INCIDENT DURATION MODEL AND SECONDARY INCIDENT CAUSATION MODEL BASED ON ARCHIVED TRAFFIC MANAGEMENT CENTER DATA

By

Nithin Michael Gomez

Thesis

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

MASTER OF SCIENCE

in

Civil Engineering

May, 2005

Nashville, Tennessee

Approved:

Professor Robert E. Stammer, Jr.

Professor Malcolm E. Baird

To my parents and my sister, ever supportive and ever loving

ACKNOWLEDGEMENTS

This work would not have been possible without the constant help and encouragement from my advisor, Dr. Robert E Stammer, Jr. I owe him special thanks for his extensive technical guidance and valuable insights throughout the course of my graduate study. He has been an outstanding teacher and an excellent mentor.

I would also like to thank Dr. Malcolm E Baird for helping me get the data for my research. His constructive suggestions and attentive review was very helpful in completing this study.

I would like to extend my gratitude to the Nashville Traffic Management Center for providing me with the data for use in this research.

I am grateful to all my fellow graduate colleagues and my friends for making my life at Vanderbilt very enjoyable. Last but not the least; I would like to extend a heart-felt thanks to my parents and my sister for their love and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

1.1 Background Information

1.1.1 Congestion: The Problem

Congestion on highways and roads is one of the most distressing problems the American public is facing. Congestion has grown everywhere in areas of all sizes, occurs for longer hours of the day and delays more travelers and goods than ever before. This trend is well illustrated by Figure 1.1.

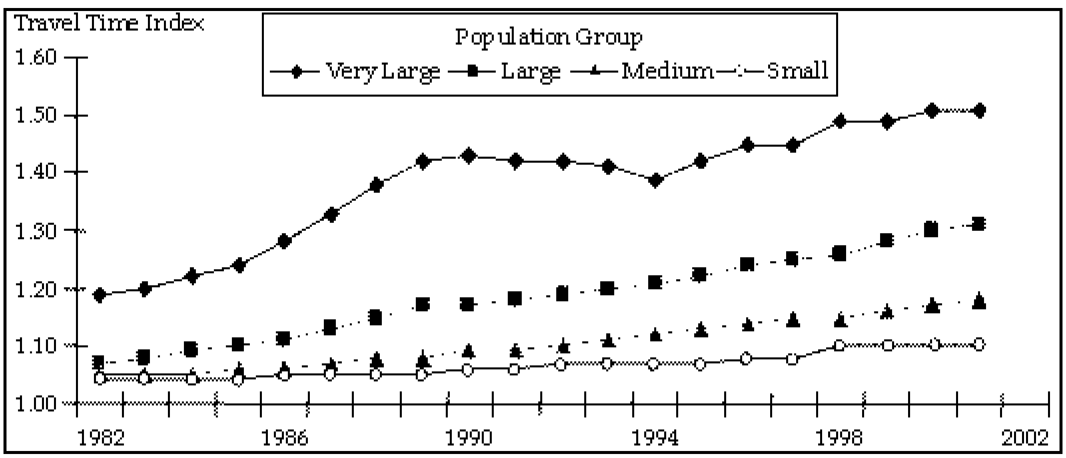

Figure 1.1: Peak-Period Congestion (Travel Time Index) Trends by U.S. Population Group

Source: Schrank, D. and Lomax, T., *2003 Annual Urban Mobility Report*, Texas Transportation Institute.

Note: The Travel Time Index is a measure of the total amount of congestion and is defined as the ratio of the weekday peak-period travel time to the travel time under ideal conditions. A Travel Time Index value of 1.3 indicates 30 percent longer travel time curing peak-period than under ideal conditions. Population groups are: Very Large (greater than three million); Large (one to three million); Medium (500 thousand to one million); Small (less than 500 thousand).

As is clear from Figure 1.1 congestion is increasing and is also expected to further increase in the future. Further information verifying the congestion expansion is evident from the changes in key indicators of the national roads and highways as depicted in Table 1.1.

Table 1.1: Changes in Indices Related to Surface Transportation in the US

Index	Percent Change from 1992 to 2002
US Population	12.9
Licensed Drivers	10.4
State Registered Motor Vehicles	21.1
Highway Motor Fuel Usage	26.2
Vehicle Miles of Travel	27.5

Sources: Highway Statistics 2002 (Office of Highway Policy Information, FHWA, US DOT.) and Highway Statistics 1992 (Office of Highway Policy Information, FHWA, US DOT.)

This congestion results in wasteful fuel consumption, contributes to increased atmospheric pollution (associated with the burning of fossil fuels), causes thousands of unproductive hours affecting the nation's economy, disrupts transportation schedules and causes distress to drivers and passengers leading to accidents and other mishaps. According to the 2004 Urban Mobility Report conducted by the Texas Transportation Institute, congestion in the US caused:

- a total delay of 3.5 billion hours,
- total fuel wastage of 5.7 billion gallons, and
- wastage of 63.2 billion dollars (base year: 2002).

As a matter of fact, in most metropolitan cities congestion is so bad and prolonged over multiple hours during both morning and evening periods, making the idea of a single "rush hour" obsolete. For the year 2002, 67 percent of the peak travel was congested with 58 percent of the total road system experiencing congestion spread over 7.1 hours a day, according to the 2004 Urban Mobility Report.

1.1.2 The Solution

Strategies to counter congestion can be simply classified as follows:

- Increasing the capacity of the existing transportation infrastructure, and
- Operating the existing infrastructure more efficiently.

Over the years, experience in the transportation world has shown the latter being a more economically viable and “intelligent” option for congestion mitigation and for the past twenty five years there has not been more than 1% increase in the total road and street mileage in the US (Source: Highway Statistics 2002). The strategies falling under the second category are collectively termed as ITS (Intelligent Transportation Systems). ITS can be defined as those techniques which using the help of electronics and telecommunications, ensures smooth, safe, fast and economic transportation of people and goods.

One of the strategies under the ITS umbrella is ‘Traffic Incident Management’ which is defined as the systematic, planned, and coordinated use of human, institutional, mechanical, and technical resources to reduce the duration and impact of incidents, and improve the safety of motorists, crash victims, and incident responders. These resources are also used to increase the operating efficiency, safety, and mobility of the highway by systematically reducing the time to detect and verify an incident occurrence; implementing the appropriate response; and safely clearing the incident, while managing the affected flow until full capacity is restored (Source: Traffic Incident Management Handbook, FHWA).

Traffic incident management is still a nascent technique with lots of research possibilities. This study focuses on building an understanding of traffic incidents and their corresponding durations. Towards this end, this thesis involves developing models to explain the characteristics of incidents affecting their durations. This research also investigates the effect of incident durations in causing secondary incidents. (Secondary incidents are those occurring partly or entirely as a result of an earlier incident; the primary incident causing the secondary incident either directly or indirectly.)

1.2 Motivation

Traffic congestion is the result of many different interacting factors, all of which can be summed up into two categories:

- Traffic volumes exceeding the physical capacity of the road system
- Traffic-influencing events like crashes and bad weather

The level of congestion on a roadway is determined by the interaction of physical capacity with events taking place at a given time. Nationally, a composite estimate of how much each of these sources contributes to total congestion is depicted in Figure 1.2.

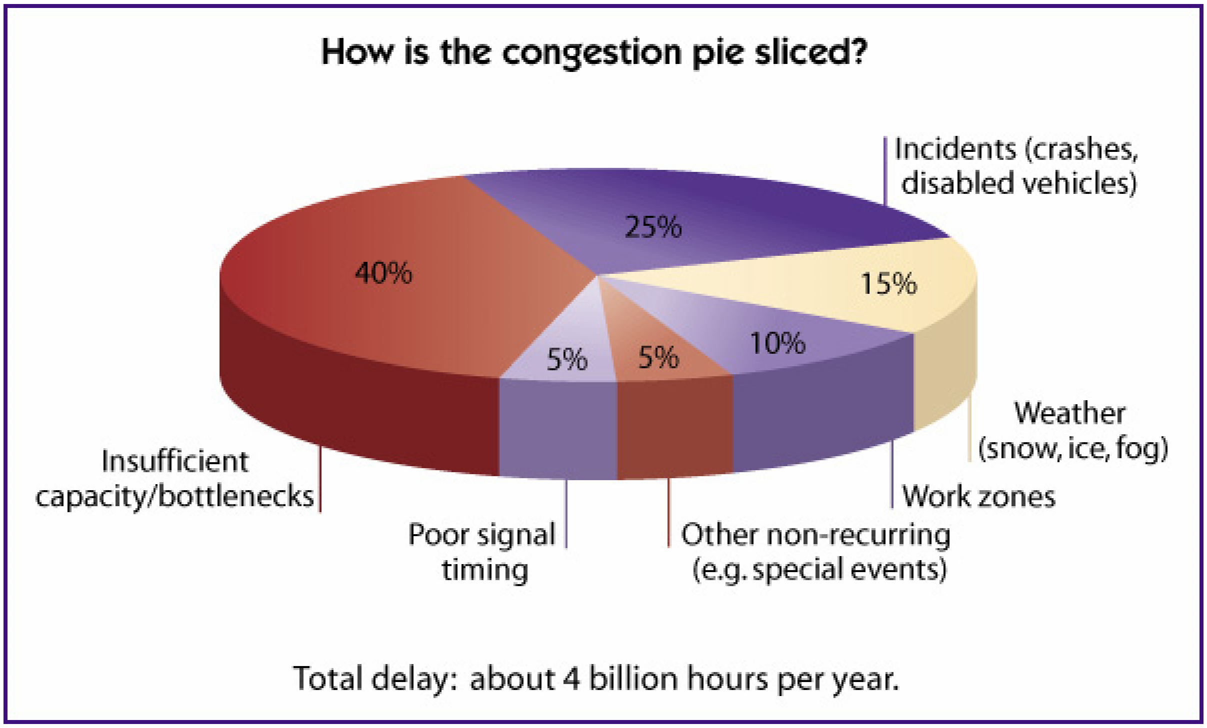

Figure 1.2: Sources of Congestion: National Summary

Source: *Congestion Mitigation: 21st Century Operations Using 21st Century Technologies*, Office of Operations, FHWA, US DOT.

Figure 1.2 illustrates a sizeable impact on traffic congestion due to highway incidents. As a result, several state Departments of Transportation (DOTs) have adopted traffic incident management as one of their prime operations seeking to reduce incident durations and incident-induced traffic congestion. These efforts have proved very beneficial in terms of the returns on the invested capital. According to the 2004 Urban Mobility Report, operational treatments have saved 335 million hours of delay amounting to six billion dollars saved on congestion costs.

With the incident management programs proving to be a very beneficial operation, there is a growing need to evaluate these programs. Statistical analysis of highway incident durations has become relevant to understand the impacts of traffic incidents on traffic congestion. Understanding the characteristics of incidents affecting incident duration can be very helpful in planning efficient incident management policies and strategies. In addition, primary incident duration is the prime-most factor in the study of secondary incident occurrence rates. Secondary incidents, factors which contribute to their occurrence, their effect on regional traffic are areas which require extensive research. Considering the statistic that 15 to 20 percent of all incidents in the US are secondary in nature, the study of secondary incidents demand much more attention than what currently exists.

Therefore, the chief motivation for this study is to develop a framework predicting incident durations depending on incident characteristics and probability of a secondary incident occurring given the primary incident duration.

1.3 Objectives and Overview of the Methodology

Within the context of the events and needs presented earlier, this study seeks to carry out the following goals:

- Propose a model to predict incident durations. The model takes into consideration the factors affecting the incident duration and uses regression techniques to develop a framework for predicting incident durations.
- Propose a model to predict secondary incident occurrences. This model is also based on a regression approach considering primary incident duration as the main factor.

1.4 Organization of the Research

The remainder of the thesis is organized as follows. Chapter 2 provides an introduction to traffic incident management programs followed by a comprehensive literature review on incident duration studies and secondary incident studies. In Chapter 3, a brief description about the research site and data acquired for this study is presented. Chapter 4 elaborates the research methodology adopted and also elucidates the data preparation required for the analysis. The incident duration model and the secondary incident causation model are presented in the subsequent two chapters. Finally the thesis is summarized with the findings, conclusions and directions for further research.

CHAPTER II

LITERATURE REVIEW

The most important lesson to be learned from the numerous congestion related studies conducted in the US is that the congestion problem cannot be solved by building more and more highways. Where increasing roadway capacity as a measure to reduce traffic congestion fails due to enormous quantities of resources required such as land, money, fuel and labor; other alternatives like land use changes, roadway operations improvement, transit, travel demand management, incident management, weather information management and work zone management have been suggested as feasible and effective. Among these options traffic incident management is perhaps the most promising short-term measure to alleviate congestion problems on freeways and urban arterials. Traffic incident management measures on freeways have received considerable attention from motorists and traffic management officials alike. The widespread success of such programs has led to more and more cities adopting traffic incident management as a viable step towards improving the city's transportation reliability and safety.

2.1 Traffic Incident Management

2.1.1 Introduction

Traffic incident management is a collective and coordinated effort by different agencies to respond to highway traffic disruptions, yielding significant benefits through reduced vehicle delays and enhanced safety to motorists through reducing incident frequency and improving response and clearance times. The major disciplines constituting an efficient traffic incident management program are state and local departments of transportation (DOTs), law enforcement agencies, fire companies, rescue agencies, tow operators, traveler information providers like the media, HAZMAT cleanup services, and other agencies supporting these major players.

Since an effective incident management program involves coordinating the operations of many of these agencies to respond to incidents, traffic incident management poses a significant institutional and management challenge. Many cities now have traffic management centers helping to integrate the communication systems of these disparate agencies and thereby achieving a synergic effort towards rapid and efficient incident management.

2.1.2 Traffic Management Centers

Traffic management centers (TMCs) can be considered as the core of a transportation management system, where information about the transportation network is collected, processed, and disseminated. The TMCs link various elements of the transportation system such as variable message signs, closed circuit video equipments, roadside count stations, and other elements enabling decision makers to identify and react to an incident in a timely manner based on real-time data.

In addition to being the focal point of a traffic incident management program, most of the TMCs also function as a repository of archived traffic data. Data relating to traffic incidents, specifically the incident type, temporal and spatial characteristics, weather-related data and other related miscellaneous information, are stored at the TMC. The main purpose of archiving such traffic data is to measure the performance of the TMC over long periods of time. Analysis of these data can indicate areas where improvements or changes are required. These data are also used by researchers in studying incident characteristics and trends.

Incident data obtained from the Nashville TMC are used to study incident duration properties, factors affecting incident clearance times and probability of an incident causing a secondary incident. Incident duration studies and secondary incident studies based on archived incident data are popular research areas due to their nascency and extensive research potential. Though the site-specific nature of these studies because of the data-dependency can be viewed as a shortcoming, they can provide very valuable insights into overall improvement of the corresponding city's incident management program.

In the following sections the framework and salient features of incident duration and secondary crash occurrence studies are presented. Essential characteristics of the study and an outline of the approaches adopted by various researchers are discussed, while highlighting their prominent advantages and limitations. The importance of mathematical modeling in such studies is also presented along with the advantages and disadvantages of the different modeling applications.

2.2 Incident Duration Studies

2.2.1 Introduction

One of the chief objectives of a traffic incident management program is to reduce the impact of an incident on regional travel and travelers. The most obvious way to achieve this is to clear the incident scene as quickly as possible. The duration of an incident can be defined as the time elapsed between the start of the incident and when normal travel conditions are restored. The timeline of a typical incident management process and the different constituents of incident duration are illustrated in Figure 2.1.

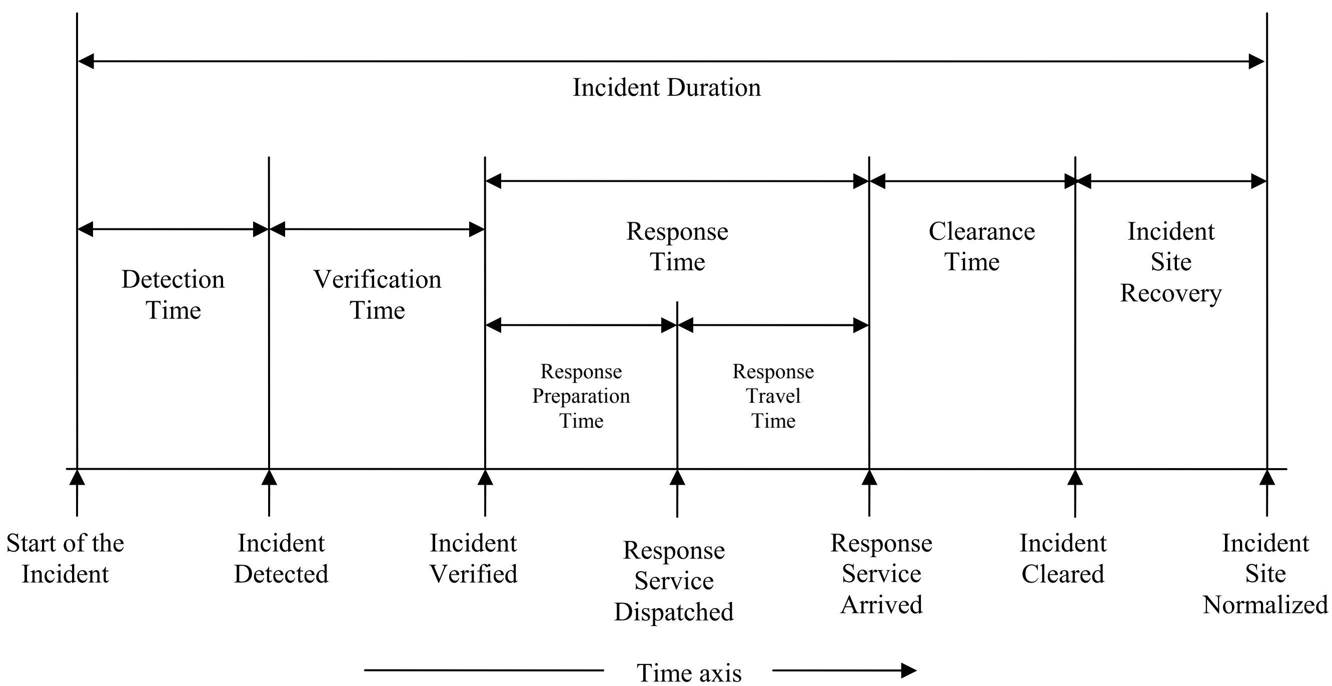

Fig 2.1: Traffic Incident Management Timeline

(Note: space intervals do not correspond to time of each event)

The total incident duration can be subdivided as follows:

- incident detection time (time between the incident occurrence and detection),
- incident verification time (time required to establish the presence of the incident),
- incident response preparation time (time between notifying the incident response team and their dispatch),

- incident response travel time (time required by the incident response team to reach the incident site after being informed about the incident),
- incident clearance time (time required to clear the incident site) and
- incident recovery time (time required for the system to reach normalcy after incident clearance).

The clearance process (involving the safe and timely removal of any stalled vehicles, wreckage, and debris from the roadway or shoulders and restoring the roadway for unimpeded flow of traffic) is usually the most time consuming step in the incident management program, and hence reducing incident clearance times has the greatest benefit on improving overall incident management time. Typically the whole incident clearance process takes at least twice the duration of other steps in incident management process. Reducing clearance times has the greatest potential effect (benefit) for improving overall incident management durations. (Source: Incident Management: "Successful Practices – A Cross Cutting Study", Intelligent Transportation Systems, US DOT). Documenting incident clearance times and understanding their properties will allow for better incident clearance strategies in the future. Towards this end, this research focuses on investigating incident clearance time properties and factors affecting them.

2.2.2 Incident Duration Models

Research work based on predictive techniques applied to incident duration was conducted as early as 1987. Though a lot of research has been done on this topic, there is little agreement between studies due to the site-specific nature of the data adopted for such studies. Almost each of these studies uses different data, different predictive variables and different statistical modeling techniques.

One of the earliest studies was conducted by Golob et al. in 1987 to analyze freeway accidents involving trucks. Using statistical tests like chi-squared test and Kolmogorov-Smirnov test, they were able to model the total duration of an incident according to a lognormal distribution. The main advantage of using a statistical distribution to represent incident durations is the straightforward manner of calculating the probability of an incident lasting for a particular duration of time. The proportion of incidents lasting less than a particular duration of time can also be calculated with the help of the cumulative distribution function of the incident duration.

Two properties of incident duration values, specifically positive values and having larger proportions of short-duration incidents, makes statistical distributions like lognormal, log-logistic and Weibull suitable for representing them. In fact, research studies by Giuliano in 1989, Garib et al. in 1997, and Sullivan in 1997 have supported the use of a lognormal distribution to describe freeway incident duration. Jones et al. used the log-logistic distribution in 1991 to a specific data set from the Seattle area. Nam and Mannering in 2000 found that the Weibull distribution could also be used to describe some incident data.

Predictive models based on regression have also been developed over the past. These models can be used to examine the influence of incident characteristics on the duration. In 1997, Garib et al. developed a model based on linear regression. Based on the analysis consisting of 205 incidents over a two-month period from Oakland, California, the authors developed a multiple linear regression model with six statistically significant variables: number of lanes affected (X_1), number of vehicles involved (X_2), binary variable for truck involvement (X_3), binary variable for time of day (X_4), natural logarithm of the police response time (X_5), and a binary variable for weather conditions (X_6). The log-based regression model is given by:

$$\text{Log (Duration)} = 0.87 + 0.027\, X_1\, X_2 + 0.2\, X_3 - 0.17\, X_4 + 0.68\, X_5 - 0.24\, X6$$

Khattak et al. (1995) argues that most incident duration prediction models have no operational value since they require knowledge about all incident variables whereas in the field, accident information is acquired sequentially and this progression should be reflected in the model. In their paper, the authors identified different stages of an incident based on the information available and developed truncated regression models for each of those stages. As more and more pieces of information about the incident become available the model adds on more variables to predict the incident duration depending on the information available at that stage. This study was based on a small sample of incidents and was intended to demonstrate a new methodology. This model has neither been validated since then nor applied to any future study on forecasting incident duration.

The primary drawback of linear regression models is the bulkiness of the predictive equation due to the categorical nature of independent variables resulting in a lot of dummy variables. Another disadvantage of using linear models is in assuming a ‘simplifying’ linear relationship between the dependent variable and the predictor variables.

Models based on conditional probabilities have also been developed over the past decade. These are models predicting incident duration probabilities given that the incident has already lasted a given time span. Jones et al. proposed a model based on conditional probabilities in 1991. Nam and Mannering followed up on the concept by applying hazard-based models (using conditional probabilities to find the likelihood that an incident will end in the next short time period given its continuing duration) to develop incident duration models (Nam and Mannering, 2000).

In this research a new model based on logistic regression technique is developed. To date, there has been no published study of the use of logistic regression for predicting incident duration. The fundamental approach involved in such models and advantages of such a method over linear regression models are discussed in detail in Chapter 5.

2.3 Secondary Incident Studies

2.3.1 Introduction

A secondary incident is an incident occurring because of the congestion or distraction from a prior incident. National averages in the US reveal almost 15 to 20 percent of all incidents being secondary incidents (Source: Traffic Incident Management, Office of Operations, FHWA). Secondary crashes often can be more serious than the original crash, especially if they occur at the boundary between free-flowing, highway speed traffic and stopped traffic.

Incident duration is a very important factor determining the likelihood of a primary incident giving rise to a secondary incident. The quicker the original incident is cleared, the lesser is the time motorists and response personnel are exposed to traffic hazards, and the possibility of secondary collision is greatly reduced.

2.3.2 Identifying a Secondary Incident

Identifying a secondary incident and the corresponding primary crash is the most important and difficult step in a secondary incident study. Most of the incident databases or crash reports rarely include information about any incidents giving rise to secondary incidents.

Parameters linking a primary incident and a secondary incident were first suggested by Raub (1997) based on a space-time surface. He argued that an incident has to occur within the ‘time of effect’ of an earlier incident to be considered secondary (‘Time of effect’ of an incident is the time for which the effects of that incident can be felt on the regional traffic flow which was calculated as 15 minutes more than the incident clearance time). The secondary incident should also lie within one mile (upstream) of the primary incident (Raub, 1997).

These set of assumed spatial and temporal criteria were updated in subsequent studies depending on on-site observations of secondary incidents. In order to account for secondary incidents caused on opposite direction of travel due to rubbernecking (or the gawking-effect) the spatial criteria was extended to downstream too in the opposite direction of travel for that route. For example, CHART (Coordinated Highways Action Response Team, Maryland ITS program) evaluation used the criteria that secondary incidents are those lying within 2 miles upstream and 2 hours of the primary incident in the same direction and within 0.5 miles downstream and 0.5 hours of the primary incident for the opposite direction

2.3.3 Secondary Incident Models

The earliest of secondary incident models was developed by Raub (1997). This was simply a secondary accident (a vehicular crash) rate model predicting the number of secondary accidents per incident. The research concluded more than 15 percent of all crashes were by an earlier event. Moore et al. (2004) developed a secondary accident rate model based on a different criterion for identifying secondary accidents than Raub (1997). The authors argued that only those incidents giving rise to a queue can cause a secondary accident. This may not be correct always as secondary accidents can be caused due to other factors of the primary incident (like debris left on the road). The authors also exclude chain reaction accidents (accidents occurring within few seconds of a primary incident and in immediate reaction) from secondary accidents. The space-time criteria that the authors followed were that the secondary accident had to be within 2 miles and 2 hours of the primary incident in either direction. The authors also argued that an accident occurring after an incident in the same direction, but downstream cannot be a secondary incident. This seems to a logical claim because the probability of a secondary accident happening beyond the primary incident site in the same direction is very small. Based on these assumptions Moore et al. concluded that for every incident there are 0.7 to 1.3 percent secondary accidents, much less than what Raub (1997) had predicted.

In a different approach to secondary crash modeling, Karlaftis et al. studied different primary incident characteristics likely to influence the occurrence of secondary crashes based on logistic regression. The authors employed logistic regression techniques to fit incident data from the Indiana's Hoosier Helper freeway service patrol system to determine the effects of several primary incident descriptors (clearance time, season, weekday/weekend, type of vehicle involved, lateral location etc.) on the probability of secondary crash occurrence. Based on the analysis the authors concluded that for every minute increase in clearance time the likelihood of a secondary crash increases by 2.8 percent.

In all the above discussed models, the subject of research was secondary accidents/crashes. But congestion and traffic delay on highways can be caused not only by secondary crashes but also due to other secondary incidents like engine stalls, traffic stops, overheating, running out of fuel and any other incident that might be a result of another incident. (Source: Traffic Incident Management, Office of Operations, FHWA). Even though the severity of an incident involving crashes is greater than other incident cases, from a traffic operations standpoint all secondary incidents are problematic. Hence in this research all secondary incidents are considered in the analysis and model development. Towards this end, a secondary incident causation model based on logistic regression is developed to investigate the influence of primary incident characteristics on causing secondary incidents and not just secondary crashes.

CHAPTER III

RESEARCH SITE AND DATA COLLECTION

The database used in this research was obtained from the Nashville Transportation Management Center (TMC) which houses the Tennessee Department of Transportation's intelligent transportation system for the Nashville area, the TDOT SmartWay. Mainly concerned with traffic incident management, the TDOT SmartWay combines roadway traffic sensors, video surveillance cameras and dynamic message signs to detect, verify and respond to incidents in an efficient manner and manage traffic conditions around the incident site in a safe and secure manner. TDOT's freeway service patrol, known as HELP, is also located in the TMC, and the HELP dispatches work in the TMC control room. Furthermore, the TDOT SmartWay coordinates with law enforcement agencies, fire companies, rescue agencies, tow operators, traveler information providers like the media, and other agencies for expeditious restoration of the incident scene.

3.1 Study Site

This research is based on the freeways in the city of Nashville which have been monitored by the Nashville TMC as a part of the above-mentioned TDOT SmartWay project. The highways covered under this project include all or segments of I-40, I-440, I-65, I-24, Ellington Parkway, parts of Briley Parkway and parts of Vietnam Veterans Parkway with the help of 56 traffic cameras and 150 roadway traffic sensors. Figure 3.1 shows the spatial extent of this project, the freeways monitored and also the location of the cameras and the Dynamic Message Signs.

In this study the data collected from July 2, 2003 to May 26, 2004 are used amounting to a total of three thousand, eight hundred and six recorded incidents.

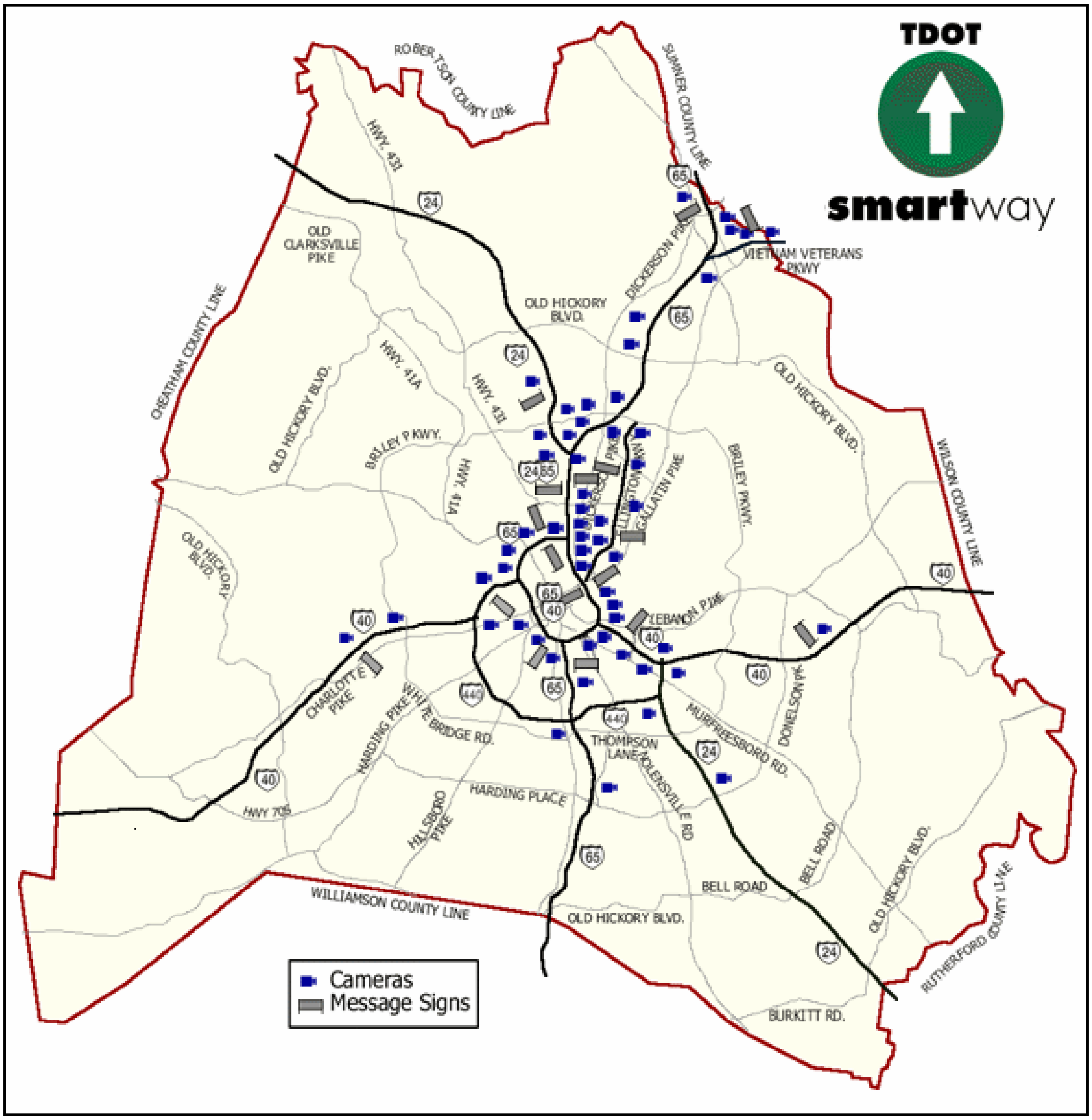

Figure 3.1: Study Site

Source: Tennessee Dept. of Transportation (TDOT)

3.2 Incident Data

The incident data is a part of the archived database at the TMC and is comprised of the incident description, spatial, temporal and environmental characteristics, and also descriptive information about the services rendered at the incident site. A detailed description of all the records in the incident data is included in the Appendix A. Following is a generic description of the records pertinent to this study:

- Incident Description: This consists of general characteristic features of an incident like the cause of the incident, the number of vehicles involved, truck involvement and whether or not that incident is a secondary incident.
- Spatial Characteristics: The approximate location of the incident is documented by recording the travel route, direction and the nearest mile-marker.
- Temporal Characteristics: The time at which the incident was notified to the TMC, the time the various services was rendered, the time at which the incident site was cleared and brought back to normalcy sums up all the temporal characteristics of the incident at hand.
- Environmental Characteristics: This consists of the prevailing weather conditions at the time of the incident.
- Services Rendered: The information regarding the incident detection, various services at the site catering to the incident clearance are also documented.

CHAPTER IV

RESEARCH METHODOLOGY AND DATA PREPARATION

In this chapter the logistic regression methodology for developing the incident duration model and the secondary incident causation model are introduced. The advantages of choosing logistic regression over a linear regression model are also discussed in detail.

The first section of this chapter deals with the fundamentals behind logistic regression models. A narrative follows explaining how the raw data from the TMC database is prepared for final analysis. Finally the variables that are considered for developing the model are discussed in detail.

4.1 Logistic Regression

Logistic regression is used when the dependent (response) variable is categorical in nature and the independent (input) variables are continuous, categorical, or both. The logistic regression approach may be further classified as follows depending on the nature response variable:

› Binary Logistic Regression: where the response variable is dichotomous (i.e., taking only two values, usually representing the occurrence or non-occurrence of some outcome event and usually coded as 0 or 1)

› Ordinal Logistic Regression: where the response variable is polytomous and ordered (i.e., coded as three or more ordered categorical levels; for example, the response may be like least, lesser than average, average, more than average and most)

› Nominal Logistic Regression: where the response variable is polytomous and un-ordered (i.e., coded as three or more un-ordered categorical levels; for example, a response like cloudy, sunny, rainy, snowing etc.)

Logistic regression is thus in contrast with ordinary linear regression where the response variable is continuous in nature and unlike ordinary linear regression, logistic regression does not assume that the relationship between the independent variables and the dependent variable is a linear one, nor does this technique assume that the dependent variable or the error terms are distributed normally (Source: Hosmer, David W. and Lemeshow, Stanley., "Applied Logistic Regression").

The general form of logistic regression models used in this study will be discussed in detail in the Chapter 5 and Chapter 6. Essentially these models involve regression constants associated with each of the response variables (which relates to the probability of that response happening) and regression coefficients associated with each predictor variable (which relates to the effect each of those predictor variables have on changing the probabilities of the various responses happening).

4.2 Why Choose Logistic Regression over Linear Regression?

For both the models (Incident Duration model and Secondary Incident Causation model) in hand, most of the independent variables analyzed are categorical in nature. Statistical modeling using linear regression would involve a lot of dummy variables making the model bulky and cumbersome for further usage. Also, the response variable can easily be coded into a categorical variable. In the case of the Incident Duration model the response variable can be coded as a polytomous response: shortest, shorter than medium, medium, longer than medium, longest. The response variable for the Secondary Incident Causation model is merely a dichotomous one: yes (caused a secondary incident) or no (did not cause a secondary incident). The relationship between the response and the predictor variable need not be restricted to a linear one which would have been the case if a linear regression model were to be used. Also encouraging the use of logistic regression is the generation of “odds ratio” for each predictor variable, with which the effect of the predictor variable on the response variable can be understood in a concise and correct manner. The odds ratio for a predictor is defined as the relative amount by which the odds of an outcome increase (odds ratio greater than 1.0) or decrease (odds ratio less than 1.0) for each unit change in the predictor variable (for a covariate variable case). In the case where the predictor variable is a factor ‘unit change’ refers to a comparison of a certain level to the reference level. The "odds" of an event is defined as the probability of the outcome event occurring divided by the probability of the event not occurring. (Source: *MINITAB®* Reference Guide)

4.3 Data Preparation

Before actual analysis, the incident data from the TMC had to be inspected for errors, false entries, duplicate records, irreconcilable records and missing fields within records to ensure adequate data quality. This was done mainly by the help of applications written in Java™ or by the built-in functions within *MS-Access*®.

Note: Based on these inspections about 12.1% (459 out of 3806) of the available recorded incidents in the database was found not fit for the analysis. Hence all the subsequent analysis and discussion is based on the incident population containing 3347 incidents.

Following is an account describing the preparation and further usage of various fields required for the analysis.

› Incident-ID: Incident-ID serves as an index for each incident in the database. This is an automatically generated field in the database and uniquely represents each record. This is therefore used to identify each record within the database for comparisons (to eliminate duplicities) and corrections in the associated fields.

› Start-Time and Date: This denotes the time and date on which the incident was reported and recorded into the database. This field is used as to denote the starting time of the incident for the lack of a better estimate, and is further used to calculate the incident clearance time and also to classify whether the incident happened during peak time or not and also whether on a weekday or weekend.

› Clearance-Time: The incident clearance time is calculated as a difference between the start and clear time of that incident. Corrections are made to the start-time/clear-time whenever a negative clearance time results from the calculations.

› Cause-Type: This classifies the incidents based on what caused the incident. Information regarding the number of vehicles involved, and whether the incident occurred in a construction zone or not is obtained from this field.

› Number of Vehicles Involved: The number of vehicles involved in the incident is stored in this field.

- Route: The name and direction of the route for each incident occurrence is inferred from this field. Duplicities in the name of the route are corrected using built-in applications within *MS-Access®*.
- Mile-Marker: The spatial location of the incident is determined using this field. Errors and empty entries in this field are rectified using other descriptive spatial information documented for each incident recorded. The mile-marker field is used in subsequent analyses for determining whether a particular incident resulted in any secondary incident or not. The location of an incident as to whether situated in an area covered well by the traffic cameras or not is also judged using the mile-marker information.
- Detected-By: This field stores the information according to whom the incident was reported. Corrections are made to do away with errors and missing entries to restrict the entries to ITS Operator, HELP Operator, Metro Services and Other Callers.
- Weather: This field is used to find out whether the incident happened during rainy conditions or not.
- Camera-Coverage: This field denotes the extent of camera coverage (the linear density of camera placements) at the vicinity of the incident. This information is determined using the mile-marker information and the information on camera locations. With reference to the Fig: 3.1, the routes can be grouped into three classes of varying camera coverage by mere visual inspection. Accordingly the classes are as follows:
    - Scarce: if the cameras are placed much less frequently than one per mile
        - all of Briley Parkway except between mile-markers 14.0 and 19.0 (14.0 and 19.0 excluded);
        - all of Ellington Parkway;
        - I-24: mile-markers greater than 54.0 (54.0 excluded);
        - all of I-40 except between mile-markers 206.0 and 213.5,
        - all of I-440;
        - I-65: mile-markers equal to and lesser than 82.0
    - Good: if the cameras are placed such that there is approximately an average of one per mile
        - Briley Parkway between mile-markers 14.0 and 19.0 (14.0 and 19.0 included);

        - I-24: mile-markers between 52.0 and 54.0 (54.0 excluded) and between 40.0 and 45.0 (45.0 excluded);
        - I-40: mile-markers equal to and between 206.0 and 211.0 and equal to and between 213.0 and 213.5;
        - I-65: mile-markers equal to and between 90.0 and 98.0
    - Very Good: if the cameras are placed much frequently than one per mile
        - I-24: mile-markers equal to and between 47.0 and 52.0;
        - I-40: mile-markers equal to and between 211.0 and 213.5;
        - I-65: mile-markers equal to and between 84.0 and 90.0
- Truck-ID: The involvement of a truck in an incident is indicated using this field.
- Roadwork: This field stores the information indicating roadwork near the location of the incident.

4.4 Variables Considered for Model Building

The different characteristics of an incident were investigated for their effect on incident duration and on the probability of that incident causing a secondary incident. Table 4.1 illustrates various factors that were used in the analysis and model development.

Table 4.1: Variables Used in Developing the Models

Variable	Description
Clearance Time	Time difference between start time and the clearance time of an incident and is a surrogate for incident duration.
Incident Duration	Depending on the clearance time, incident duration becomes one of the following five values; shortest, shorter than medium, medium, longer than medium, longest. (Explained in detail in Chapter 5).
Weekday/Weekend	Denotes whether an incident happened on a weekday or a weekend. Values: Weekday, Weekend
Peak/Non-Peak	Denotes whether an incident occurred during a peak time (Morning Peak: between 7:30 AM and 9:30 AM, Afternoon Peak: between 3:00 PM and 6:00 PM) or a non-peak time. Values: Morning Peak, Afternoon Peak, Non Peak
Detected-By	Denotes the personnel who reported the particular incident. Values: ITS Operator, HELP, Metro, Other Caller
Number of Vehicles Involved	Denotes the number of motor vehicles involved in the incident in case the incident involved a motor vehicle.
Roadwork	Denotes the presence of roadwork in the incident site Values: Yes, No
Truck	Denotes truck involvement in the incident Values: Yes, No
Weather Condition	Denotes the weather condition at the time of the incident Values: Clear, Cloudy, Rain, Fog, Snow
Rain	Denotes presence of rain Values: Yes, No
Camera-Coverage	Denotes the extent of camera coverage at the incident site Values: Scarce, Good, Very Good
Secondary Causing	Denotes whether an incident caused a secondary incident or not Values: Yes, No

CHAPTER V

INCIDENT DURATION MODEL

This model investigates the effect of various factors influencing the clearance time of an incident using an ordinal logistic regression approach. Ordinal logistic regression is used when the response variable is in a categorical form that has three or more possible levels with a natural ordering (such as strongly disagree, disagree, neutral, agree, and strongly agree). In the case of the incident duration model the categories involved in the response variable are shortest, shorter than medium, medium, longer than medium, longest.

5.1 Introduction

The general form of an ordinal logistic regression model with K distinct categories is as follows:

$$\boxed{P(y \leq k) = \frac{e^{\theta_k + x'\beta}}{1+e^{\theta_k + x'\beta}}} \tag{5.1}$$

where:

K is the number of distinct categories

k is the category number, taking values 1, 2,…., K-1

P(y ≤ k) is the probability that the response falls into category k or below

θ_k is the constant associated with the k^{th} response category

x' is the vector of predictor variables

β is the vector of coefficients associated with the predictor variables

The regression constants and coefficients are calculated using a logit link function by linking the probabilities to a linear combination of the predictor variables as shown below:

$$\boxed{LOGIT[P(y \le k)] = \log_e\left(\frac{P(y \le k)}{1 - P(y \le k)}\right) = \theta_k + x'\beta} \tag{5.2}$$

The coefficients are then estimated by a method equivalent to the maximum likelihood estimation procedure. Once the coefficients are evaluated, the cumulative probabilities and individual response probabilities can be calculated as follows:

› Cumulative probability of the first response category: P(y ≤ 1) = $\dfrac{e^{\theta_1 + x'\beta}}{1 + e^{\theta_1 + x'\beta}}$

› Cumulative probability of the second response category: P(y ≤ 2) = $\dfrac{e^{\theta_2 + x'\beta}}{1 + e^{\theta_2 + x'\beta}}$

and so on …

› For the last response category, Cumulative probability: P(y ≤ K) = 1.0

From above,

› Probability of first response category:

$$\boxed{P(y = 1) = \frac{e^{\theta_1 + x'\beta}}{1 + e^{\theta_1 + x'\beta}}} \tag{5.3}$$

› Probability of second response category:

$$\boxed{P(y = 2) = \frac{e^{\theta_2 + x'\beta}}{1 + e^{\theta_2 + x'\beta}} - \frac{e^{\theta_1 + x'\beta}}{1 + e^{\theta_1 + x'\beta}}} \tag{5.4}$$

and so on …

The probabilities calculated using equations 5.3 and 5.4 depend on the predictor variable pattern ‘x’. Therefore probabilities of responses due to different predictor variable scenarios can be calculated by simply varying the predictor vector. This in turn helps to find out individual effects of different predictor variables as explained later in this chapter.

5.2 Use of Statistical Software for Computation

Most of the commercially available statistical softwares are capable of estimating these regression constants and coefficients. All statistical computations for this research work were done using *MINITAB*®.

5.3 Response Variable Coding

To be used in the ordinal logistic regression model, the response variable (incident duration) for a particular incident is coded as follows, depending on the pentile range of the incident clearance time distribution within which the clearance time of that incident falls:

› Shortest – first pentile range
› Shorter than Medium – second pentile range
› Medium – third pentile range
› Longer than Medium – fourth pentile range
› Longest – fifth pentile range

The above step is accomplished by first fitting the clearance time values to a standard statistical distribution. As a first step the histogram of the incident clearance time is constructed and different distributions are fitted using *MINITAB*®. The best fitting distribution is found out by computing the Anderson-Darling test value and the corresponding p-value. The distribution which gives the lowest value for the Anderson-Darling test is chosen.

Note: The Anderson-Darling (A-D) test is a goodness-of-fit test used to inspect if a sample of data comes from a specific distribution. This test is a modification of the commonly used Kolmogorov-Smirnov (K-S) test and gives more weight to the tails of the distribution than does the K-S test (Source: NIST/SEMATECH e-Handbook of Statistical Methods). This is particularly important and suitable because in the case of incident duration values there is a large proportion of low values and negligible proportion of high values.

A three parameter log-logistic distribution is chosen based on the analysis using *MINITAB*®. The associated p-value is found to be almost equal to zero confirming that this is a good fit. The procedure is illustrated by the following figures.

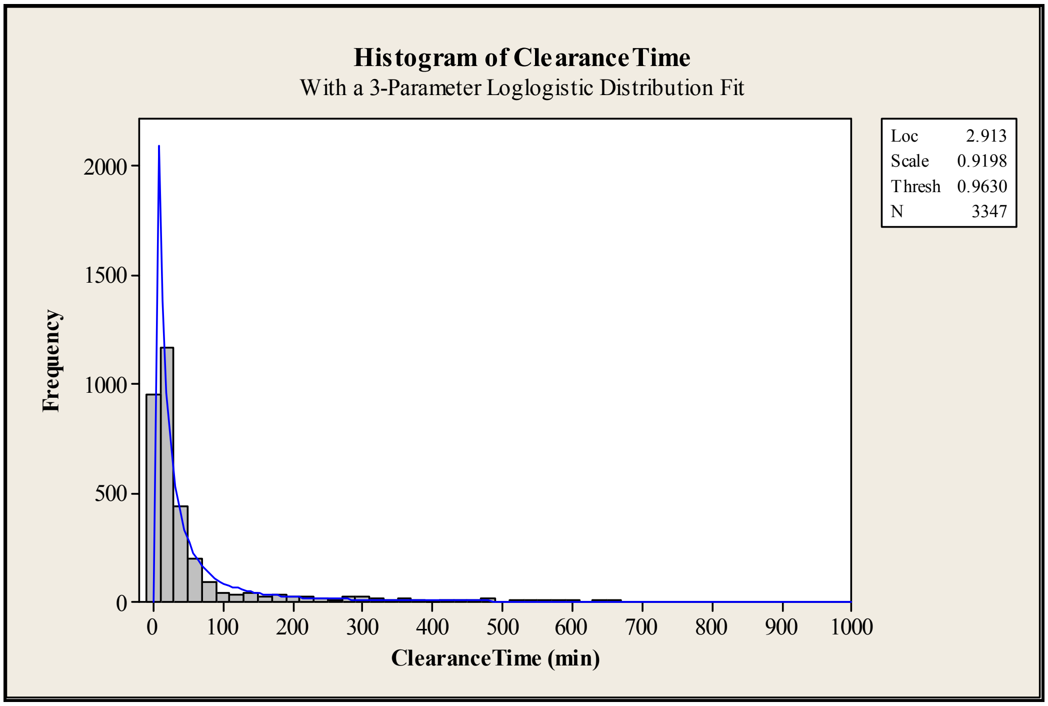

Figure 5.1: Histogram of Clearance Time with a Distribution Fit

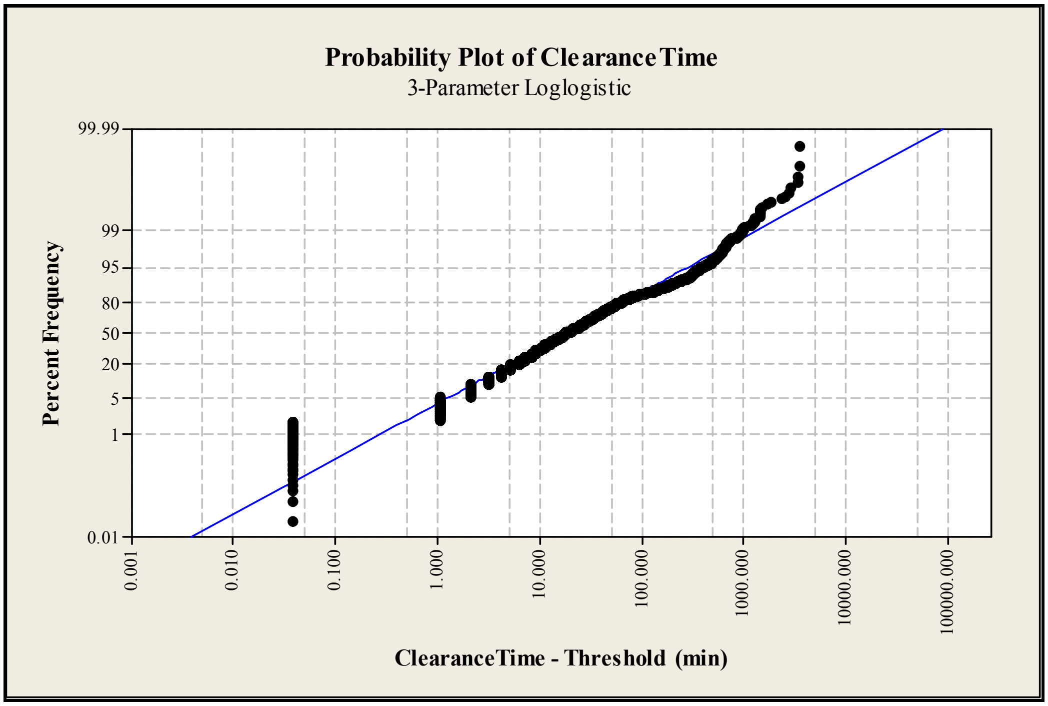

Figure 5.2: Probability Distribution Fit for Clearance Time

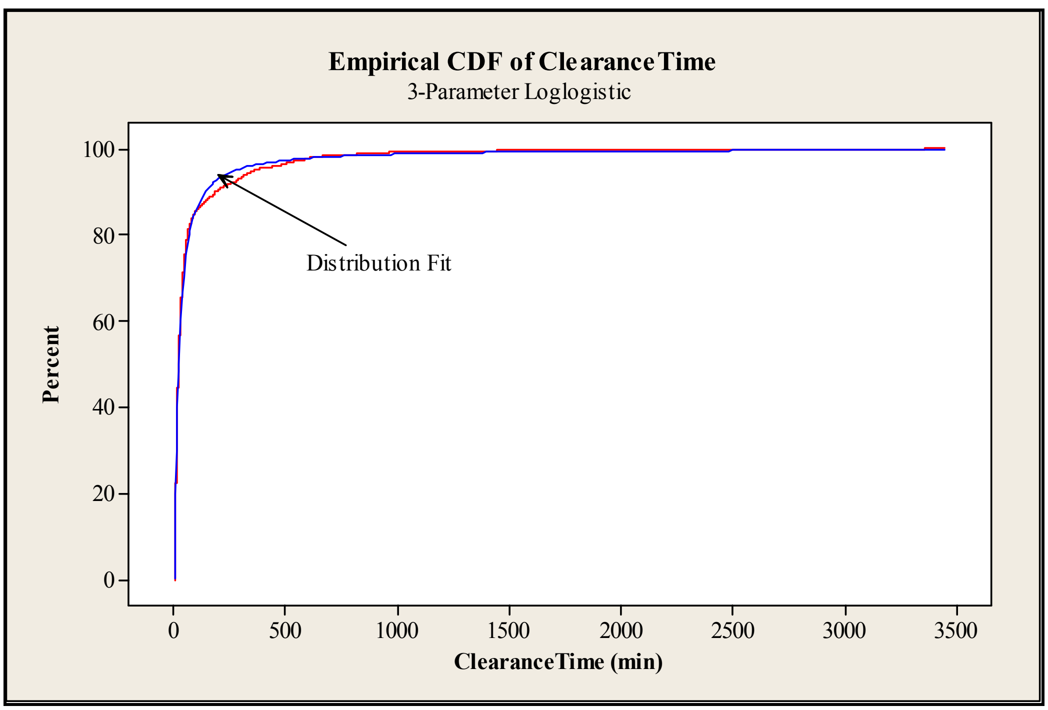

Figure 5.3: Cumulative Distribution Fit for Clearance Time

The above analysis shows that incident clearance time follows a three parameter log-logistic distribution with the following parameters:

- Location Parameter: 2.913
- Scale Parameter: 0.9198
- Threshold Parameter: 0.9630

The pentile values for the clearance time distribution can then be calculated as follows, where F denotes the Cumulative Distribution Function (CDF):

- First Pentile Value = $F^{-1}(0.2)$ = 6.1 min
- Second Pentile Value = $F^{-1}(0.4)$ = 13.6 min
- Third Pentile Value = $F^{-1}(0.6)$ = 27.7 min
- Fourth Pentile Value = $F^{-1}(0.8)$ = 66.9 min

As the clearance time is computed to the nearest minute, the incident duration coding for the regression analysis is defined as follows:

- Shortest – Clearance time less than 6 minutes (6 minutes included)
- Shorter than Medium – Clearance time greater than 6 minutes but less than or equal to 13 minutes
- Medium – Clearance time greater than 13 minutes but less than or equal to 27 minutes
- Longer than Medium – Clearance time greater than 27 minutes but less than or equal to 66 minutes
- Longest – Clearance time greater than 66 minutes

The following tables summarize the descriptive statistics of clearance time:

Table 5.1: Descriptive Statistics of Clearance Time

Variable	Mean	Std. Dev	Minimum	Median	Maximum
Clearance Time	77.0	215.8	1.0	19.0	3449.0

Table 5.2: Descriptive Statistics of Clearance Time within each Incident Duration Class

Variable	Incident Duration	Mean	Std. Dev	Minimum	Median	Maximum
Clearance Time	Shortest	3.7	1.6	1.0	4.0	6.0
	Shorter	9.9	2.0	7.0	10.0	13.0
	Medium	19.4	3.9	14.0	19.0	27.0
	Longer	41.8	10.8	28.0	40.0	66.0
	Longest*	338.9	414.9	67.0	211.0	3449.0**

* Note: The range of values in the clearance time class-'Longest' equal to 3382 minutes is an indication that the data contains a lot of outliers at the higher end.

** Note: This value is probably a recording error.

5.4 Predictor Variables

The predictor variables used in the analysis are as follows:

- Weekday/Weekend
- Peak/Non-Peak
- Detected-By
- Number of Vehicles Involved
- Roadwork-Presence
- Truck-Presence
- Weather-Condition
- Camera-Coverage

The reader is directed to Table 4.1 for descriptions of the above variables.

As a preliminary measure of the influence of these predictor variables on the response, incident clearance time medians factored by each predictor variable are compared to the median incident clearance time of the total population. Those variables which show considerable influence indicated by the percentage change in the median clearance times are then selected as independent variables in the regression analysis.

Median is used as the central tendency for comparison due to a lot of outliers present in the dataset causing the mean value to be a biased estimate for the central tendency. Median which is the 50th percentile or the mid-value is hence a more appropriate estimate for a central estimate. The median value for the incident clearance time dataset is nineteen minutes and the mean value is seventy seven minutes. This mean value corresponds to the 82nd percentile value of the incident clearance time (which can be calculated using the incident clearance distribution parameters). This is a clear indication that the incident clearance time dataset contains a lot of high extreme values.

The following tables (Table 5.3 – Table 5.10) show the corresponding values and percentage changes in incident clearance time medians caused by each factor.

Table 5.3: Shifts in Incident Clearance Time Median due to 'Weekday/Weekend'

Population Median (Min)	'Weekday/Weekend' – Median (Min)	
	Weekday	Weekend
19.0	17.0	26.0
Percentage Change (%)	-10.5	36.8

Table 5.4: Shifts in Incident Clearance Time Median due to 'Peak/Non-Peak'

Population Median (Min)	'Peak/Non-Peak' – Median (Min)		
	Non Peak	Morning Peak	Evening Peak
19.0	22.0	20.0	13.0
Percentage Change (%)	15.8	5.3	-31.6

Table 5.5: Shifts in Incident Clearance Time Median due to 'Detected-By'

Population Median (Min)	'Detected-By' – Median (Min)			
	ITS OPERATOR	HELP	METRO	OTHER CALLER
19.0	20.0	14.0	17.0	62.0
Percentage Change (%)	5.3	-26.3	-10.5	226.3

Table 5.6: Shifts in Incident Clearance Time Median due to 'Number of Vehicles Involved'

Population Median (Min)	'Number of Vehicles Involved' – Median (Min)			
	One	Two	Three	Zero
19.0	13.0	16.0	29.0	100.0
Percentage Change (%)	-31.6	-15.8	52.6	426.3

Table 5.7: Shifts in Incident Clearance Time Median due to 'Roadwork-Presence'

Population Median (Min)	'Roadwork-Presence' – Median (Min)	
	No	Yes
19.0	15.0	167.0
Percentage Change (%)	-21.1	778.9

Table 5.8: Shifts in Incident Clearance Time Median due to 'Truck-Presence'

Population Median (Min)	'Truck-Presence' – Median (Min)	
	No	Yes
19.0	19.0	21.0
Percentage Change (%)	0	10.5

Table 5.9: Shifts in Incident Clearance Time Median due to 'Weather-Condition'

Population Median (Min)	'Weather-Condition' – Median (Min)				
	CLEAR	CLOUDY	RAIN	FOG	SNOW
19.0	18.0	20.0	19.0	38.0	25.0
Percentage Change (%)	-5.3	5.3	0	100	31.6

Table 5.10: Shifts in Incident Clearance Time Median due to 'Camera-Coverage'

Population Median (Min)	'Camera-Coverage' – Median (Min)		
	Scarce Camera-Coverage	Good Camera-Coverage	Very Good Camera-Coverage
19.0	25.0	19.0	15.0
Percentage Change (%)	31.6	0	-21.1

As evident from the percentage change values from the above tables (Table 5.3 – Table 5.10) all of the predictor variables have at least one factor which impacts the clearance time. Hence all of the above predictor variables are included as independent variables for analysis into the logistic regression model.

5.5 Results of the Logistic Regression

The results of the logistic regression are summarized in the Table 5.11 which consists of the regression constants and the regression coefficients of the predictor variables.

The regression constants corresponding to the different response categories (θ_1, θ_2, θ_3, θ_4) are calculated by assuming default factor levels for all the predictor variables. Hence *MINITAB*® does not calculate any separate coefficients corresponding to these default factors. For the other predictor variable factors the regression coefficient calculated denote the change in the logit link functions of each response categories (Equation 5.2). A positive regression coefficient indicates reduction in the incident clearance time due to the corresponding factor, whereas a positive regression coefficient indicates an increment in the incident clearance time.

The odds-ratios corresponding to each predictor variable factors is computed as the inverse natural logarithm of their regression coefficients. Odds-ratios and their implications are further discussed in Section 5.7.

MINITAB® also computes the associated p-values for the regression constants and coefficients. A 5% confidence level is assumed throughout this research for ascertaining the significance of the various predictor variable factors. Hence for a particular predictor variable factor to be significant the corresponding p-value has to be less than 0.005.

Table 5.11: Logistic Regression Results

Predictor Variable	Predictor Variable Factors	Regression-Coefficient(β)	p-Value	Odds-Ratio
Const(1) -- θ_1		-1.157	0.000	n/a
Const(2) -- θ_2		-0.038	0.716	n/a
Const(3) -- θ_3		1.023	0.000	n/a
Const(4) -- θ_4		2.496	0.000	n/a
Weekday/Weekend	Default Factor: Weekday	n/a	n/a	n/a
	Weekend	-0.306	0.001	0.740
Peak/Non-Peak	Default Factor: Non-Peak	n/a	n/a	n/a
	Morning Peak	0.010	0.915	1.010
	Afternoon Peak	0.338	0.000	1.400
Detected-By	Default Factor: ITS Operator	n/a	n/a	n/a
	HELP	0.206	0.023	1.230
	METRO	-0.138	0.081	0.870
	Other Caller	-0.191	0.172	0.830
Number of Vehicles Involved	Default Factor: One	n/a	n/a	n/a
	Two	-0.224	0.006	0.800
	Three	-1.127	0.000	0.320
	Zero	-0.310	0.024	0.730
Roadwork-Presence	Default Factor: No	n/a	n/a	n/a
	Yes	-2.754	0.000	0.060
Truck-Presence	Default Factor: No	n/a	n/a	n/a
	Yes	-0.692	0.000	0.500
Weather-Condition	Default Factor: Clear	n/a	n/a	n/a
	Cloudy	-0.063	0.455	0.940
	Rain	-0.157	0.087	0.850
	Fog	-0.926	0.002	0.400
	Snow	-0.087	0.828	0.920
Camera-Coverage	Default Factor: Scarce Camera-Coverage	n/a	n/a	n/a
	Good Camera-Coverage	0.113	0.177	1.120
	Very Good Camera-Coverage	4.480	0.000	1.760

5.6 Discussion

Based on the regression constants and coefficients obtained by the regression analysis, the probabilities of different responses (incident duration class) can be calculated for any predictor variable scenario. For any particular predictor variable situation the response probabilities can be calculated as explained earlier in section 5.1 by using the corresponding regression constant (θ, depending on the response probability sought) and appropriate regression coefficients (β, depending on the predictor variable situation).

The effect of each predictor variable in impacting the probabilities of the responses can be studied by changing the regression coefficients of that variable only and keeping all other coefficients the same. After presenting a base predictor vector corresponding to normal conditions, the procedure for comparing response probabilities of different predictor variable scenarios are explained in the following section.

5.6.1 Response Probabilities based on Base Predictor Vector

The base predictor vector is defined as a predictor set corresponding to normal circumstances or base conditions. For use in the case at hand the base predictor vector is constituted by assuming the following predictor variable values:

- Weekday/Weekend – Weekday
- Peak/Non-Peak - Non Peak
- Detected-By - ITS Operator
- Number of Vehicles Involved - One
- Roadwork-Presence - No
- Truck-Presence - No
- Weather-Condition - Clear
- Camera-Coverage - Scarce Camera-Coverage

Based on this set of base predictor conditions the response probabilities can be calculated as follows (as explained earlier in section 5.1):

Probability of Incident Duration being 'Shortest':

$$= \frac{e^{\theta_1 + x'\beta}}{1 + e^{\theta_1 + x'\beta}} = \frac{e^{-1.157}}{1 + e^{-1.157}} = 0.24$$

Probability of Incident Duration being 'Shorter than Medium':

$$= \frac{e^{\theta_2 + x'\beta}}{1 + e^{\theta_2 + x'\beta}} - \frac{e^{\theta_1 + x'\beta}}{1 + e^{\theta_1 + x'\beta}} = \frac{e^{-0.038}}{1 + e^{-.038}} - \frac{e^{-1.157}}{1 + e^{-1.157}} = 0.25$$

Similarly,

Probability of Incident Duration being 'Medium': = 0.25

Probability of Incident Duration being 'Longer than Medium': = 0.19

Probability of Incident Duration being 'Longest': = 0.08

Examining the above probability values leads to the conclusion that for incidents characterized by the 'defined' base conditions the clearance time has the highest probability to lie in the incident-duration class 'Shorter than Medium'. Almost 75% (0.239 + 0.251 + 0.245 ≈ 75%) of the incidents happening under the base conditions fall under the incident duration categories 'Shortest', 'Shorter than Medium' or 'Medium', which in other words means having less than 27 minutes (Refer section 5.3) of clearance time.

In the following sections, the impact of each predictor variable is examined by calculating the response probabilities using the corresponding regression coefficients and then comparing them to the base response probabilities.

5.6.2 Weekday/Weekend

The regression coefficient associated with 'Weekend' is negative (Refer Table 5.11) which means that under this condition (i.e. for incidents happening on weekends) the response probabilities computed will predict larger proportions of incidents having longer clearance times when compared to incidents happening under normal conditions. There is sufficient evidence from the data corroborating this behavior as indicated by the fact that this coefficient is significant at 5% confidence level.

This behavior may be due to the fact that during weekdays there is more urgency for an incident to be cleared and normal conditions restored than on a weekend. Also the frequency and number of freeway service patrols may be lower on a weekend than on a weekday.

Following are the steps for calculating the response probabilities when the predictor variable 'Weekday/Weekend' takes the value 'Weekend' and all other predictor variables remaining the same. Similar calculations apply for computing response probabilities for other predictor variable scenarios.

Regression coefficient associated with the predictor variable factor 'Weekend' = -0.306

This implies that the change in the logit link function corresponding to each response category is equal to -0.306.

Probability of Incident Duration being 'Shortest' during a Weekend:

$$= \frac{e^{\theta_1 + x'\beta}}{1 + e^{\theta_1 + x'\beta}} \quad = \frac{e^{-1.157-0.306}}{1 + e^{-1.157-0.306}} = 0.19$$

Probability of Incident Duration being 'Shorter than Medium':

$$= \frac{e^{\theta_2 + x'\beta}}{1 + e^{\theta_2 + x'\beta}} - \frac{e^{\theta_1 + x'\beta}}{1 + e^{\theta_1 + x'\beta}} \quad = \frac{e^{-0.038-0.306}}{1 + e^{-.038-0.306}} - \frac{e^{-1.157-0.306}}{1 + e^{-1.157-0.306}} = 0.23$$

Similarly,

Probability of Incident Duration being 'Medium': = 0.26

Probability of Incident Duration being 'Longer than Medium': = 0.23

Probability of Incident Duration being 'Longest': = 0.10

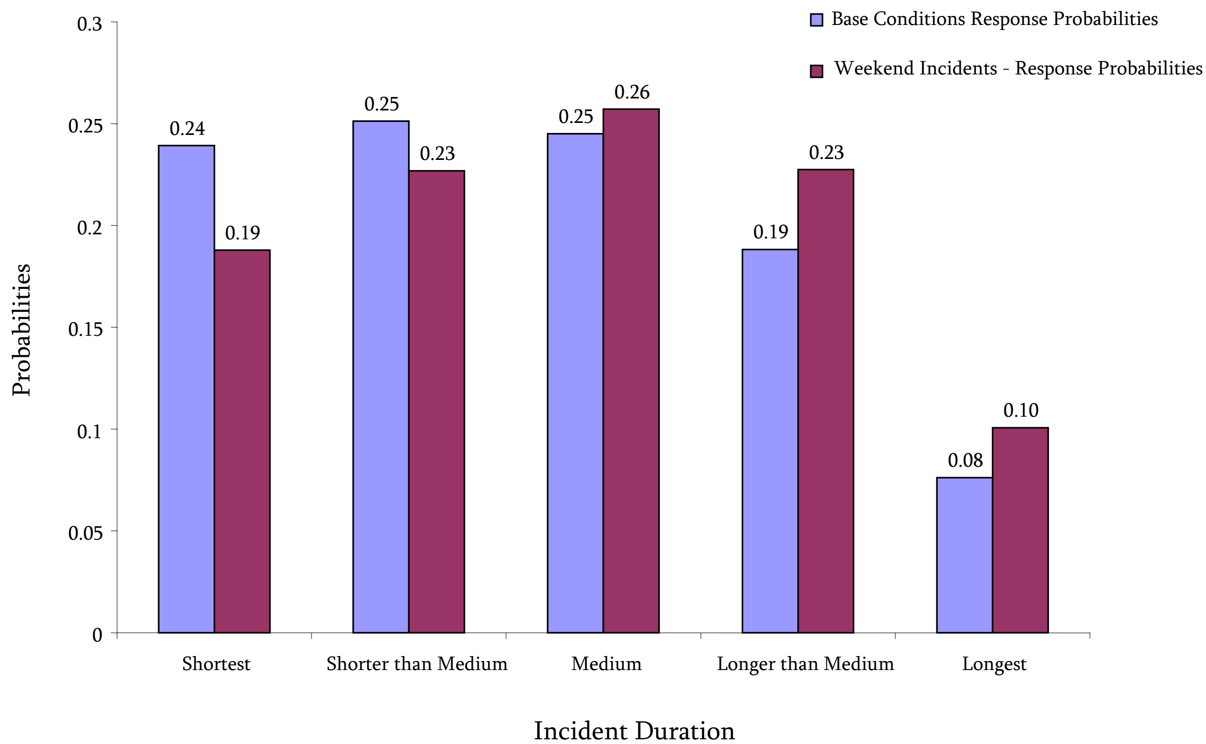

Figure 5.4: Response Probabilities for Incidents During 'Weekend' Compared to Base Response Probabilities

5.6.3 Peak/Non-Peak

The coefficients associated with both 'Morning Peak' and 'Afternoon Peak' are positive (Refer Table 5.11) indicating that during peak period the proportion of shorter incidents are more than longer ones compared to non-peak times. As evident from the corresponding p-values, only the coefficient associated with 'Afternoon Peak' is statistically significant at 5% confidence level. Hence, no comment can be made about the effect on incident durations during morning peak hours other than the fact that the data does not support the claim that duration of incidents during the 'Morning Peak' are different from those during non-peak hours.

This behavior can be explained from the fact that there is more urgency to clear the incident during peak hours than in non-peak hours. Another interesting observation is that the coefficient associated with afternoon peaks is much greater than that of morning peaks, indicating that incidents during afternoon peaks are cleared much faster than incidents happening during morning peaks.

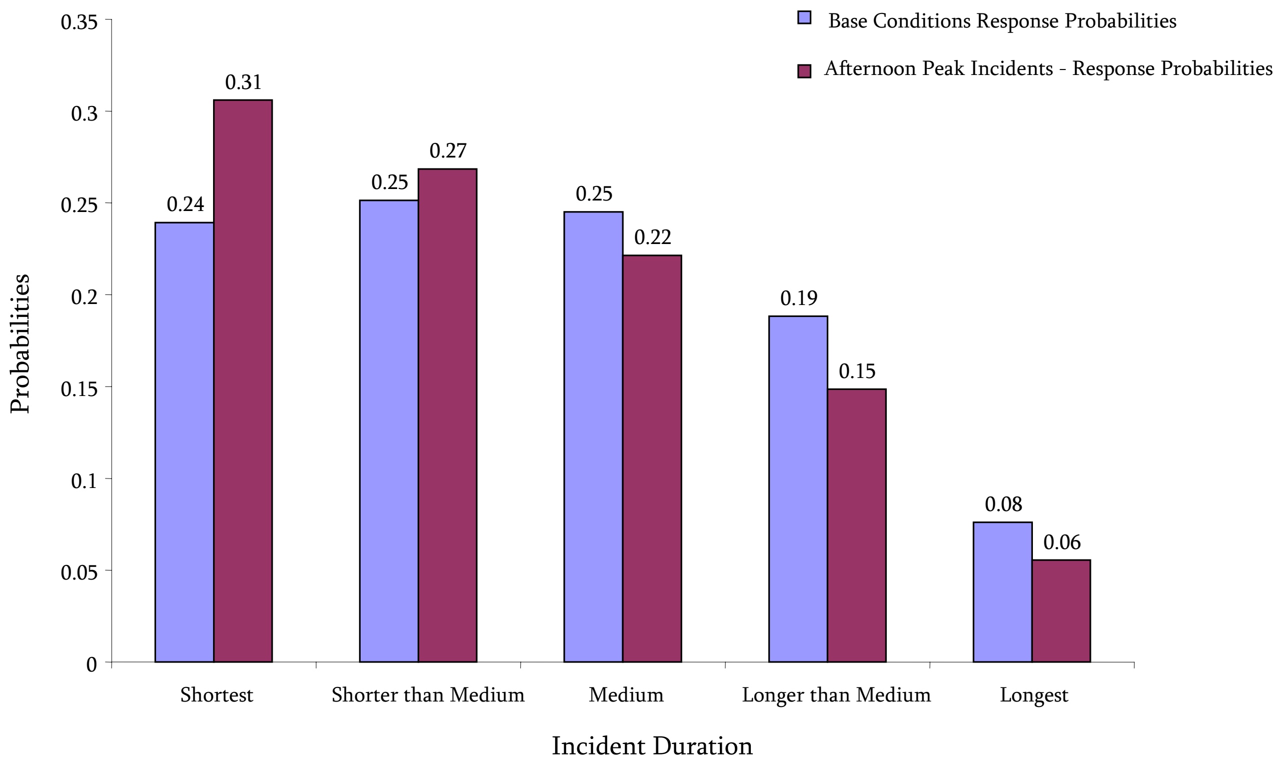

Figure 5.5: Response Probabilities for Incidents During 'Afternoon Peak' Compared to Base Response Probabilities

5.6.4 Detected-By

The coefficient associated with 'HELP' is positive, whereas those associated with 'METRO' and 'Other Caller' are negative. This implies that those incidents reported by 'HELP' have higher proportions of shorter durations (are cleared faster) than those incidents reported by an 'ITS Operator" (the base case). The opposite applies to those incidents reported by 'METRO' or by 'Other Callers'. In this case the coefficients associated with 'METRO' and 'Other Callers' are not significant at 5% confidence level which indicates that the data does not support the aforementioned behavior.

This behavior is only logical because one of the prime services that assist in incidence clearance is the 'HELP'- Tennessee's freeway service patrol (FSP), and hence those incidents detected and reported by 'HELP' will obviously be shorter in durations than those incidents reported by any other agency; where the incident is notified by some other agency other than

HELP, the incident duration will also include the time to notify the FSP and the time for the FSP to reach the incident site.

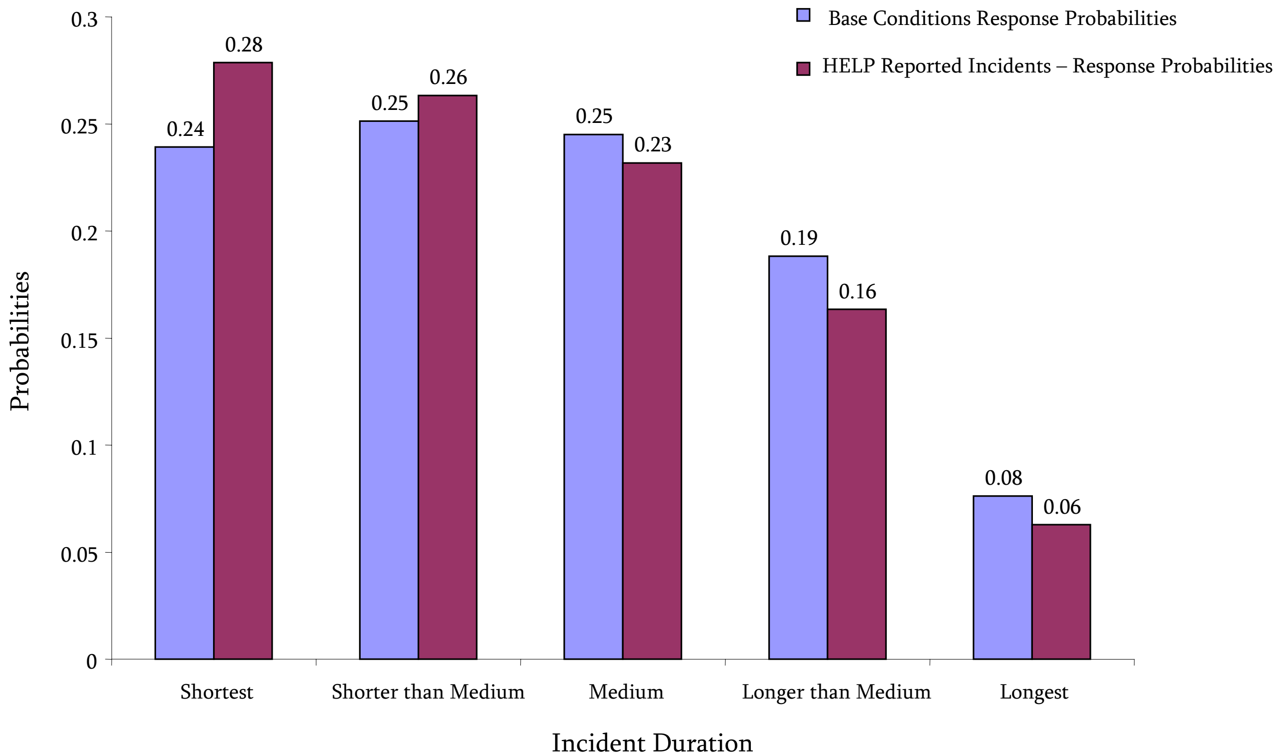

Figure 5.6: Response Probabilities for HELP Reported Incidents Compared to Base Response Probabilities

5.6.5 Number of Vehicles Involved

The coefficients associated with ‘Two’ and ‘Three’ are negative and are both significant at 5% confidence level. This implies that incidents involving more than one vehicle are probable to take more time to get cleared, which is expected. Another worthy observation is that the magnitude of the coefficient associated with ‘Three’ is greater in magnitude than that of ‘Two’, implying that a three vehicle incident is prone to be more time consuming to clear than a two vehicle incident which again is as expected.

The factor class ‘Zero’ denotes those incidents which do not involve a vehicle; like ‘Debris on the road’, ‘Spillage’ etc. This factor class also has a negative coefficient which is statistically significant at 5% confidence level. The magnitude of the coefficient of this factor class is greater than that of the factor class 'Two' indicating that the incidents belonging to this class generally takes more time to clear than an incident involving one or two vehicles. This is probably due to the reason that incidents involving spillage, debris requires services of agencies like HAZMAT, roadway cleaning which can increase the clearance times as these services are typically dispatch-on-order services rather than continuous patrolling services like the FSPs.

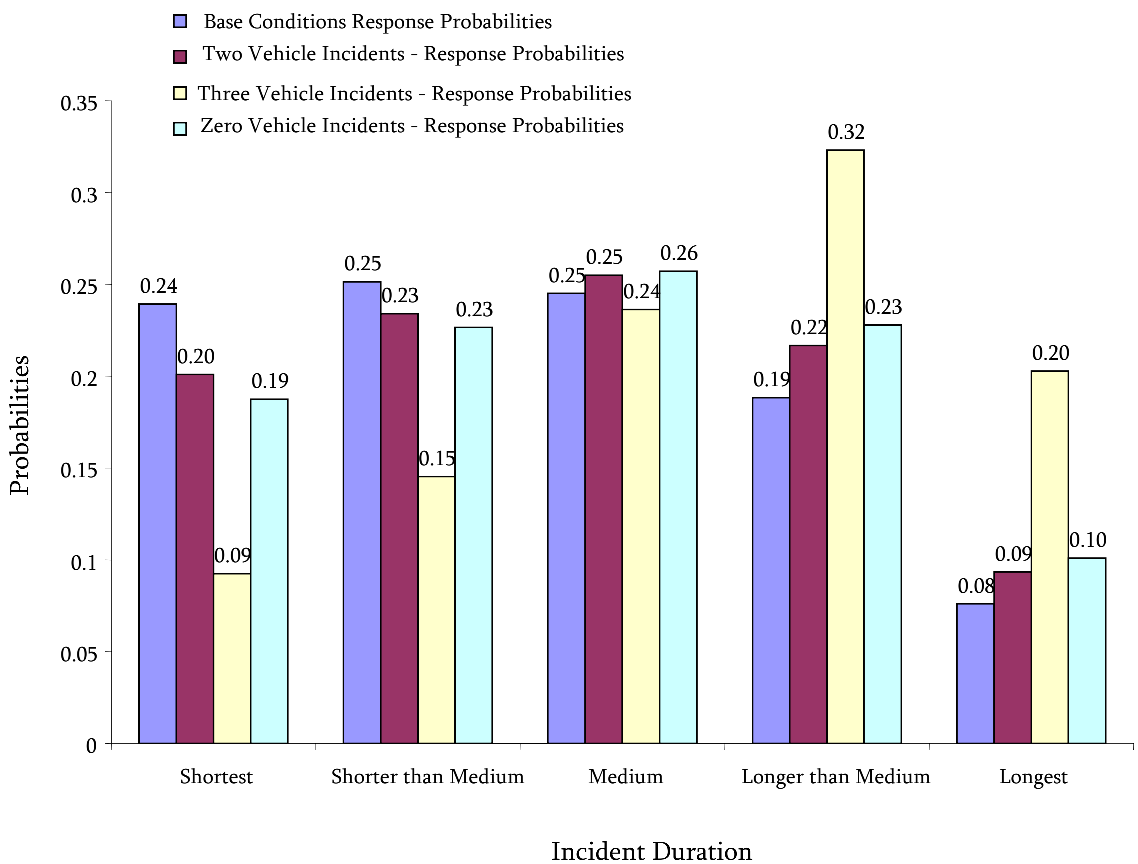

Figure 5.7: Response Probabilities for 'Number of Vehicles Involved' Compared to Base Response Probabilities

5.6.6 Roadwork Presence

The coefficient for the factor class denoting roadwork is negative, indicating that when an incident happens near a roadwork site the odds are that such incidents will tend to be longer than another incident which happens away from a roadwork site. The p-value also suggests that the coefficient is significant at a 5% confidence level.

This behavior can be explained by the congested nature of a roadwork site, which can hamper the speedy access of response services and hence lead to longer clearance times.

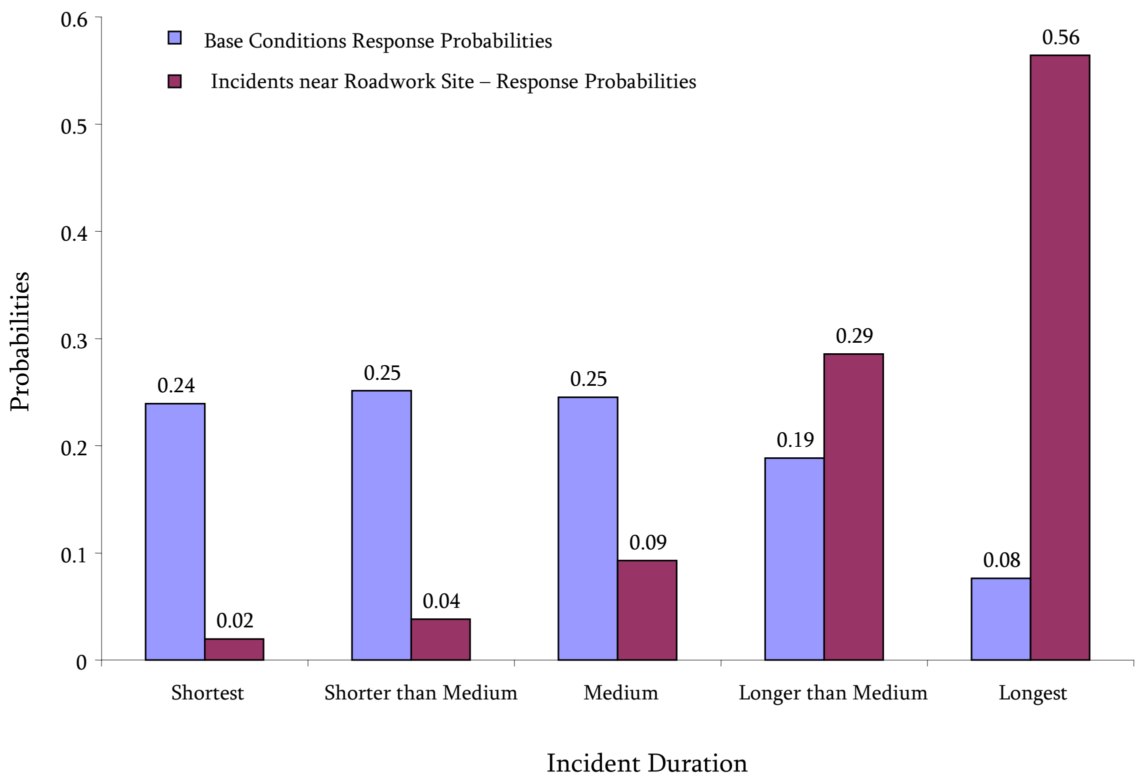

Figure 5.8: Response Probabilities for Incidents at Roadwork Sites Compared to Base Response Probabilities

5.6.7 Truck Presence

The coefficient for the factor class denoting truck involvement in an incident is negative, indicating that an incident involving a truck takes more time to be cleared than otherwise. The p-value also suggests that the coefficient is significant at a 5% confidence level.

This is as expected because the severity of an incident involving a truck is typically worse than an incident involving no trucks. The personnel and equipments needed to clear such an incident involving trucks is also typically more than when no trucks are involved.

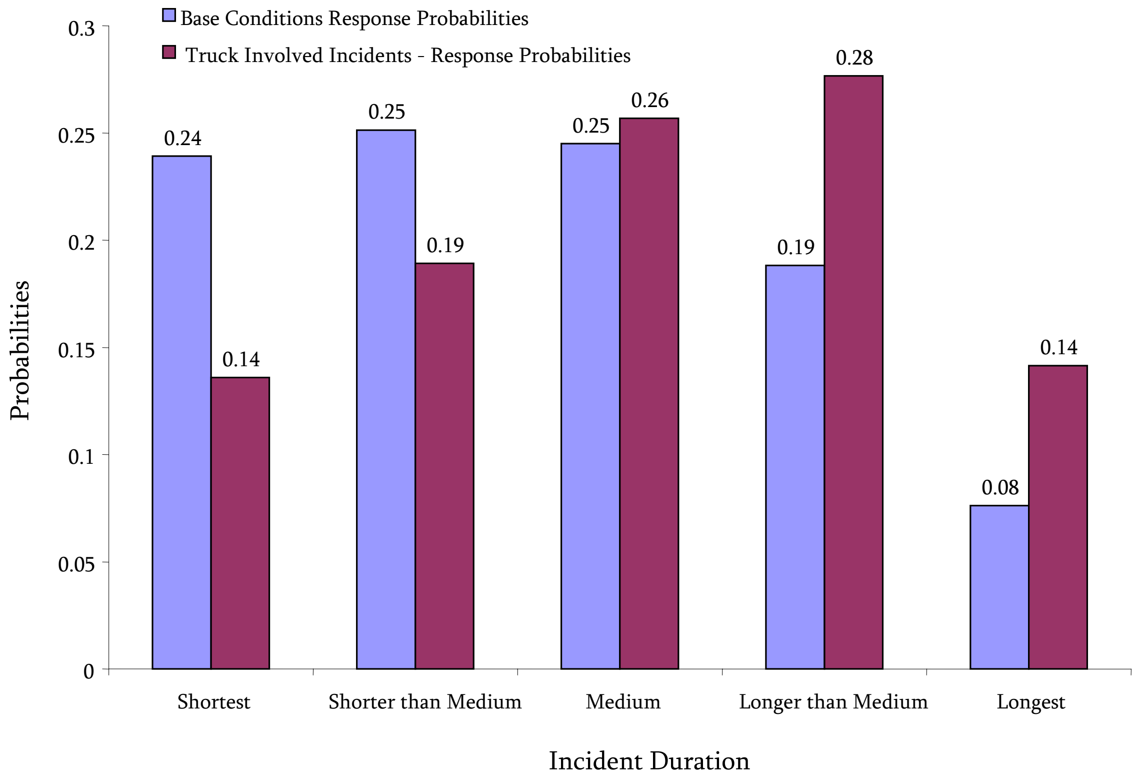

Figure 5.9: Response Probabilities for Incidents Involving Trucks Compared to Base Response Probabilities

5.6.8 Weather Condition

All coefficients for factor classes other than ‘Clear’ (which is the base condition) are negative showing that incidents tend to take longer to clear in case of adverse weather. Although the p-value associated with each of the levels shows that the coefficients are not significant enough to support this claim. Hence the conclusion is that the data does not provide enough proof to state that adverse weather prolongs an incident any more than an incident that happened during normal weather.

5.6.9 Camera-Coverage

The coefficients associated with both ‘Good Camera-Coverage’ and ‘Very Good Camera-Coverage’ camera-coverage are positive demonstrating that better camera coverage does decrease the proportion of longer incidents. Also remarkable is the fact that the coefficient of ‘Very Good Camera-Coverage’ coverage is greater than that for ‘Good Camera-Coverage’, denoting that the better the coverage the greater the odds of incidents being shorter in duration. For the factor class 'Good Camera-Coverage' the coefficient is not significant at 5% confidence level and hence the data does not seem to support such a claim.

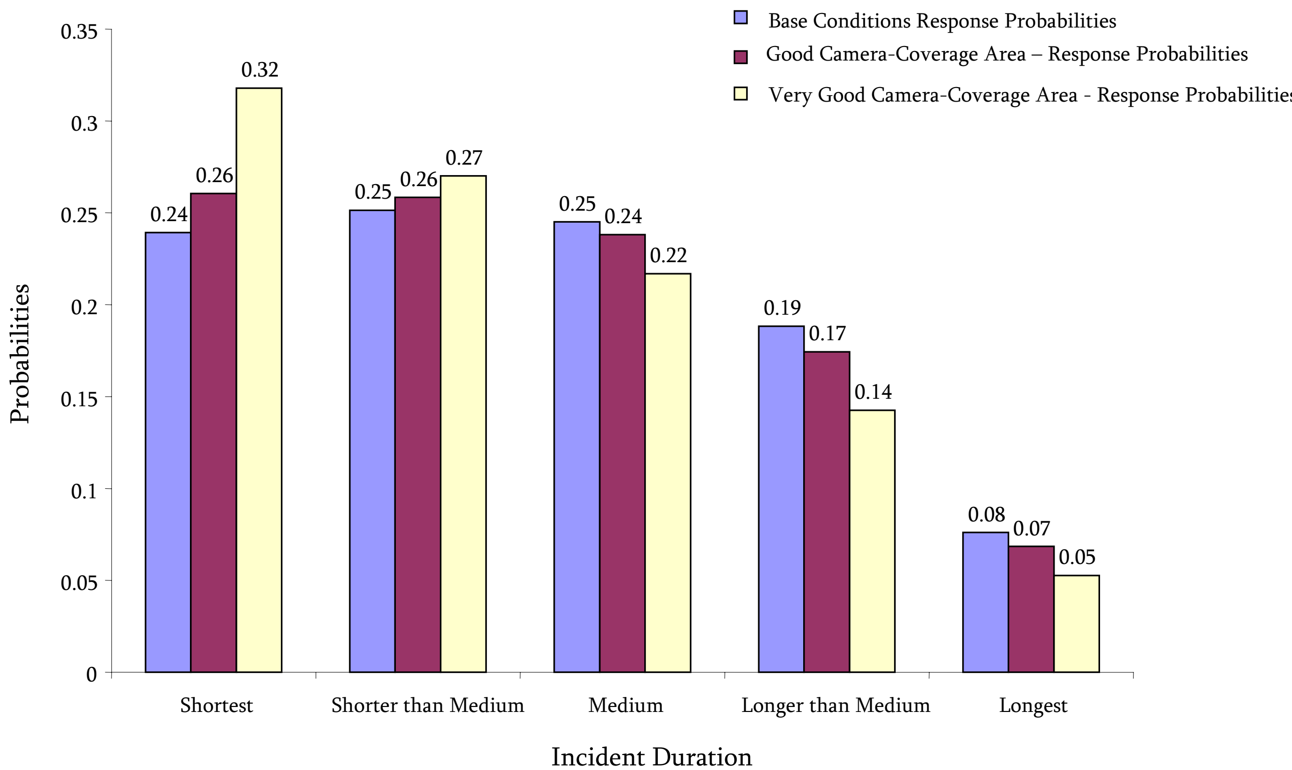

Figure 5.10: Response Probabilities for 'Camera-Coverage' compared to Base Response Probabilities

5.7 Comparing the Effects of the Various Predictors

The relative effects of the various predictors on the response probabilities can be assessed by comparing the odds ratio corresponding to the factors of those predictors. The more the odds ratio deviates from unity, the greater is the impact of that factor of that predictor variable in affecting the response probabilities. Odds ratio less than one indicates greater proportions of incidents having high clearance times, in other words the corresponding predictor variable factor causes longer clearance times. Similarly an odds ratio greater than one denotes greater proportions of incidents having low clearance times due to the corresponding predictor variable factor.

For example, the odds ratio corresponding to the predictor variable factor ‘weekend’ is 0.74 means that proportion of incidents having longer clearance times during weekends is higher than proportion of incidents having longer clearance times during weekdays. In other words the probability of an incident having higher clearance times is more on weekends than on weekdays. In the case of the predictor variable factor ‘Afternoon Peak’, the odds ratio is 1.40 indicating that proportion of incidents having longer clearance times during an afternoon peak hour is lesser than proportion of incidents having longer clearance times during any other time of the day. In other words the probability of an incident having higher clearance times is less during afternoon peak hours compared to any other time of the day.

Based on the odds ratio values from the Table 5.11 the following inferences can be made:

› Odds ratios corresponding to predictor variable factors ‘Morning Peak’, ‘Afternoon Peak’, ‘HELP’, ‘Good Camera-Coverage’, and ‘Very Good Camera-Coverage’ are greater than unity indicating that under these conditions incidents are cleared much faster than when other conditions prevail.

› Odds ratio corresponding to ‘Very Good Camera-Coverage’ is the highest which indicates that this condition has the highest effect on reducing incident clearance times.

› Prominent among the incident conditions which increase clearance times are three vehicle involved incidents, incidents involving trucks, incidents which happen at roadwork sites, and incidents happening during foggy conditions.

› The odds ratio associated with ‘Roadwork-Presence’ is remarkably lower than all other values. Though roadwork presence at an incident site can increase the clearance time, this value is probably due to some recording errors in the database.

5.8 Validity of the Model

This section deals with the statistical significance of the model, given the data. In the following sub-sections issues like multicollinearity of the predictor variables and goodness-of-fit statistics for the model are discussed.

5.8.1 Multicollinearity of Predictor Variables

The following is the correlation table for the predictor variables used in the model. The variables with nominal categorical levels are excluded because application of the correlation coefficient is inappropriate when the data is clearly nominal categorical with more than two levels. (Stockburger, W David., *Introductory Statistics: Concepts, Models and Applications*).

Table 5.12 demonstrates that none of the variables have significant correlation with any other variable and hence the model can be safely assumed to be devoid of any issues related to multicollinearity of predictor variables.

Table 5.12: Correlation Table for the Prediction Variables

	Weekday/Weekend	Roadwork-Presence	Truck-Presence
Weekday/Weekend	1.00		
Roadwork-Presence	0.088	1	
Truck-Presence	-0.046	-0.158	1

5.8.2 Goodness-of-fit Test and Measures of Association

The goodness-of-fit test based on Deviance residuals was computed with the help of *MINITAB*®. The associated p-value of this test is 0.371, which means that at a confidence level of 5%, there is insufficient data to conclude that the model does not fit the data adequately. *MINITAB*® also calculates 69.7 percent predictive-ability of the model. With these two measures the model can be assumed to be a satisfactory fit for the given data.

CHAPTER VI

SECONDARY INCIDENT CAUSATION MODEL

In this chapter a secondary incident causation model based on binary logistic regression is developed, for estimating the probability that an incident induces a secondary incident. The concepts involved in binary logistic regression are discussed in the first section. The following sections present the steps involved in developing the model and the analysis involved. The chapter is concluded by presenting the results and inferences.

6.1 Introduction

Binary logistic regression as the name suggests, is used when the response variable is dichotomous in nature. For the secondary incident causation model, the response variable is simply a yes/no (causing/not causing a secondary incident). The general form of a binary logistic regression model is as follows:

$$\boxed{P(event) = \frac{e^{g(x)}}{1 + e^{g(x)}}} \tag{6.1}$$

where:

P (event) denotes the probability of the *event* happening,

g(x) is the logit link function (which is a linear combination of the predictor variables)

$$\boxed{g(x) = \beta_0 + x'\beta} \tag{6.2}$$

β_0 is the regression constant,

β is the regression coefficients associated with the predictor variables,

x denotes the vector of predictor variables.

The regression constant and coefficients are obtained from *MINITAB*® which uses an equivalent maximum likelihood method. The probability of the event occurrence can then be computed using the equation 6.1. Also the relative effects of the individual predictor variables can be inferred from the regression coefficients associated with them.

6.2 Response Variable

6.2.1 Description

The response variable for this model is coded as a yes/no, indicating whether or not an incident has caused a secondary incident. Though the database from Nashville Transportation Management Center (TMC) has a field indicating whether or not a particular incident is secondary, there is no indication towards the primary incident which has caused a particular secondary incident. Also the percentage of incidents registered as secondary incidents (2.1%) are too low compared to average national levels observed (15% - 20%), which is a strong indication that only very few secondary incidents are actually recorded. Hence in this research, secondary incidents and secondary-causing incidents are identified by examining the database using a search algorithm coded using the programming language *JAVA*™.

6.2.2 Secondary Incident Identification

As discussed in Chapter 2, identifying a secondary incident and finding the primary incident which can be linked to this secondary incident is accomplished by space-time based search on the incident database. The two parameters associated with the primary incident to be defined for this search are the 'time of effect' and 'spatial influence'. Time of effect of an incident is the time for which the effects of that incident can be felt on the regional traffic flow. Similarly 'spatial influence' of an incident refers to the incident neighborhood where the effect of that incident is felt. This influence is denoted by the upstream linear distance in the same direction (due to queues and shock-waves) or the downstream linear distance in the opposite direction (due to rubbernecking) along the route of the primary incident. Both these parameters are based on available literature and regional experiences. In this research the time of influence of an incident for both upstream and downstream travel is considered to be the same and is assumed to be the total incident clearance time, subject to a minimum of two hours. No distinction is made for upstream and downstream travel because the influence of an incident on the regional travel is bound to remain until the incident is cleared. On the contrary, the spatial influence of an incident is felt for a longer distance in the upstream side on the same direction of route than in the downstream side on the opposite direction of the same route. This is because, in the former case (upstream side on the same direction of route) the secondary incident is caused due to the formation of queues and shock waves which can extent for longer distances (as long as two to three miles) than the phenomenon of rubbernecking which is the cause of secondary incidents in the downstream side on the opposite direction of the same route. For rubbernecking to happen the drivers on the opposite direction of travel have to see the incident site and hence will only happen from short distances, like within half a mile. The probability of a secondary incident happening downstream of an incident on the same direction or upstream of the incident on the opposite direction is much less and hence is not considered in this research. However the spatial influence limits are relaxed by 0.2 miles to accommodate any recording errors on the mile-marker field because any location on a freeway can be referenced by the mile-marker ahead or behind that incident. Based on these assumptions, a pair of incidents is termed as 'secondary-causing' and 'secondary' if they concur to one of the following criteria:

i. Both the incidents take place along the same route and direction of travel and the second incident occurs within 2 miles upstream of the first and no later than 2 hours after the first incident.
ii. Both the incidents take place along the same route but in opposite directions of travel and the second incident occur within 0.5 miles downstream of the first and no later than 2 hours after the first incident.

Note: In a case where the first incident clearance time is more than 2 hours, the time interval within which a secondary incident may occur is the clearance time itself instead of 2 hours as stated in the above criteria.

Based on the above search criteria, the database is found to consist of 15.2% (508 out of 3347) secondary incidents and 11.2 (375 out of 3347) secondary-causing incidents. Also interesting to note is that almost 21% of these secondary-causing incidents cause multiple secondary incidents.

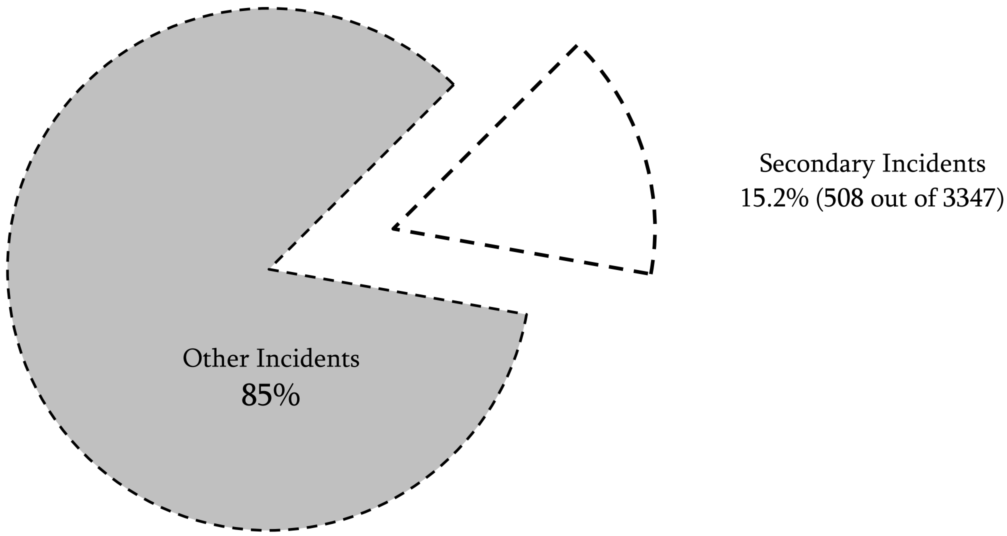

Figure 6.1: Pie-Chart Illustrating Percentages of Secondary Incidents

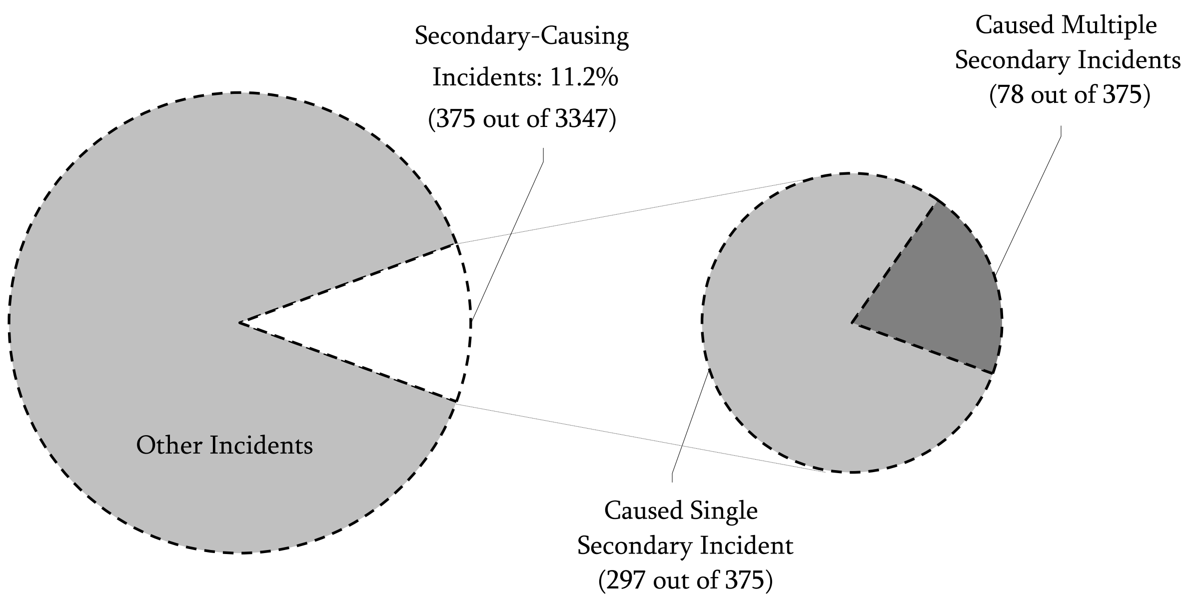

Figure 6.2: Pie-Chart Illustrating Percentages of Secondary-Causing Incidents

6.3 Predictor Variables

This model attempts to estimate the probability of an incident causing a secondary incident based on the incident characteristics. Several incident descriptors like clearance time, weekday/weekend, time of the day, type of vehicle involved, etc. could be studied to determine the effects on the probability of secondary crash occurrences. Among this the most important factor is the incident clearance time which denotes the exposure time of the incident (total time for which the regional traffic is exposed to the incident). As established in Chapter 5, incident descriptors such as weekday/weekend, time of the day and type of vehicle involved have a substantial effect on incident clearance time. Hence including clearance time and the other incident descriptors as predictor variables in the same model would lead to deceptive results. Therefore the only predictor variable used in this model is incident clearance time.

In contrast to the incident duration model, incident clearance time is modeled as a continuous variable. As a preliminary step, the relationship between the secondary incident causation probability and incident clearance time is investigated by comparing the median incident durations of those incidents which caused a secondary incident with those which did not cause a secondary incident. The results of the same are presented in the table below:

Table 6.1: Median and Mean Clearance Times for Secondary Incident Causing Incidents

	Median Incident Clearance Time (minutes)	Mean Incident Clearance Time (minutes)
All Incidents	19.0	77.0
Secondary Incident Causing Incidents	23.0	155.9
Multiple Secondary Incident Causing Incidents	36.5	247.6

6.4 Results of the Logistic Regression

The results of the binary logistic regression are summarized as follows:

Table 6.2: Binary Logistic Regression Results for the Secondary Incident Causation Model Using the Continuous Incident Duration Variable

Predictor	Regression-Coefficient (β)	P-Value	Odds-Ratio
Const -- (β_0)	-2.183	0.00	
Incident Duration	0.0012	0.00	~1.00

The regression coefficient associated with incident duration is 0.0012 and is significant at 5% confidence level as shown by the computed p-value. Though the coefficient being greater than zero demonstrates a positive dependence of secondary causation probability on incident duration, the relationship is very weak as shown by the computed odds-ratio. The odd-ratio of 1.00 implies that given the data, there is no appreciable change in secondary incident causation probabilities as incident duration varies.

As this result is not expected from a logical stand point, a further analysis was conducted using a categorical incident duration variable coded as: shortest, shorter than medium, medium, longer than medium and longest. The reader is directed to section 5.3 for further details on how the categorical coding is carried out. The results based on this variable are presented as follows:

Table 6.3: Binary Logistic Regression Results for the Secondary Incident Causation Model Using the Continuous Incident Duration Variable

Predictor	Incident Duration Level	Regression-Coefficient (β)	p-Value	Odds-Ratio
Const -- (β_0)		-2.098	0.000	
Incident Duration Reference Level: Medium				
	Shortest	-0.052	0.767	0.950
	Shorter	-0.129	0.466	0.880
	Longer	-0.125	0.477	0.880
	Longest	0.414	0.012	1.510

As shown in the above table the regression coefficient associated with the incident duration class 'longest' (which are incidents with duration greater than 66.9 minutes) is the only one significant at a 5% confidence level. Also the associated odds-ratio suggests that those incidents belonging to the incident-duration class 'longest' has almost 50% more chance of causing a secondary incident than an incident belonging to the incident duration class 'medium'.

6.5 Final Comment

A secondary incident causation model based on binary logistic regression was presented in this chapter. Two cases where the incident duration is modeled as a continuous variable and as a categorical variable were explored. In the case where a continuous incident duration variable is used, the model predicts that incident duration does not have a significant effect on changing secondary incident causation probabilities. On the other hand, when incident duration is modeled as a categorical variable, the regression model predicts that those incidents belonging to the incident duration class 'longest' (duration greater than 66.9 minutes) have almost 50% more chance of causing a secondary incident than an incident belonging to the incident duration class 'medium'. Hence the data suggests that the secondary incident causation probabilities are affected by incident duration only if the incident belongs to the level 'longest' among the incident duration classes.

CHAPTER VII

CONCLUSIONS

7.1 Overview

This study has examined two main incident characteristics, namely incident duration and secondary incident probability. Archived incident data from the Nashville Traffic Management Center was used to develop statistical models to study these characteristics.

Based on the review of existing literature, this study recognizes the need for developing new models relaxing certain restrictive assumptions imposed in many previous studies. In this research the relationship between incident characteristics and its duration and probability to cause a secondary incident is not assumed to be linear. The possible non-linear relationship is accounted for by employing logistic regression rather than linear regression to develop the models.

Most of the existing literature on secondary incidents deals with just secondary crashes (or secondary accidents) and their corresponding occurrence probabilities. But from a traffic operations standpoint all secondary incidents are of concern. Hence in this research the secondary incident causation model developed investigates the influence of primary incident characteristics on causing secondary incidents and not just secondary crashes. Several incident features like descriptive, spatial, temporal and environmental were investigated for plausible relationships with the duration of the incident and secondary incident causation probability. Mathematical computations required for model building were carried out using the statistical software *MINITAB*^{TM}.

These two new models provide a framework to make reasonable judgments about both an incident's duration and its probability to cause a secondary incident. The models predict these two properties by using the incident characteristics which are available to a user at the time of incident detection. This information can be of help for improved freeway incident management and decision making.

7.1.1 Incident Duration Model

The incident duration model investigates the relationship between freeway incident clearance times and the following incident characteristics:

- Weekday/Weekend
- Peak/Non-Peak
- Detected-By
- Number of Vehicles Involved
- Roadwork-Presence
- Truck-Presence
- Weather-Condition
- Camera-Coverage

The response variable (i.e., incident clearance time) was coded into five different duration classes and the influence of the aforementioned characteristics in causing an incident to fall into each of these five duration classes was estimated based on the logistic regression model. The extent of influence of the different incident characteristics were established by examining the odds-ratios corresponding to each of these incident characteristics. This model revealed the importance of several incident management utilities (i.e., closed circuit televisions, freeway service patrols) and the effect of many incident characteristics in affecting freeway incident clearance times.

7.1.2 Secondary Incident Causation Model

In the secondary incident causation model, the probability of an incident causing a secondary incident is investigated. This study recognizes incident exposure time (time for which the regional environment is exposed to the incident or the incident site) as the most important factor influencing secondary incident occurrences. Incident clearance time (used as a surrogate for incident duration) is the sole explanatory variable used in this model. All other incident characteristics are shown (in the incident duration model) to considerably influence incident clearance time and are hence excluded from this model (because including both incident clearance time and other incident characteristics as predictor variables can lead to unreliable results).

Two scenarios, where the incident clearance time was modeled as a continuous variable and as a categorical variable, were explored. The model did not predict any significant influence on secondary incident causation probability when incident clearance time is modeled as a continuous variable. On the contrary when incident clearance time is modeled as a categorical variable, the regression model predicted incidents with clearance time greater than 66.9 minutes have almost 50% more chance of causing a secondary incident than an incident with medium clearance time. Thus the model revealed a substantial influence of incident clearance time on secondary incident causation probabilities for very high values of incident clearance times.

7.2 Findings and Conclusions

Based on the analyses and modeling results, some of the salient findings and conclusions are presented below:

- Incidents during weekdays and peak hours (especially afternoon peak hours) are shown to have smaller clearance times compared to incidents happening during weekends and non-peak hours respectively. Although this is perceivably because of the urgency in clearing incident sites during weekdays and peak hours, this finding shows the higher efficacy of the several incident clearance services during these prime hours.
- Incidents happening at sites well monitored by closed circuit cameras are shown to have smaller clearance times than the incidents at other sites. There is a possibility that some minor incidents (with lower clearance time) happening in areas which are scarcely monitored by closed circuit cameras are cleared without being recorded in the TMC database. But the number of such unrecorded incidents is less in number compared to the total number of incidents. Hence this is assumed not to bias the claim that good camera coverage aids in quick incident clearance and incident site recovery.
- The odds-ratio value corresponding to the predictor variable factor 'Very Good Camera-Coverage' is significantly larger than the odds-ratio value of the predictor variable factor 'Good Camera-Coverage'. This supports the claim that good camera

coverage aids in quick incident clearance. This finding although perhaps not surprising is very pronounced and encourages installing closed circuit cameras in areas experiencing very high incident clearance times.

› The study reveals that freeway service patrols (FSPs) play a very significant role in reducing incident clearance times. Those incidents which are detected by FSPs are cleared faster than those incidents which are detected by other means.

› Incidents involving a greater number of vehicles were shown to have higher clearance times. This can justify having improved incident clearance services (like more FSPs) so that multiple types of services can attend such incidents, especially if the frequency of such incidents involving two or more vehicles are high (based on the available historical data).

› Roadwork presence at an incident site is shown to delay incident clearance services. This finding indicates that strategies like work zone management can have significant impacts on speedy incident clearance procedures.

› The study-indicated very strong relationship between roadwork presence at an incident site and incident clearance times can in part be also due to some database recording errors. Some of the recorded roadwork-involving incidents may not be incidents in the strict sense but just roadwork schedules or construction works which were misinterpreted as incidents. This is evident from the fact that there are several incidents (almost 25 in number) of clearance time greater than 1000 minutes which are recorded as roadwork-involving incidents. A large share of these can be just 'scheduled roadworks' and not incidents as such.

› The strong influence of truck involvement in causing higher clearance times indicates that advanced clearance strategies have to be adopted (like heavy duty clearance vehicles, cranes) and in a quicker manner in case of incidents involving trucks.

› Incidents having very high clearance times (e.g., an hour or more) are shown to exhibit high probabilities of secondary incident causation than incidents with median clearance times.

7.3 Recommendations and Directions for Future Research

This study has presented a useful framework for analyzing the influence of incident characteristics on incident clearance times and secondary incident causation probabilities. The models developed are totally based on the data collected from the Nashville TMC and hence are bound to have spatial and temporal limitations. They are unlikely to predict incident properties accurately for any city other than Nashville due to the differences in traffic behaviors (both on the user side and the service side). But the relationships discovered could be representative of those existing in other cities.

Further validation of these models with subsequent available data is very important to ensure acceptable predictions during future usage. Further analyses should be conducted to investigate the influence of traffic characteristics such as volume, speed, vehicular characteristics (% trucks). These models could be expanded to include predictor variables like different seasons, roadway geometry and highway infrastructure (e.g., bridges, median barriers). Comparative studies involving different locations investigating incident clearance times and secondary incident causation probabilities based on this modeling framework can be helpful in understanding site dependency of such studies. Another exciting research possibility is to investigate the causation of secondary incidents due to rubbernecking (especially the incidents happening on the opposite direction of travel) by modifying the secondary incident causation model presented here. These are some research directives which can be pursued to expand this modeling framework.

APPENDIX A

TMC Incident Database Records

APPENDIX A

Table A-1: TMC Incident Database Records

Field Name	Description	% Records Filled
1. IncidentIDNumber	Incident Identification Number	100
2. IncidentDescription	Short Description of the Incident	22.7
3. Operators	Name of the Operator on duty	98.7
4. StartDate	Start date of incident	100
5. StartTime	Start time of incident	100
6. ClearDate	Travel lane clear date	100
7. ClearTime	Travel lane clear time	~100
8. NormalDate	Queue clear date	69.9
9. NormalTime	Queue clear time	61.9
10. ShldrDate	Shoulder clear date	12.8
11. ShldrTime	Shoulder clear time	13.3
12. HelpTime	Help arrival date	59.2
13. HelpDate	Help arrival time	59.1
14. LawEnforceTime	Law enforcement arrival date	38.2
15. LawEnforceDate	Law enforcement arrival time	38.0
16. FireTime	Fire arrival date	14.9
17. FireDate	Fire arrival time	14.8
18. EMSTime	EMS arrival date	12.3
19. EMSDate	EMS arrival time	12.3
20. TowTime	Tow truck arrival date	11.2
21. TowDate	Tow truck arrival time	11.1
22. MaintTime	Maintenance arrival date	2.65

Table A-1, continued

23. MaintDate	Maintenance arrival time	2.65
24. NotificationTime	The time Metro, Fire, Media, etc. were notified.	11.9
25. RecvBy	Incident reported by	97.7
26. ConfirmBy	Incident confirmed by	97.1
27. Route	Route and direction	~100
28. City	Town or city	~100
29. LocationArticle	At, before, or Past the Location	77.0
30. Location	Nearest exit/cross street	~100
31. Milemarker	Closest mile marker	~100
32. TotalLanes	Total number of lanes on route/direction	~100
33. ClosedLanes	Number of lanes closed	~100
34. LanesClosed	Description of lanes that are blocked	83.9
35. LtShldrClosed	Left shoulder blocked time	12.2
36. LtShldrOpen	Left shoulder open time	10.8
37. 1LaneClosed	Lane 1 blocked time	36.8
38. 1LaneOpen	Lane 1 open time	35.8
39. 2LaneClosed	Lane 2 blocked time	25.6
40. 2LaneOpen	Lane 2 open time	24.9
41. 3LaneClosed	Lane 3 blocked time	35.4
42. 3LaneOpen	Lane 3 open time	34.5
43. 4LaneClosed	Lane 4 blocked time	11.8
44. 4LaneOpen	Lane 4 open time	11.3
45. 5LaneClosed	Lane 5 blocked time	Not used at all
46. 5LaneOpen	Lane 5 open time	Not used at all

Table A-1, continued

47. RtShldrClosed	Right shoulder blocked time	16.1
48. RtShldrOpen	Right shoulder open time	12.1
49. RampClosed	Ramp blocked time	Not used at all
50. RampOpen	Ramp open time	Not used at all
51. TotalDMS	Number of DMSs used	~100
52. CauseType	Cause of closure	100
53. VehicleNumber	Number of vehicles involved	100
54. TruckID	Truck involved indicator	N/A
55. Hazmat	Hazmat called?	N/A
56. SuperNotified	Supervisor notified?	N/A
57. ConstZone	Construction zone?	N/A
58. WeatherCond	Weather conditions	98.6
59. SecondaryIncident	Was this a secondary incident?	N/A
60. HAR	Was HAR used?	N/A

BIBLIOGRAPHY

1. Agrawal, N. (1995) Incident management: An integrated systems approach. M.S. Thesis. Vanderbilt University.

2. Bertini, R.L. and McGill, G.E. (2002) Analysis of Oregon's rural incident response program using archived intelligent transportation systems data. *Ninth World Congress on Intelligent Transport Systems*, Chicago, Illinois.

3. Bickerstaff, R. (2003) Measuring the effect of a freeway service patrol on non-recurrent congestion due to crash events. M.S. Thesis. Vanderbilt University.

4. Chang, G.L., Liu, Y., Lin, P.W. and Zou, N. (2003) Performance evaluation of CHART: Year 2002.

5. *CHART Incident response evaluation final report by COMSIS Corporation*, 1996.

6. Chayanan, S. (2000) Incident management: Secondary crash characteristics. M.S. Thesis. Vanderbilt University.

7. Federal Highway Administration. (2000) *Incident Management Successful Practices: A Cross-Cutting Study.* http:// ww.itsdocs.fhwa.dot.gov/jpodocs/repts_te/8V001!.pdf

8. Federal Highway Administration. (2000) Traffic Incident Management Handbook. http://www.itsdocs.fhwa.dot.gov//jpodocs/rept_mis /@9201!.pdf

9. Federal Highway Administration. (2001) *Regional Traffic Incident Management Programs: Implementation Guide*. http://www.itsdocs.fhwa.dot.gov/jpodocs/repts_te/@5901!.pdf

10. Garib, A., Radwan, A. E. and Al-Deek, H. (1996) Estimating Magnitude and Duration of Incident Delays. *Journal of Transportation Engineering*, **123**, 459-466.

11. Giuliano, G. (1989) Incident Characteristics, Frequency, and Duration on a High Volume Urban Freeway. *Transportation Research Part A*, **23**, 387-396.

12. Green, W.H. (2000) *Econometric Analysis* (4^{th} ed.). Prentice Hall, New Jersey.

13. Hosmer, D.W. and Lemeshow, S. (1989) *Applied Logistic Regression*. John Wiley & Sons, New York.

14. Karlaftis, M.G., Latoski, S.P., Richards, N.J. and Sinha, K.C. (1999) ITS impacts on safety and traffic management: An investigation of secondary crash causes. *ITS Journal*, **5**, 39-52.

15. Khattak, A. J., Joseph L. S. and Wang. M. H. (1995) A Simple Time Sequential Procedure for Predicting Freeway Incident Duration. *IVHS Journal*, **2**, 113-138.

16. Latoski, S.P., Pal, R. and Sinha, K.C. (2000) Cost-effectiveness evaluation of Hoosier helper freeway service patrol. *Journal of Transportation Engineering*, **125**, 429-438.

17. Lomax, T. and Schrank, D. (2002) The *2004 Urban Mobility Report*. Texas Transportation Institute.

18. Minitab Inc. (2003). MINITAB Statistical Software, Release 14 for Windows, State College, Pennsylvania.

19. Mobility 2020, Congestion Mitigation Strategies.
http://www.dfwinfo.com/trans/mtp/previous/mobility2020/congestion.html

20. Moore, J.E., Giuliano, G. and Cho, S. (2004) Secondary accident rates on Los Angeles freeways. *Journal of Transportation* Engineering, **130**, 280-285.

21. Nam, D. and Mannering, F. (2000) An exploratory hazard-based analysis of highway incident duration. *Transportation Research Part A*, **34**, 85-102.

22. *NIST/SEMATECH e-Handbook of Statistical Methods.*
http://www.itl.nist.gov/div898/handbook/

23. Office of Highway Policy Information, Federal Highway Administration, U.S. Department of Transportation. Highway Statistics 1992.
http://www.fhwa.dot.gov/ohim/hs92/hs92page.htm

24. Office of Highway Policy Information, Federal Highway Administration, U.S. Department of Transportation. Highway Statistics 2002.
http://www.fhwa.dot.gov/policy/ohim/hs02/index.htm

25. Office of Operations, Federal Highway Administration, U.S. Department of Transportation. *Congestion Mitigation: 21st Century Operations Using 21st Century Technologies*.
http://ops.fhwa.dot.gov/aboutus/one_pagers/cong_mitigation.pdf

26. Office of Operations, Federal Highway Administration, U.S. Department of Transportation. *Traffic Incident Management: 21st Century Operations Using 21st Century Technologies*.
http://ops.fhwa.dot.gov/aboutus/one_pagers/tim.pdf

27. Pampel, F.C. (2000) *Logistic Regression: A Primer*. Sage University Papers Series on Quantitative Applications in the Social Sciences, series no. 07-132. Thousand Oaks, CA: Sage.

28. Qi, Y. (2002) Incident delay analysis and estimation. PhD Dissertation. University of Virginia.

29. Raub, R.A. (1997) Secondary crashes: An important component of roadway incident management. *Transportation Quarterly*, **51**, 93-104.

30. Sinha, K.C. and Pal, R. (2002) Simulation model for evaluating and improving effectiveness of freeway service patrol programs. *Journal of Transportation* Engineering, **128**, 355-365.

31. Smith, B.L., Qin, L. and Venkatanarayana, R. (2003) Characterization of freeway capacity reduction resulting from traffic accidents. *Journal of Transportation* Engineering, **129**, 362-368.

32. Smith, K. and Smith, B.L. (2001) *Forecasting the clearance time of freeway accidents*. Center for Transportation Studies, University of Virginia.

33. Sreedevi, I. and Picado, R. (2001) *Incident Clearance* http://www.calccit.org/itsdecision/serv_and_tech/Incident_management/Incident_clearance/clearrep_print.htm

34. Stockburger, D. W. (1996) *Introductory Statistics: Concepts, Models, and Application s.* http://www.psychstat.smsu.edu/introbook/sbk00.htm

35. Subramaniam, R. (1995) Modeling incident durations for advanced traffic management systems. M.S. Thesis. Vanderbilt University.

36. Sullivan, E. (1997) New model for predicting freeway incidents and incident delays. *Journal of Transportation Engineering*, **123**, 267–275.

37. Transportation Research Board. (2004) *Measuring and Communicating the Effects of Traffic Incident Management Improvements*. National Cooperative Highway Research Program (NCHRP), Research Results Digest 289.

38. U.S. Department of Transportation. *Version 5.0 of the National ITS Architecture*. http://www.iteris.com/itsarch/