.9757	0.73	0.3932

Page 3

firstbulletxhuman

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
MVA14	2	1	49.1443	118499.2	-232205	232303.3	0.00	0.9997
MVA14	3	0	0.0000	0.0000	0.0000	0.0000	.	
MVA16	1	1	1.3372	0.4003	0.5526	2.1218	11.16	0.0008
MVA16	2	0	1.1184	0.0000	1.1184	1.1184	.	
MVA16	3	1	24.4387	80480.40	-157714	157763.1	0.00	0.9998
MVA16	4	0	0.0000	0.0000	0.0000	0.0000	.	
MVA16	6	1	-20.4213	167583.2	-328477	328436.5	0.00	0.9999
MVA16	8	0	0.0000	0.0000	0.0000	0.0000	.	
MVA18	1	1	22.7021	0.5321	21.6593	23.7450	1820.51	<.0001

The SAS System 00:43 Thursday, August 28, 2003 79

The GENMOD Procedure

Analysis Of Parameter Estimates

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
MVA18	2	0	22.5095	0.0000	22.5095	22.5095	.	
MVA18	3	1	1.6943	83809.78	-164262	164265.8	0.00	1.0000
MVA18	4	0	0.0000	0.0000	0.0000	0.0000	.	
MVA20	1	0	22.4731	0.0000	22.4731	22.4731	.	
MVA20	2	1	-0.5284	55173.53	-108139	108137.6	0.00	1.0000
MVA20	3	0	0.0000	0.0000	0.0000	0.0000	.	
MVA20	4	0	0.0000	0.0000	0.0000	0.0000	.	
MVA32	1	0	0.0000	0.0000	0.0000	0.0000	.	
MVA32	2	0	0.0000	0.0000	0.0000	0.0000	.	
MVA47	1	1	-23.8726	0.4621	-24.7784	-22.9669	2668.42	<.0001
MVA47	2	1	-24.3179	129195.0	-253242	253193.2	0.00	0.9998
MVA47	4	0	-22.3587	0.0000	-22.3587	-22.3587	.	
MVA47	7	0	0.0000	0.0000	0.0000	0.0000	.	
Scale		0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

Page 4

finalbulletxhuman
Final (10th) Model for Bullet Extra-Human Factors
The SAS System 00:43 Thursday, August 28, 2003 83

The GENMOD Procedure

Model Information

Data Set	WORK.TANDBT
Distribution	Binomial
Link Function	Logit
Dependent Variable	D2F
Observations Used	454
Missing Values	310

Class Level Information

Class	Levels	Values
MVA10	5	1 2 3 4 5

Response Profile

Ordered Value	D2F	Total Frequency
1	1	141
2	0	313

PROC GENMOD is modeling the probability that D2F='1'.

Parameter Information

Parameter	Effect	MVA10
Prm1	Intercept	
Prm2	R2MAX	
Prm3	V2APSPD	
Prm4	MVA10	1
Prm5	MVA10	2
Prm6	MVA10	3
Prm7	MVA10	4
Prm8	MVA10	5

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	447	507.3270	1.1350
Scaled Deviance	447	507.3270	1.1350
Pearson Chi-Square	447	464.5806	1.0393
Scaled Pearson X2	447	464.5806	1.0393

The SAS System 00:43 Thursday, August 28, 2003 84

The GENMOD Procedure

Criteria For Assessing Goodness Of Fit
Page 1

finalbulletxhuman

Criterion	DF	Value	Value/DF
Log Likelihood		-253.6635	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept		1	19.6613	1.0675	17.5691	21.7535	339.24	<.0001
R2MAX		1	0.0499	0.0174	0.0158	0.0841	8.21	0.0042
V2APSPD		1	-0.0396	0.0064	-0.0521	-0.0272	38.92	<.0001
MVA10	1	1	-22.2518	0.8960	-24.0080	-20.4956	616.73	<.0001
MVA10	2	1	-21.4829	0.3925	-22.2522	-20.7136	2995.74	<.0001
MVA10	3	1	-22.7452	0.7429	-24.2013	-21.2891	937.35	<.0001
MVA10	4	0	-20.9672	0.0000	-20.9672	-20.9672	.	
MVA10	5	0	0.0000	0.0000	0.0000	0.0000	.	
Scale		0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

LR Statistics For Type 3 Analysis

Source	DF	Chi-Square	Pr > ChiSq
R2MAX	1	8.65	0.0033
V2APSPD	1	42.38	<.0001
MVA10	4	9.33	0.0534

Contrast Estimate Results

Label	Estimate	Standard Error	Alpha	Confidence Limits		Chi-Square	Pr > ChiSq
2 vs 1	0.7689	0.8204	0.05	-0.8390	2.3769	0.88	0.3486
3 vs 1	-0.4934	1.0349	0.05	-2.5218	1.5350	0.23	0.6336
4 vs 1	1.2846	0.8960	0.05	-0.4716	3.0408	2.06	0.1517
5 vs 1	22.2518	0.8960	0.05	20.4956	24.0080	616.73	<.0001

The SAS System 00:43 Thursday, August 28, 2003 85

Page 2

finalbulletXhuman

The GENMOD Procedure

Contrast Results

Contrast	DF	Chi-Square	Pr > ChiSq	Type
2 vs 1	1	1.00	0.3175	LR
3 vs 1	1	0.22	0.6376	LR
4 vs 1	1	2.36	0.1248	LR
5 vs 1	1	2.18	0.1395	LR

Page 3

firsttargethuman

First Model For Target Human Factors

The SAS System 00:43 Thursday, August 28, 2003 86

The GENMOD Procedure

Model Information

Data Set	WORK.TANDBT
Distribution	Binomial
Link Function	Logit
Dependent Variable	D1F
Observations Used	161
Missing Values	603

Class Level Information

Class	Levels	Values
MVA33	8	1 3 4 5 6 7 8 12
MVA35	5	1 2 3 6 8
D1SEX	2	F M
V1P1SEX	2	F M
D1LIC	5	G G1 G2 GM_any other

Response Profile

Ordered Value	D1F	Total Frequency
1	1	109
2	0	52

PROC GENMOD is modeling the probability that D1F='1'.

Parameter Information

Parameter	Effect	MVA33	MVA35	D1SEX	V1P1SEX	D1LIC
Prm1	Intercept					
Prm2	MVA33	1				
Prm3	MVA33	3				
Prm4	MVA33	4				
Prm5	MVA33	5				
Prm6	MVA33	6				
Prm7	MVA33	7				
Prm8	MVA33	8				
Prm9	MVA33	12				
Prm10	MVA35		1			
Prm11	MVA35		2			
Prm12	MVA35		3			
Prm13	MVA35		6			
Prm14	MVA35		8			

The SAS System 00:43 Thursday, August 28, 2003 87

The GENMOD Procedure

Parameter Information

Page 1

firsttargethuman

Parameter	Effect	MVA33	MVA35	D1SEX	V1P1SEX	D1LIC
Prm15	D1SEX			F		
Prm16	D1SEX			M		
Prm17	D1AGE					
Prm18	V1P1AGE					
Prm19	V1OCCUP					
Prm20	V1P1SEX				F	
Prm21	V1P1SEX				M	
Prm22	D1LIC					G
Prm23	D1LIC					G1
Prm24	D1LIC					G2
Prm25	D1LIC					GM_any
Prm26	D1LIC					other

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	141	117.5959	0.8340
Scaled Deviance	141	117.5959	0.8340
Pearson Chi-Square	141	165.3981	1.1730
Scaled Pearson X2	141	165.3981	1.1730
Log Likelihood		-58.7979	

WARNING: Negative of Hessian not positive definite.

Analysis Of Parameter Estimates

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept		1	25.5076	1.8979	21.7878	29.2274	180.63	<.0001
MVA33	1	1	-27.1247	0.6543	-28.4072	-25.8422	1718.41	<.0001
MVA33	3	1	-0.2003	322113.8	-631332	631331.2	0.00	1.0000
MVA33	4	1	-51.1214	322114.2	-631383	631281.1	0.00	0.9999
MVA33	5	0	0.0000	0.0000	0.0000	0.0000	.	.
MVA33	6	1	-23.5046	0.7912	-25.0553	-21.9539	882.52	<.0001
MVA33	7	1	-25.5241	0.9682	-27.4218	-23.6265	694.96	<.0001
MVA33	8	0	-23.4121	0.0000	-23.4121	-23.4121	.	.
MVA33	12	0	0.0000	0.0000	0.0000	0.0000	.	.
MVA35	1	1	-0.2042	0.8853	-1.9392	1.5309	0.05	0.8176
MVA35	2	1	23.0464	227540.7	-445948	445994.5	0.00	0.9999
MVA35	3	1	23.5530	322113.8	-631308	631355.0	0.00	0.9999

Page 2

firsttargethuman
The SAS System 00:43 Thursday, August 28, 2003 88

The GENMOD Procedure

Analysis Of Parameter Estimates

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
MVA35	6	1	26.1228	322113.6	-631305	631357.6	0.00	0.9999
MVA35	8	0	0.0000	0.0000	0.0000	0.0000	.	.
D1SEX	F	1	-0.0163	0.5048	-1.0058	0.9731	0.00	0.9742
D1SEX	M	0	0.0000	0.0000	0.0000	0.0000	.	.
D1AGE		1	-0.0004	0.0196	-0.0389	0.0381	0.00	0.9846
V1P1AGE		1	-0.0124	0.0148	-0.0414	0.0166	0.70	0.4032
V1OCCUP		1	0.5900	0.3489	-0.0940	1.2739	2.86	0.0909
V1P1SEX	F	1	-0.1956	0.5320	-1.2384	0.8472	0.14	0.7131
V1P1SEX	M	0	0.0000	0.0000	0.0000	0.0000	.	
D1LIC	G	1	-0.3954	1.3181	-2.9788	2.1881	0.09	0.7642
D1LIC	G1	1	0.4324	1.7652	-3.0272	3.8921	0.06	0.8065
D1LIC	G2	1	0.1620	1.4209	-2.6228	2.9468	0.01	0.9092
D1LIC	GM_any	1	-1.2450	2.1016	-5.3640	2.8739	0.35	0.5536
D1LIC	other	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale		0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

Page 3

finaltargethuman

Final (6th) Model for Target Human Factors

The SAS System 00:43 Thursday, August 28, 2003 89

The GENMOD Procedure

Model Information

Data Set	WORK.TANDBT
Distribution	Binomial
Link Function	Logit
Dependent Variable	DIF
Observations Used	168
Missing Values	596

Class Level Information

Class	Levels	Values
MVA33	8	1 3 4 5 6 7 8 12

Response Profile

Ordered Value	DIF	Total Frequency
1	1	113
2	0	55

PROC GENMOD is modeling the probability that DIF='1'.

Parameter Information

Parameter	Effect	MVA33
Prm1	Intercept	
Prm2	MVA33	1
Prm3	MVA33	3
Prm4	MVA33	4
Prm5	MVA33	5
Prm6	MVA33	6
Prm7	MVA33	7
Prm8	MVA33	8
Prm9	MVA33	12
Prm10	V1P1AGE	
Prm11	V1OCCUP	

The SAS System 00:43 Thursday, August 28, 2003 90

The GENMOD Procedure

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	159	123.9885	0.7798
Scaled Deviance	159	123.9885	0.7798

finaltargethuman

Criterion	DF	Value	Value/DF
Pearson Chi-Square	159	164.1541	1.0324
Scaled Pearson X2	159	164.1541	1.0324
Log Likelihood		-61.9942	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept		1	23.6231	0.9944	21.6741	25.5722	564.31	<.0001
MVA33	1	1	-26.1462	0.6012	-27.3246	-24.9679	1891.32	<.0001
MVA33	3	1	-0.0975	138098.3	-270668	270667.6	0.00	1.0000
MVA33	4	1	-48.7304	195372.1	-382971	382873.6	0.00	0.9998
MVA33	5	0	0.0000	0.0000	0.0000	0.0000	.	.
MVA33	6	1	-22.3754	0.7712	-23.8870	-20.8639	841.78	<.0001
MVA33	7	1	-24.2759	0.8869	-26.0141	-22.5377	749.28	<.0001
MVA33	8	0	-22.2421	0.0000	-22.2421	-22.2421	.	.
MVA33	12	0	0.0000	0.0000	0.0000	0.0000	.	.
V1P1AGE		1	-0.0105	0.0119	-0.0338	0.0129	0.77	0.3808
V1OCCUP		1	0.6690	0.3408	0.0011	1.3370	3.85	0.0496
Scale		0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

LR Statistics For Type 3 Analysis

Source	DF	Chi-Square	Pr > ChiSq
MVA33	6	87.67	<.0001
V1P1AGE	1	0.77	0.3792
V1OCCUP	1	4.25	0.0394

The SAS System 00:43 Thursday, August 28, 2003 91

The GENMOD Procedure

Contrast Estimate Results

Label	Estimate	Standard Error	Alpha	Confidence Limits	Chi-Square	Pr > ChiSq

finaltargethuman

3 vs 1	26.0487	138098.3	0.05	-270642	270693.7	0.00	0.9998
4 vs 1	-22.5842	195372.1	0.05	-382945	382899.8	0.00	0.9999
5 vs 1	Non-est	.		.	.	.	.
6 vs 1	3.7708	0.7177	0.05	2.3641	5.1775	27.60	<.0001
7 vs 1	1.8703	0.8289	0.05	0.2458	3.4948	5.09	0.0240
8 vs 1	3.9041	0.6012	0.05	2.7257	5.0824	42.17	<.0001
12 vs 1	26.1462	0.6012	0.05	24.9679	27.3246	1891.3	<.0001

Non-Estimable Rows

Contrast	Row
5 vs 1	1

Contrast Results

Contrast	DF	Chi-Square	Pr > ChiSq	Type
3 vs 1	1	6.93	0.0085	LR
4 vs 1	1	0.31	0.5799	LR
5 vs 1	0	.	.	LR
6 vs 1	1	.	.	LR
7 vs 1	1	5.34	0.0209	LR
8 vs 1	1	.	.	LR
12 vs 1	1	3.71	0.0542	LR

Page 3

firsttargetXhuman
First Model for Target Extra-Human Factors
The SAS System 00:43 Thursday, August 28, 2003 92

The GENMOD Procedure

Model Information

Data Set	WORK.TANDBT
Distribution	Binomial
Link Function	Logit
Dependent Variable	DIF
Observations Used	241
Missing Values	523

Class Level Information

Class	Levels	Values
V1BODY	4	2DR 3DR 4DR LTTRUCK
V1COL	9	BLACK BLUE BROWN GOLD GREEN MAROON RED SILVER WHITE
MVA9	5	1 2 3 4 5
MVA11	2	1 2
MVA13	3	1 2 3
MVA15	5	1 2 3 4 5
MVA17	4	1 2 3 4
MVA19	5	1 2 3 4 11
MVA31	1	1
MVA46	4	1 3 4 10

Response Profile

Ordered Value	DIF	Total Frequency
1	1	143
2	0	98

PROC GENMOD is modeling the probability that DIF='1'.

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	206	216.7952	1.0524
Scaled Deviance	206	216.7952	1.0524
Pearson Chi-Square	206	223.4408	1.0847
Scaled Pearson X2	206	223.4408	1.0847
Log Likelihood		-108.3976	

The SAS System 00:43 Thursday, August 28, 2003 93

The GENMOD Procedure

WARNING: Negative of Hessian not positive definite.

Page 1

firsttargetXhuman
Analysis Of Parameter Estimates

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept		1	107.9037	89.3754	-67.2688	283.0761	1.46	0.2273
R1LANES		1	0.0765	0.1441	-0.2060	0.3589	0.28	0.5958
R1MAX		1	0.0722	0.0307	0.0119	0.1324	5.51	0.0189
V1APSPD		1	-0.0151	0.0119	-0.0384	0.0082	1.61	0.2051
V1BODY	2DR	1	0.5665	0.6752	-0.7568	1.8898	0.70	0.4015
V1BODY	3DR	1	26.3238	195372.1	-382896	382948.5	0.00	0.9999
V1BODY	4DR	1	0.7031	0.5738	-0.4216	1.8278	1.50	0.2205
V1BODY	LTTRUCK	0	0.0000	0.0000	0.0000	0.0000	.	.
V1YEAR		1	-0 0596	0.0449	-0.1476	0.0283	1.76	0.1840
V1COL	BLACK	1	1 3948	0.7230	-0.0223	2.8119	3.72	0.0537
V1COL	BLUE	1	1.6074	0.6473	0.3387	2.8760	6.17	0.0130
V1COL	BROWN	1	1 6168	0.8286	-0.0073	3.2409	3.81	0.0510
V1COL	GOLD	1	24.4189	195372.1	-382898	382946.6	0.00	0.9999
V1COL	GREEN	1	2 1526	0.7786	0.6265	3.6786	7.64	0.0057
V1COL	MAROON	1	1.6403	1.5328	-1.3640	4.6445	1.15	0.2846
V1COL	RED	1	1.5219	0.6299	0.2873	2.7565	5.84	0.0157
V1COL	SILVER	1	1.3919	0.6280	0.1611	2.6227	4.91	0.0267
V1COL	WHITE	0	0.0000	0.0000	0.0000	0.0000	.	.
MVA9	1	1	26.9465	1.1970	24.6004	29.2927	506.75	<.0001
MVA9	2	1	27.0284	0.6514	25.7517	28.3051	1721.61	<.0001
MVA9	3	1	27.0125	1.0396	24.9750	29.0500	675.19	<.0001
MVA9	4	0	25.4312	0.0000	25.4312	25.4312	.	.
MVA9	5	0	0.0000	0.0000	0.0000	0.0000	.	.
MVA11	1	1	-0.1063	1.9720	-3.9712	3.7587	0.00	0.9570
MVA11	2	0	0.0000	0.0000	0.0000	0.0000	.	.
MVA13	1	0	-22.1883	0.0000	-22.1883	-22.1883	.	.
MVA13	2	1	2.1069	195372.1	-382920	382924.3	0.00	1.0000
MVA13	3	0	0.0000	0.0000	0.0000	0.0000	.	.
MVA15	1	1	2.2637	0.4621	1.3581	3.1694	24.00	

Page 2

<.0001

firsttargetxhuman

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
MVA15	2	0	2.7400	0.0000	2.7400	2.7400	.	.
MVA15	3	1	-20.1695	195372.1	-382943	382902.2	0.00	0.9999
MVA15	4	0	0.0000	0.0000	0.0000	0.0000	.	.
MVA15	5	0	0.0000	0.0000	0.0000	0.0000	.	.
MVA17	1	1	-0.4618	1.6011	-3.5998	2.6762	0.08	0.7730
MVA17	2	1	0.9337	1.6969	-2.3922	4.2596	0.30	0.5822
MVA17	3	1	-1.1782	2.5982	-6.2705	3.9141	0.21	0.6502
MVA17	4	0	0.0000	0.0000	0.0000	0.0000	.	.
MVA19	1	1	-25.2708	1.8503	-28.8973	-21.6442	186.53	<.0001
MVA19	2	0	-21.8874	0.0000	-21.8874	-21.8874	.	.
MVA19	3	0	0.0000	0.0000	0.0000	0.0000	.	.
MVA19	4	1	-49.5129	195372.1	-382972	382872.9	0.00	0.9998
MVA19	11	0	0.0000	0.0000	0.0000	0.0000	.	.

28, 2003 94 The SAS System 00:43 Thursday, August

The GENMOD Procedure

Analysis Of Parameter Estimates

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
MVA31	1	0	0.0000	0.0000	0.0000	0.0000	.	.
MVA46	1	1	21.9906	0.5201	20.9712	23.0100	1787.52	<.0001
MVA46	3	1	-1.5327	195372.1	-382924	382920.8	0.00	1.0000
MVA46	4	0	24.6542	0.0000	24.6542	24.6542	.	.
MVA46	10	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale		0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

Page 3

finaltarget\human
Final (10th) Model for Target Extra-Human Factors
The SAS System 00:43 Thursday, August 28, 2003 95

The GENMOD Procedure

Model Information

Data Set	WORK.TANDBT
Distribution	Binomial
Link Function	Logit
Dependent Variable	D1F
Observations Used	448
Missing Values	316

Class Level Information

Class	Levels	Values
MVA9	5	1 2 3 4 5

Response Profile

Ordered Value	D1F	Total Frequency
1	1	313
2	0	135

PROC GENMOD is modeling the probability that D1F='1'.

Parameter Information

Parameter	Effect	MVA9
Prm1	Intercept	
Prm2	R1LANES	
Prm3	V1APSPD	
Prm4	MVA9	1
Prm5	MVA9	2
Prm6	MVA9	3
Prm7	MVA9	4
Prm8	MVA9	5

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	441	498.2350	1.1298
Scaled Deviance	441	498.2350	1.1298
Pearson Chi-Square	441	460.0866	1.0433
Scaled Pearson X2	441	460.0866	1.0433

The SAS System 00:43 Thursday, August 28, 2003 96

The GENMOD Procedure

Criteria For Assessing Goodness Of Fit
Page 1

finaltarget\human

Criterion	DF	Value	Value/DF
Log Likelihood		-249.1175	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept		1	-1.9025	1.1865	-4.2280	0.4231	2.57	0.1088
R1LANES		1	0.2570	0.0815	0.0973	0.4168	9.95	0.0016
V1APSPD		1	-0.0375	0.0065	-0.0501	-0.0248	33.52	<.0001
MVA9	1	1	2.5929	1.2748	0.0942	5.0915	4.14	0.0420
MVA9	2	1	2.6697	1.1468	0.4221	4.9173	5.42	0.0199
MVA9	3	1	2.7476	1.2675	0.2634	5.2319	4.70	0.0302
MVA9	4	1	2.1278	1.1954	-0.2152	4.4707	3.17	0.0751
MVA9	5	0	0.0000	0.0000	0.0000	0.0000	.	
Scale		0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

LR Statistics For Type 3 Analysis

Source	DF	Chi-Square	Pr > ChiSq
R1LANES	1	10.40	0.0013
V1APSPD	1	35.38	<.0001
MVA9	4	8.90	0.0637

Contrast Estimate Results

Label	Estimate	Standard Error	Alpha	Confidence Limits		Chi-Square	Pr > ChiSq
2 vs 1	0.0768	0.5822	0.05	-1.0642	1.2178	0.02	0.8950
3 vs 1	0.1548	0.7949	0.05	-1.4032	1.7127	0.04	0.8456
4 vs 1	-0.4651	0 6781	0.05	-1.7942	0.8640	0.47	0.4928
5 vs 1	-2.5929	1.2748	0.05	-5.0915	-0.0942	4.14	0 0420

The SAS System 00:43 Thursday, August 28, 2003 97

Page 2

finaltargetXhuman

The GENMOD Procedure

Contrast Results

Contrast	DF	Chi-Square	Pr > ChiSq	Type
2 vs 1	1	0.02	0.8954	LR
3 vs 1	1	0.04	0.8456	LR
4 vs 1	1	0.48	0.4884	LR
5 vs 1	1	5.15	0.0233	LR

Page 3

firstintersection
First Model for Intersection Factors

The SAS System 00:43 Thursday, August 28, 2003 98

The GENMOD Procedure

Model Information

Data Set	WORK.INTERSCTN
Distribution	Binomial
Link Function	Logit
Dependent Variable	Inj_fat_PDO
Observations Used	1043
Missing Values	792

Class Level Information

Class	Levels	Values
System	3	MTSS SCOOT other
FAGINT	2	no yes
FDW	2	no yes
Num_Legs	3	3 4 99
Staggered	2	no yes
City	6	ET EY NY SC TO YO
Any_LA	2	no yes
N	5	COLLCTR LOCAL MAJART MINART OTHER
S	5	COLLCTR LOCAL MAJART MINART OTHER
W	5	COLLCTR LOCAL MAJART MINART OTHER
E	5	COLLCTR LOCAL MAJART MINART OTHER

Response Profile

Ordered Value	Inj_fat_PDO	Total Frequency
1	1	882
2	0	161

PROC GENMOD is modeling the probability that Inj_fat_PDO='1'.

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	1010	654.7390	0.6483
Scaled Deviance	1010	654.7390	0.6483
Pearson Chi-Square	1010	1072.4396	1.0618
Scaled Pearson X2	1010	1072.4396	1.0618
Log Likelihood		-327.3695	

The SAS System 00:43 Thursday, August 28, 2003 99

The GENMOD Procedure

Algorithm converged.

Page 1

firstintersection
Analysis Of Parameter Estimates

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept		1	-26.8910	1.8589	-30.5343	-23.2477	209.28	<.0001
System	MTSS	1	-0.1710	0.8123	-1.7632	1.4211	0.04	0.8333
System	SCOOT	1	-0.8612	0.8888	-2.6032	0.8808	0.94	0.3326
System	other	0	0.0000	0.0000	0.0000	0.0000	.	.
FAGINT	no	1	-1.0037	0.3533	-1.6962	-0.3112	8.07	0.0045
FAGINT	yes	0	0.0000	0.0000	0.0000	0.0000	.	.
FDW	no	1	0.4133	0.2184	-0.0147	0.8413	3.58	0.0584
FDW	yes	0	0.0000	0.0000	0.0000	0.0000	.	.
Num_Legs	3	1	19.0796	0.5492	18.0031	20.1561	1206.74	<.0001
Num_Legs	4	0	20.3901	0.0000	20.3901	20.3901	.	.
Num_Legs	99	0	0.0000	0.0000	0.0000	0.0000	.	.
Staggered	no	1	1.2535	0.7079	-0.1338	2.6409	3.14	0.0766
Staggered	yes	0	0.0000	0.0000	0.0000	0.0000	.	.
City	ET	1	-0.6638	0.6144	-1.8680	0.5405	1.17	0.2800
City	EY	1	-0.0193	0.7696	-1.5276	1.4891	0.00	0.9800
City	NY	1	-0.7072	0.6175	-1.9174	0.5030	1.31	0.2521
City	SC	1	-0.8927	0.6114	-2.0909	0.3055	2.13	0.1442
City	TO	1	-1.0416	0.5768	-2.1721	0.0889	3.26	0.0710
City	YO	0	0.0000	0.0000	0.0000	0.0000	.	.
Q_major		1	0.0000	0.0000	-0.0000	0.0001	3.07	0.0797
Q_minor		1	0.0000	0.0000	-0.0000	0.0001	1.36	0.2437
DstInt		1	-0.0000	0.0005	-0.0010	0.0010	0.01	0.9383
Any_LA	no	1	0.3583	0.5578	-0.7350	1.4516	0.41	0.5207
Any_LA	yes	0	0.0000	0.0000	0.0000	0.0000	.	.
N	COLLCTR	1	0.8025	1.0356	-1.2273	2.8323	0.60	0.4384
N	LOCAL	1	0.0382	1.0417	-2.0035	2.0800	0.00	0.9707
N	MAJART	1	1.9468	1.1994	-0.4040	4.2976	2.63	0.1046
N	MINART	1	1.8923	1.1197	-0.3023	4.0869	2.86	0.0910
N	OTHER	0	0.0000	0.0000	0.0000	0.0000	.	.

Page 2

firstintersection

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
S	COLLCTR	1	2.0254	0.8741	0.3122	3.7386	5.37	0.0205
S	LOCAL	1	1.4419	0.8828	-0.2884	3.1721	2.67	0.1024
S	MAJART	1	2.7047	1.0112	0.7227	4.6867	7.15	0.0075
S	MINART	1	1.8895	0.9301	0.0666	3.7124	4.13	0.0422
S	OTHER	0	0.0000	0.0000	0.0000	0.0000	.	.
W	COLLCTR	1	1.2949	0.9218	-0.5119	3.1016	1.97	0.1601
W	LOCAL	1	0.1057	0.9165	-1.6906	1.9020	0.01	0.9082
W	MAJART	1	3.7356	1.3157	1.1569	6.3143	8.06	0.0045
W	MINART	1	2.1122	1.0503	0.0536	4.1709	4.04	0.0443
W	OTHER	0	0.0000	0.0000	0.0000	0.0000		.
E	COLLCTR	1	2.1816	0.9909	0.2395	4.1236	4.85	0.0277
E	LOCAL	1	2.0242	0.9775	0.1083	3.9401	4.29	0.0384
E	MAJART	1	0.9967	1.3369	-1.6235	3.6170	0.56	0.4559

28, 2003 100 The SAS System 00:43 Thursday, August

The GENMOD Procedure

Analysis Of Parameter Estimates

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
E	MINART	1	2.2660	1.1044	0.1014	4.4306	4.21	0.0402
E	OTHER	0	0.0000	0.0000	0.0000	0.0000	.	
Scale		0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

LR Statistics For Type 3 Analysis

Source	DF	Chi-Square	Pr > ChiSq
System	2	3.54	0.1707
FAGINT	1	9.42	0.0021
FDw	1	3.64	0.0565
Num_Legs	2	5.78	0.0557
Staggered	1	2.99	0.0835
City	5	6.97	0.2227
Q_major	1	3.15	0.0759
Q_minor	1	1.39	0.2389
DstInt	1	0.01	0.9382
Any_LA	1	0.40	0.5281

Page 3

firstintersection

N	4	14.78	0.0052
S	4	9.63	0.0472
W	4	19.98	0.0005
E	4	7.31	0.1202

Page 4

finalintersection
Final (8th) Model for Intersection Factors
The SAS System 00:43 Thursday, August 28, 2003 101

The GENMOD Procedure

Model Information

Data Set	WORK.INTERSCTN
Distribution	Binomial
Link Function	Logit
Dependent Variable	Inj_fat_PDO
Observations Used	1142
Missing Values	693

Class Level Information

Class	Levels	Values
System	3	MTSS SCOOT other
FAGINT	2	no yes
Num_Legs	3	3 4 99
N	5	COLLCTR LOCAL MAJART MINART OTHER
S	5	COLLCTR LOCAL MAJART MINART OTHER
W	5	COLLCTR LOCAL MAJART MINART OTHER

Response Profile

Ordered Value	Inj_fat_PDO	Total Frequency
1	1	949
2	0	193

PROC GENMOD is modeling the probability that Inj_fat_PDO='1'.

Parameter Information

Parameter	Effect	System	FAGINT	Num_Legs	N	S	W
Prm1	Intercept						
Prm2	System	MTSS					
Prm3	System	SCOOT					
Prm4	System	other					
Prm5	FAGINT		no				
Prm6	FAGINT		yes				
Prm7	Num_Legs			3			
Prm8	Num_Legs			4			
Prm9	Num_Legs			99			
Prm10	Q_major						
Prm11	N				COLLCTR		
Prm12	N				LOCAL		

The SAS System 00:43 Thursday, August 28, 2003 102

The GENMOD Procedure

Page 1

finalintersection
Parameter Information

Parameter	Effect	System	FAGINT	Num_Legs	N	S	W
Prm13	N				MAJART		
Prm14	N				MINART		
Prm15	N				OTHER		
Prm16	S					COLLCTR	
Prm17	S					LOCAL	
Prm18	S					MAJART	
Prm19	S					MINART	
Prm20	S					OTHER	
Prm21	W						COLLCTR
Prm22	W						LOCAL
Prm23	W						MAJART
Prm24	W						MINART
Prm25	W						OTHER

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	1123	775.5366	0.6906
Scaled Deviance	1123	775.5366	0.6906
Pearson Chi-Square	1123	1254.6514	1.1172
Scaled Pearson X2	1123	1254.6514	1.1172
Log Likelihood		-387.7683	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept		1	-24.7784	1.1300	-26.9931	-22.5637	480.85	<.0001
System	MTSS	1	0.1662	0.6444	-1.0968	1.4292	0.07	0.7965
System	SCOOT	1	-0.6454	0.7099	-2.0369	0.7460	0.83	0.3633
System	other	0	0.0000	0.0000	0.0000	0.0000	.	.
FAGINT	no	1	-1.1240	0.2968	-1.7057	-0.5422	14.34	0.0002
FAGINT	yes	0	0.0000	0.0000	0.0000	0.0000	.	.
Num_Legs	3	1	19.3750	0.3148	18.7580	19.9920	3788.02	<.0001
Num_Legs	4	0	20.2705	0.0000	20.2705	20.2705	.	.
Num_Legs	99	0	0.0000	0.0000	0.0000	0.0000	.	.

Page 2

finalintersection

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
Q_major		1	0.0000	0.0000	0.0000	0.0001	12.06	0.0005
N	COLLCTR	1	0.6024	1.0114	-1.3799	2.5846	0.35	0.5514
N	LOCAL	1	-0.0987	1.0134	-2.0851	1.8876	0.01	0.9224

The SAS System 00:43 Thursday, August 28, 2003 103

The GENMOD Procedure

Analysis Of Parameter Estimates

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square	Pr > ChiSq
N	MAJART	1	1.1713	1.1181	-1.0202	3.3627	1.10	0.2949
N	MINART	1	1.4001	1.0744	-0.7057	3.5059	1.70	0.1925
N	OTHER	0	0.0000	0.0000	0.0000	0.0000	.	.
S	COLLCTR	1	2.0912	0.8569	0.4116	3.7708	5.96	0.0147
S	LOCAL	1	1.5065	0.8643	-0.1874	3.2005	3.04	0.0813
S	MAJART	1	2.8501	0.9745	0.9402	4.7600	8.55	0.0034
S	MINART	1	2.0228	0.9081	0.2429	3.8027	4.96	0.0259
S	OTHER	0	0.0000	0.0000	0.0000	0.0000	.	.
W	COLLCTR	1	2.6356	0.5139	1.6285	3.6428	26.31	<.0001
W	LOCAL	1	1.2130	0.4970	0.2388	2.1871	5.96	0.0147
W	MAJART	1	3.8872	0.5454	2.8184	4.9561	50.81	<.0001
W	MINART	1	3.3069	0.5427	2.2432	4.3707	37.13	<.0001
W	OTHER	0	0.0000	0.0000	0.0000	0.0000	.	
Scale		0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed

LR Statistics For Type 3 Analysis

Source	DF	Chi-Square	Pr > ChiSq
System	2	6.67	0 0366
FAGINT	1	16.96	<.0001
Num_Legs	2	8.28	0.0159
Q_major	1	13.03	0.0003
N	4	10.80	0.0289
S	4	10.86	0.0282
W	4	93.94	<.0001

finalintersection

Contrast Estimate Results

Label	Estimate	Standard Error	Alpha
3 vs 4 lanes	-0.8954	0.3148	0.05
other vs 4 lanes	-20.2705	0.0000	0.05
Northbound Collector vs Major Arterial	-0.5689	0.6901	0.05
Northbound Local vs Major Arterial	-1.2700	0.6950	0.05
Northbound Minor vs Major Arterial	0.2289	0.6655	0.05
Northbound All other road classes vs Major Arterial	-1.1713	1.1181	0.05
Southbound Collector vs Major Arterial	-0.7589	0.7225	0.05
Southbound Local vs Major Arterial	-1.3436	0.7294	0.05

The SAS System 00:43 Thursday, August 28, 2003 104

The GENMOD Procedure

Contrast Estimate Results

Label	Estimate	Standard Error	Alpha
Southbound Minor vs Major Arterial	-0.8273	0.6820	0.05
Southbound All other road classes vs Major Arterial	-2.8501	0.9745	0.05
Westbound Collector vs Major Arterial	-1.2516	0.3426	0.05
Westbound Local vs Major Arterial	-2.6743	0.3548	0.05
Westbound Minor vs Major Arterial	-0.5803	0.2969	0.05
Westbound All other road classes vs Major Arterial	-3.8872	0.5454	0.05

Contrast Estimate Results

Label	Confidence Limits		Chi-Square	Pr > ChiSq
3 vs 4 lanes	-1.5124	-0.2784	8.09	0.0044
other vs 4 lanes	-20.2705	-20.2705	.	
Northbound Collector vs Major Arterial	-1.9215	0.7837	0.68	0.4097
Northbound Local vs Major Arterial	-2.6322	0.0922	3.34	0.0676
Northbound Minor vs Major Arterial	-1.0755	1.5332	0.12	0.7309

finalintersection

Northbound All other road classes vs Major Arterial	-3.3627	1.0202	1.10	0.2949
Southbound Collector vs Major Arterial	-2.1719	0.6571	1.10	0.2935
Southbound Local vs Major Arterial	-2.7731	0.0860	3.39	0.0655
Southbound Minor vs Major Arterial	-2.1611	0.5095	1.47	0.2251
Southbound All other road classes vs Major Arterial	-4.7600	-0.9402	8.55	0.0034
Westbound Collector vs Major Arterial	-1.9232	-0.5800	13.34	0.0003
Westbound Local vs Major Arterial	-3.3697	-1.9788	56.81	<.0001
Westbound Minor vs Major Arterial	-1.1623	0.0017	3.82	0.0507
Westbound All other road classes vs Major Arterial	-4.9561	-2.8184	50.81	<.0001

Contrast Results

Contrast	Type	DF	Chi-Square	Pr > ChiSq
3 vs 4 lanes	LR	1	7.88	0.0050
other vs 4 lanes	LR	1	0.55	0.4596
Northbound Collector vs Major Arterial	LR	1	0.70	0.4017
Northbound Local vs Major Arterial	LR	1	3.67	0.0555
Northbound Minor vs Major Arterial	LR	1	0.12	0.7327
Northbound All other road classes vs Major Arterial	LR	1	1.10	0.2936
Southbound Collector vs Major Arterial	LR	1	1.09	0.2975
Southbound Local vs Major Arterial	LR	1	3.30	0.0691
Southbound Minor vs Major Arterial	LR	1	1.43	0.2310
Southbound All other road classes vs Major Arterial	LR	1	8.01	0.0046
Westbound Collector vs Major Arterial	LR	1	11.02	0.0002
Westbound Local vs Major Arterial	LR	1	65.62	<.0001

The GENMOD Procedure

Contrast Results

Contrast	Type	DF	Chi-Square	Pr > ChiSq
Westbound Minor vs Major Arterial		1	3.83	0.0503

finalintersection

Contrast	Type	DF	Chi-Square	Pr > ChiSq
	LR			
Westbound All other road classes vs Major Arterial	LR	1	48.79	<.0001

Appendix F: Data Entry Guide

City of Toronto Collision Data Entry Instructions

- Target vehicles are V1 traveling on R1. Bullet vehicles are V2 on R2. Data entry must be accommodated for this. Many fields are dependant on this being coded in proper sequence to this. When it is determined that police have indicated target and bullet as vehicle 2, 1, as opposed to 1, 2, then many fields are reversed (w.r.t.the "&" sign), including:
    - MVA 9 & 10
    - MVA 11 & 12
    - MVA 13 & 14
    - MVA 15 & 16
    - MVA 17 & 18
    - MVA 19 & 20
    - MVA 21 & 22
    - MVA 23 & 24
    - MVA 33 & 34
    - MVA 35 & 36
    - MVA 43 & 44
    - MVA 46 & 47
    - MVA 48, 49, 50, 51, 52, 53 & 54, 55, 56, 57, 58, 59
    - MVA 60 & 61
    - MVA 62,63 & 64,65

This may require reversal of involved persons (drivers and passengers) gender, age, injuries, position, etc.

- Unless otherwise stated, yes = 1, no = 2. When unknown, observation is left as blank.
- The failed to remain checkbox is located near the top of the collision report and is simply a yes = 1 or no = 2 or unknown = blank field.
- The intersection perpendicular field is determined from the sketch(es) provided in the collision report(s). It is a yes or no field.
- The number of lanes in R1 and R2 is usually recorded by police in boxes nearest the sketch. On a two-way roadway with 3 lanes in each way, that would be 6 lanes recorded.
- The color of lights field is indicative of any note that points to any of the following collision characteristics:

  1 - Involvement of Amber/Red
  2 - End of LT priority phase
  3 - Totally red light running, no amber
  4 - Green

Otherwise, it is assumed to be either other or unknown and is left blank. This information is often indicated in the sketch notes or diagram itself.

- Position of passengers is to be in accordance with the manner specified on the coded template page. Do not follow the police recorded style for this field.
- Only use the middle seating row seats (3 and 4) when a middle row exists. Otherwise, use the last row. I.e., in the case of a 4 door sedan, use the front most seats (D, 2, 1) and the rearmost seats (5,6,7) and never the middle row seats (3 and 4).
- Using this manner of coding passengers, there will be many blank passenger fields where no passenger exists.

References

1 Insurance Institute for Highway Safety. *Frequently Asked Questions*. Highway Loss Data Institute. http://www.highwaysafety.org/faq.htm, June 2003.

2 The Auto Accident Helpline. *The Medical and Legal Resource for Auto Accident Victims*. http://www.autoaccidenthelpline.com/SideImpact.html, June 12, 2003.

3 Ontario Office of Road Safety. *Ontario Road Safety Annual Report 2000*. Ontario Ministry of Transportation, Toronto, Canada, February 2003.

4 Persaud, B., Chipman, M., Lyon, C. Methods for Assessing Collision Causes and Remedies from a Multidisciplinary Perspective. Proceedings of The Canadian Multidisciplinary Road Safety Conference XIII; Banff, Alberta, June 8-11, 2003.

5 Methodology for Reviewing Evidence: Methodology Chapter in Guide to Clinical Preventive Services: Report of the U.S. Preventive Services Task Force, Second Edition. Williams & Wilkins Publishers, Baltimore, 1996.

6 Kirk, A., Stamatiadis, N. Crash Rates And Traffic Maneuvers Of Younger Drivers. Transportation Research Board. 80th Annual Meeting, Washington D.C., January 7-11, 2001.

7 Preusser, F. D., Williams, F. A., Ferguson, S. A., Ulmer, R. G., Weinstein, H. B. Fatal Crash Rish For Older Drivers At Intersections. Accident Analysis and Prevention, Volume 30, No. 2, 151-159, Pergamon, 1998.

8 Stamatiadis, N., Deacon, J., A. Quasi-Induced Exposure: Methodology and Insight. Accident Analysis and Prevention; Volume 50, No. 1, 37-52, 1997.

9 Zhang, J., Lindsay, J., Clarke, K., Robbins, G., Mao, Y. Factors Affecting The Severity Of Motor Vehicle Traffic Crashes Involving Elderly Drivers In Ontario. Accident Analysis and Prevention 32, 117-125, 2000.

10 Gebers, M.A. Exploratory Multivariable Analysis of California Driver Record Accident Rates. Transportation Research Record, Volume 1635, 72 –80, 1998.

11 Pulugurtha, S. S., Nambisan, S. S., Pulugurtha., S. M. Analysis of Crashes By Traffic Control Device In Clark County, Nevada. TRB, Volume 80, Washington D.C., January 2001.

12 McGinnis, R.G., Wissinger, L.M., Kelly, R.T., Acuna, C.O. Estimating the Influences of Driver, Highway, and Environmental Factors on Run-off-Road Crashes Using Logistic Regression. TRB, Washington D.C. Revised October 31, 1998.

13 Marija N. J. SPSS Professional Statistics 7.5, SPSS Inc., Chicago, 1997.

14 An Evaluation of Severity And Outcome Of Injury By Type Of Object Struck (First Object Struck Only) for Motor Vehicle Crashes in Connecticut. National Highway Traffic Safety Administration. http://www.nhtsa.dot.gov/people/injury/research/connecticut/severityconnecticut.html June 2003.

15 Chipman, M. L. Side Impact Crashes - A Review of the Evidence for Factors Affecting Incidence and Severity. Proceedings of The Canadian Multidisciplinary Road Safety Conference XIII, Banff, Alberta, June 8-11, 2003.

16 Do Seatbelts Really Work? Canadian Health Network, www.canadian-health-network.ca, January 1999.

17 Brehmer, C.L., Klacir, K.C., Noyce, D.A., Manser, M.P. National Cooperative Highway Research Program (NCHRP) Report 493. Evaluation of Traffic Signal Displays for Protected/ Permissive Left-Turn Control. Transportation Research Board. http://gulliver.trb.org/publications/nchrp/nchrp_rpt_493.pdf.

18 Hauer, E. Observational Before-After Studies in Road Safety: Estimating The Effect Of Highway and Traffic Engineering Measures on Road Safety. Pergamon. New York. 1997.

19 Central East Health Information Partnership. Toronto. http://www.cehip.org/DataArea/ DataInfo/MTO%20Accident%20Form.pdf, 2002.

20 Motor Vehicle Traffic Collisions Injuries: Chronic Disease and Injuries - Injury Prevention and Substance Abuse Prevention. http://www.cehip.org/apheo/indicators/pages/indicators/ind04c07.html#national, Revision January 7, 2003.

21 Self Reporting Collision Centres. Toronto Police. http://www.torontopolice.on.ca, 2003.

22 Ministry of Ontario. Highway Traffic Act. 1990. Publications Ontario http://192.75.156.68/DBLaws/Statutes/English/90h08_e.htm

23 Ontario Road Safety Annual Report. (ORSAR) Ministry of Transportation. Publications Ontario. Toronto. February 2003. http://www.mto.gov.on.ca/english/safety/orsar/orsar01/ors_01.pdf.

24 Hauer, E. Observational Before-After Studies in Road Safety. Permagon Press, 1997.

25 Pain Centre. How the Body Controls Pain. Arthritis Foundation. http://www.arthritis.org/conditions/pain_center/default.asp, 2002.

26 Interpreting Odds Ratios in Logistic Regression. UCLA Academic Technology Services. 2003. http://www.ats.ucla.edu/stat/stata/faq

27 Stokes, M.E., Davis, C.S., Koch, G.G. Categorical Data Analysis Using the SAS System. Using the GENMOD Procedure, 208-213, Cary, North Carolina, 1995.

28 Stokes, M.E., Davis, C.S., Koch, G.G. Categorical Data Analysis Using the SAS System. Using the GENMOD Procedure, 208-213, Cary, North Carolina, 2001.

29 Insurance Act: Fault Determination Rules. Regulation 668. Made available through: Canadian Legal Information Institute. http://www.canlii.org/on/regu /cron/20030205/r.r.o.1990reg.668/whole.html, July 3, 2003.

30 Ontario Auto Insurance Tips: Reports and Articles. But It Wasn't My Fault. Charles E. Boyd Limited. http://www.autoinsurancetips.com/fault.htm, 2001.

31 Garber, N. J., Zhao, M. Distribution and Characteristics of Crashes At Different Locations Within Work Zones In Virginia. TRB. National Research Council. Washington D.C., 2001.

32 Persaud., B. N. Accident Potential Models for Ontario Road Section. Final Report Submitted to the Research and Development Branch Ministry of Transportation, Ontario. Professor Department of Civil Engineering Ryerson Polytechnic University, page 4.

33 Guisan, A., T. C. Edwards, Jr., and T. Hastie. In press. Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecological Modelling. June 2000.

34 SAS Online Version 8. Type 3 Analysis. SAS Institute Inc., Cary, NC. http://www.id.unizh.ch/software/unix/statmath/sas/sasdoc/stat/chap29/sect31.htm, 1999.

35 Dickman, P. Using PROC GENMOD for logistic regression (SAS version 6): Advanced SAS. Department of Medical Epidemiology at Karolinska Institute. Stockholm. http://www.pauldickman.com/teaching/sas/genmod_logistic.html.

36 Weber, B. Regression Modeling: Stepwise Regression. Managerial Economics and Decision Sciences, Kellogg School of Management. http://www.kellogg.nwu.edu /faculty/weber/emp/Modeling.htm

37 Marshall, M., Mercer, B. Estimating the Presence of Alcohol and Drug Impairment in Traffic Crashes and their Costs to Canadians: A Discussion Paper. Applied Research and Evaluation Services (ARES). http://www.madd.ca/library/ magnitude_report2k2.pdf December 2002.

38 Terhune, K.W. An Evaluation of Crash Culpability to Assess Alcohol and Drug Impairment Effects. 26th Annual Proceedings, AAAM, Ottawa, Canada, 329-348, October 4-6, 1982.

39 Altman, D.G., Andersen, P. K. Bootstrap Investigation of the Stability of a Cox Regression Model. *Statistics in Medicine 8*: 771-783. 1989.

40 Insurance Act: Fault Determination Rules. Regulation 668. Made available through: Canadian Legal Information Institute. http://www.canlii.org/on/regu /cron/20030205 /r.r.o.1990reg.668/whole.html, July 3, 2003.

Glossary of Terms and Abbreviations

Ability Impaired Alcohol
Driving while one's ability is impaired by alcohol or driving with a blood alcohol concentration exceeding 80 milligrams in 100 millilitres of blood.

Alcohol Involved
This category includes both drivers reported as ability impaired by alcohol and drivers reported as "had been drinking."

Driver
Unless specified otherwise, any person, whether licensed or not, considered to be in care and control of a vehicle at the time of a collision.

Fatal Collision
A motor vehicle collision in which at lest one person sustains bodily injuries resulting in death. Since January 1, 1982, fatal collision statistics include only deaths within 30 days of the collision.

Had Been Drinking
Driving after having drunk an amount of alcohol not considered sufficient to be legally impaired or with a measured blood alcohol count of greater than zero but less than 80 milligrams.

Highway
A common and public highway, street, avenue, etc, any part of which is intended for public use or used by the general public for the passage of vehicles and including the area between property lines.

Major Injury
A non-fatal injury severe enough to require that the injured person be admitted to hospital, even if for observation only.

Minimal Injury
A non-fatal injury, including minor abrasions and bruises, which does not necessitate the injured person going to a hospital.

Minor Injury
A non-fatal injury requiring medical treatment at a hospital emergency room, but not requiring hospitalization of the involved person.

Motor Vehicle Collision
Any incident in which bodily injury or damage to property is sustained as a result of the movement of a motor vehicle, or of its load while a motor vehicle is in motion.

Off-Highway Collision
An off-highway collision involving any of the motorized vehicles which are covered by legislation under the Highway Traffic Act, the Motorized Snow Vehicles Act, and the Off-Road Vehicles Act.

On-Highway Collision
A motor vehicle collision which occurs on the highway between the property lines.

Pedestrian
Any person not riding in or on a vehicle involved in a motor vehicle collision.

Personal Injury Collision
A motor vehicle collision in which at least one person involved sustains bodily injuries not resulting in death.

Property Damage Collision
A motor vehicle collision in which no person sustains bodily injury, but in which there is damage to any public property or damage to private property including damage to the motor vehicle or its load.

Reportable Collision
Any fatal or injury collision, or any collision in which there is any damage to public property in excess of a monetary value prescribed in law. The minimum reportable level for property damage only collision rose from $200 to $400 on January 1, 1978 and rose again to $700 on January 1, 1985. As of January 1, 1998, the minimum reportable level for property damage only collision is $1,000.

Self-Reporting of a Collision
Under a new section of the Highway Traffic Act [S199(1.1)], when one is in a collision in which there is only property damage (no injury or death, and, among other conditions, no criminal activities such as impaired driving) the involved person(s) may report the collision immediately by proceeding with one's vehicle to a Collision Reporting Centre. Self-Reporting of a collision was introduced on January 1, 1997.

Suspension
Withdrawal of a driver's privilege to operate a motor vehicle for a prescribed period of time. [40]

ADT	Average Daily Traffic
CO	Crude Odds
COR	Crude Odds Ratio
CRC	Collision Reporting Centre
FAG	Flashing Advance Green

FDW	Flashing Don't Walk
LA	Left Turn Arrow
LOS	(Statistical) Level of Significance
MTO	Ministry of Transportation
MVA	Motor Vehicle Accident Report
OR	Odds Ratio
PDO	Property Damage Only
RAIR	Relative Collision Involvement Rates
SAS	Statistical Analysis Software
SCOOT	Split Cycle Offset Optimization Technique
MTSS	Main Traffic Signal System
MVTC	Motor Vehicle Traffic Crashes
VMT	Vehicle Miles Traveled
VSRC	Vehicle Safety Research Centre