Modelling Traffic Accidents Using Duration Analysis Techniques: A Case Study of Abu Dhabi

SCHOOL OF CIVIL ENGINEERING AND GEOSCIENCES

NEWCASTLE UNIVERSITY

By

Abdulla Mohammed Saeed Al Kaabi

A thesis submitted for the degree of Doctor of Philosophy in Transport at the University of Newcastle upon Tyne

June 2013

Abstract

One of the main aims of Traffic Incident Management (TIM) is to reduce the duration of the disruption to traffic caused by an accident. Several approaches have been applied in the past in order to analyse and predict this. Incident duration can be broken down into four time intervals: reporting, response, clearance and recovery. Accurate models of each interval allow traffic controllers to deploy resources efficiently, thereby minimising an accident's effect on traffic flow and congestion. This may, in turn, lead to a reduction in other adverse impacts of traffic accidents such as air pollution, fuel consumption and secondary crashes.

A new approach to this problem, based on the accidents' characteristics, was developed using a fully parametric hazard based modelling technique to predict accident durations. The road network around Abu Dhabi, capital of the UAE, was used as a case study. Data was obtained from the UAE Federal Traffic Statistics System (FTSS) and the Abu Dhabi Serious Collision Investigation Section (ASCIS). These data included the start and end of each time interval, the total accident duration, temporal, geographical, environmental and other accident characteristics. To analyse the total duration, the analysis was conducted using three time intervals. Accordingly, fully parametric Accelerated Failure Time (AFT) models were created for the purpose of reporting time, response time, and clearance time (all urban roads) and response time (rural freeways), depending on the data available.

Analysis showed that the time intervals had different distributions. In addition, there was no similarity in the variable that affected each interval.

The results also revealed that weaknesses exist in the current practices of TIM in Abu Dhabi. The results of the analysis were used to create decision trees to aid traffic controllers with decisions regarding traffic diversion and disseminating traffic information to travellers.

Acknowledgements

First of all, I would like to express my sincere appreciation to my supervisors Dr Dilum Dissanayake, Lecturer in Transport Modelling, and Mr Roger Bird, Senior Lecturer in Highway Engineering, who offered genuine support with their helpful comments and feedback during all stages of my research and did not hesitate to provide any advice I required from them.

Secondly, a very special thanks is due to the members of the Transport Operations Research Group (TORG) for their constructive comments and suggestions concerning my work, especially Professor Margaret Bell, Science City Chair of Transport and Environment, Dr Anil Namdeo, Senior Lecturer in Transport and Sustainability, and Dr Neil Thorpe, Lecturer in Transport Studies.

I am also very grateful to Lt. General H.H Sheikh Saif bin Zayed Al Nahyan, Deputy Prime Minister and Minister of Interior in the United Arab Emirates, who gave me this opportunity to do my PhD in the United Kingdom. His motivations and incentives are the main source of any success achieved by any police officer work for the Ministry of Interior.

Thanks are also extended to Major General Nasser Lekhrebani Al Nuaimi, Secretary-General of the Office of HH Deputy Prime Minister and Minister of Interior. His direct and continuous supervision had a great impact on the completion of this research.

Above all, I would like to express my deepest appreciation to my parents, my wife and children, and my friends, who all stood beside me and supported me with their patience, knowledge and opinions during my study period.

Table of Contents Page

List of Figures Page

List of Tables Page

1 Introduction

1.1 Background

Many countries regard traffic incidents as being one of the high priority problems that needs to be addressed. The reason for this is that such incidents have the potential to generate adverse effects such as increasing the possibility of secondary accidents and increasing traffic congestion levels, air pollution and fuel consumption, as well as reducing roadway capacity. According to (Charles *et al.*, 2003), a traffic incident is “any event that degrades safety and slows traffic”. Thus, the term ‘traffic incident’ can refer to occurrences of many types such as demobilised vehicles, traffic accidents, spilled loads and debris.

Previous studies have shown that secondary incidents account for about a quarter of all incidents and congestion, although figures vary from place to place. For example, secondary crashes were found to comprise 20% of all incidents in the US (FHWA, 2004) and 20-30% of the total traffic incidents in Europe (Versavel, 2004). One of the most cited studies estimated that incident-related congestion accounts for 25% of total congestion in the United States (FHWA, 2005), as in the UK, where traffic incidents are responsible for 25% of total traffic congestion (Rillie and Byard, 2006).

In response to the severe consequences of traffic incidents, great attention has been focussed on improving the effectiveness of Traffic Incident Management (TIM). TIM can be defined as applying available resources to reducing the impacts of traffic incidents and incident duration (Farradyne, 2000). TIM can be manifest in any initiative or programme in the fields of legislation, operation or technology, which aims to reduce the harmful impacts of traffic incidents. As the name suggests, reducing incident duration is one of the main targets of TIM. Thus, improving the efficiency of the TIM process requires a clear understanding of the factors affecting incident duration.

Several studies have defined incident duration in various ways. One of the earliest studies measured incident duration as the time an incident remained on the travel lane (Derose, 1964). Another study defined accident duration as the time difference between a police officer receiving a call to responding to an accident, including accident clearance time (Jones *et al.*, 1991). This definition clearly excludes reporting time as a component of incident duration. However, most researchers defined incident duration as the time difference between incident occurrence time and incident clearance time (Nam and Mannering, 2000a; Shin, 2003; Wang *et al.*, 2005a; Wei and Lee, 2007).

Accordingly, the total incident duration can be divided into several phases or interval times. The Highway Capacity Manual (TRB, 1994), breaks down the total incident duration into the following four phases:

1. Detection time: the time between the incident occurrence and incident reporting time.
2. Response time: the time between incident reporting time and the arrival time of the first responder at the scene.
3. Clearance time: the time between the arrival of the first responder at the scene and the moment when the incident has been cleared from the highway.
4. Recovery time: the time taken for traffic flow to return to normal after the incident has been cleared.

Traffic incident duration has been influenced by various factors, such as temporal characteristics, incident characteristics and environmental characteristics. Since the main purpose of TIM is to reduce the adverse impacts of traffic incidents and achieve a decrease in incident duration, investigating these factors and understanding how they affect incident duration is important as far as reaching a high level of TIM effectiveness is concerned. Such an investigation can be carried out for the total incident duration or for every single interval time (reporting, response, clearance and recovery time) of the total duration. Different kinds of data are required to determine the start and the end points of each interval time and to investigate the effects of incident characteristics and traffic flow data at these times.

Over the last few decades, many researchers have investigated traffic incident duration for the purpose of gaining more insight into the nature of incident characteristics that have an influence on incident duration. Among several approaches that have been applied to investigate incident duration, duration analysis has been found to be an effective method. Also, duration analysis is known by a variety of terms in different disciplines as will be demonstrated in section 3.2 of Chapter 3; however, in this thesis the duration analysis is referred to as Hazard-Based Duration Models (HBDMs). Furthermore, most of the previous research in this area was based on developed countries, including the USA, and almost no attention has been given to developing countries.

In addition, most of the earlier analysis focused on the total duration of the incident and not on the single interval times. This approach may not gain sufficient insight into the effects on each interval time of the total incident duration. Also, it may not help traffic operators to evaluate and review TIM programmes with due attention to each interval time. Moreover, the analysis based on urban areas was limited in the literature. Therefore, investigating each interval time of the total incident duration in developing countries, with attention to urban areas, will be a significant addition to the current literature and knowledge of incident duration analysis.

1.2 Aim and Objectives

The aim of this PhD project is to investigate the effects of traffic accident characteristics on each interval time (reporting, response and clearance time) of the urban and highway traffic accidents in Abu Dhabi, the capital of the United Arab Emirates (UAE), using fully parametric HBDMs with emphasis on the Accelerated Failure-Time (AFT) metric.

This aim will be achieved by attaining the following objectives:

1. To conduct a Literature Review in order to elaborate upon research gaps, formulate research questions and create a sound methodology for the study.
2. To investigate existing traffic incident data collection and reporting systems in Abu Dhabi.
3. To design a data collection process in order to be able to gather data as required for this PhD study.
4. To collect characteristics and duration data related to traffic accidents.
5. To develop and estimate a sub-model, applying HBDMs to analyse each interval time of the total incident duration.
6. To critically investigate accident characteristics that affect incident duration, considering various interval times that exist in the TIM process.
7. To interpret the effects of accident characteristics on each interval time.
8. To develop a decision making tool that can be used to predict traffic accident durations.

In this research, recovery time of both urban and highway accidents could not be modelled due to the difficulties of collecting traffic flow data from traffic sensors. The absence of detailed reporting time and clearance time of highway traffic accidents led to modelling only the response time of such accidents. Thus, four sub-models are developed in this research, where three sub-models are related to urban accidents (based on reporting, response and clearance times) and there is one sub-model for highway accidents (based on response time).

1.3 Thesis Outline

The introductory chapter concentrates on the general background of the importance of incident duration analysis, as well as addressing the aim and the objectives of this study.

Chapter 2 deals with state-of-the-art literature on traffic incident impacts, traffic incident management definition and process, and incident duration definition. Also, more emphasis is placed upon the most common approaches to analysing traffic incident duration.

Chapter 3 describes the basic concepts of HBDMs. Details of the advantages and the applications of the three approaches of HBDMs are given. Subsequently, modelling concerns and selection criteria are presented.

Chapter 4 illustrates the methodology of analysing each interval time of the total incident duration. It describes the initial and revised case study areas considered in this research. Then, it covers all the process of data collection and preparation through data analysis and interpretation. Finally, it demonstrates the use of the model results to develop a decision tree as a prediction tool of accident duration.

In Chapter 5, the details of the databases used to extract the required data for both accident characteristics and accident duration are explained. Descriptive analysis of accident duration and accident characteristics is shown.

Chapter 6 starts with a preliminary analysis to decide the analytical method. Following that, it presents the results of estimating three models for urban accidents (reporting time, response time, clearance time) and one model for highway accidents (response time). The best-fit distribution for each one of these models is described in more detail with a discussion of the significant variables. Furthermore, duration prediction and accuracy are performed followed by developing a decision tree for each model based on the significant variables.

Chapter 7 concludes by summarising the aim, methodology and the main findings of this research. Furthermore, recommendations regarding the improvement of the TIM process in Abu Dhabi and future work are illustrated.

2 Literature Review: Traffic Incident Duration

2.1 Introduction

This chapter begins by presenting the definition of traffic incidents and their related impacts. This is followed by a review of the aims and process of TIM. Subsequently, the definition and components of incident duration concepts are presented. Then, issues related to traffic incident data are explained. Finally, the last section reviews various methods of modelling traffic incident duration.

2.2 Traffic Incidents and Related Impacts

Traffic incidents can be defined as any events that encumber normal traffic flow and cause a reduction in roadway capacity. These events can be categorized in two groups: planned and unplanned events. Examples of the former could be highway construction schemes or special events; the latter could include traffic accidents, vehicle breakdown and debris (Farradyne, 2000; Abdel-Rahim, 2004).

Incident impacts may generate a variety of problems, including increasing the risk for incident responders and road users' safety, increasing the likelihood of secondary crashes, fuel consumption and delays to motorists. The extent of these adverse effects varies from place to place. For example, secondary crashes were found to account for 13% of total peak-period crashes in the US, whereas they represent 20-30% of the total traffic incidents in Europe (Minnesota, 1982; Versavel, 2004).

Another impact of traffic incidents is incident-related congestion. One of the most cited studies estimated that incident-related congestion comprises 61% of the total congestion in the United States (Lindley, 1986). According to (Kay and Kinnersley, 2002), incident-related congestion constitutes between 13% and 30% of the traffic congestion during the peak period. In the UK, traffic incidents account for 25% of the total traffic congestion (Rillie and Byard, 2006). A study conducted in Brisbane, Australia, showed that 50% of congestion delays are incident-related (Charles *et al.*, 2003). Thus, it is clear that there is a possibility of incidents occurring due to traffic congestion and the impact of this may vary from one location to another.

Furthermore, a study conducted by (Farradyne, 2000) to measure the impacts of traffic incidents on freeway capacity proved that freeway capacity can be affected more than the actual physical reduction caused by an incident. It found that blocking one lane of a three-lane freeway can cause a 50% reduction in freeway capacity, as shown in Table 2-1. Higher figures were found in another research carried out in the Hampton Roads region of Virginia, USA (Smith *et al.*, 2003). The results showed significant reductions in the capacity of a three-lane freeway when an accident blocked one or two lanes (63% and 77% respectively).

Table 2-1 Percentage of freeway capacity available under incident conditions (Farradyne, 2000)

No. Freeway Lanes in Each Direction	**Shoulder Disablement**	**Shoulder Accident**	**Lanes Blocked**		
			One	**Two**	**Three**
2	0.95	0.81	0.35	0.00	N/A
3	0.99	0.83	0.49	0.17	0.00
4	0.99	0.85	0.58	0.25	0.13
5	0.99	0.87	0.65	0.40	0.20
6	0.99	0.89	0.71	0.50	0.25
7	0.99	0.91	0.75	0.57	0.36
8	0.99	0.93	0.78	0.63	0.41

Following the presentation of incident-related impacts, it is widely recognized that traffic incidents have an adverse effect on road networks. Thus, minimizing these effects is the biggest challenge facing road network operators through improving TIM procedures. The following section explains TIM in more detail.

2.3 Traffic Incident Management (TIM)

In response to the severe consequences of traffic incidents, great attention has been focussed on improving the effectiveness of TIM. TIM can be defined as applying available resources to reduce the impacts of traffic incidents and incident duration (Farradyne, 2000). Thus, it is clear that TIM aims to reduce the adverse consequences of traffic incidents and to minimize incident duration by creating more co-operation and co-ordination between incident responders and traffic operators as far as using available resources is concerned. Another study pointed out that the main objectives of TIM are saving life and property, reducing incident duration and collecting the required evidence for investigation (Raub and Schofer, 2007).

TIM has various kinds of practices and programmes. When talking about such programmes, it is necessary to clarify their meaning. A TIM programme can be defined as any initiative or scheme developed to improve the effectiveness of TIM processes (Austroads, 2007). Some of these programmes are designed to improve the operational part of TIM, such as introducing a quick clearance policy, training, co-ordination and communication among TIM stakeholders. Other programmes aim to cover the technological part of TIM, such as applying automatic incident detection, freeway surveillance systems and laser scanning for accident scene clearance. Furthermore, there are several initiatives covering the institutional part of TIM. For instance, creating national coalitions and increasing TIM resources may contribute greatly to improving the quality of the TIM process.

A TIM process may include several phases. According to Farradyne (2000) and Margiotta et al. (2004), a TIM process consists of the following seven phases:

1- Detection: The process by which a traffic incident becomes known to the transport agency or other responsible agencies. Several tools can be utilised to detect incidents, such as mobile phones, traffic patrol, CCTV systems, emergency telephone systems and traffic reporting services.

2- Verification: The key requirement of the verification stage is to collect further information about the incident, starting by confirming an incident occurrence. Other information, like the location and other relevant details, is important with regard to determining the proper initial response. Incident verification can be conducted by means of various methods, such as using CCTV systems or dispatching police patrols to the scene.

3- Motorist Information: This activity is necessary in order to inform motorists about the incident-related information or any route diversion. Disseminating information can be carried out by highway advisory radio, internet, variable message signs or other media services. However, this information has the potential to cause a harmful impact on the traffic if it is inaccurate or not updated (UK, 2002). Furthermore, the length of this activity depends upon the incident severity and the impact on traffic flow.

4- Response: After confirming incident occurrence and obtaining incident details, the response process will be initiated. The appropriate personnel and equipment will be dispatched to the incident scene with an effective communication link between the responders and the activation of motorist information.

5- Site Management: The focus of this activity is to manage the available resources at the incident scene in order to maintain the safety of responders, incident victims and other motorists. Many functions occur during this phase, such as maintaining good co-ordination with incident responders, conducting incident assessment and assisting with injuries.

6- Traffic Management: This involves the utilization of traffic control measures around the incident scene to reduce the impact of the incident on traffic flow and to ensure the safety of responders. A possible way of performing this phase is by controlling roadway space, either by managing road lanes or creating a traffic diversion around the scene.

7- Clearance: This is the final phase of TIM, aiming to restore normal traffic flow by removing any obstacles such as vehicles, debris or wreckage before reopening the roadway to traffic. The length of this process depends upon using the appropriate equipment and technologies which should be made available on the scene, based on an accurate scene assessment carried out in the site management stage.

Following the presentation of TIM phases, it should be noted that the sequence of these phases may assume a different shape and more than one phase could be performed simultaneously. For example, motorist information is usually updated throughout the duration of other phases, such as clearance (Margiotta *et al.*, 2004). Another example is that some minor incidents can be dealt with by the people involved and may not result in any traffic flow reduction. Such incidents may not exhibit both response phase and recovery phase (Smith and Smith, 2001).

On the other hand, there are many agencies involved in TIM phases, including law enforcement, fire and rescue, emergency medical services, transportation agencies, towing and recovery services, information providers and hazardous materials handlers. The appearance of these agencies at the scene is based on the incident type and level of severity. However, not all of these agencies will participate in all TIM phases. Some may be involved in more than one phase. For example, law enforcement and transportation agencies are able to contribute to detection, traffic management and clearance activity. Others can participate in only one activity, such as towing and recovery services (Austroads, 2007).

2.4 Traffic Incident Duration

As mentioned in section 2.3, reducing incident duration is one of the main targets of TIM, so measuring the reduction in incident duration can be considered as a performance measurement of a TIM programme. Consequently, this section is devoted to explaining the previous literature on the definitions and distributions of total incident duration.

Almost invariably, the definition of incident duration differs from one study to another. One of the earliest studies measured incident duration as the time an incident remained on the travel lane (Derose, 1964). Thus, if the vehicle was moved from the travel lane to the shoulder, response, clearance and recovery times would not be measured. Margiotta et al. (2006) defined incident duration as “The time elapsed from the notification of an incident to when the last responder has left the incident scene”. Thus, the reporting (detection) time was not included as part of this definition. However, most researchers defined incident duration as the time difference between incident occurrence time and incident clearance time (Nam and Mannering, 2000b; Shin, 2003; Wang *et al.*, 2005a; Wei and Lee, 2007). According to this view, the total incident duration is divided into several interval times. Consistent with the Highway Capacity Manual (TRB, 1994), the total incident duration consists of the following four phases (as shown in Figure 2-1):

1. Detection time: the time between the incident occurrence and incident reporting time.
2. Response time: the time between the incident reporting time and the arrival time of first responder at the scene.
3. Clearance time: the time between the arrival of the first responder at the scene and the moment when the incident has been cleared from the highway.
4. Recovery time: the time taken for traffic flow to return to normal after the incident has been cleared.

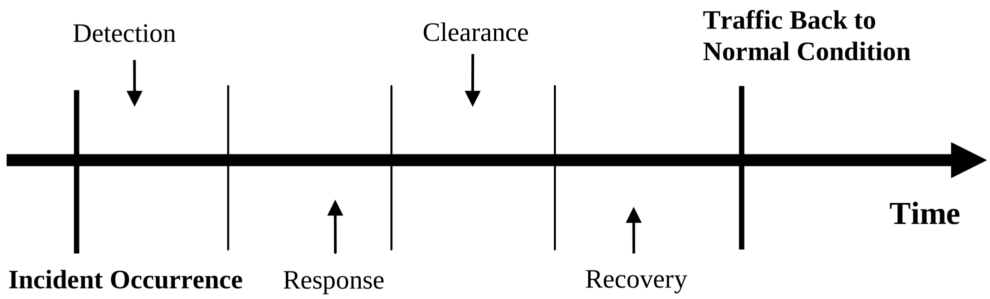


Figure 2-1 Incident duration interval times

Another study (Wang *et al.*, 2005b) considered verification time as being part of the total incident duration. This time occurs after the detection time and before response time. Verification time can be defined as the time taken to confirm incident occurrence by means of different methods, such as CCTV systems (Florida, 2005). However, other studies incorporated this time in the detection time (Shin, 2003) or in the response time (Nam and Mannering, 2000b).

On the other hand, some researchers excluded recovery time from the total incident duration interval times, because this time depends on the traffic demand at the incident scene and is difficult to measure (Garib *et al.*, 1997). In 1991, Jones et al. defined accident duration as "the length between the times a police officer receives a report of an accident until he/she leaves the accident scene". As far as this definition is concerned, it is clear that recovery time is not included as a component of incident duration.

Through reviewing the earlier studies of incident duration, it became clear that the focus of previous studies was on investigating three interval times of the total incident duration, including detection (reporting) time, response time, and clearance time. Verification time was included in the response time and detection time in some studies; however, recovery time was out of their scope. In addition, the difference in definitions of incident duration can be justified due to the varying purposes of different studies.

2.5 Traffic Incident Data

It is widely recognized that the success of any study conducted to analyse incident duration or evaluate any TIM programme is highly dependent on the quality and accuracy of traffic incident data. Thus, collecting incident data became one of the key steps in performing such studies.

When investigating previous works in this area, it can be observed that each phase of the total incident duration has different data that can be collected by diverse methods. For example, the data of detection phase can be retrieved from CCTV and loop detectors (Austroads, 2007). Also, the data of response time and clearance time can be obtained from the responders, and each responder may have a different dataset, depending on the incident type. For instance, Washington State Incident Response Team typically responds to the major incidents, thus this team has the response and clearance data for this kind of incident (Nam and Mannering, 2000b). For more details regarding some traffic incidents, such as traffic accidents, there is a specific form to collect the characteristics of such incidents. In the UK, Stats 19 is in use to collect traffic accident data. It has an accident record, a vehicle record to be filled in for each vehicle involved and casualty records in case of injury in an accident (Anderson, 2003).

In addition, many technologies have been introduced in this area and have contributed to reducing the time of some interval times. Laser scanning is one of these technologies, mainly focused on gathering accident-related evidence in a short time. Research conducted by Washington State Department of Transportation showed that Laser scanning technology has the ability to reduce investigation time by more than 50% compared with the co-ordinate method (Jacobson *et al.*, 1992). A further technology is Photogrammetry, used to gather crash evidence in a shorter time than traditional methods. This technology requires only one person to take several photos of a traffic accident within 20 to 30 minutes of occurrence (Balke and Cooner, 2000). All of these technologies can play an active role in reducing clearance time and consequently opening the road to the traffic more quickly.

Following incident data collection methods, attention turned to the type and comprehensiveness of the incident data required for conducting incident duration analysis. In fact, determining the nature of the required data is one of the main purposes of the study. When research aims to evaluate a TIM programme, it can be seen that the data are collected from the logs of the programme (Carson, 1999). If the purpose is to predict total incident duration or some interval times, it can be observed that data are collected from different sources. For instance, in 2006, Knibbe et al. conducted a study in the Netherlands to estimate incident duration using Rijkswaterstaat Verkeerscentrum

Nederland database. An example of databases utilised in this research is the National Incident Management Centre.

Furthermore, it was found that the majority of the previous studies used reported data that were recorded by people other than the researchers themselves. However, these studies were restricted in one way or another by insufficient or limited data (Golob *et al.*, 1987; Jones *et al.*, 1991). These studies could not draw a clear conclusion and concluded with recommendations for more accurate and comprehensive data. A few studies used a data collection form and conducted field surveys to obtain the required data (Garib *et al.*, 1997; Ozbay and Kachroo, 1999). This way of collecting data has the advantage of overcoming the problem of inaccurate data, and collecting accurate data that can fulfil research objectives. Moreover, due to limitations with regard to incident duration data, some researchers added a fixed time to represent detection and response times, in order to be able to estimate incident duration (Sullivan, 1997). Unfortunately, this addition had the potential to create a bias to the model of estimating incident duration and its results.

After carefully examining the research on traffic incident duration, it can be said that when attempting to analyse traffic incident duration, it is important to address certain issues in relation to incident data. First of all, since incident duration consists of several interval times, each time may have associated data that may not exist at other times. Secondly, traffic incident data are spread out among different sources such as incident responders and traffic operation centres; thus, to obtain a comprehensive database, it is necessary to bring together these sources of data in an integrated database. Also, to avoid inaccurate or limited data, it is better to design a collection tool and conduct a field survey rather than employ data that has been recorded by other sources. Finally, the amount and type of data required to analyse incident duration or any TIM programme is dependent on the number of incident interval times that the particular programme covers. For example, if the study aims to evaluate the performance of a technology introduced to reduce the amount of clearance time, then there is no need to obtain any information regarding the response or reporting time.

2.6 Traffic Incident Duration Modelling: State-of-the-Art

In the last few decades, several methods have been employed to study traffic incident duration. Although these methods provided valuable results, it is difficult to draw a comparison between them due to the variation of data sources and datasets. This chapter is dedicated to reviewing previous methods that have been used to study incident duration. The first section presents the literature focused on estimating the probable distributions of incident duration. The subsequent sections describe the most common methods employed for such a purpose, including Linear Regression Models, Time Sequential Models, Nonparametric Regression Method, Decision Tree and Classification Trees, Bayesian Networks (BNs), Discrete Choice Models, Fuzzy Logic (FL), Artificial Neural Network (ANN), and Hazard-Based Duration Models (HBDMs). Table 2-2 summarises previous work on applying the most common methods of studying traffic accident duration.

Table 2-2 Summary of previous research into incident duration analysis

Author(s)	Details	Sample Size	Modelling Method	Independent Variables
Khattak, et al., 1995	- USA - Freeway incidents - Forecasting duration	109	Time Sequential Procedure Regression Model	Incident type, vehicle type, number of vehicles involved, injuries and fatalities, property damage, response time, number of responders, weather conditions, incident location, seasonal factors, flow conditions and motorist information.
Golob et al., 1987	- USA - Freeway accidents - Find out the probable distribution	525	Probabilistic Distributions of Incident Duration	Number of lanes or ramps closed, date, time of day, location and collision type.
Giuliano, 1989	- USA - Freeway incidents - Estimating duration	512	Linear Regression Model	Injuries, number of vehicles involved, or lane closures, time of day, day of week, location and accident type.
Wang, 1991	- USA - Freeway incidents - Predict incident clearance time	121	Linear Regression Model	Heavy wrecker, sand/salt, pavement operations, assistance from other response agencies, heavy loading, number of heavy vehicles involved, severe injuries in vehicles, liquid or uncovered broken loading in heavy vehicles, extreme weather conditions, freeway facility damaged caused by incident, response time and incident report.
Garib et al., 1997	- USA - Freeway incidents - Predict incident duration	205	Linear Regression Model	Number of lanes affected, number of vehicles involved, binary variable for truck involvement, binary variable for time of day, natural logarithm of the police response time and a binary variable for weather conditions.
Peeta et al., 2000	- USA - Freeway incidents - Predict incident clearance time	2011	Linear Regression Model	Incident characteristics, traffic characteristics, environmental characteristics and operational characteristics.

Table 2-2 (Continued)

Author(s)	Details	Sample Size	Modelling Method	Independent Variables
Valenti et al., 2010	- Italy - Motorway incidents - Predict incident duration	237	Linear Regression Model and Decision Tree model	Emergency medical services at the scene, heavy duty vehicles involved, peak hour, infrastructure damage, number of lanes and vehicle fire incident.
Choi, 1996	- USA - Freeway incidents - Forecasting duration	2981	Fuzzy Logic Model	Vehicle problem, vehicle assistance type and location.
Kim and Choi, 2001	- USA - Freeway incidents - Forecasting incident response time	2457	Fuzzy Logic Model	Type of incident, type of vehicle and location of incident vehicle.
Wang et al., 2005	- UK - Motorway incidents - Forecasting vehicle breakdown duration	213	Fuzzy Logic Model and Artificial Neural Network	Time of day, vehicle type, location and reporting mechanism.
Wei and Lee, 2007	- Taiwan - Freeway incidents - Forecasting duration	24	Artificial Neural Network	Incident characteristics, geometry characteristics, special relationship and time relationship.
Sethi, et al., 1994	- USA - Freeway incidents - Estimating duration	801	Decision Tree Model	Roadway type, incident type, incident severity and traffic condition.
Ozbay and Kachroo, 1999	- USA - Freeway and non freeway incidents - Estimating duration	650	Decision Tree Model	Incident type, severity factor, operational factor, location, lane closure and environmental factor.

Table 2-2 (Continued)

Author(s)	Details	Sample Size	Modelling Method	Independent Variables
Knibbe, et al., 2006	- Netherlands - Highway incidents - Estimating duration	1853	Classification Tree Model	Incident type, kind of vehicles involved, number of vehicles involved, weekday or weekend, during peak hour or not during peak hour and type of responder.
Kim et al., 2008	- USA - Highway incidents - Predict incident duration	6765	Classification and Regression Tree CART and developed Rule-Based Tree Model (RBTM)	Spatial variable, out of peak period, lane closure and wet pavement.
Smith and Smith, 2001	- USA - Freeway incident - Forecasting clearance time	6828	Classification Tree Model, Stochastic, Nonparametric Regression Model	Time of day, day of the week, weather condition, number of vehicles involved, vehicle type and response type.
Smith and Smith, 2000	- USA - Highway incidents - Forecasting duration	2798	Nonparametric Regression Model	Incident type, time and date, location and number of lanes closed.
Ozbay and Noyan, 2006	- USA - Freeway incidents - Estimate incident clearance time	600	Bayesian Networks	Type of incident, number of police vehicles, number of ambulances, number of fire engines, number of injuries, number of trucks involved, number of cars involved, total number of lanes and type of roadway.

Table 2-2 (Continued)

Author(s)	Details	Sample Size	Modelling Method	Independent Variables
Yang et al. , 2008	- Netherlands - Highway incidents - Estimating incident duration	1853	Bayesian Networks	Police, truck, roadway closure, rescue, inspector, damage and workday.
Lin, et al., 2003	- USA - Highway incidents - Estimating duration	22495	Discrete Choice Model and Rule-Based Model	Incident or accident type, number of lanes blocked, incident time, truck involved, number of vehicles involved, weather condition and visibility, use of a heavy wrecker, county name and detection source.
Jones et al., 1991	- USA - Freeway incidents - Estimating duration	2156	Hazard-Based Duration Models	Season, time of day, driver characteristics, vehicle characteristics, severity, location and special event.
Nam and Mannering, 2000	- USA - Highway incidents - Estimating duration	681	Hazard-Based Duration Models	Temporal characteristics, environmental characteristics, geographic characteristics and incident characteristics.
Lee and Fazio, 2005	- USA - Freeway incidents - Estimating duration	5708	Hazard-Based Duration Models	Severity, day of week, number of vehicles involved, light condition, number of lanes, posted speed limit, road condition, heavy vehicle involvement, weather condition and urban or rural area.
Jovanis and Chang, 1989	- USA - Freeway incidents - Develop a model of accident occurrence	1200	Hazard-Based Duration Models	Vehicle (type of truck, weight of cargo, age of tractor), driver (age, experience, accident record, off-duty), roadway (type of roadway, width of lane, number of lanes) and environment (weather, lighting, traffic volume, topography, night or daytime).

Table 2-2 (Continued)

Author(s)	Details	Sample Size	Modelling Method	Independent Variables
Qi and Teng, 2004	- USA - Expressway incidents - Estimating incident duration	858	Hazard-Based Duration Models	Weather characteristics, temporal characteristics and incident characteristics.
Qi and Teng, 2008	- USA - Roadway incidents - Estimating incident duration	1660	Hazard-Based Duration Models	Weather characteristics, temporal characteristics, incident characteristics, involved vehicle characteristics, geographic information and incident clearance agency.
Chung, 2009	- Korea - Freeway incidents - Estimating accident duration	2369	Hazard-Based Duration Models	Number of vehicles involved, number of injured, fatality, vehicle type, accident type, accident location, reporter type and accident time.
Songchitruksa et al., 2009	- USA - Roadway incidents - Estimating incident duration	30971	Hazard-Based Duration Models	Incident type, detection method, verification method, severity level, weather condition, type of vehicles involved, time of day and responder type.

2.6.1 *Probabilistic Distributions of Incident Duration*

This approach aims to investigate the probability distribution of incident duration by viewing this duration as a random variable. The key idea of this approach is to give the traffic operator an idea about the mean and variance of incident duration which might be useful for the purpose of predicting incident duration. Golob et al. (1987) analysed freeway accidents involving trucks in California, USA, to find out the probable distribution of such accidents' duration. In this analysis, a dataset of 332 highway accidents and 193 ramp accidents were retrieved from the California Highway Patrol dispatch record logs over a two-year period on a freeway in the greater Los Angeles area. These data were divided into 16 groups based on accident type and lane closure. Researchers claimed that each accident could include several stages, including: "1) detection, 2) initial response, 3) injury attention (if required), 4) emergency vehicle response (if required), 5) accident investigation, 6) debris removal, 7) clean-up, and 8) recovery". In addition, they assumed that the length of each stage has a direct impact on the length of the subsequent stage. Based on this assumption, they theorized that the distribution of total incident duration could be fitted to log-normal. To check the validity of this theory, the Kolmogorov-Smirnov test was applied. The result showed that incident duration confirmed log-normal distribution for all groups. However, the limitation of this study was the assumption that each stage of incident duration is time-dependent on the previous stage and could not be investigated because the dataset included the total incident duration only. Also, the sample size for each group was relatively small (17-57). However, the results of this research were supported by several subsequent studies (Giuliano, 1989; Garib *et al.*, 1997; Sullivan, 1997).

In 1991, Jones et al. found that a log-logistic distribution fitted their study dataset for Seattle freeway accidents. Also, Ozbay and Kachroo (1999) found that the shape of their study dataset of incident duration in North Virginia, USA, was similar to log-normal distribution. However, after running a number of statistical significance tests such distribution was rejected, and they found that incident duration could match normal distribution with a divided dataset based on incident type and severity. This finding could be considered as an important support for the theory which proposes that incident duration is a random variable.

Nam and Mannering (2000) analysed a two-year dataset of highway incidents in Washington State, USA. They broke down total incident duration into three interval times, including detection (reporting) time, response time, and clearance time. Both detection and response times' datasets were found to fit in a Weibull distribution. However, clearance time data fitted in a log-logistic distribution. Similarly, Wang et al. (2005b) found that the dataset of vehicle breakdown duration on the M4, UK, corresponded to a Weibull distribution. In a review of previous literature, it is clear that incident duration can fit different distributions.

2.6.2 Linear Regression Models

Linear regression is another approach used to predict incident duration. The key aspect of this method is to include several binary variables that represent incident characteristics to model their effects on accident duration (dependent variable) by fitting a linear equation (Fox, 1997; Johnson and Wichern, 2003). In 1989, Giuliano developed a model to estimate incident duration based on incident characteristics. A dataset of 512 incidents was obtained from the California Highway Patrol (CHP) for I-10 freeway in Los Angeles, California, USA. Due to data limitations, the researcher could not develop a fully specified model for the entire sample, so two separate models were developed. The first model was developed for all incidents and the other for accidents. In both models a number of variables were identified as being statistically significant variables, including incident types, lane closures, time of day and the presence of trucks.

In an unpublished study, (Wang, 1991) developed a regression model to predict incident clearance time. 121 incidents' records were used from the Illinois Department of Transport logs. After conducting statistical assessment tests for the used variables, only 9 variables out of 22 were found to be statistically significant, including heavy wrecker (WRECKER), sand/salt, pavement operations (SAND), assistance from other response agencies (OTHER), heavy loading (HEAVY), number of heavy vehicles involved (NTRUCK), severe injuries in vehicles (SEVIN), liquid or uncovered broken loading in heavy vehicles (NONCON), extreme weather conditions (WX) and freeway facility damaged caused by incident (RDSIDE). Also, other variables such as response time (RESP) and incident report (HAR) were used, even though they were not statistically significant. As a result of that, the final form of predicting clearance time was as follows:

Clearance Time = 14.03 + 35.57(HEAVY) + 16.47(WX) + 18.84(SAND) – 2.31(HAR) + 0.69(RESP) + 27.97(OTHER) + 35.81(RDSIDE) + 18.44(NTRUCK) + 32.76(NONCON) + 22.90(SEVINJ) + 8.34(WRECKER).

In 1997, Garib et al. utilized incident characteristics and traffic characteristics on interstate I-880 in Alameda County, Oakland, California to develop two regression models for predicting incident delay and one model for predicting incident duration. 205 incidents were collected by a group of observers from on/off ramp loop detectors. Based on statistical significance tests, 6 variables were found to be significant, including number of lanes affected (X_1), number of vehicles involved (X_2), binary variable for truck involvement (X_5), binary variable for time of day (X_6), natural logarithm of the police response time (X_7) and a binary variable for weather conditions (X_8). The form of incident duration model was as follows:

$$\text{Log (Duration in minutes)} = 0.87 + 0.027\, X_1\, X_2 + 0.2\, X_5 - 0.17\, X_6 + 0.68\, X_7 - 0.24\, X_8$$

Furthermore, they applied the R-square test to check the validity of this regression model. The result was 0.81, which means that this model can predict 81% of incident duration based on the formulated function (Garib *et al.*, 1997).

Peeta et al. (2000) applied a linear regression model to predict incident clearance time for incidents in Borman Expressway in North Indiana, USA. 2011 incident records were extracted from the Hoosier Helpers (Freeway Service Patrol on the Borman Expressway) Database to develop the model. Each record had several variables from major categories, including incident characteristics, traffic characteristics, environmental characteristics and operational characteristics. The model results demonstrated that the significant variables that affect incident clearance time are number of vehicles involved, severity of the incident, ramp, night time, temperature, rain and snow (Peeta *et al.*, 2000).

More recently, the Multiple Linear Regression (MLR) method was used to predict incident duration in Italy (Lange *et al.*, 1989). 237 incident records were extracted from the Autostrade per l’Italia Spa incident database to achieve the aim of the study. The results showed that the MLR model underestimated the duration of some incidents due to insufficient sample size.

2.6.3 Time Sequential Models

The key feature of Time Sequential Models is to develop a prediction model of incident duration at an early stage of the accident when only a small amount of information about the accident has become available. A series of updates to the initial model can then be conducted when further information is available (Khattak *et al.*, 1995). The main motivation of this approach is that a better prediction of incident duration can be made based on the accurate data gathered about the accident instead of historical data.

Khattak et al. (1995) stated that the models for predicting incident duration have no operational value, because the variables that are required for incident duration prediction will not be available until the incident is cleared. Thus developing a model to predict incident duration at the early stages of an incident could support incident management even if it may not have high accuracy. Based on that, they used records of 109 incidents collected from the Illinois Department of Transport on the Chicago freeway to introduce a time sequential model. Several variables were used in this model, including incident type, vehicle type, vehicle number, injuries and fatalities, property damage, response time, number of responders, weather conditions, incident location, seasonal factors, flow conditions and motorist information. Based on the availability of information, this model divided incident duration into 10 stages, each of which had a separate truncated regression model and more information than the previous stage. Although this approach has a sound justification, researchers could not accept the validity of such a method because of the small sample size of incidents.

2.6.4 Nonparametric Regression Method

Nonparametric regression is a common technique employed for the purpose of predicting traffic flow. The main idea of this method is to utilize past experience to make a current decision for a similar experience. It is based on the data which explain the relationship between dependent and independent variables. The advantage of this method is that no specific assumption is required to explain the relationship between the dependent and independent variables (Smith *et al.*, 2002).

Generally, few researchers have applied this method to study incident duration. However, Smith and Smith (2000) did make use of this approach to predict traffic incident duration. They analysed 2798 past incidents from I-95 Corridor for the period from 1997 to 1999 to predict the current incident duration time. Several independent variables were employed, including incident type, location, incident time and date, and number of lanes closed. All of these data were provided by I-95 Corridor Coalition's Information Exchange Network Database. The results show a 100% difference between the actual incident duration and the estimated incident duration, with the average incident duration being 73 minutes. This error was explained as a matter of inadequate selection of independent variables and dataset (Smith and Smith, 2000).

In 2001, the same authors conducted another study to forecast the clearance time of freeway accidents in Virginia. They used three different forecasting models: stochastic, nonparametric regression and classification tree. A dataset of 6,828 accidents provided by Smart Travel Lab was used for the period between January 1997 and December 2000. These data included many variables such as time of day, day of the week, weather conditions, number of vehicles involved and vehicle type. After developing these methods, none of the developed models produced an accurate prediction. This was interpreted as a function of poor data quality and model selection (Smith and Smith, 2001). Thus, it is apparent that this approach could be improved and may be developed to obtain an accurate result in further studies in the future.

2.6.5 Decision Tree and Classification Tree

Another approach found in the literature is the decision tree. This method is a nonparametric model designed to find out patterns in a certain dataset without any assumption regarding the underlying probabilistic distribution. It works through a repeated process of splitting the dataset into subgroups until termination, based on the significant explanatory variables. Also, it should be stressed that this method consists of a series of decision variables and the outcome represents the average incident duration of a specified dataset (Smith and Smith, 2001).

In 1994, Seith et al. applied this approach to predict incident duration in the USA. 801 incidents were used with variables of roadway type, incident type, incident severity and traffic condition. The results showed that the average duration was 21 minutes. However, the researchers recommended that further details of incident characteristics were required to increase the accuracy of the prediction.

Another investigation using this approach was carried out by Ozbay and Kachroo (1999) in the Northern Virginia region, USA. They began their study by using a linear regression model; however, the results show a low R-square value (approximately 0.35) due to the wide difference in the data. Also, the distribution of incident duration could not be fitted into either log-normal distribution or log-logistic distribution. As a result of this, they developed a decision tree model. After testing 77 incidents to check the accuracy of this model, they found that 57.14% of the these incidents were within 10 minutes of prediction error (Ozbay and Kachroo, 1999).

In 2001, Smith and Smith developed a classification tree to forecast incident clearance time in Virginia, USA. A classification tree is a special type of decision tree and works by allocating a class rather than a value. Several independent variables were used, such as time of the day, date, type of vehicle and response agency. The classification tree followed a top-to-bottom format and the outcomes were classified into three levels: including short time (1-15 minutes), medium (16-30 minutes) and long (30+ minutes). Accuracy tests showed that the accuracy level of this model was 58% for the tested accidents. They concluded that this model could perform better with higher data quality (Smith and Smith, 2001).

Furthermore, Knibble et al. (2006) applied a classification tree to forecast incident clearance time in Utrecht (the Netherlands). A dataset of 1853 incidents with independent variables of incident type, vehicle type and casualties was used to construct this model. They found that only 1 out of 3 incidents were accurately predicted. This low level of accuracy can be attributed to their use of a low detailed dataset (Knibbe *et al.*, 2006).

A Classification and Regression Tree (CART) approach was adopted to predict incident duration in China (Zhao *et al.*, 2009). CART is a statistical software package based on the methodology of a decision tree and is considered to be a nonparametric approach. The advantage of CART is that it can be easily interpreted; however, the difficulty of determining the threshold for continuous variables is one of the drawbacks of this approach. Data of 65000 incidents were collected from the Beijing Transportation Management Bureau to develop the model, along with another 8000 incidents for validation purposes. A validation test showed that the model gave an acceptable prediction with an average error of 29.5%.

In 2008, Kim et al. redesigned CART and developed a Rule-Based Tree Model (RBTM) for the purposes of identifying the significant variables that influence incident duration and to predict incident duration for highway incidents in Maryland, USA. A four-year dataset was extracted from Maryland State Highway (MDSHA) database to achieve the aim of this research. The results showed that several factors affect incident duration, including spatial variable, out of peak period, lane closure and wet pavement. Also, RBTM was found to have an advantage over CART for the given conditions (Kim *et al.*, 2008).

One of the latest studies (Lange *et al.*, 1989), developed a decision tree model to predict incident duration in Italy. Around 237 incident records were extracted from the Autostrade per l'Italia Spa incident database to achieve the aim of this study. Validation test results demonstrated that the Decision Tree model performed well in giving a satisfactory estimation of incident duration.

Additional details of the presented methods can be found in many references (Kass, 1980; Brieman *et al.*, 1984; Biggs *et al.*, 1991; Loh and Shin, 1997).

2.6.6 Bayesian Networks (BNs)

The Bayesian Network (BN) approach has been used as a classification tree for many tasks, including document classification (Brücher *et al.*, 2002) and dialogue act recognition (Keizer *et al.*, 2002). BNs are simply graphs where stochastic variables are used for nodes and the dependencies between these variables are represented by arcs (Jensen, 2001). The BN method has three advantages that give this approach merit over other classification methods such as decision tree: bi-directional induction, incorporation of missing variables and probabilistic inference. Ozbay and Noyan (2006) applied this approach to estimate incident clearance time in Virginia, USA. Their aim was to support decision makers (e.g. traffic operators and traffic engineers) in making real-time decisions. The assessment of the BN results for incident clearance time showed that this method was able to present the stochastic nature of incidents (Ozbay and Noyan, 2006).

Yang et al. (2008) applied this method to estimate incident durations in the Netherlands. To achieve the aim of this study, around 1853 incidents were collected by Rijkswaterstaat Verkeerscentrum from 1st May to 13th September 2005. The results illustrated that this model performs well compared with the classification tree method (even with incomplete information).

2.6.7 Discrete Choice Models

Discrete choice models are statistical measures that enable a choice to be made from a fixed set of alternatives. To achieve this, all of the possible alternatives need to be included in the set and the choice should be from the alternative set, only one alternative can be chosen and the set should have finite alternatives (Ben-Akiva and Lerman, 1985).

Line et al. (2003) developed an integrated approach of the discrete choice model and the rule-based model to predict incident duration in Maryland. Based on the needs of control centre operators, they divided the incident duration sample into several interval times with 5 minutes increment. Then they applied an order probit model to calibrate the model. The results showed that a discrete choice model was reliable for incidents with a duration of less than 60 minutes. To enhance the effectiveness of this approach, a rule-based model was developed to estimate incidents with a duration of more than 60 minutes (Lin *et al.*, 2003).

2.6.8 Fuzzy Logic (FL)

Fuzzy Logic (FL) is a multi-valued logic that maps an input data into a scalar crisp output by allowing for intermediate values. The basic concepts of FL are linguistic variable, fuzzy if-then rule and calculus of fuzzy rules. In order to develop an FL system, four components are required: input, process structure and output flow concept, as well as sufficient expert knowledge (Mendel, 1995).

FL has been shown to be a suitable approach for modelling transportation and traffic processes. In terms of modelling incident duration, this method has the advantage of allowing for the input of linguistic or category variables (Teodorovic, 1999). The first trial of this method to predict incident duration was carried out by Choi (1996) in Los Angeles, USA. In this study, the focus was on disabled vehicle incidents and the author used three variables, including vehicle problem, location, and type of assistance. It concluded that this method was appropriate to estimate such types of incidents (Choi, 1996).

In 2001, Kim and Choi used the FL method to develop an incident response model. The aim was to approve the suitability of this approach to judge incident operating processes. In this study, nearly 2457 freeway incidents' records were collected from the Los Angeles area by the Freeway Service Patrol with only three variables, namely type of incident, type of vehicle and location of incident vehicle. The results showed that the estimation times of FL are very similar to the actual time. As a result, it demonstrated that the fuzzy system is an appropriate approach for freeway incident management. However, the authors stated that the limitation of this study was the low number of variables used to develop the model. The structure of FL models needs to consider more parts than those used (Kim and Choi, 2001).

Recently, Wang et al. (2005) used this method to model vehicle breakdown duration in the UK. In this study, the authors aimed to analyze the available characteristics of vehicle breakdown incidents and developed two models using FL and Artificial Neural Networks (ANN). For this purpose, a dataset of 213 breakdown vehicle incidents, which occurred on the M4 motorway from May 2000 to April 2001, were obtained from the Road Network Master Database (RNMD). This dataset has many variables such as time of day, vehicle type, location and reporting mechanism. After collecting these records, a test of incident duration distribution (Kologorov-Smirnov) was carried out and the results demonstrated that the incident duration fitted into a Weibull distribution. Following a distribution test, two models were developed using FL and ANN. Based on the results of statistical parameters such as root mean square error and R^2, it concluded that ANN outperforms FL.

2.6.9 Artificial Neural Network (ANN)

An Artificial Neural Network (ANN) model is a relatively a new method that can be used to analyse the effects of independent variables on the dependent variable. The main assumption of the ANN method is that the effect on the output variable is not drawn directly by the input variables and there are hidden variables in the middle that influence the output variable (Lange *et al.*, 1989; Cross *et al.*, 1995). Thus, this approach has the advantage of facilitating the examination of complex relationships between the independent variables and the output variable. However, it is too difficult to recognise the effect of an individual independent variable on the dependent variable due to the main assumption of this model (Clark *et al.*, 2003). For more details on this approach, see Bishop (1995) and (Bishop, 1995; Haykin, 1999).

ANN has been widely applied to issues in the field of transportation. Some of these applications include pavement maintenance, driver behaviour, traffic control and traffic forecasting. These methods are based on collecting data from different sources and computing them using an internal “transfer function” (Dougherty, 1995). Some of the early studies applied this method for incident duration analysis and were focused on incident detection analysis (Teng and Oi, 2003; Yuan and Cheu, 2003). Also, as mentioned in the previous section, this method was used to model vehicle breakdown duration in the UK, showing good results compared with FL (Wang *et al.*, 2005b).

Recently, Wei and Lee (2007) developed two ANN models to forecast car accident duration in Taiwan. The first model was based on real-time data at the time of incident notification. In the second model, a series of updates were made to the preliminary forecasting in the first model, using incident data. They used 24 incidents’ records from the national freeway in Taiwan for the period from November, 2004 to April, 2005. These records had many variables, such as incident characteristics, geometry characteristics, special relationship and time relationship. Six experiments were conducted for the two models to study the relationship between the actual duration of car accidents and the predicted duration. All results showed the existence of a linear relationship between the actual duration and the predicted duration. Also, the correlation coefficient was found to be over 0.72. Furthermore, the accuracy of these models was tested by many criteria, including mean absolute percentage error, mean absolute error and root mean square error. The mean absolute percentage error was found to be less than 40% at each forecasting point in the second model. As a result of this, researchers concluded that these models are appropriate to forecast incident duration and are viable in Intelligent Transportation Systems (Wei and Lee, 2007).

2.6.10 Hazard-Based Duration Models (HBDMs)

Hazard-Based Duration Models (HBDMs) are a group of analysis methods that aim to analyse the time until an event occurs, having said that the event has not occurred for a specific time (Hensher and Mannering, 1994; Kiefer, 1988). These methods have three forms: non-parametric, semi-parametric (Cox Model) and full-parametric. Full details of these methods’ concept and modelling concerns are explained in the following chapter.

In terms of the application of these methods in accident duration studies, one of the earliest studies was by Jovanis and Chang (1989), whose paper described the application of the Cox model to model the occurrence of truck highway accidents. In their study, the end of survival time was determined by the occurrence of accident and the length measured by hours. Also, truck trips that ended without an accident were considered to be right censored data in this model. The model estimation results show that driver characteristics have the highest significant effects on accident occurrence (Jovanis and Chang, 1989).

Furthermore, the time between the occurrence of the accident and the end of clearance activity was examined by many research studies. One of the early studies, using log-logistic distribution to investigate incident duration, was carried out by Jones et al. (1991). They aimed to examine the hazard function of traffic accidents in Seattle, USA, modelling such incidents' duration throughout. 2156 accidents records were used from the State Patrol to model this duration. The results showed that accident data fitted in a log-logistic distribution more than a log-normal distribution and that the hazard function decreases throughout which means that the longer an accident lasts, the less likely it is to end soon. However, the study found that there are some factors which have a significant influence on accident duration, but these factors, such as alcohol detected and driver age, are difficult to collect. Thus, this study emphasised that there is a great need for more a accurate and appropriate dataset to obtain a clear conclusion (Jones *et al.*, 1991).

In 2000, Nam and Mannering applied the proportional hazard approach to analyse highway incidents in Washington State. However, this study applied another approach by developing a sub-model for each interval time of the incident duration, namely incident detection/reporting time, response time and clearance time. Their aim was to approve the suitability of such a method for the assessment of TIM programmes and to determine the factors affecting incident duration. The data of around 681 incidents were collected from the Washington State Incident Response Teams (IRTs) for the years 1994 and 1995. The results of analysis showed that Weibull distribution with gamma heterogeneity resulted as the best fit distribution for both reporting time and response time, whereas log-logistic distribution fitted clearance time the best. On the other hand, significant variables in one model were not found to be significant in the other. This difference in significant variable among the three models provided an excellent demonstration of the suitability of using sub-models when the interest is to evaluate each part of the total incident duration (Nam and Mannering, 2000b).

Another study that fitted log-logistic distribution to model incident duration was on the Gownus Expressway in New York City, USA. Several variables representing different categories were collected to develop this model, including weather characteristics, temporal characteristics and incident characteristics. The results showed that the variables found to significantly affect incident duration were snow, night, injury and fatal incidents (Qi and Teng, 2004).

In 2005, Lee and Fazio applied a proportional hazard-based Cox-regression model to analyse the influence on the Emergency Management Services (EMS) response and clearance times of the independent variables on Ohio State major freeways. Crash severity was found to influence response time significantly, whereas environmental variables had only a slight effect. Also, clearance time was affected by several independent variables, such as response time, location and number of vehicles involved (Lee and Fazio, 2005). These findings showed that there is a possible relationship between the length of one interval time of the total incident duration and the length of the subsequent stage.

In 2008, Qi and Teng used a parametric model to analyse incident duration data from New York State, USA. Data were collected from 11 roadways from December 6, 1995 to February 29, 1996. According to the goodness-of-fit tests, a log-logistic distribution was determined as the best fit for the data. Fitting this model demonstrated that different variables from the main categories (weather characteristics, temporal characteristics, incident characteristics, involved vehicle characteristics, geographical information and incident clearance agency) affect different interval times of incident duration (Qi and Teng, 2008).

Recently, a log-logistic accelerated failure time AFT model was used to develop a prediction model for accident duration in Korea. Although this model covers the whole incident duration, the prediction of accident duration is continuous from the accident occurrence time until the responders' vehicle departure. In incident management, modelling the whole duration might not meet the needs of operators and responders who are interested in improving various elements of the incident management process. This approach would lose information in both estimation and interpretation of results (Chung, 2009).

Another study utilised fully parametric HBDMs to predict incident duration in Houston, USA. Data collected from Houston's TranStar for the period 2004-2007 were used to develop four sub-models based on accident type. Results showed that a Weibull distribution was the best fit distribution for these models. Also, several variables from different categories were found to significantly affect incident duration, including incident type, detection method, verification method, severity level, weather condition, type of vehicles involved, time of day and responder type (Songchitruksa *et al.*, 2009).

Following the presentation of the previous works that applied HBDMs, it can be concluded that these models have many advantages over others, in that they can give more insight into duration dependence through hazard function and they have the ability to deal with censored observations. HBDMs focus not only on what and how factors affect accident duration, but also on the likelihood that the accident duration will end soon, given that it has lasted for a specific period of time. These advantages led to the selection of these models to model traffic accident duration in this research.

2.7 Summary

In summary, traffic incidents are the cause of a great deal of harmful impacts on safety and traffic. To mitigate these impacts, TIM has played a crucial role for several agencies worldwide. It aims to reduce incident duration and minimise incident impacts through applying different kinds of programmes or initiatives.

Also, an investigation of the data used to study incident duration shows that several variables have been used to examine factors affecting incident duration. Among these factors are incident type, location, number of lanes affected, weather condition, incident time and number of vehicles involved.

Finally, several approaches have been applied to model traffic incident durations. It is widely recognized that incident duration could fit at any distribution, and the assumption that assumes log-normal distribution can be rejected. Also, it should be stressed that poor data quality was a common problem in previous studies. Finally, it is worth mentioning that conducting comparisons of previous work is extremely difficult, because almost every study has applied a different method, used a different dataset, had different aims and used different data collection methods. Furthermore, each approach has its strengths and weaknesses, and as a result of this there is no best method that applies under all conditions.

With respect to the use of HBDMs to investigate incident duration, many more parametric and semi-parametric models were applied compared to nonparametric models. Furthermore, it was observed that the vast majority of the available literature considered the whole duration as one time in their analysis, with the exception of the studies of Lee and Fazio (2005) and Nam and Mannering (2000), which developed a sub-model for each interval time of the total incident duration. This separation showed that the same independent variable may have different effects at different interval times. Thus, due to the limited literature that proved from empirical evidence that the effect of the independent variables may vary from one interval time to another (such as response, clearance), this study is going to contribute in this area by developing a sub-model for each interval time of the total accident duration. The following chapter focuses on HBDMs used to study accident duration in this research.

3 Literature Review: Hazard-Based Duration Models (HBDMs)

3.1 Introduction

Following the introduction of a variety of approaches that have been applied to model incident duration, this chapter presents a review of the modelling approach in this study in more detail. In the first section, the basic concept of HBDMs and mathematical components are presented. This is followed by describing various kinds of model estimation in this approach. Subsequently, modelling concerns and model selection criteria are detailed before displaying the application of HBDMs in transportation in the final section.

3.2 Basic Concept of HBDMs and Mathematical Components

HBDMs are sometimes referred to as Time-to-Event modelling, survival analysis, event history analysis or lifetime analysis. Analysis of HBDMs refers to them as a collection of statistical methods for analysing time until an event occurs. In other words, the purpose of HBDMs is to investigate the effects of a number of explanatory variables on event occurrence. An event may present any change or transition from one state to another, whereas the time could be measured in any scale such as minute, day, week, month, or year (Alison, 1984).

HBDMs have been applied in a wide range of areas such as social sciences, criminal sciences and political sciences. However, it should be stressed that the initial area that applied such approaches was medical sciences, where the event of interest was usually death or any negative experience. Thus, most of the terms used by HBDMs such as survival time or hazard rate (which will be explained in more detail later) are explained by the initial application of these methods in health sciences (Alison, 1984).

When applying HBDMs, it is most important to be aware of the meaning of the terms used in order to facilitate understanding their related concepts. Time is known as a survival time because it means that the observation remains (survives) in the current stage over some time and does not experience the event of interest (which determines the termination of the survival time). Also, in the initial application in the medical sciences, event refers to a failure. Thus, to be able to measure survival time precisely, there are several requirements, including unambiguous time origin, clear meaning of failure, clear definition of event and an appropriate time scale. Failing to fulfil these requirements may cause a problem called censoring (Cox and Oakes, 1984; Kleinbaum and Klein, 2005).

As mentioned earlier, HBDMs aim to study the relationship between a dependent variable (time) and some explanatory variables. This can be done by using regression models (Allison, 1995). There are some similarities and differences between HBDMs and other regression approaches. For example, regression analysis includes many estimation methods such as Ordinary Least Squares (OLS) and Maximum Likelihood Estimation (MLE). Also, it is widely recognised that regression modelling encompasses a variety of error distributions and, therefore, the decisions have to be made based on the structure of the data and the method of analysis with regard to which approach is best (Lange *et al.*, 1989; Washington *et al.*, 2003). According to Cleves *et al.*, (2004) and Golder (2005) there are four reasons why OLS should not be used for duration data analysis. The first reason is that the survival time is usually censored, which means that the start or end point of interest will occur outside the observation period. The second reason is concerned with time-varying covariates, which means that the value of the independent variable changes over time. The third reason is that there is a possibility of generating negative predicted values and this is not acceptable when analysing duration data. The fourth reason is that the error term is assumed to be normally distributed when applying OLS. This normality assumption is not appropriate when estimating duration models because the distribution of duration data can take any shape due to the length of each observation time. Thus, applying OLS would cause a bias in the interpretation of results.

Another important point is that compared to the regression approach, which specifies unconditional probability for time data distributions, HBDMs are based on an important concept: the conditional probability of a duration ending at some time, given that the duration has continued for some specific time. This concept is important in duration study because in many instances the probability of a time duration ending depends on the length of time the duration has lasted. This probability may increase, decrease, or remain constant. For example, the probability of a new driver being involved in an accident may vary over time due to the experience and skills gained over time. Thus, conditional probability is an essential concept to take into consideration when studying duration data.

Furthermore, applying HBDMs approaches can deal easily with some issues relating to duration analysis, including censoring and time-vary variables. More details of these issues will be presented later in this chapter; however, in brief, censoring refers to incomplete observed duration, whereas time-vary variables occur when the value of the independent variable changes over time (Hensher and Mannering, 1994; Bhat, 2000). To summarize, the application of HBDMs has both methodological and conceptual advantages over regression methods.

On the other hand, developing HBDMs is based on four functions (Figure 3-1), which can be expressed mathematically in the following equations.

The first function is the cumulative distribution function:

$$F\ (t)=P\ (T<t) \tag{1}$$

Where P indicates probability, T refers to a random time and t is some specified time. This function gives the probability of having an event before some specific time t. The second function is the corresponding density function, which is the first derivative value of the cumulative distribution with respect to time:

$$f\ (t)=\frac{d}{dt}F\ (t) \tag{2}$$

The third function of HBDMs is the survival function. This function presents the probability that the duration is greater than or equal to some specified time *t*. In other words, it can be expressed as an opposite to the cumulative distribution function:

$$S\ (t)=P(T\geq t) \tag{3}$$

$$=1\text{-}F(t)$$

Finally, the hazard function is:

$$h(t)=\frac{f\ (t)}{[1\text{-}F(t)]} \tag{4}$$

$$=\frac{f\ (t)}{S\ (t)}$$

Where h (t) presents the conditional probability of an event happening between time t and t + dt, given that the event has not occurred up to some specific time, t. In other words, the hazard function is the product of probability per time unit. Thus, it gives the rate rather than the probability in the case of a survival function (Hensher and Mannering, 1994; Kleinbaum and Klein, 2005).

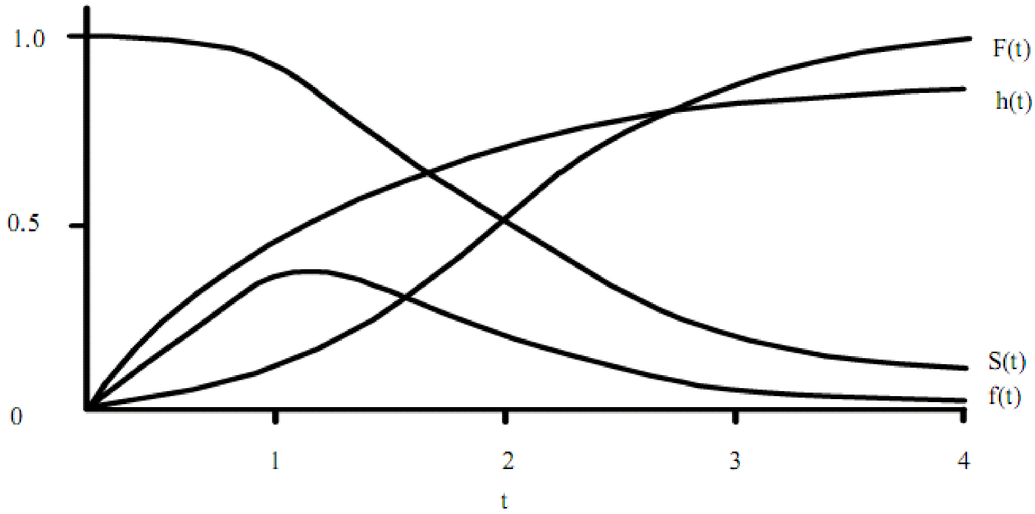


Figure 3-1 Basic functions graph (Mannering, 2007)

These four functions are related to each other. So, if one of the mentioned functions is known, any of the others can be obtained. Mathematically, this relationship can be presented by the following functions:

$$S(t)= 1-F(t)= 1- \int_0^t f(t)dt \ =EXP[-h(t)] \tag{5}$$

$$f(t)= \frac{d}{dt} F(t)=h(t)EXP[-H(t)]= - \frac{d}{dt} S(t) \tag{6}$$

$$H(t)= \int_0^t h(t)dt= -LN[s(T)] \tag{7}$$

$$h\ (t)= \frac{f(t)}{S(t)} = \frac{f(t)}{1-F(t)} = \frac{d}{dt} H(t) \tag{8}$$

Additionally, the shape of the hazard function slope has a significant inference in terms of duration dependence. Thus, the different shape of this slope has a different interpretation regarding duration dependence. As shown in Figure 3-2, the first hazard function slope (h_1) has $\frac{dh_1(t)}{dt}$<0 for all t. This means that the hazard function is monotonically decreasing in duration and indicates that the longer observations go without facing the event of interest, the less likely they will face it soon. The second hazard function slope (h_2) has both $\frac{dh_2(t)}{dt}$>0 and $\frac{dh_2(t)}{dt}$<0 for t, which implies that the probability of having the event increase and decrease depends on the duration length. The third hazard function slope (h_3) has $\frac{dh_3(t)}{dt}$>0 for all t. In this case, the hazard function is monotonically increasing, meaning that the longer observations go without facing the event of interest, the more likely they will face it soon. Finally, the fourth hazard function slope has $\frac{dh_4(t)}{dt}$=0, as shown in $h_4(t)$. This implies that the probability of facing the event of interest is independent of duration (Washington *et al.*, 2003).

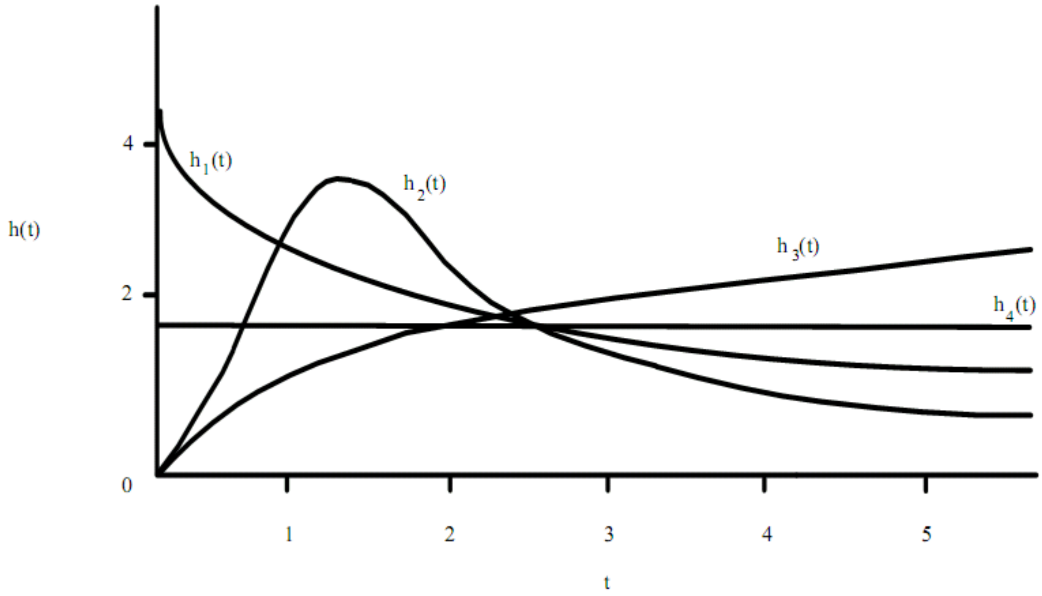


Figure 3-2 Alternative hazard function slopes (Mannering, 2007)

Further details of HBDMs can be found in specific books on the subject of survival analysis (Collett, 1994; Parmer and Machin, 1995; Klembaum, 1996; Piantadosi, 1997; Kalbfleisch and Prentice, 2002; Altman, 2003; Greene, 2003).

3.3 Models Estimation

Estimation of HBDMs can be mainly divided into three approaches: nonparametric models, semi-parametric models and fully parametric models. In this section, the differences between these approaches will be discussed, in addition to their advantages and disadvantages.

3.3.1 Nonparametric Models

Nonparametric models recognize duration as the only variable and do not consider the effects of any other variables on the occurrence of the event of interest. As the name suggests, these estimation approaches have no parameters and thus no assumption is made about the distribution of the survival time or the effect of exogenous variables (Cleves et al., 2004). Thus, the application of nonparametric models is a useful approach when there is no independent variable and no specific distribution for the survival data due to lack of theoretical support.

In general, there are two approaches for nonparametric models. The first approach is the product-limit (PL), developed by Kaplan and Meier and known as Kaplan-Meier (KM) estimator (Kaplan and Meier, 1958). KM uses individual survival times as continuous variables to indicate the probability of surviving after time t. KM estimation of the survival function is given by the product of the number of cases 'at risk' (n_j) for the event at time t_j, minus the number of cases which experience the event (d_j) at time t_j, divided by the number of cases 'at risk' (n_j) for the event at time t_j.

$$\widehat{s(t)} = \prod_{j|t_j \leq t} \left(\frac{n_j - d_j}{n_j}\right) \tag{9}$$

The second approach is based on categorising survival data into intervals and is known as a life table. Thus, in contrast to the previous estimator, this method deals with grouped survival data. The groups are presented in a table where for each group several quantities are presented, including the number of cases entered in this group, the number of cases that faced the event of interest, and the number of cases missing (Kaplan and Meier, 1958; Washington *et al.*, 2003). These quantities can be used to calculate several functions, including the survival function.

According to Lee (1992), the KM approach is considered to be the most widely used nonparametric approach because it gives an appropriate estimation of survival probability and an adequate graphical presentation of the survival distribution. However, this method has a limitation of use, especially when dealing with censoring. For example, in the case of having right censoring observation the estimation of KM will be impossible. Also, if more than half of the observation is censored, the estimation of median survival time will be difficult. Moreover, nonparametric models cannot deal properly with other issues such as continuous data and quantitative covariates (Cox and Oakes, 1984; Mannering *et al.*, 1990). Thus, such estimation models have limited application in the transportation field; however, this does not mean that these models are not useful. For instance, they can be used as preliminary assessment tools that can help to select a distribution for parametric models.

3.3.2 Semi-Parametric Models

The second estimation approach of HBDMs is the semi-parametric models. A well known example of semi-parametric models was developed by Cox in 1972 and is known as the Cox model. Compared to nonparametric models, these models are similar in that they do not make any assumption regarding hazard function shape; however, they do have a parametric assumption of the influence of explanatory variables on the hazard function. In particular, proportional hazards are assumed to be the parametric assumption of variables' effects (Cox, 1972; Box-Steffensmeier and Jones, 2004; Cleves *et al.*, 2004)). Thus, these models are referred to as semi-parametric, because as far as duration is concerned, time distribution is not specified; they are nonparametric. However, they are parametric in terms of assuming the parametric assumption of the covariate influence.

The Cox model assumes that there is a baseline hazard $h_o(t)$, and that the explanatory variables move this hazard up or down. Mathematically, the Cox model can be calculated as:

$$h_i(t)= h_o(t)e^{x_i \beta} \tag{10}$$

Where $h_i(t)$ is the hazard function, x_i is the covariate, and β is the covariate coefficient (Cox, 1972). The parameters of the covariates are estimated by means of the partial likelihood estimation method, which does not consider the baseline hazard in the estimation. The baseline hazard is estimated non-parametrically in the Cox model.

The advantage of this approach is that it is appropriate in the case of having slight or no awareness of the hazard functional form. This flexibility makes the Cox model the most frequently chosen method in survival analysis (Washington *et al.*, 2003; Garson, 2009). On the other hand, it has some drawbacks, such as some studies reporting that this approach displays a slight loss in the efficiency of estimating the parameters for exogenous variables compared to parametric models when the right distribution of hazard function is selected (Washington et al., 2003).

3.3.3 Fully Parametric Models

As the name indicates, these models require a specification of parametric assumptions in advance of both the hazard function slope and the effect of external covariates. With fully parametric models, different alternative distributions can be selected from a hazard function slope, such as exponential, Weibull, log-logistic, log-normal, and Gompertz (Table 3-1). The preference of any distribution needs to be justified by statistical assessment or on a theoretical basis. This preference has essential implications relating to the efficiency of the estimated parameters and the slope of the baseline hazard (Washington *et al.*, 2003).

Table 3-1 Alternative hazard function distributions (Box-Steffensmeier and Jones, 2004)

Distribution	Hazard Function
Exponential	$h(t) = \lambda$
Weibull	$h(t) = \lambda P(\lambda t)^{P-1}$
Log-logistic	$h(t) = \frac{\lambda \mathrm{P}(\lambda \mathrm{t})^{\mathrm{P}-1}}{1 + (\lambda \mathrm{t})^{\mathrm{P}}}$
Log-normal	$h(t) = \frac{f(t)}{\mathrm{S}(\mathrm{t})}$
Gompertz	$h(t) = (P)EXP^{\lambda t}$

Also, as mentioned in the previous section, one of the main concepts of HBDMs is that the probability of a time duration ending depends upon the length the time duration lasted. This concept is known as duration dependence (Garson, 2009). Thus, the key issue when applying fully parametric models is to choose a parametric model that provides proper duration dependence. As a result of this, these models have an advantage over nonparametric models and semi-parametric models in that they provide precise estimated parameters if the right distribution is selected (Gloder, 2008; Garson, 2009).

Among these distributions, the first four are common and will be summarized in this section. Exponential distribution is considered to be the easiest distribution to use and interpret. It is based on rejecting the concept of duration dependence (Hensher and Button, 2000). So, the hazard rate for this distribution is constant for the duration. Thus, the exponential hazard is given by:

$$h(t) = \lambda \tag{11}$$

where λ is a positive constant. Furthermore, the survivor function, density function and the mean duration are defined from equations (12), (13), and (14) respectively as:

$$S(t) = e^{-\lambda(t)} \tag{12}$$

$$f(t) = \lambda(t)e^{-\lambda(t)} \tag{13}$$

$$E(T) = \frac{1}{\lambda} \tag{14}$$

Since this distribution does not allow consideration of any type of duration dependence, it has been seen as a restrictive assumption (Hensher and Mannering, 1994; Hensher and Button, 2000).

The second distribution is Weibull. This distribution is known as a general form of exponential distribution because it allows for three types of duration dependence (Hensher and Mannering, 1994; Hensher and Button, 2000). The first type is positive duration dependence, which means the hazard rate is monotonically increasing in the duration ($h_3(t)$ in Figure 3-2). The second type is negative duration dependence, indicating that that hazard rate is monotonically decreasing in duration ($h_1(t)$ in Figure 3-2). Finally, there is a no duration dependence type, wherein the hazard rate is considered to be constant in duration ($h_4(t)$ in Figure 3-2).

The Weibull hazard is expressed as:

$$h(t) = \lambda P(\lambda t)^{P-1} \tag{15}$$

where λ is a positive scale parameter and P is known as the shape parameter. From equation (15), the hazard will be positive if P > 1, negative if P < 1, and has no duration dependence (constant hazard) when P = 1. The latter case will lead to an exponential distribution (Hensher and Button, 2000). The survivor function, density function and the mean duration for the Weibull are expressed as:

$$S(t) = e^{(-\lambda t)^P} \tag{16}$$

$$f(t) = \lambda \mathrm{p}(\lambda \mathrm{t})^{\mathrm{P}-1}\mathrm{e}^{-(\lambda \mathrm{t})^{\mathrm{P}}} \tag{17}$$

$$E(T) = \frac{\Gamma\left(1 + \frac{1}{P}\right)}{\lambda} \tag{18}$$

where Γ denotes the gamma function. Although this distribution is more flexible than an exponential distribution, it only allows for a monotonic hazard. Thus, a non-monotonic hazard may be a good alternative choice.

The third distribution is log-logistic. This distribution has an advantage over the previous distributions in that it allows a non-monotonic hazard. Thus, the hazard rate for log-logistic is:

$$h(t) = \frac{\lambda \mathrm{P}(\lambda \mathrm{t})^{\mathrm{P}-1}}{1 + (\lambda \mathrm{t})^{\mathrm{P}}} \tag{19}$$

where λ is a scale parameter and P is known as the shape parameter. From this equation, the hazard rate is monotonically decreasing if $P < 1$, monotonically decreasing from parameter λ if $P = 1$, having a non-monotonic slope and increasing from zero to a maximum at time $t = [(P - 1)^{1/P}] / \lambda$ and decreasing thereafter if $P > 1$ (Hensher and Mannering, 1994; Hensher and Button, 2000).

The survivor function for the log-logistic model is given by:

$$S(t) = \frac{1}{1 + (\lambda t)^P} \tag{20}$$

Whereas the probability density function is given by:

$$f(t) = \frac{\lambda p(\lambda t)^{p-1}}{(1 + (\lambda t)^p)^2} \tag{21}$$

Finally, the mean duration is given by:

$$E(T) = \frac{1}{\lambda} \frac{\frac{\pi}{P}}{\sin\left(\frac{\pi}{P}\right)} \tag{22}$$

The fourth distribution is log-normal. This distribution is similar to log-logistic distribution in terms of avoiding the assumption of monotonic hazard rate, which means the hazard rate rises first and then falls. The survivor function for the log-normal model can be expressed as:

$$S(t) = 1 - \Phi\left(\frac{\log(t) - \beta' x}{\sigma}\right) \tag{23}$$

where Φ is the cumulative distribution function for the standard normal distribution and $\beta' x$ are the covariates and parameter vector from the log-linear duration model. The probability density function for the log-normal is given by:

$$f(t) = \frac{1}{\sigma\sqrt{(2\pi)}} t^{-1} exp\left[-\frac{1}{2}\left(\frac{\log(t) - \beta' x}{\sigma}\right)^2\right] \tag{24}$$

where σ is the scale parameter. Then, the hazard function for the log-normal is expressed as:

$$h(t) = \frac{f(t)}{\mathrm{S(t)}} \tag{25}$$

Finally, the mean duration is given by:

$$E(T) = \exp\left(\beta + \frac{\sigma^2}{2}\right) \tag{26}$$

In order to investigate the effects of explanatory variables using HBDMs, two alternate parametric approaches can be used: the Proportional Hazard model (PH) and Accelerated Failure Time model (AFT) (Mannering, 2007). The PH model assumes that the effect of a change in covariate value on the hazard is constant over time. Also, it assumes that covariates operate multiplicatively on a baseline hazard (when all covariates vectors are assumed to be zero) due to the change in covariate value from zero to another value. Thus, in the PH model the hazard rate is written as follows (Mannering, 2007):

$$\mathrm{h(t|x_i\ )= h_0(t)e^{x\beta}} \tag{27}$$

where $h_0(t)$ is the baseline hazard and β is the estimated coefficient of a vector.

Furthermore, when applying this approach, an assumption has to be made about the baseline hazard (Steele, 2005). So, the results of this approach can be compared to the results of Cox PH models (where the assumption of covariate effects is similar and the shape of the baseline hazard left unspecified) in order to verify whether the baseline hazard has the right parameterization.

On the other hand, the AFT model assumes that the external variables rescale (accelerate) time scale. In other words, this approach assumes that covariates act multiplicatively on time scale (equation 28) rather than on baseline hazard, as is the case in PH (Washington *et al.*, 2003; Steele, 2005).

$$\mathrm{In}(t_i) = x_i\beta + \varepsilon_i \tag{28}$$

where t donates a specific value of clearance time, x is a vector of covariates, β is the vector of the estimated coefficients and ϵ is an error term. Furthermore, parametric assumption in the AFT model needs to be made about the error term (Garson, 2009). Thus, it is clear that the AFT model is a linear regression model. In other words, the AFT model regresses the logarithm of the survival time over the covariates. However, the only difference between the AFT model and the usual linear regression models is log transformation of the dependent variable (t). This is to ensure that the predicted values of time are positive (Wei, 1992; Allison, 1995; Kalbfleisch and Prentice, 2002).

Comparing AFT to PH, it can be seen that AFT focuses mainly on time analysis, whereas PH is used to gain more insight into the hazard of facing the event of interest and how this hazard changes with different covariate values. This difference yields opposite signs for the estimated parameters in both approaches; however, the significance of parameter estimate for each covariate should be the same in both metrics (Cleves *et al.*, 2004; Garson, 2009).

Finally, after explaining the possible forms of handling the effects of the explanatory variables, it should be emphasized that not all parametric distributions could be written in both PH and AFT approaches (Cleves *et al.*, 2004; Garson, 2009). Some distributions can be written in one metric only; whereas others can be written by both of them (Table 3-2).

Table 3-2 Alternative metrics of parametric distributions (Cleves *et al.*, 2004; Garson, 2009)

Distribution	Metric
Exponential	PH, AFT
Weibull	PH, AFT
Gompertz	PH
Log-logistic	AFT
Log-normal	AFT

3.4 Modelling Concerns

Following the explanation of the basic concept and estimation methods of HBDMs, some of the potential problems that may appear when applying such models must be considered. The first problem that may arise in duration data, due to its nature, is censoring. Censoring occurs when some individuals' duration data falls before or after the observation period. To illustrate this problem, consider duration data collected from seven drivers from February to October to study the time taken from the day of obtaining a driving licence until the occurrence of their first accident. As demonstrated in Figure 3-3 below, there are many types of censoring. The first type is left censoring (driver D). This type occurs if the duration's starting time for an individual (the day of having the driving licence) began before the observation period. This type of censoring causes many difficulties when applying hazard-based models because it makes the likelihood function more complex (Washington *et al.*, 2003).

The second form of censoring is right censoring (drivers B, C). It happens if the duration end of an individual lies after the end of observation time. In other words, this type of censoring occurs if the driver has an accident after the end of the study period. Compared to left censoring, right censoring is easier to handle by making a small adjustment to the likelihood function and continuing the estimation by applying standard maximum likelihood methods. Other types may combine both left and right censoring if the individual duration began and ended outside the observation period (Washington *et al.*, 2003).

Furthermore, censoring may occur due to reasons other than incomplete observation. These reasons could include the difficulty of follow-up of some individuals due to numerous causes such as withdrawal (driver E) or loss (driver F) (Washington *et al.*, 2003). Also, in some cases, the event of interest which determines the end point of time (occurrence of an accident in this example) is not clearly defined.

In an attempt to handle the problem of censoring, one of the possible solutions is to make sure that all individuals' duration data are within the time period of observation. This could be achieved by fulfilling three requirements. First of all, both the start and end points of the study period must be unambiguously identified. The second requirement is that an appropriate time scale needs to be selected. Finally, the event of interest must be clearly defined (Alison, 1984; Cox and Oakes, 1984).

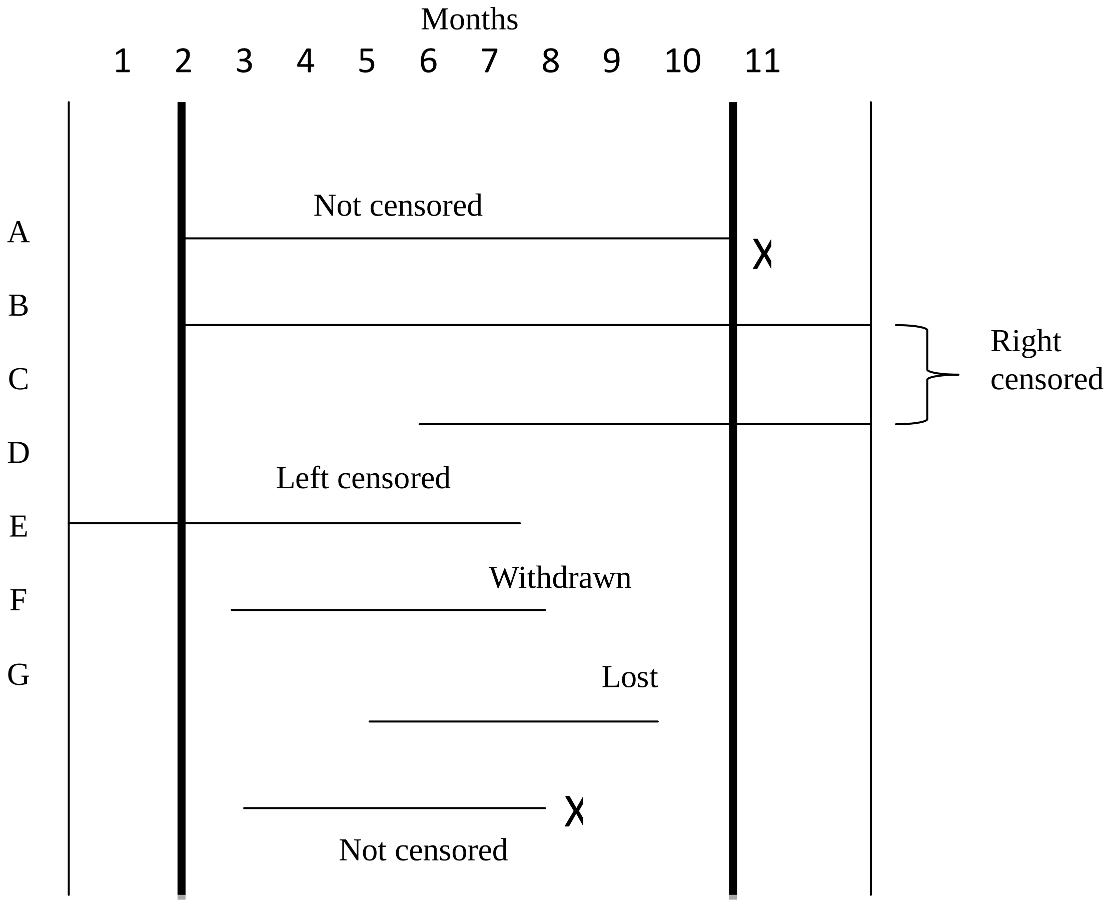


Figure 3-3 Illustrations of duration data

Another problem that may arise when modelling duration data is time-varying variables. Using the same example of studying the time until a first accident from the day of possessing a driving licence, one or more covariates could be changed over the study period, such as vehicle type. If this change has not been considered in the model, the estimated parameter could be biased. Moreover, although there are some possible ways of incorporating this problem in HBDMs, the interpretation of duration effects will remain difficult (Washington *et al.*, 2003).

Furthermore, another area of concern is unobserved heterogeneity (frailty). When using HBDMs, an implicit assumption is made that the survival distribution needs to be homogenous across all observations. In other words, it is assumed that a covariate vector captures all deviations in the time duration. This homogeneity will not appear if there are unobserved factors affecting the duration and not included in the covariate vector, causing what is known as unobserved heterogeneity (Mannering et al., 1990). Some of the reasons for not including relevant covariates can be the difficulty of measuring them or they could even be unobservable. In some occasions the analyst may not be aware that a particular covariate is a suitable one to be included in the model. As a result, failure to control for unobserved heterogeneity may yield severe problems such as incoherent estimation of coefficient and standard error, incorrect inference of hazard function shape and wrong estimation of covariate effects (Heckman and Singer, 1984; Box-Steffensmeier and Jones, 2004).

To investigate the appearance of this problem, a common approach used is introducing a heterogeneity term to capture unobserved effects in the model. The role of the heterogeneity term is to incorporate an error term into the model specification. These models are referred to as frailty models. Also, it should be noted that in traditional regression modelling, the error term shows how the expectation of duration depends on the covariates; however, in duration models the error term shows how the distribution of duration depends on the covariates (Blossfeld *et al.*, 2007). Thus, it can be seen that the focus in traditional regression modelling is different to duration modelling.

In the PH model, heterogeneity is introduced as follows:

$$h(t|x_i\ ) = h_0(t)e^{x\beta+w} \tag{29}$$

where w denotes the unobserved heterogeneity term, β denotes an unknown parameter, and x denotes the independent variable. Furthermore, this term is assumed to have a certain distribution over the population such as gamma and inverse Gaussian. Among these distributions, gamma distribution is widely adopted. However, there is rarely any justification behind the selection of a distribution. Also, it should be stressed that the selection of a certain distribution has an impact on the estimation of the model and identification of key parameters (Heckman and Singer, 1984).

In the general form of AFT models, unobserved heterogeneity cannot be introduced. This can be due to the incorporation problem. According to equation (28), it can be seen that there is an error term in this log-linear equation of the accelerated lifetime model. So, it is not possible to add another error term to the equation. This means that the heterogeneity term is not incorporated in general AFT models. However, in specific distributions, including exponential or Weibull, the heterogeneity term can be incorporated because both distributions can be written in PH and AFT metrics (Bhat, 2000). In addition, the prediction of the mean duration following the fitting of frailty model would not be possible (Golder, 2012). Review of previous research shows that there is no evidence of calculations of predicted durations once frailty (unobserved heterogeneity) has been added into their models. Also, when using Stata software to predict the mean duration following frailty model, the following error appears: "unconditional mean predictions for frailty models currently unavailable" (StataCorp, 2007). Thus, for the purpose of achieving Objective 8 of this research, unobserved heterogeneity was not considered in the models.

Finally, to minimize the appearance of unobserved heterogeneity, data collection and data analysis (specifically variable selection) should be carefully performed (Mannering *et al.*, 1990; Hensher and Mannering, 1994; Washington *et al.*, 2003).

3.5 Model Selection Specifications

As stated in section 3.2, HBDMs comprise a collection of statistical methods for analysing time until an event occurs. So, how is an appropriate approach selected for the available study data? The answer to this question is guided by some specifications that draw clear distinctions between different methods of analysing time until an event occurs. Some of these specifications are discussed in this section.

1. *Discrete versus continuous time*

Discrete time methods assume that time to event is measured in groups. Usually such models have a large time scale such as a month or year. On the other hand, models which suppose that time could be measured exactly are referred to as continuous time models (Alison, 1984). This specification in approach to measuring has an important implication when selecting the forms of basic HBDMs functions. For example, if assuming continuous time, the hazard rate is written as follows:

$$h(t)=\frac{f(t)}{s(t)}=\lim_{\Delta t\to 0}\frac{\Pr(t\leq T\leq t+\Delta t|T\geq t)}{\Delta t} \qquad (30)$$

Whereas in discrete time, it has the following form:

$$h(t)=\Pr(T=t|T\geq t) \qquad (31)$$

Thus, it can be seen that in discrete time models, hazard rate can be expressed in terms of probability; however, in continuous time models, it is the rate (Alison, 1984; Chatterjee and Ma, 2008).

2. Parametric versus semi-parametric and nonparametric methods.

The choice between these approaches plays a major role in data analysis. This choice can be justified by the level of knowledge with regard to the underlying hazard distribution. If hazard distribution is known based on statistical test or theory, then fully parametric methods are the proper approach. In contrast, when there is a lack of knowledge, a nonparametric approach is used. Between these two methods, a semi-parametric approach is considered the best when little is known about the underlying hazard distribution. Furthermore, due to the advantages and drawbacks of each method, there is a substantial amount of work in the literature devoted to each method (Alison, 1984; Washington *et al.*, 2003).

3. Proportional Hazard (PH) versus Accelerated Failure Time (AFT).

In HBDMs, the common approaches to account for the external covariates' effects are PH and AFT. Although both methods aim to investigate the covariates' effects, specifying one of them is necessary because of the effects of covariates being different between them. In a PH model, the effects of covariates are to multiply the hazard rate by a factor; however, in an AFT model the effects of covariates are to multiply the mean duration by a factor (Jenkins, 2005; Kleinbaum and Klein, 2005). Also, when specifying any method, consideration needs to be given to the fact that not all distributions can be written by means of both methods, as explained in Table 3-2 (Chatterjee and Ma, 2008).

4. *Baseline hazard function shape.*

In fully parametric models, there are many assumptions that could be made regarding the distribution of the baseline hazard. Selecting the best distribution is critical to yield proper estimated coefficients for the covariates and model parameters (Garson, 2009). Thus, failing to do so will create a bias in the interpretation of the estimated coefficients. Additionally, selecting the best distribution could be aided by running some goodness-of-fit tests such as a likelihood ratio test and Akaike Information Criterion (AIC) (Cleves *et al.*, 2004; Garson, 2009). More details of these tests are presented in Chapter Four.

3.6 Application of HBDMs in Transport

A considerable amount of literature has been published on the application of HBDMs in the field of transportation. These applications are varied in their objectives and methodologies. In addition to the application of these methods for accident duration analysis, these methods were applied in different transportation areas. The Weibull PH model was applied in order to examine the relationship between drivers' gender characteristics and accident risk (Mannering, 1993). Mannering's study of University of Washington drivers yielded significant results relating to different gender characteristics and accident risk. Further work applied the Cox PH model to examine what factors affect the occurrence of mini-bus accidents (Hamed *et al.*, 1998). Data were collected from 438 mini-bus drivers in Jordan. This paper demonstrated that the more time that passes without having an accident, the less likely an accident will occur soon.

Several studies have analysed the time between households' vehicle purchases. In 1992, Gilbert published a paper in which he aimed to estimate what affects automobile length ownership using a Weibull model. Information concerning car, household and macroeconomic characteristics of 7500 households was utilized to estimate this model. The length of ownership was measured in months and three events of interest were used to determine the end of ownership duration, including replacing with a new vehicle, disposal and replacing with a used vehicle (Gilbert, 1992). Further research fitted a Weibull model with and without gamma heterogeneity and time-vary covariates using 2745 observations to model vehicle holding duration (Jong, 1996). Moreover, a non-parametric approach was used to examine the holding duration of household vehicles. The results show that many of the independent variables used in this research have a significant effect in vehicle holding duration (Yamamoto and Kitamura, 2000).

In addition, numerous studies have attempted to study the time between trips by applying different HBDMs (Mannering and Hamed, 1990; Hamed and Mannering, 1993; Ettema *et al.*, 1995; Bhat, 1996; Niemeier and Morita, 1996; Wang, 1996; Kitamura *et al.*, 1997; Kharoufeh and Goulias, 2002; Bhat *et al.*, 2004).

Furthermore, other transportation topics were examined using HBDMs, such as safety of motor carrier operation (Lin *et al.*, 1993), predicting vehicular delay (Wei, 1992), rhythms of travel (Schönfelder and Axhausen, 2000), estimating congestion duration (Stathopoulos and Karlaftis, 2002), household vehicle transaction behaviour (Jiezhi Qi, 2009), household weekend activity (Zhong and Hunt, 2005), the response duration to new transport service (Chatterjee and Ma, 2008), examining repeat offences (Lapham *et al.*, 2006), investigating the factors that influence the duration of driving under the influence of alcohol accidents (Fu, 2008), survival of motor-vehicle and motorcycle riders after a crash (Bradburn *et al.*, 0000), predicting the risk associated with overtaking behaviour (Heckman and Singer, 1984), daily travel time (Reid, 1994) and discharge headway at signalized intersections (Liu *et al.*, 2011).

3.7 Summary

This chapter has reviewed the basic theoretical and mathematical concepts of HBDMs. The approaches of three HBDMs were illustrated in terms of their advantages and disadvantages of application. Furthermore, different issues related to modelling were demonstrated, such as censoring and time vary variables. Selection criteria between the approaches of HBDMs were explained in more detail. Finally, the applications of HBDMs in the transportation field were presented. This showed that the applications of HBDMs emerged as promising methods for use in transportation studies.

4 Research Methodology

4.1 Introduction

The Literature Review presented an overview of the different methods of modelling traffic accident duration. Among the available methods, HBDMs were found to be a promising approach for conducting this study. This chapter covers the details regarding the case study area and the research methodology.

4.2 Study Area

The initial plan of this research was to analyse traffic accident duration in the Newcastle urban area. However, due to some difficulties in accessing the required data, it was decided to consider Abu Dhabi, the capital of the UAE, as a case study area for this research. This section presents an explanation of the initial study area and the revised study area.

4.2.1 Initial Study Area

This section explains the initial approach used to collect traffic accident duration data and accident characteristics from the Newcastle urban area. Newcastle City Council CCTV System, Northumbria Police logs, and Traffic and Accident Data Unit (TADU) records were to be used to develop a traffic accident duration database for this research.

Before explaining the proposed methodology of developing incident duration data, it is worth mentioning the general difficulties of collecting such data. Deciding upon the exact time of incident occurrence is not an easy task unless video recording evidence is available. In most cases, this data is gathered by police interview with drivers involved in the incident. To overcome this problem, the accuracy of the length of these data was to be investigated using the available CCTV System of Newcastle City Council.

On the other hand, accident characteristics data were to be retrieved from the TADU, which collects all data related to traffic accidents occurring in the Tyne and Wear area. The selection of such characteristics was to be based on the Literature Review and data availability.

For the purpose of checking the validity of the data collection approach, a pilot study was carried out in the Newcastle City area during January and February, 2009. It began by receiving initial information from Northumbria Police relating to the latest accidents in the Newcastle area, such as time of the day, day of the month, area code and location. The data were then used to review CCTV System recording and traffic flow data from Newcastle City Council. In this investigation, it was not possible to measure the length of incident duration or collect incident characteristics because of several problems. First of all, some of the available cameras were not working at all, while others returned live video, without recording. Thus, only a few cameras were suitable for use in reviewing traffic accidents. Another problem was that location data provided by Northumbria Police did not have the exact co-ordinates of accident scenes and were limited to street names. Therefore, since all cameras and traffic count sites are located on certain road segments, it was difficult to determine at which road segment the accident had occurred in order to select the adjacent camera or traffic count site. Furthermore, some traffic flow data were missing due to a failure on the traffic count sites. Finally, most of the data provided were related to accidents that happened on roads where no nearby camera or traffic count site was available. These issues made checking the validity of police information related to traffic accidents extremely problematic and an alternative approach had to be determined.

In response to this shortage of incident duration data, a visit to Northumbria Police was arranged on February 16th, 2009, to check the possibility of obtaining more data regarding incident occurrence time, responders' arrival times and end of clearance time. This meeting revealed that all of the required data would be accessible from the Incident Logs. However, permission from the Operations Department is required to gain access to these logs.

As a result of the meeting with Northumbria Police, several communications were made with the Operations Department Chief Inspector and the Chief Constable of Northumbria Police. The request was sent to the Disclosure and Data Protection Unit for advice. They stated that the request would not be granted under the provisions of the Freedom of Information Act:

> 'This section does not oblige a public authority to comply with a request for information if the authority estimated that the cost of complying with the request would exceed the appropriate limit of 18 hours, equating to £450.00'

Thus, unfortunately, it was not possible to get any further information on accident duration from Northumbria Police. As a result of this, the decision was made to change the study area and to start collecting data from Abu Dhabi. Full details of the Abu Dhabi area and data availability are presented in the following subsection.

4.2.2 Revised Study Area

Abu Dhabi is the capital of the United Arab Emirates, with a population of 1,463,491 in 2006 (Authority, 2009). It comprises 80% of the total land area of UAE, with only 30% of the population. Abu Dhabi has a modern road network with dual three-lane carriageways linking the main cities such as Al Ain and Tarif. In the urban areas, there are a large number of roundabouts and signalised intersections designed and constructed to high standards (Al kathairi *et al.*, 2001).

Although Abu Dhabi has a good road infrastructure, road traffic accidents are still growing as one of the main concerns of public health. According to the Abu Dhabi Health Authority, road traffic injuries caused 68% of the total injury-related deaths in 2008 (Abu Dhabi Health Authority, 2009); see Figure 4-1. A study comparing road traffic accidents in the UAE to some regional and Western countries in 2002 found that the UAE had the highest rates (Figure 4-2) (Bener and Crundall, 2005). Clearly, these facts necessitate more efforts and co-operation from the local authorities to reduce these dangerous rates and to reduce the adverse impacts of road traffic accidents. One of the possible ways to do so is by improving the efficiency of the TIM process.

4.2.2.1 Introduction to the Abu Dhabi Serious Collision Investigation Section (ASCIS)

The Abu Dhabi Serious Collision Investigation Section (ASCIS) is a division of the Abu Dhabi Police Traffic and Patrols Department. The main role of ASCIS is to investigate all kinds of serious collisions that resulted in fatalities or casualties and occurred within the boundaries of the Abu Dhabi Emirate. This section has four branches that cover the Abu Dhabi City urban area (1), Abu Dhabi Highways (2), Eastern Region of Abu Dhabi (3), and Western Region of Abu Dhabi (4) (Figure 4-3).

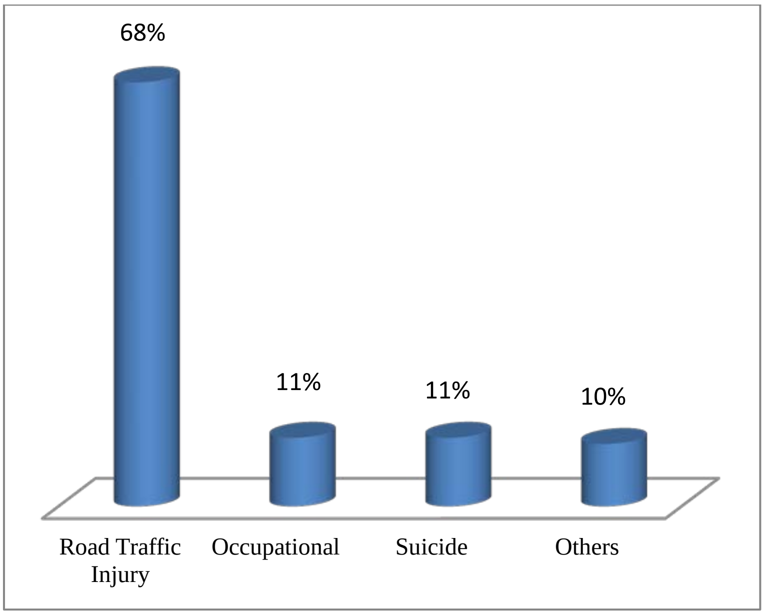


Figure 4-1 Abu Dhabi injury-related deaths in 2008 (Abu Dhabi Health Authority, 2009)

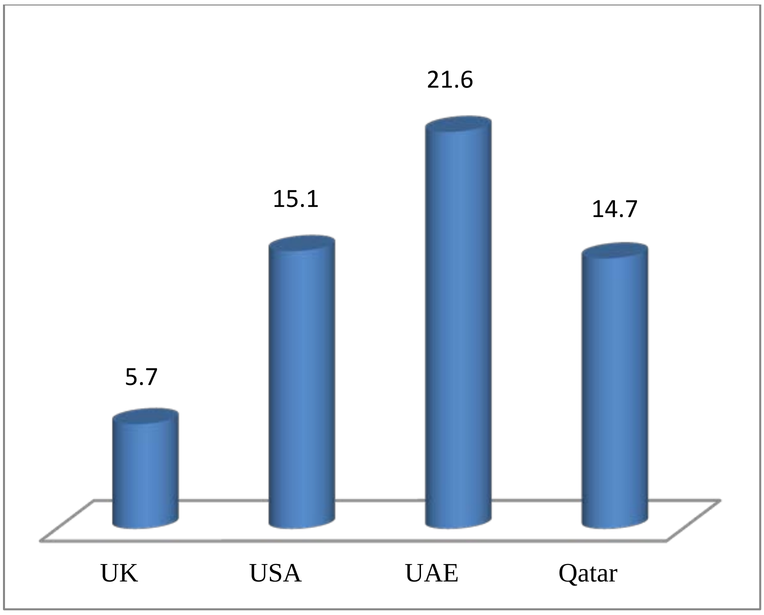


Figure 4-2 Fatality rate per 100,000 population (Bener and Crundall, 2005)

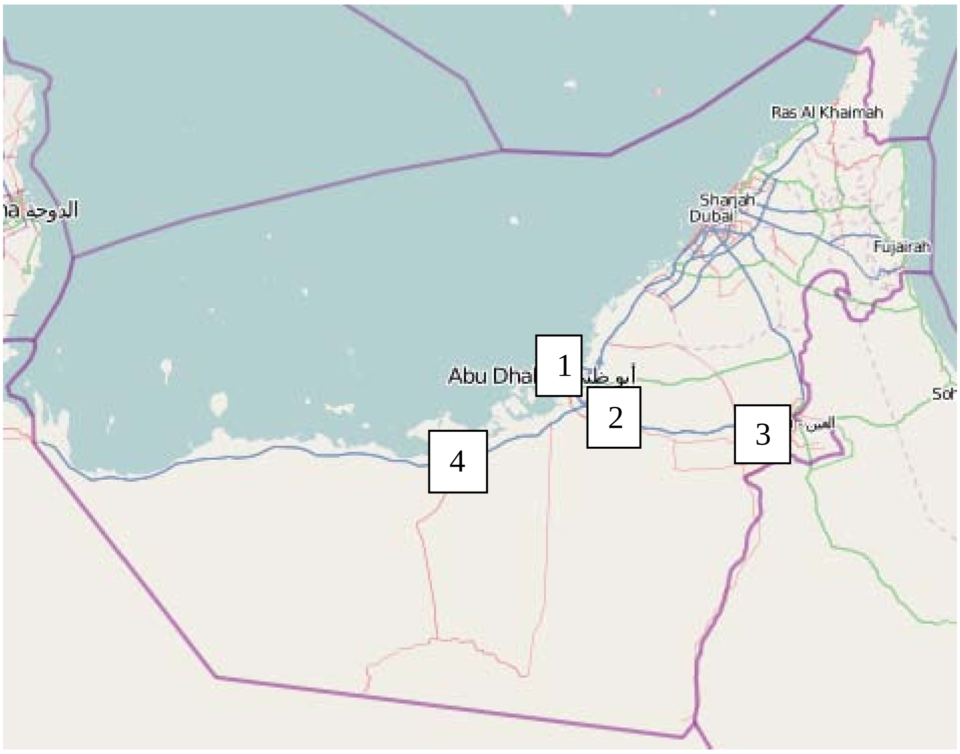


Figure 4-3 Location of ASCIS branches

The operational process of Traffic Accident Management has many stages (Figure 4-4). The process starts when the Police Operational Centre is notified of a traffic accident. A comprehensive police patrol is then assigned to assess accident severity based on the accident type. If it is a property damage or 'damage only' accident, a private company called SAAED will be contacted to carry out the investigation. However, if the accident is found to be serious (such as an injury or fatal accident), ASCIS will move to the scene to begin a comprehensive investigation. Upon arrival at the scene, the collision investigator is required to perform several duties, including preserving the accident scene, detaining suspects if known, securing evidence, drawing a sketch map of the final accident scene, recording witness statements and filling in an accident report. In addition to ASCIS, different types of responders will be dispatched to the accident scene, including the Ambulance and Rescue Service to move injuries, and the Traffic Control Centre to observe the traffic flow and apply traffic diversions when necessary. The Crime Scene Department deals with suspicious cases and the Fire Department clears the debris.

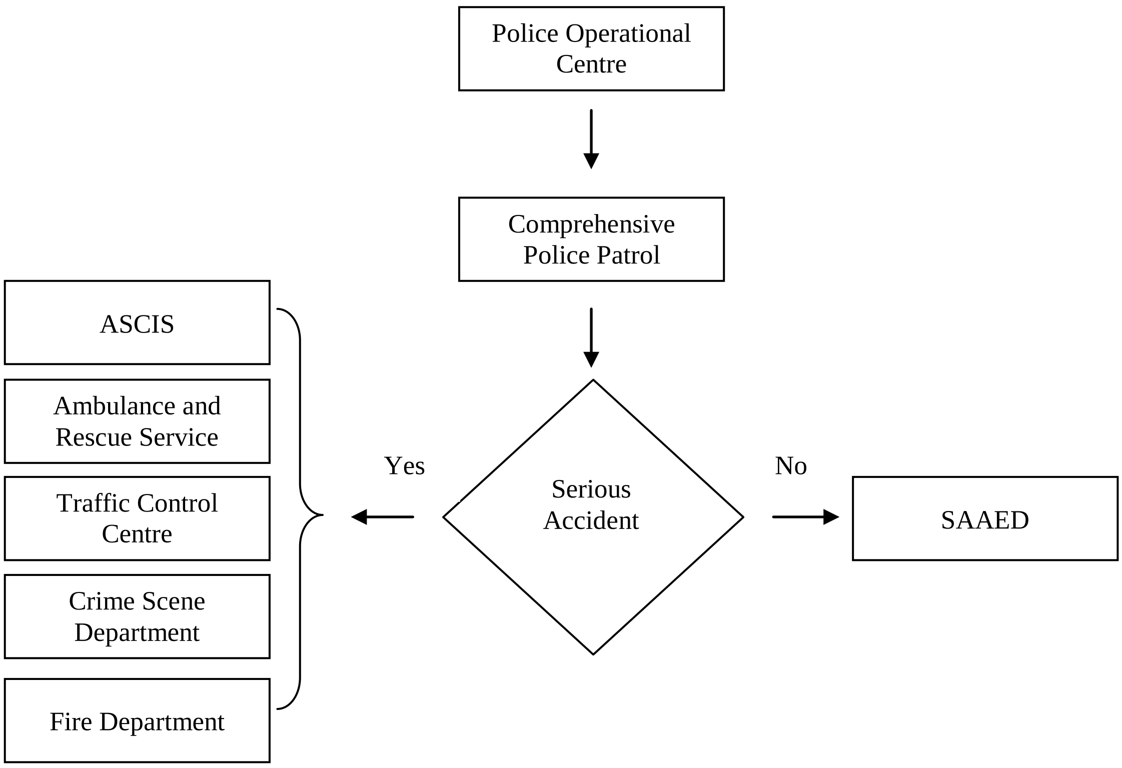


Figure 4-4 Operational process of traffic accident management

4.2.2.2 Traffic Accident Databases

Two types of data were required to conduct this research, including accident characteristics and accident duration. The required data were extracted from two databases, the UAE Federal Traffic Statistics System (FTSS) and the Abu Dhabi Serious Collision Investigation Section (ASCIS). The FTSS is operated by the Ministry of the Interior (MOI), which is located in Abu Dhabi and has comprehensive data of traffic accidents that have occurred all over the country. Each Emirate's Traffic Department sends the details of traffic accidents that occurred within their regional sector to FTSS on a monthly basis. Examples of such data include personal details of drivers and injuries, as well as temporal, geographical, environmental and accident characteristics. Further details of the extracted explanatory variables of this database are expounded in Chapter 5.

On the other hand, accident time was collected from the records of two branches of ASCIS including the Abu Dhabi Urban Collision Investigation Branch (AUCIB) and the Abu Dhabi Highway Collision Investigation Branch (AHCIB). These records contain the values of some of the accident interval times. However, it should be noted that for urban traffic accidents, it was possible to collect three interval times, including reporting time, response time and clearance time, whereas for traffic accidents on the major highways only response time was available. Unfortunately, this is because only response time was required by ASCIS staff for the purpose of performance evaluation. Thus, an accident data collection form (Appendix 1) was developed and given to ASCIS staff. However, only the Abu AUCIB staff were supportive in using the form to collect time interval data, and not the AHCIB staff.

4.3 Data Collection and Preparation

4.3.1 Data Collection

Research methodology in this research consists of four stages, including data collection and preparation, preliminary analysis, data analysis, and duration prediction (Figure 4-5). Data collection is a vital part of the research approach in this study. In general, data required to develop HBDMs can be classified into two categories: dependent variable and explanatory (independent) variables. The dependent variable will be determined based on the study aim, whereas the explanatory variables will be selected based on previous studies and data availability.

In HBDMs, the dependent variable refers to the time variable of incident, in other words duration variable. As mentioned earlier in section 2.4, the definition of incident duration was found to vary from one study to another, based on the particular aims and objectives of different research projects. Thus, this is highly dependent on the purpose of the study.

In this study the total incident duration is divided into three time intervals:

1. Reporting time: the time in minutes between the incident occurrence and the responder receiving the call to respond to the incident.

2. Response time: the time in minutes between receiving the call and responding to the incident, and the arrival of the first responder (accident investigator) at the scene.
3. Clearance time: the time in minutes between the arrival of the first responder (accident investigator) at the scene and the last responder (accident investigator) departing from the scene.

It is important to collect the reporting, response and clearance times using an accurate approach to avoid censoring issues. This could be achieved as discussed in section 3.4 by clearly defining the start point and end point of each interval time with an appropriate time scale. For the purpose of this study, the start point, the end point and the time scale are clearly stated in the definitions of the interval times.

Another issue is that the event of interest, which determines the end of the interval time and the beginning of the following interval time, needs to be clearly defined. As mentioned in the definitions of the interval times, the event of interest for reporting time is receiving the call regarding the incident; the event of interest for response time is the arrival of the first responder; and the event of interest for clearance time is the departure of the last responder.

The explanatory data generally includes traffic accident characteristics. These characteristics may include variables such as time, location, severity, reporting mechanism and investigation mechanism of incidents. In addition, there are common characteristics that can be used for each interval time of the total incident duration, namely: location-specific data, time of day and day of the week. However, to obtain a clear insight into duration dependence, some interval times require specific information. For example, reporting mechanism may have an effect at reporting time interval. Also, clearance time could be affected by the investigation method and related mechanisms used for clearing incident sites. Thus, collecting interval related information is necessary to avoid misinterpretation of the model results.

Previous research has shown that numerous variables can be used to estimate or analyse incident duration. It was found that there are some common variables that have been used in previous studies, including incident type, location, the number of affected lanes, weather conditions, incident time, and the number of vehicles involved. Table 4-1 presents a summary of explanatory variables used by previous research in the area of incident duration.

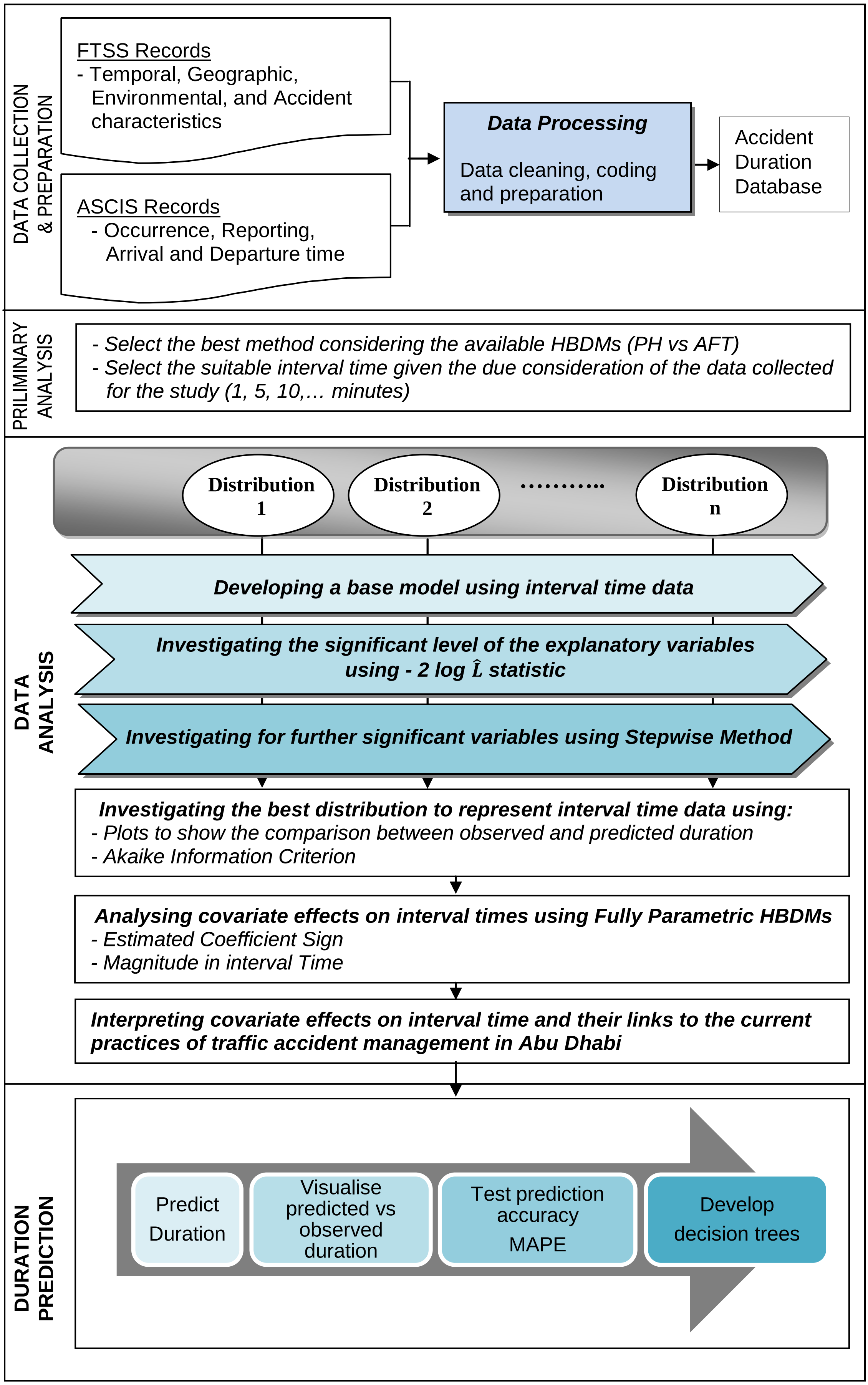


Figure 4-5 Methodology framework

Table 4-1 Explanatory variables of incident duration

Study	**Explanatory Variables**
(Sullivan, 1997)	Freeway characteristics, incident type, times, location and traffic volumes.
(Garib *et al.*, 1997)	Time of the day, police response time, weather and number of vehicles involved.
(Jones *et al.*, 1991)	Season, time of day, special events, driver and vehicle characteristics, accident severity measures and location.
(Khattak *et al.*, 1995)	Incident type, vehicle type, number of vehicles involved, injuries and fatalities, property damage, response time, number of responders, weather condition, incident location, seasonal factors, flow conditions and motorist information.
(Nam and Mannering, 2000b)	Temporal characteristics, environmental characteristics, geographic information, incident characteristics and lead agency information (clearance time only).
(Sethi *et al.*, 1994)	Roadway type, incident type, incident severity and traffic conditions.
(Lee and Fazio, 2005)	Crash severity, average daily traffic, day of week, number of vehicles involved in crash, light conditions, number of lanes, on- or off-freeway location, posted speed limit, road condition, work zone present, heavy vehicle involvement, urban or rural area and weather conditions.
(Wei and Lee, 2007)	Incident characteristics, geometry characteristics, special relationship and time relationship.
(Smith and Smith, 2001)	Time of day, day of the week, weather condition, number of vehicles involved, vehicle type and response type.
(Kim and Choi, 2001)	Type of incident vehicle, incident service time, type of vehicle and location of incident vehicle.
Pal et al, 1998	Type of incident, the position of the incident (lane, ramp or shoulder) and the time of day

Considering the explanatory variables used in previous studies, this study recognises the importance of using some common variables in the analysis. The appropriate variables were determined after examining the available data of traffic accidents in the study area for both urban accidents and highway accidents. Upon completion of the data collection process, the data were entered into a database before moving on to the data preparation stage.

4.3.2 Data Preparation for the Analysis

Data preparation aims to organise the data in a way that facilitates conducting survival analysis. It consists of three steps, beginning with data coding, followed by data declaration and finally, data examination. Before explaining these steps in detail, it is worth mentioning that data preparation and data analysis are going to be performed using Stata 10 software.

For the purpose of entering data, it is necessary to develop a coding system. However, prior to explaining this system, it should be noted that this system was only applied to the explanatory (independent) variables. This is because the dependent variable (time or duration) is measured in a continuous scale, therefore coding is not required. Also, it is worth mentioning that each accident was recorded from the beginning of each interval time until the end of that time. Thus, no censoring problem exists in the dependent variable, which makes recording survival data easier in one variable.

Furthermore, some explanatory variables were separated into sub-groups in order to investigate how they affect accident duration. For example, time of day was divided into three periods in Abu Dhabi: (1) morning, 12:01am-12:00pm; (2) afternoon, 12:01pm-04:00pm; and (3) evening, 04:01pm-12:00am. Also, three additional sub groups were included in the database to find out whether the accident occurred either within the peak periods (AM peak: 06:00-08:00, PM peak: 14:00-16:00) or in the off peak periods.

The first step was to develop a coding system using Stata 10. This process starts by assigning a variable name to each independent variable. Then, the process of labelling the values of each independent variable was carried out in two phases. The first phase is to save the text and its value. This is known as labelling mapping, which consists of two texts with their values in this study, including ‘Yes=1’ and ‘No=0’. The second phase is to allocate labelling mappings to each independent variable.

Following data coding, the second step is data declaration, which is to clarify the dependent variable that represents the survival time in Stata 10. This stage is important in order to avoid repeating this process when making any survival command. It may consequently save a considerable amount of time when analysing data. The last step of data preparation is data examination. This step aims to check the suitability of the data for analysis. More details of applying HBDMs in Stata 10 are presented in section 4.7.

4.4 Preliminary Analysis

As mentioned in the Literature Review chapter, there are two approaches to investigate the effects of explanatory variables in HBDMs: Proportional Hazard (PH) and Accelerated Failure Time (AFT) models. So, in the second stage of the methodology, a preliminary analysis will be conducted to select the most appropriate analytical method to model accident duration (see Figure 4-5). Also, preliminary analysis focuses on selecting the suitable interval time for the study. To achieve this, a dataset of highway accidents in Abu Dhabi will be used.

4.5 Data Analysis

The third stage of the research methodology is data analysis. During this stage four steps were conducted, including developing the base model, investigating the best distribution, analysing covariate effects and interpreting covariate effects. This section explains these steps in more detail.

Data analysis commenced by developing a base model for each interval time of the total accident duration, including reporting time, response time, and clearance time. Separate models were used because this research aims to determine the factors that influenced each interval time of accident duration. The analytical approach proposed here is important to assist accident responders to obtain more insight into what affects each interval time of the total accident duration, which may assist in improving various elements of the TIM process. Although it is possible to look at the whole duration as one, this approach would lose information in both estimation and interpretation of results (Nam and Mannering, 2000a).

A fully parametric approach was especially selected in this study to investigate accident duration. The merit of using a fully parametric approach, as opposed to semi-parametric, is that the latter does not produce a parameter that tells the shape of the baseline hazard, making it difficult to determine duration effects (Nam and Mannering, 2000a).

Before developing HBDMs, it is necessary to select explanatory variables for each model (reporting time model, response time model and clearance time model). To identify the most relevant variables, three steps, as shown below, were conducted (Collett, 2003):

1. The first step was to analyse the models considering Weibull, Log-normal and Log-logistic distribution without any explanatory variables in order to check the value of the log-likelihood before convergence. This is referred to as a null or base model.
2. The second step aims to identify which explanatory variables significantly reduce $-2\ log\hat{L}$ statistic. This is done by considering one variable at a time. In this step, all variables that are significant at the level of 85% were selected.
3. The third step aims to check whether any of the excluded variables in step 2 are significant in the model. Since any of the excluded explanatory variables from the initial model could be significant when put back into the model, the models of the significant variables were fitted with one of the excluded variables at a time. In this step, all variables significant at the level of 90% were selected.

As a result of this, the structure of model development was as follows:

Model 1: Reporting Time

The dependent variable in this model is the time until an accident is reported to the collision investigator. In other words, the model relates to the time until the reporting of an accident, given that the accident has not been reported up to time *t*. The hazard function is defined as the rate at which the accident is being reported at time *t*, given that no accident has been reported until time *t*.

Model 2: Response Time

The dependent variable in this model is the time until an accident is responded to by a collision investigator. In other words, the model relates to the time until the response to the accident, given that the accident has not been responded to up to time *t*. The hazard function is defined as the rate at which the accident is responded to at time *t*, given that no accident has been responded to until time *t*.

Model 3: Clearance Time

The dependent variable in this model is the time until an accident is cleared. So, the model relates to the time until the clearance of the accident, given that the accident has not been cleared up to time *t*. The hazard function is defined as the rate at which the accident is being cleared at time *t*, given that no accident has been cleared until time *t*.

Once the models have been estimated, the plots that compare observed and predicted durations for the incidents in the dataset will be presented for each interval time. These comparisons will be conducted for the three distributions used in this study. Following that, a goodness-of-fit test should be carried out to select the best fit distribution prior to the interpretation of results. In fact, several tests can be conducted, such as a Likelihood-ratio test, Wald test and Akaike Information Criterion (AIC). The first two tests are appropriate when the models are nested, such as Weibull against Exponential, whereas the last test can be selected when the models are not nested. Because the models used in this study are not nested, AIC was used. Akaike (1974) proposed 'penalizing each model's log-likelihood to reflect the number of parameters being estimated and then comparing them'. This test can be written as the following:

$$\text{AIC}=-2\text{InL}+2(k+c) \tag{32}$$

Where In *L* refers to the model's log-likelihood at convergence, k denotes the number of covariates in the model, and c is the number of distribution parameters. The criteria is to select the distribution that has the lowest value of AIC (Akaike, 1974; Cleves *et al.*, 2004). It should also be noted that this test should be conducted independently for each sub-model.

Following the fitting of the best-fit distribution, the outcomes of the selected model are interpreted based on the sign of the estimated coefficient and the percentage change in duration. In the AFT model, the sign of the coefficient specifies how the variable affects the interval time duration. For example, a positive coefficient in the clearance time model means that the variable increases the clearance time duration (Gloder, 2008).

On the other hand, the percentage change in the interval time by each of the explanatory variables can also be calculated. This could be done by taking the exponent of the estimated coefficient of the significant variable (Chung, 2009; Nam and Mannering, 2000). Generally, when the exponent of the estimated coefficient is greater than 1.0, the relevant explanatory variable adds more time to the respective accident interval time (reporting, response, and clearance) and vice-versa. Finally, the estimated results were interpreted in terms of the covariate effects on the interval time and their relation to the current practices of the TIM process in Abu Dhabi.

4.6 Duration Prediction

In order to achieve objective 8, the fitted models will be used to predict the mean duration of each interval time for any given set of values of the covariates. Then, plots will be developed to visualise the predicted duration with the observed duration. Furthermore, to measure how accurate the models are in predicting the incident duration the Mean Absolute Percentage Error (MAPE) will be used. MAPE is a widely used approach to measure prediction accuracy. MAPE can be calculated using this equation

$$\mathrm{MAPE} = \frac{1}{\mathrm{M(t)}} \sum_{\mathrm{K}=1}^{\mathrm{M(t)}} \left| \frac{\hat{\mathrm{x}}(\mathrm{k}) - \mathrm{x}(\mathrm{k})}{\mathrm{x}(\mathrm{k})} \right| \times 100\% \tag{33}$$

where M = number of samples (e.g.560 accidents), $\hat{x}$ forecasted (predicted) incident duration, and x is actual incident duration. The scale rates of accuracy evaluation in MAPE are illustrated in Table 4-2 (Lewis, 1982; Wei and Lee, 2007).

Table 4-2 MAPE levels of accuracy prediction

MAPE	Assessment
Less than 10 %	Highly accurate forecasting
11% - 20%	Good forecasting
21% - 50%	Reasonable forecasting
51% or more	Inaccurate forecasting

Following the MAPE test, a decision tree will be developed as a decision-making tool to predict accident duration. As mentioned in section 2.6.5, the advantage of this method is that no assumption of probable distribution is required. Therefore, this can be identified as a useful approach to predict accident duration in comparison to other approaches. Also, this method consists of a series of decision variables and the outcome represents the average incident duration of a specified dataset (Smith and Smith, 2001). In addition to MAPE, model fit was assessed through R^2 and root mean square error (RMSE).

Decision trees were developed with respect to each interval time (for instance reporting, response and clearance times) of the total duration, considering the significant variables that affect each interval time. Thus, the size of decision trees was based on the number of significant variables. In other words, the number of levels per decision tree was made according to the number of significant variable categories. As the significant variables that affect each interval time vary, the decision tree for each interval may have considerable differences. It is expected that the operator selects the relevant decision tree based on the interval time and details of the accident.

4.7 Selection of Computer Software for Modelling

Several software packages can be used to estimate traffic accident duration in HBDMs, such as R, SPSS, STATA, SAS, LIMDEP, SUDAAN, MLwiN, and S-Plus (Box-Steffensmeier and Jones, 2004). In this research, the software package Stata 10 was used to estimate HBDMs because of the availability of this software in the school. This section intends to illustrate some of the main commands used in Stata for the purpose of data declaration, data preparation, data examination, and the application of a fully parametric AFT approach.

In Stata, the data declaration is conducted using '*stset*' command. When all accidents faced the event of interest by the end of the interval time, the command that will be used is '*stset variable*'. Here the '*variable*' refers to the dependent variable of the model for each interval time, such as reporting time, response time and clearance time.

After running '*stset*' command, four variables will appear in the dataset, including *_t0*, *_t*, *_d*, and *_st*. The variables *_t0* and *_t* record the start time and end time for each accident in minutes. Each accident interval time starts at *_t0* and concludes at *_t*. The variable *_d* denotes the outcome at the end of each interval time, where 1 represents the interval time that ends in failure (meet the event of interest), while 0 represents when it does not end. The variable *_st* reports whether an accident is relevant to the analysis. For each accident, the variable contains 1 if the accident is to be used and 0 if it is to be ignored. Finally, all variables *_t0*, *_t*, *_d*, and *_st* were used to run the analysis while ignoring initial variables such as time and event (Acock, 2006).

In the data preparation step, it is necessary to apply several commands prior to commencing data analysis. The first command is '*stsum*'. The purpose of this command is to give a summary of the survival time. In the output, several types of information could be extracted, including time at risk, incident rate, number of subjects, and the 25th, 50th, and 75th percentiles of survival time (see Table 4-3).

Table 4-3 Summary of the survival time

	Time at risk	Incident Rate	Number of Subjects	[…………Survival Time……….]		
				25%	50%	75%
Total	1596	0.2274436	363	3	4	6

The second command is '*stdes*'. This command describes the survival time that has been set and detects whether any gap exists (Table 4-4). Thus, the most important result of the outcome is if subjects with gap are shown.

Table 4-4 Description of the survival time

Category	total	mean	min	median	Max
no. of subjects	363				
no. of records	363	1	1	1	1
(first) entry time		0	0	0	0
(final) exit time		4.396694	2	4	7
subjects with gap	0				
time on gap, if gap	0				
time at risk	1596	4.396694	2	4	7
Failures	363	1	1	1	1

It can be seen from the output that no subjects with gap were found. This indicates that the survival time of each accident was recorded from the beginning of each interval time until facing the event of interest. So, the same result should appear in the outcome, because all accidents were followed from the start point and end point of each interval time. However, if the results show that there are some subjects with gap, then this indicates an error in data entry, and this should be corrected before conducting the analysis.

The last command of data examination is '*stvary*'. This command reports whether the explanatory variables are changing over time and displays whether there are any missing values of any independent variable (Table 4-5).

Table 4-5 Variation of the explanatory variables

Variable	Constant	Varying	Never Missing	Always Missing	Sometimes Missing
Road condition	363	0	363	0	0
Severity	363	0	363	0	0
Visibility	363	0	363	0	0

Out of the three independent variables used in this example, it is clear that no accident has a variable that changes over time, and no accident in the dataset has missing values for these variables. These results indicate that complete data of these variables were collected and entered for each accident. Since the collected independent variables in this study are constant over time, the varying column should be 0 in all independent variables. Also, because all values of independent variables were collected completely from the accident report, none of the output should show 'always missing'. However, if any variable is found to be missing, it is necessary to check for the missing value of the dataset and edit it before carrying out the analysis (Acock, 2006).

Using a fully parametric AFT approach means that there are many alternative distributions available to assume the shape of the distribution of the times about the mean, such as Exponential, Weibull, Log-logistic, and Log-normal. In Stata, the selection of these distributions can be done by means of the '*streg'* command followed by any distribution. For example, if the distribution assumed is to be Weibull, the command is '*streg var1, dist(Weibull) nohr*', where '*var1*' denotes the explanatory variable, and '*nohr*' denotes the use of coefficients instead of hazard ratios. Furthermore, since most of these distributions have a scale parameter and shape parameter, the '*streg*' command is used to calculate these parameters' values. All of these values are computed using maximum likelihood estimation, which is the main role of the '*streg*' command. Finally , the command " predict t_mean, time mean" is used to predict the mean duration with consideration of the significant variables at each interval time (StataCorp, 2007). The results will be used to develop a decision tree.

4.8 Summary

This chapter begins by describing the initial and revised study areas. Subsequently, it presents the methodology of modelling each interval time of the total traffic accident duration, starting by collecting the required data from FTSS and ASCIS databases. In addition, the approaches to selecting the best-fit distribution and result interpretation are illustrated. Finally, the methodological approach in developing an accident duration prediction tool using the decision tree is presented. Further details of data description and the study area are discussed in Chapter 5.

5 Data Description

5.1 Introduction

All traffic accident data in this research were collected from the Federal Traffic Statistics System (FTSS) and Abu Dhabi Serious Collision Investigation Section (ASCIS) in Abu Dhabi, the capital of the United Arab Emirates (UAE). These data were collected from both the urban network and highways. This chapter begins by describing the urban traffic accidents data in section 5.2. Then, section 5.3 will present an in-depth description of highway traffic accidents data, followed by a summary of the chapter in section 5.4.

5.2 Urban Traffic Accidents Data

This study is based on the metropolitan network in the city of Abu Dhabi. Roadways covered in this research have one of three classifications: Primary Roads (freeways, expressways); Secondary Roads (arterials, collectors); and Local Roads. All the main intersections are signalized in the metropolitan network (Abu Dhabi Municipality, 2009).

For the purpose of analysing urban traffic accident duration, a dataset containing a total of 525 accidents was collected over a one-year period starting from May 2009. As described above, the main database used to extract accident data was the FTSS database. To achieve the aim of this study, considerable information was used as candidate variables for the AFT model. For example, geographical information included the location of the accident in terms of the street, intersection, region, place and road layout. These variables were included as they may influence the travel time of responders and the investigation time. Another example is accident type. This variable was included because each kind of accident may require different clearance equipment, possibly affecting clearance time.

Also, some data were divided into more variables to understand the relationship between the interval time and data. For instance, time of the day was divided into 6 variables, including morning, afternoon, evening, AM peak, PM peak and out of peak. Also, day of the week was divided into 7 variables (i.e. the days of the week), and month of the year was divided into the 12 months. This was done because duration often tends to vary according to the traffic flow condition, which obviously varies according to these temporal variables.

5.2.1 Accident Duration

Three traffic accident interval times were collected for the urban accidents, including reporting time, response time and clearance time. All of these times were collected from the Abu Dhabi Urban Collision Investigation Branch (AUCIB) records and were measured using collision investigators' personal watches. Each interval time has a start and end point, measured in minutes. A common method of describing these times in survival analysis is called the Life Table Method (LTM). The first column of the LTM presents the time spent by an accident at a specific state (e.g. not being responded to). The second column (Beg. Total) is the total number of accidents at risk of failure (being responded to) at the time shown in the first column. The third column (Fail) shows the number of accidents that have been responded to at each time. The fourth column (Net Lost) gives the number of accidents censored. The remaining columns estimate the survivor function, standard error and 95% confidence interval.

Table 5-1 shows the reporting time of urban traffic accidents. As can be seen, nearly 83% of the accidents were reported to AUCIB in 10 minutes or less. In contrast, less than 10% of the accidents went over 15 minutes without being reported to AUCIB.

The second time is response time. Table 5-2 demonstrates that approximately 61% of the accidents were responded to by collision investigators in 5 minutes or less. On the other hand, around 10% of the accidents spent 12 minutes or more on the scene without being responded to by collision investigators.

The last Table (5-3) illustrates accident clearance time. Nearly 50% of the accidents were cleared in 20 minutes or less, whereas only 10% of the accidents spent 50 minutes or over before collision investigators cleared them.

Following the descriptions of the three interval times, it should be stressed that the durations show an increase in the failure time at each 5 minute interval. This might be interpreted as an obvious rounding of accident duration to values that are multiples of 5 minutes. So, in order to avoid any effect of this rounding, an investigation of 5 minutes interval times in some data will be conducted under the preliminary analysis subheading in the Analysis chapter. Then the results will be compared to a 1 minute interval time model. After that, a decision will be taken on whether to use 5 minutes interval time or 1 minute interval time.

Table 5-1 Urban traffic accidents - Reporting time

Interval Time (Min)	Beginning Total	Fail	Net Lost	Survival	Standard Error	95% Confidence Interval	
1 – 2	525	80	0	0.8476	0.0157	0.8139	0.8757
2 – 3	445	89	0	0.6781	0.0204	0.6363	0.7162
3 – 4	356	40	0	0.6019	0.0214	0.5587	0.6423
4 – 5	316	15	0	0.5733	0.0216	0.5298	0.6144
5 – 6	301	143	0	0.3010	0.0200	0.2622	0.3405
6 – 7	158	5	0	0.2914	0.0198	0.2531	0.3307
7 – 8	153	10	0	0.2724	0.0194	0.2350	0.3110
8 – 9	143	15	0	0.2438	0.0187	0.2080	0.2813
9 – 10	128	5	0	0.2343	0.0185	0.1990	0.2713
10 – 11	123	34	0	0.1695	0.0164	0.1388	0.2029
11 – 12	89	4	0	0.1619	0.0161	0.1318	0.1947
12 – 13	85	8	0	0.1467	0.0154	0.1180	0.1784
13 – 14	77	5	0	0.1371	0.0150	0.1093	0.1681
14 – 15	72	2	0	0.1333	0.0148	0.1059	0.1639
15 – 16	70	23	0	0.0895	0.0125	0.0671	0.1159
16 – 17	47	1	0	0.0876	0.0123	0.0654	0.1137
18 – 19	46	1	0	0.0857	0.0122	0.0638	0.1116
19 – 20	45	1	0	0.0838	0.0121	0.0621	0.1095
20 – 21	44	7	0	0.0705	0.0112	0.0507	0.0945
22 – 23	37	1	0	0.0686	0.0110	0.0491	0.0923
23 – 24	36	3	0	0.0629	0.0106	0.0443	0.0858
24 – 25	33	2	0	0.0590	0.0103	0.0411	0.0815
25 – 26	31	4	0	0.0514	0.0096	0.0348	0.0727
26 – 27	27	1	0	0.0495	0.0095	0.0332	0.0704
27 – 28	26	1	0	0.0476	0.0093	0.0317	0.0682
29 – 30	25	1	0	0.0457	0.0091	0.0301	0.0660
30 – 31	24	6	0	0.0343	0.0079	0.0211	0.0524
32 – 33	18	1	0	0.0324	0.0077	0.0196	0.0501
33 – 34	17	1	0	0.0305	0.0075	0.0182	0.0478
35 – 36	16	1	0	0.0286	0.0073	0.0167	0.0455
40 – 41	15	4	0	0.0210	0.0063	0.0111	0.0360
45 – 46	11	6	0	0.0095	0.0042	0.0036	0.0211
65 – 66	5	1	0	0.0076	0.0038	0.0026	0.0185
85 – 86	4	1	0	0.0057	0.0033	0.0016	0.0157
165 – 166	3	1	0	0.0038	0.0027	0.0008	0.0130
180 – 181	2	1	0	0.0019	0.0019	0.0002	0.0102
270 – 271	1	1	0	0	.	.	.

Table 5-2 Urban traffic accidents - Response time

Interval Time (Min)	Beginning Total	Fail	Net Lost	Survival	Standard Error	95% Confidence Interval	
1 – 2	525	29	0	0.9448	0.0100	0.9215	0.9613
2 – 3	496	48	0	0.8533	0.0154	0.8201	0.8809
3 – 4	448	63	0	0.7333	0.0193	0.6933	0.7690
4 – 5	385	44	0	0.6495	0.0208	0.6070	0.6886
5 – 6	341	138	0	0.3867	0.0213	0.3450	0.4281
6 – 7	203	27	0	0.3352	0.0206	0.2952	0.3757
7 – 8	176	27	0	0.2838	0.0197	0.2459	0.3228
8 – 9	149	18	0	0.2495	0.0189	0.2134	0.2872
9 – 10	131	5	0	0.2400	0.0186	0.2044	0.2773
10 – 11	126	63	0	0.1200	0.0142	0.0939	0.1494
11 – 12	63	4	0	0.1124	0.0138	0.0872	0.1411
12 – 13	59	7	0	0.0990	0.0130	0.0754	0.1264
13 – 14	52	5	0	0.0895	0.0125	0.0671	0.1159
14 – 15	47	5	0	0.0800	0.0118	0.0588	0.1052
15 – 16	42	23	0	0.0362	0.0082	0.0226	0.0547
16 – 17	19	2	0	0.0324	0.0077	0.0196	0.0501
17 – 18	17	1	0	0.0305	0.0075	0.0182	0.0478
18 – 19	16	1	0	0.0286	0.0073	0.0167	0.0455
19 – 20	15	1	0	0.0267	0.0070	0.0153	0.0432
20 – 21	14	7	0	0.0133	0.0050	0.0060	0.0262
22 – 23	7	2	0	0.0095	0.0042	0.0036	0.0211
23 – 24	5	1	0	0.0076	0.0038	0.0026	0.0185
25 – 26	4	2	0	0.0038	0.0027	0.0008	0.0130
35 – 36	2	1	0	0.0019	0.0019	0.0002	0.0102
64 – 65	1	1	0	0	.	.	.

Table 5-3 Urban traffic accidents - Clearance time

Interval Time (Min)	Beginning Total	Fail	Net Lost	Survival	Standard Error	95% Confidence Interval	
1 – 2	525	1	0	0.9981	0.0019	0.9866	0.9997
2 – 3	524	3	0	0.9924	0.0038	0.9798	0.9971
3 – 4	521	3	0	0.9867	0.005	0.9722	0.9936
4 – 5	518	5	0	0.9771	0.0065	0.9601	0.9870
5 – 6	513	12	0	0.9543	0.0091	0.9326	0.9691
6 – 7	501	1	0	0.9524	0.0093	0.9303	0.9676
7 - 8	500	4	0	0.9448	0.0100	0.9215	0.9613
8 - 9	496	5	0	0.9352	0.0107	0.9105	0.9533
9 - 10	491	5	0	0.9257	0.0114	0.8997	0.9452
10 - 11	486	37	0	0.8552	0.0154	0.8221	0.8826
11 - 12	449	2	0	0.8514	0.0155	0.8180	0.8792
12 - 13	447	11	0	0.8305	0.0164	0.7956	0.8600
13 - 14	436	9	0	0.8133	0.0170	0.7773	0.8441
14 - 15	427	6	0	0.8019	0.0174	0.7652	0.8335
15 - 16	421	75	0	0.6590	0.0207	0.6168	0.6978
16 - 17	346	8	0	0.6438	0.0209	0.6012	0.6831
17 - 18	338	8	0	0.6286	0.0211	0.5857	0.6683
18 - 19	330	13	0	0.6038	0.0213	0.5606	0.6442
19 - 20	317	8	0	0.5886	0.0215	0.5452	0.6293
20 - 21	309	48	0	0.4971	0.0218	0.4537	0.5391
21 - 22	261	8	0	0.4819	0.0218	0.4385	0.5239
22 - 23	253	5	0	0.4724	0.0218	0.4291	0.5144
23 - 24	248	10	0	0.4533	0.0217	0.4103	0.4953
24 - 25	238	4	0	0.4457	0.0217	0.4028	0.4877
25 - 26	234	43	0	0.3638	0.0210	0.3228	0.4049
26 - 27	191	4	0	0.3562	0.0209	0.3154	0.3971
27 - 28	187	8	0	0.3410	0.0207	0.3007	0.3816
28 - 29	179	8	0	0.3257	0.0205	0.2860	0.3660
29 - 30	171	5	0	0.3162	0.0203	0.2768	0.3562
30 - 31	166	28	0	0.2629	0.0192	0.2260	0.3011
31 - 32	138	3	0	0.2571	0.0191	0.2206	0.2952
33 - 34	135	3	0	0.2514	0.0189	0.2152	0.2892
34 - 35	132	3	0	0.2457	0.0188	0.2098	0.2832
35 - 36	129	19	0	0.2095	0.0178	0.1758	0.2453
36 - 37	110	1	0	0.2076	0.0177	0.1741	0.2433
37 - 38	109	2	0	0.2038	0.0176	0.1705	0.2393
38 - 39	107	1	0	0.2019	0.0175	0.1687	0.2373
39 - 40	106	1	0	0.2000	0.0175	0.1670	0.2352
40 - 41	105	20	0	0.1619	0.0161	0.1318	0.1947
42 - 43	85	3	0	0.1562	0.0158	0.1266	0.1886
43 - 44	82	2	0	0.1524	0.0157	0.1232	0.1845
44 - 45	80	2	0	0.1486	0.0155	0.1197	0.1804
45 - 46	78	8	0	0.1333	0.0148	0.1059	0.1639
46 - 47	70	2	0	0.1295	0.0147	0.1025	0.1598
47 - 48	68	4	0	0.1219	0.0143	0.0956	0.1515
48 - 49	64	8	0	0.1067	0.0135	0.0821	0.1348

49 - 50	56	2	0	0.1029	0.0133	0.0787	0.1306
50 - 51	54	10	0	0.0838	0.0121	0.0621	0.1095
51 - 52	44	1	0	0.0819	0.0120	0.0605	0.1074
52 - 53	43	2	0	0.0781	0.0117	0.0572	0.1031
53 - 54	41	2	0	0.0743	0.0114	0.0539	0.0988
54 - 55	39	1	0	0.0724	0.0113	0.0523	0.0966
55 - 56	38	4	0	0.0648	0.0107	0.0459	0.0880
56 - 57	34	1	0	0.0629	0.0106	0.0443	0.0858
57 - 58	33	1	0	0.0610	0.0104	0.0427	0.0836
58 - 59	32	1	0	0.0590	0.0103	0.0411	0.0815
59 - 60	31	2	0	0.0552	0.0100	0.0379	0.0771
60 - 61	29	4	0	0.0476	0.0093	0.0317	0.0682
61 - 62	25	1	0	0.0457	0.0091	0.0301	0.0660
62 - 63	24	2	0	0.0419	0.0087	0.0271	0.0615
63 - 64	22	1	0	0.0400	0.0086	0.0256	0.0593
64 - 65	21	2	0	0.0362	0.0082	0.0226	0.0547
65 - 66	19	1	0	0.0343	0.0079	0.0211	0.0524
67 - 68	18	2	0	0.0305	0.0075	0.0182	0.0478
69 - 70	16	1	0	0.0286	0.0073	0.0167	0.0455
70 - 71	15	3	0	0.0229	0.0065	0.0125	0.0384
71 - 72	12	1	0	0.0210	0.0063	0.0111	0.0360
75 - 76	11	2	0	0.0171	0.0057	0.0085	0.0312
76 - 77	9	1	0	0.0152	0.0053	0.0072	0.0287
86 - 87	8	1	0	0.0133	0.0050	0.0060	0.0262
90 - 91	7	2	0	0.0095	0.0042	0.0036	0.0211
95 - 96	5	1	0	0.0076	0.0038	0.0026	0.0185
105 - 106	4	1	0	0.0057	0.0033	0.0016	0.0157
109 - 110	3	1	0	0.0038	0.0027	0.0008	0.0130
120 - 121	2	1	0	0.0019	0.0019	0.0002	0.0102
130 - 131	1	1	0	0	.	.	.

5.2.2 Accident Characteristics

Table 5-4 shows the candidate variables extracted from FTSS and used in this research to analyse urban traffic accident duration.

Table 5-4 FTSS Database for urban traffic accidents

Database	Characteristics
Federal Traffic Statistics System (FTSS)	**Temporal characteristics** - Time of day (morning, afternoon, evening, AM peak, PM peak, out of peak) - Day of week (Saturday , Sunday, Monday, Tuesday, Wednesday, Thursday, Friday) - Month of year (January, February, March, April, May, June, July, August, September, October, November, December) **Geographical characteristics** - Street - Intersection - Region - Road layout - Place nature **Environmental characteristics** - Weather condition (Clear, Foggy, Rain, Windy) - Road surface condition (Dry, Wet, Sandy) - Light condition (Daylight, Darkness) **Accident characteristics** - Severity (Slight, Serious, Fatal) - Number of casualties - Number of vehicles involved - Accident type (Side Impact, Hit Pedestrian, Hit Object, Overturn, Rear-end, Head-on, Other types of Accident)

Summary statistics of the urban traffic accidents show that 79% of traffic accidents occurred out of peak period, whereas only 21% happened during the AM peak (6am-8am) and PM peak (2pm-4pm) (Figure 5-1). These figures are expected, with the exception of AM peak which has a lower number of accidents. This might be because the drivers cannot drive at a higher speed, so there is less chance of having a serious or fatal accident. About 76% of the accidents happened on weekdays (24% occurred on the weekend) (Figure 5-2). March and May have the highest number of accidents (10% each), followed by June, October and December (9% each) (Figure 5-3). These findings do not show big differences in accident numbers between months of the year.

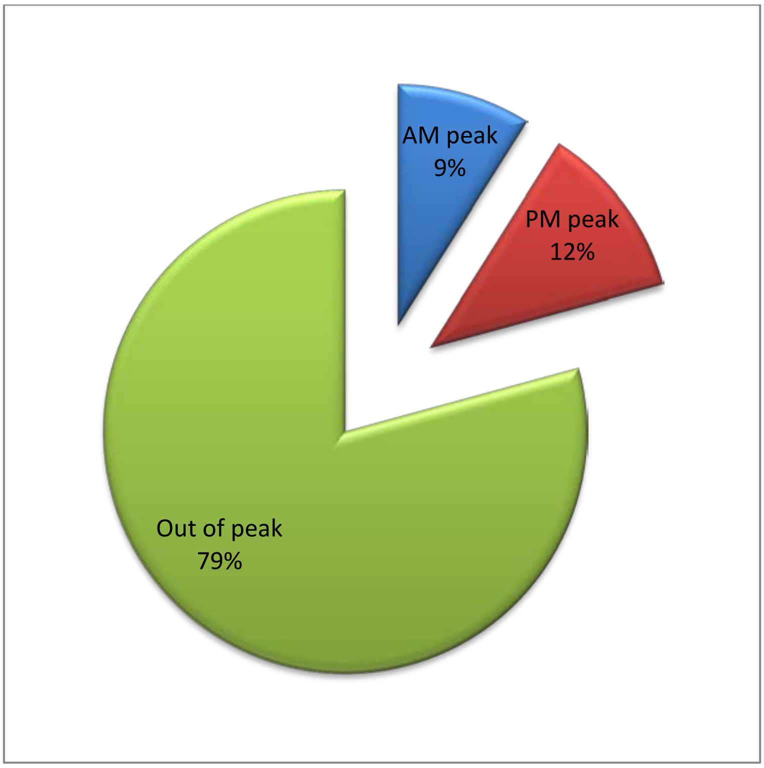


Figure 5-1 Urban traffic accidents per peak period

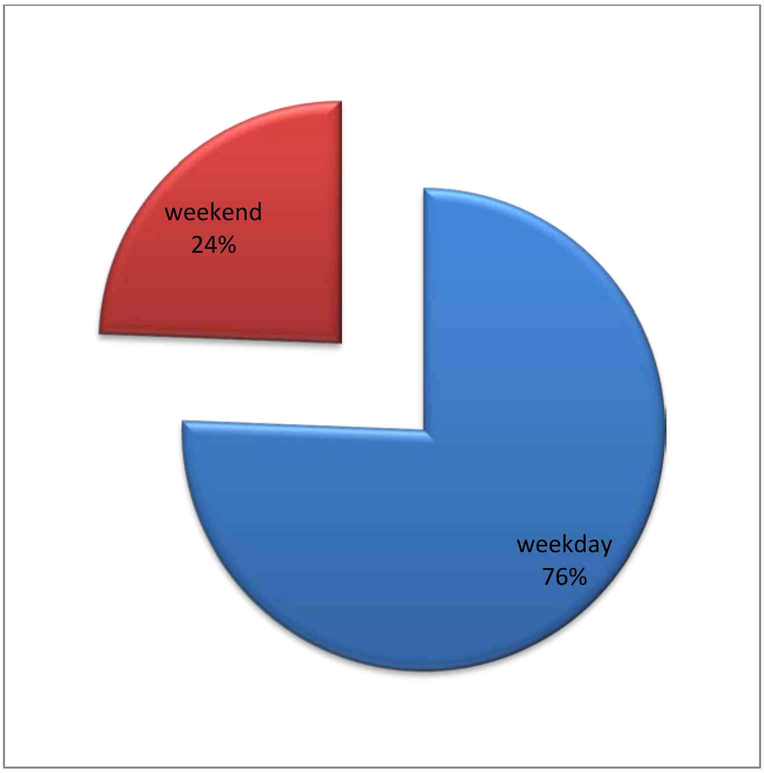


Figure 5-2 Urban traffic accidents by weekday/weekend

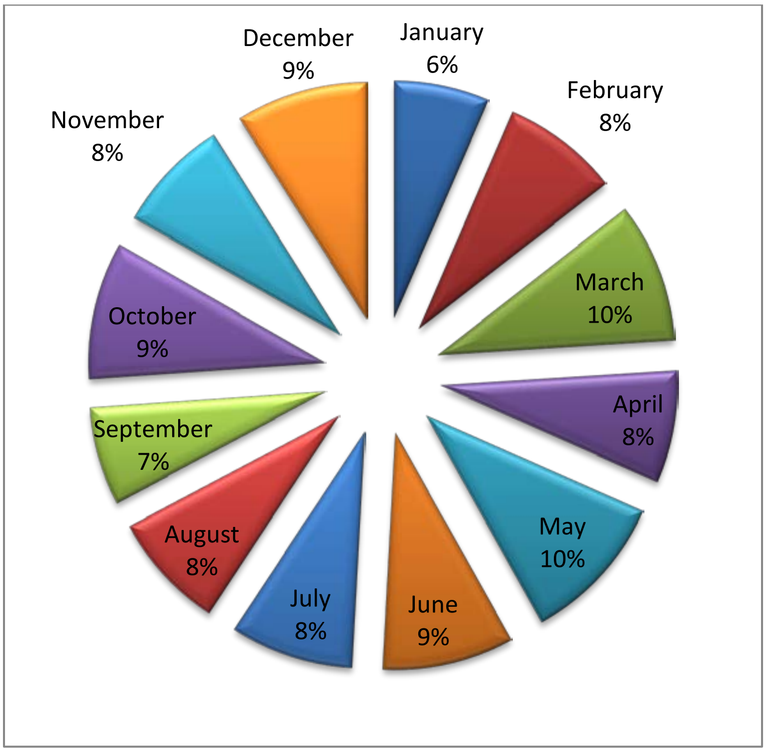


Figure 5-3 Urban traffic accidents per month

Geographically, 43% of the urban accidents occurred in commercial areas, followed by residential areas (18%) (Figure 5-4). This is expected because commercial areas have high demand compared to other study areas. In terms of the region, the highest number of accidents occurred in the Almeena area (11%) followed by the Al Mushrif area (10%), and Al Markazi area (9%) (Figure 5-5). These results reflect that the areas which have more of a commercial area, residential area or government area have more accidents.

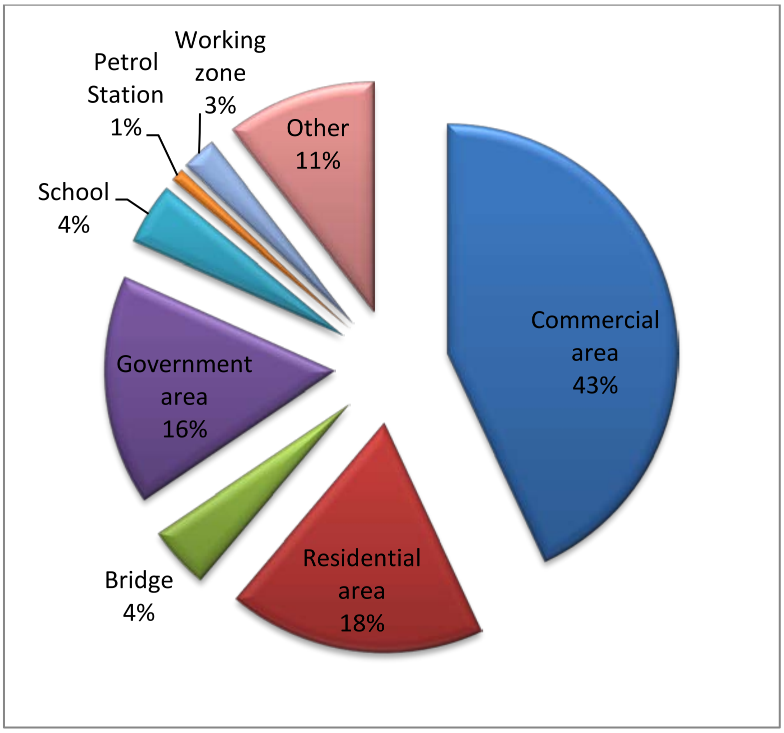


Figure 5-4 Urban traffic accidents per zone type

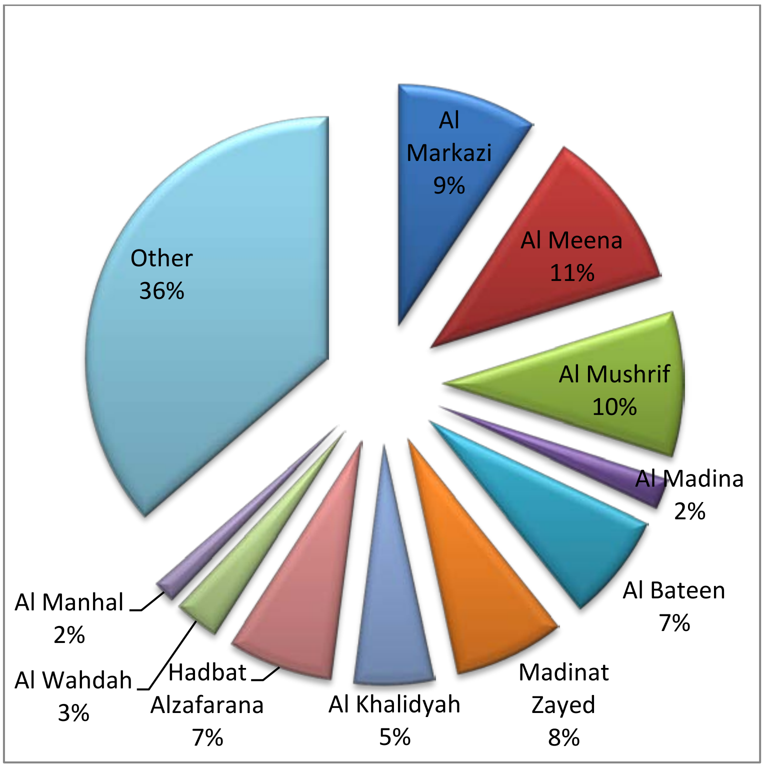


Figure 5-5 Urban traffic accidents per region

In terms of the environmental conditions of the urban accidents, nearly 97% of the accidents occurred in clear weather conditions, and the reminder occurred in fog, rain and windy conditions (Figure 5-6). Also, about 98% of accidents happened on a dry road surface, with the reminder occurring on wet or sandy road surfaces (Figure 5-7).

These findings were expected because the rain season is short and most of the urban network is surrounded by buildings.

On the other hand, 'hit pedestrian' accidents were the most common accidents (28%), followed by 'angle accidents' (23%) (Figure 5-8). This was expected because most pedestrians do not cross the road at dedicated crossing areas, so are not using the crossing facilities that have been provided. Nearly 29% of accidents were caused by failing to comply with traffic lights, whereas speeding and 'no consideration of road users' were the second most frequent cause of accidents, each with 13% (Figure 5-9). These findings provide more explanation of the higher percentage of 'hit pedestrian' accidents (28%) compared to other types of accidents.

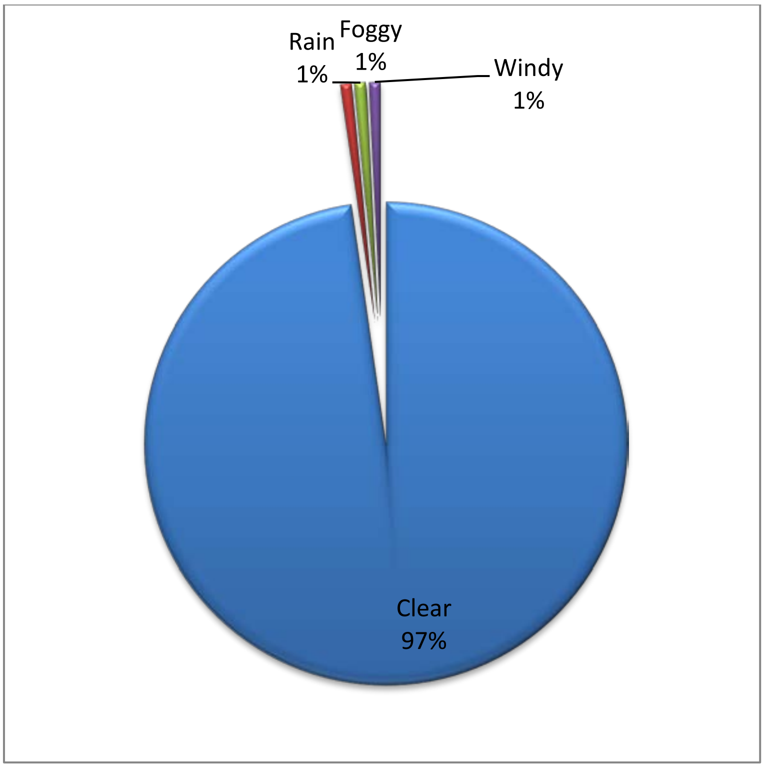


Figure 5-6 Urban traffic accidents by weather condition

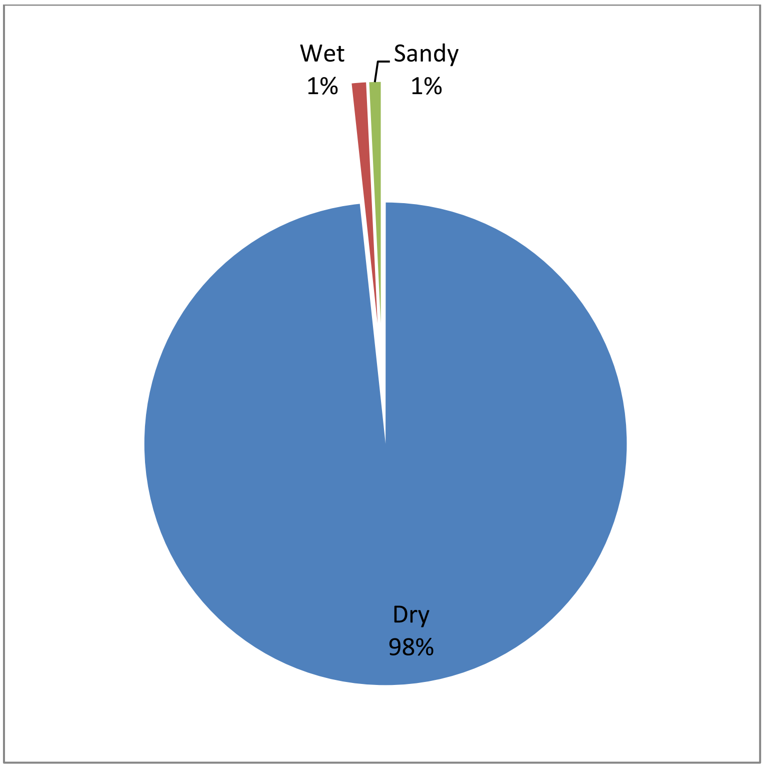


Figure 5-7 Urban traffic accidents by road surface condition

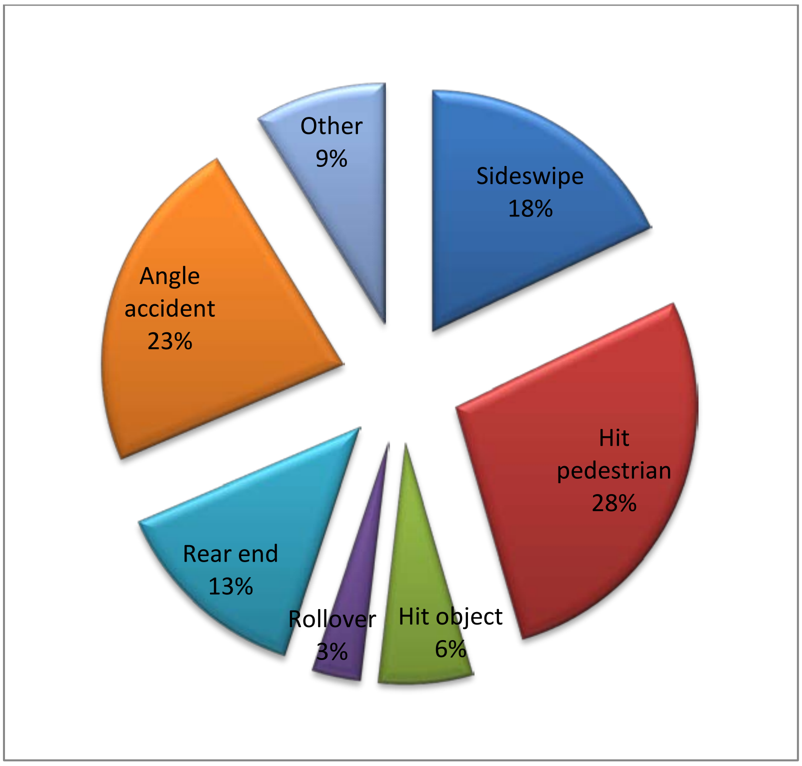


Figure 5-8 Urban traffic accidents by accident type

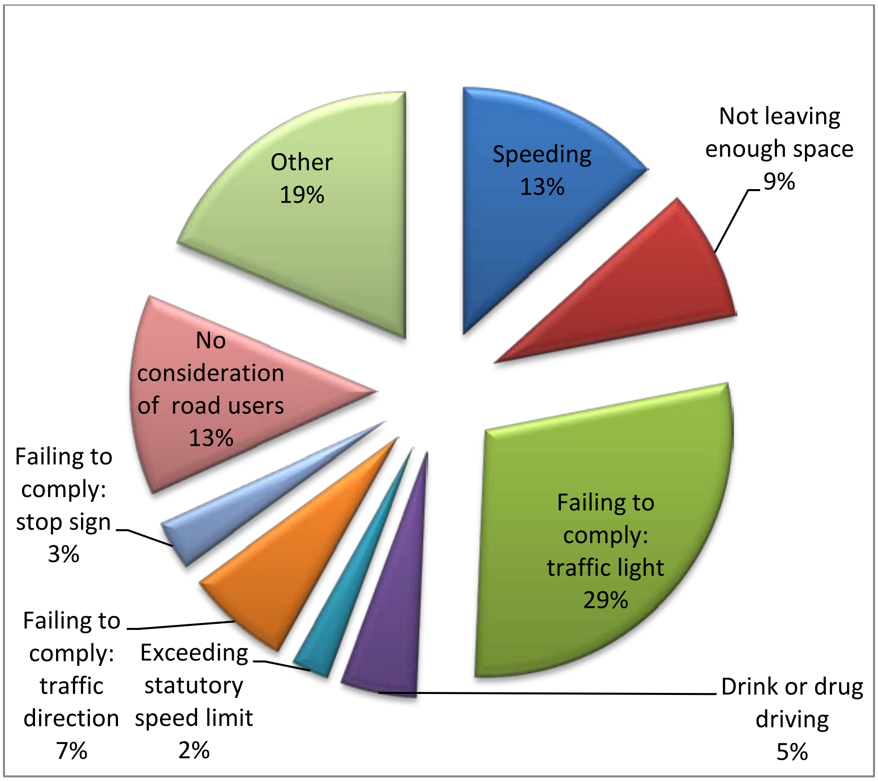


Figure 5-9 Urban traffic accidents by cause of accident

More details of the frequency and percentage of the explanatory variables are illustrated in Table 5-5 – Table 5-8 below.

Table 5-5 Urban traffic accidents - Summary statistics of the temporal characteristics

Variable	Frequency	Percentage
Peak Period		
AM Peak	47	9
PM Peak	62	12
Out of Peak	416	79
Total	525	100
Time of day		
Morning	146	28
Afternoon	144	27
Evening	235	45
Total	525	100
Day of week		
Saturday	51	10
Sunday	62	12
Monday	88	17
Tuesday	74	14
Wednesday	86	16
Thursday	87	16
Friday	77	15
Total	525	100
Weekday or Weekend		
Weekday	397	76
Weekend	128	24
Total	525	100
Month of year		
January	34	6
February	42	8
March	50	10
April	40	8
May	54	10
June	47	9
July	44	8
August	42	8
September	35	7
October	49	9
November	41	8
December	47	9
Total	525	100

Table 5-6 Urban traffic accidents - Summary statistics of the Geographical characteristics

Variable	Frequency	Percentage
Area		
Al Markazi	49	9
Al Meena	57	11
Al Mushrif	51	10
Al Madina	11	2
Al Bateen	38	7
Madinat Zayed	39	8
Al Khalidyah	28	5
Hadbat Alzafarana	37	7
Al Wahdah	16	3
Al Manhal	8	2
Other	191	36
Total	525	100
Zone Type		
Commercial area	226	43
Residential area	95	18
Bridge	23	4
Government area	85	16
School	23	4
Petrol station	5	1
Working zone	13	3
Other places	55	11
Total	525	100

Table 5-7 Urban traffic accidents - Summary statistics of the Environmental characteristics

Variable	Frequency	Percentage
Road surface condition		
Dry	516	98
Wet	5	1
Sandy	4	1
Total	525	100
Weather condition		
Clear	513	97
Foggy	4	1
Rain	4	1
Windy	4	1
Total	525	100

Table 5-8 Urban traffic accidents - Summary statistics of the Accident characteristics

Variable	Frequency	Percentage
Cause of the accident		
Speeding	70	13
Did not leave enough space	45	9
Failing to comply: traffic light	151	29
Drink or drug driving	26	5
Exceeding statutory speed limit	13	2
Failing to comply: traffic direction	36	7
Failing to comply: stop sign	17	3
No road users consideration	69	13
Other	98	19
Total	525	100
Number of Injuries		
Slight injury	462	53
Serious injury	389	44
Fatal injury	28	3
Total	879	100
Accident type		
Sideswipe	94	18
Hit pedestrian	145	28
Hit object	33	6
Rollover	17	3
Rear end	71	13
Angle accident	119	23
Other types	46	9
Total	525	100

5.3 Highway Traffic Accidents Data

Data collection for the major highway network was conducted during January to December 2009 in order to achieve the aim of this research. This study period is different from the study period for urban accidents. This is because the response duration is the only data recorded in a regular base and Abu Dhabi Highway Collision Investigation Branch (AHCIB) investigators were not encouraged to collect reporting time or clearance time. The data were collected for all injury accidents that occurred on the 9 highways for which AHCIB is responsible. These highways link Abu Dhabi's cities with other emirates such as Dubai.

In 2009, the total number of highway accidents was 860. However, AHCIB could not respond initially to all of these accidents because some accidents happened contemporaneously. During each eight-hour shift, four traffic investigators are available. When an accident occurs, two of them move to the scene together. Since a significant number of accidents may happen during a day, some accidents many not be attended to immediately by the AHCIB due to lack of resources (in this case, the limited number of investigators). Thus, in 2009, while the total number of highways accidents was 860, only 504 accidents were attended immediately by AHCIB investigators and 356 accidents were investigated later (Figure 5-10). Later investigation is carried out with support received from the highway service patrol which responded first to the accident and cleared the road. Only those receiving an immediate response are included in this study because of the availability of response time data, whereas accidents that received later investigation had no response time data.

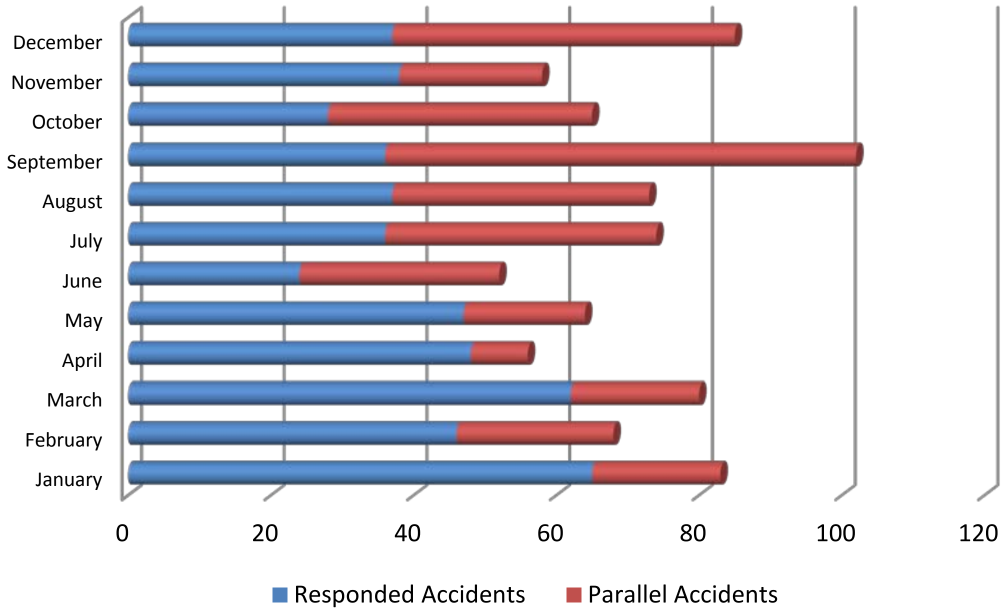


Figure 5-10 Highway traffic accidents - Frequency of total accidents per month

For this study, response time is defined as the time between AHCIB receiving the call to responding to the accident and the first collision investigator arriving at the accident scene. Furthermore, as with the urban accident dataset, some data were divided into more variables to understand the relationship between the highway accident response time and accident characteristics.

5.3.1 Accident Duration

As mentioned earlier, for the highway traffic accidents, only response time was collected. Table 5-9 shows that 50% of these accidents were responded to by AHCIB collision investigators in 13 minutes or less, whereas only 6 accidents were responded to after 35 minutes. Similar to the urban accident duration, highway accident duration seems to be rounded to the values that are multiples of 5 minutes.

Table 5-9 Highway traffic accidents -Response time

Interval Time (Min)	Beginning Total	Fail	Net Lost	Survival	Standard Error	95% Confidence Interval	
2 - 3	504	1	0	0.9980	0.0020	0.9860	0.9997
3 - 4	503	1	0	0.9960	0.0028	0.9842	0.9990
4 - 5	502	7	0	0.9821	0.0059	0.9660	0.9907
5 - 6	495	25	0	0.9325	0.0112	0.9069	0.9513
6 - 7	470	5	0	0.9226	0.0119	0.8956	0.9429
7 - 8	465	16	0	0.8909	0.0139	0.8603	0.9151
8 - 9	449	24	0	0.8433	0.0162	0.8085	0.8722
9 - 10	425	16	0	0.8115	0.0174	0.7745	0.8430
10 - 11	409	106	0	0.6012	0.0218	0.5570	0.6424
11 - 12	303	7	0	0.5873	0.0219	0.5430	0.6289
12 - 13	296	19	0	0.5496	0.0222	0.5051	0.5918
13 - 14	277	25	0	0.5000	0.0223	0.4556	0.5427
14 - 15	252	11	0	0.4782	0.0223	0.4339	0.5210
15 - 16	241	107	0	0.2659	0.0197	0.2281	0.3050
16 - 17	134	10	0	0.2460	0.0192	0.2093	0.2844
17 - 18	124	10	0	0.2262	0.0186	0.1907	0.2636
18 - 19	114	11	0	0.2044	0.0180	0.1704	0.2406
19 - 20	103	4	0	0.1964	0.0177	0.1630	0.2322
20 - 21	99	48	0	0.1012	0.0134	0.0768	0.1294
21 - 22	51	2	0	0.0972	0.0132	0.0733	0.1250
22 - 23	49	1	0	0.0952	0.0131	0.0716	0.1228
23 - 24	48	3	0	0.0893	0.0127	0.0664	0.1162
24 - 25	45	1	0	0.0873	0.0126	0.0647	0.1140
25 - 26	44	18	0	0.0516	0.0099	0.0346	0.0733
26 - 27	26	1	0	0.0496	0.0097	0.0330	0.0710
27 - 28	25	1	0	0.0476	0.0095	0.0314	0.0687
28 - 29	24	2	0	0.0437	0.0091	0.0282	0.0640
30 - 31	22	12	0	0.0198	0.0062	0.0102	0.0350
35 - 36	10	4	0	0.0119	0.0048	0.0050	0.0247
38 - 39	6	1	0	0.0099	0.0044	0.0038	0.0220
45 - 46	5	3	0	0.0040	0.0028	0.0008	0.0135
55 - 56	2	1	0	0.0020	0.0020	0.0002	0.0106
75 - 76	1	1	0	0	.	.	.

5.3.2 Accident Characteristics

Table 5-10 shows the candidate variables extracted from FTSS and used in this research to analyse highway traffic accident duration.

Table 5-10 FTSS Database for highway traffic accidents

Database	Characteristics
Federal Traffic Statistics System (FTSS)	**Temporal characteristics** - Time of day (morning, afternoon, evening) - Day of week (Saturday , Sunday, Monday, Tuesday, Wednesday, Thursday, Friday) - Month of year (January, February, March, April, May, June, July, August, September, October, November, December) **Geographical characteristics** - Highway name **Environmental characteristics** - Weather condition (Clear, Foggy, Rain, Windy) - Road surface condition (Dry, Wet, Sandy) - Light condition (Daylight, Darkness) **Accident characteristics** - Severity (Slight, Serious, Fatal) - Number of casualties - Number of vehicles involved - Accident type (Side Impact, Hit Pedestrian, Hit Object, Overturn, Rear-end, Head-on, Other types of Accidents)

Descriptive analysis of highway traffic accidents demonstrates that around 72% of traffic accidents occurred on weekdays, and the reminder (28%) occurred at the weekend (Figure 5-11). About 13% of accidents happened in January, followed by March (12%), and April (10%) (Figure 5-12). The findings show that there is no significant difference between weekdays and the weekend.

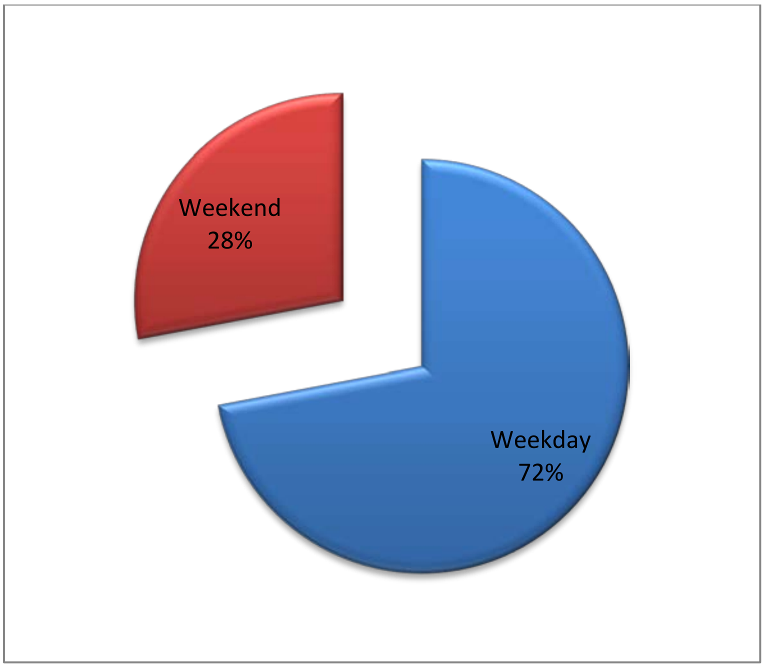


Figure 5-11 Highway traffic accidents by weekday and weekend

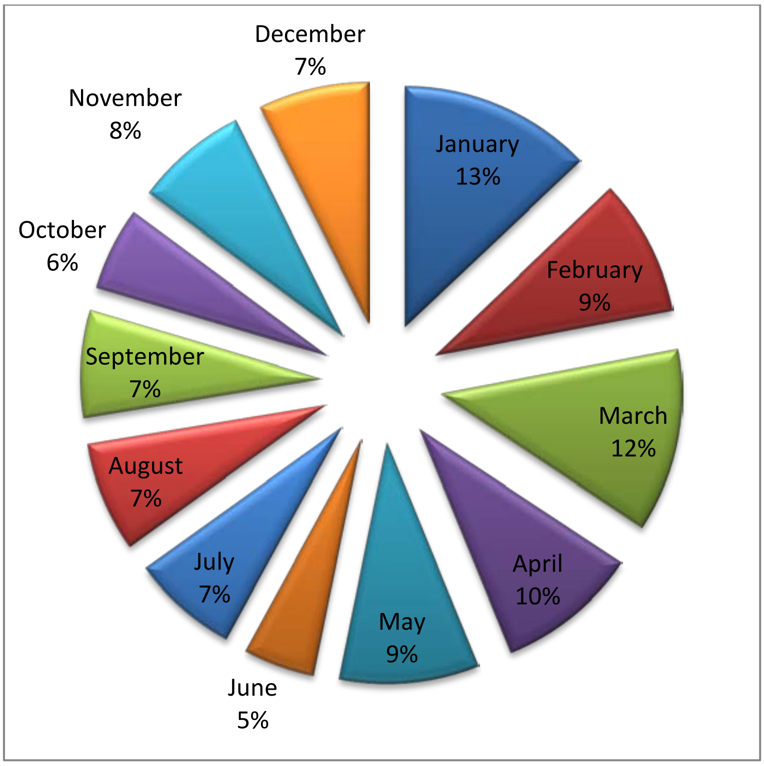


Figure 5-12 Highway traffic accidents by month of year

The distribution of accidents per highway showed that the Abu Dhabi-Tarif highway was associated with the highest number of accidents (24%), followed by the Abu Dhabi-Al Ain highway (20%) (Figure 5-13). On the other hand, the lowest number of accidents was associated with the Swihan highway (4%). These results are expected and reflect the traffic demands of these highways. The higher the traffic demand on the highway, the higher the number of accidents occurred.

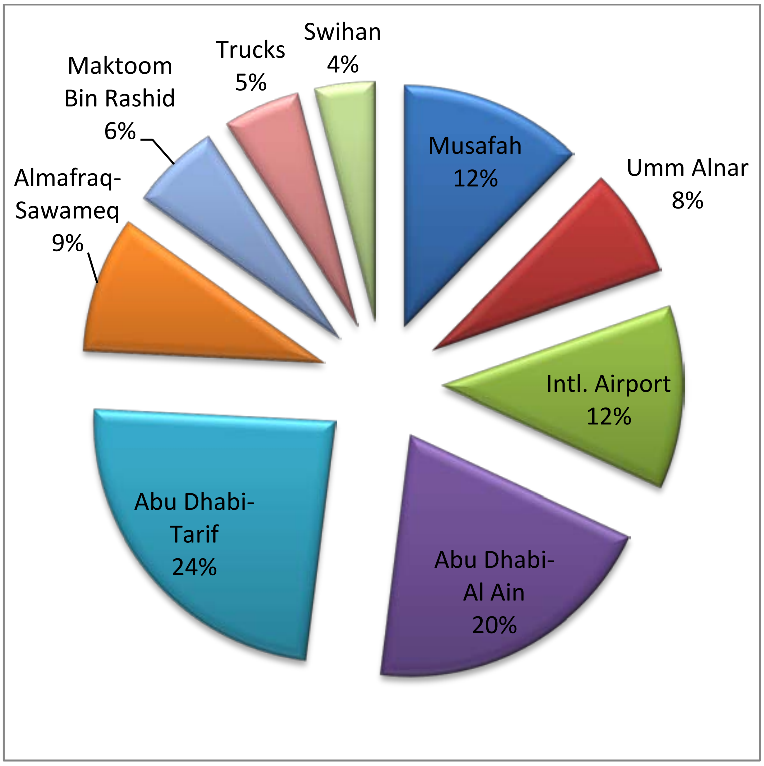


Figure 5-13 Highway traffic accidents per highway

In terms of the environmental characteristics, the majority of highway accidents occurred in clear weather conditions (94%), whereas only 6% occurred in fog, rain and windy weather (Figure 5-14). Around 91% of accidents happened on dry road surfaces; with the reminder (9%) occurring on wet or sandy road surfaces (Figure 5-15). These results are expected because there is more chance of having a wet road due to fog. Also, the chance of having a sandy road surface is higher compared to the urban area because most of these highways are surrounded by sands.

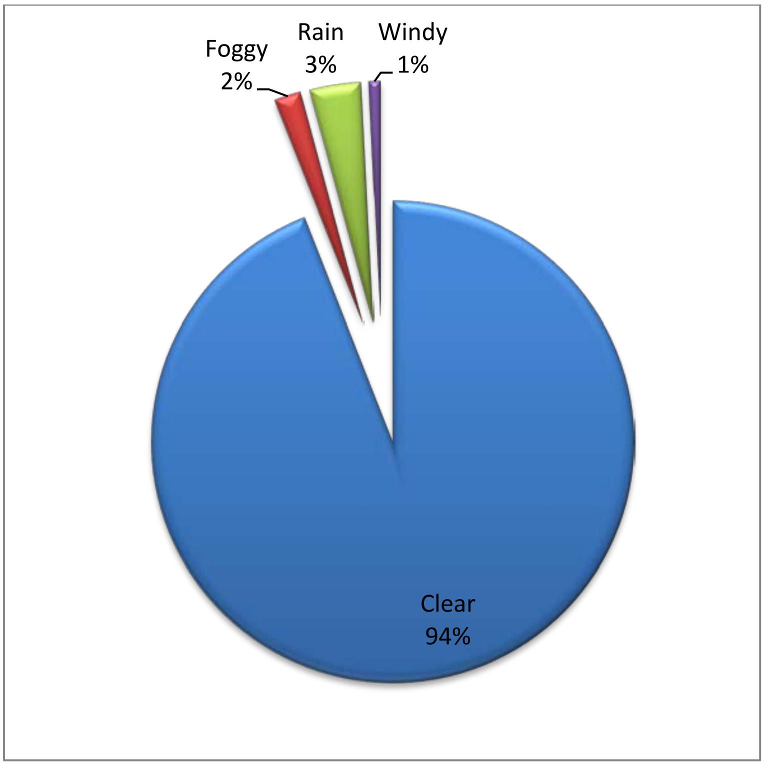


Figure 5-14 Highway traffic accidents by weather condition

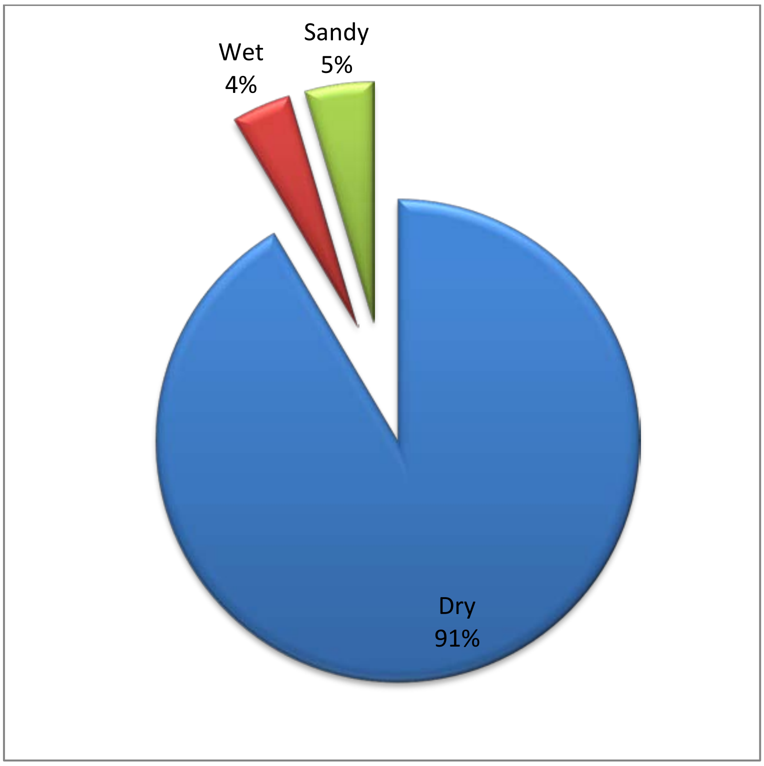


Figure 5-15 Highway traffic accidents by road surface

The severity of highway accidents shows that 68% of accidents were serious, followed by slight accidents (18%), and then fatal accidents with 14% (Figure 5-16). Additionally, the most frequent accidents were 'hit pedestrian' accidents (24%), then 'rollover' (17%) and 'hit object' (16%) (Figure 5-17). It is unexpected to find that 'hit pedestrian' is the highest type of accident because it is illegal for pedestrians to cross highways; however, the investigation reveals that pedestrians do cross highways to access shops and facilities located adjacent to them. As a result, similar to urban accidents, 'hit pedestrian' accidents are the most frequent accidents.

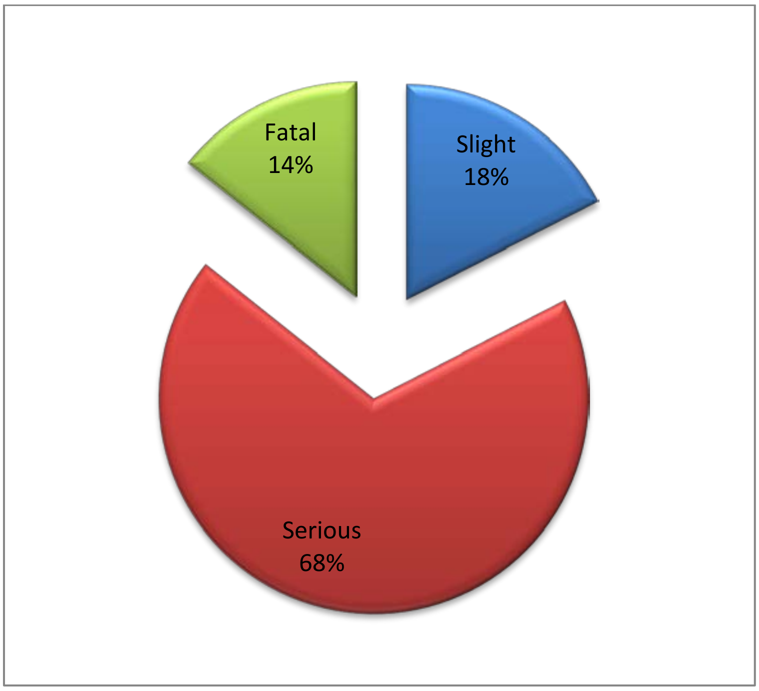


Figure 5-16 Highway traffic accidents by severity level

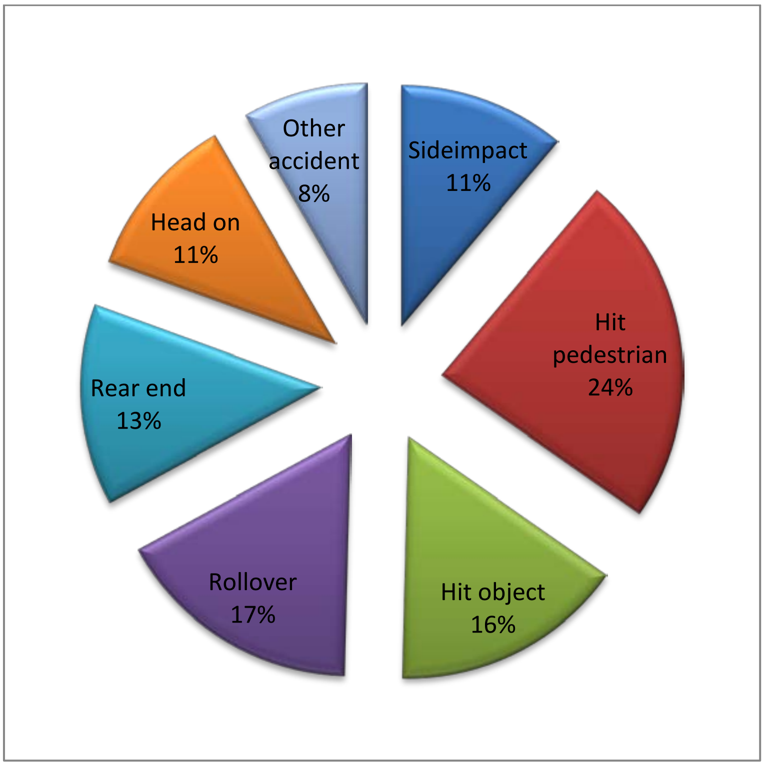


Figure 5-17 Highway traffic accidents by accident type

Further details of the frequency and percentage of accident characteristics are illustrated by Table 5-11-5-14.

Table 5-11 Highway traffic accidents -Summary statistics of the Temporal characteristics

Variable	Frequency	Percentage
Time of day		
Morning	235	47
Afternoon	123	24
Evening	146	29
Total	504	100
Day of week		
Saturday	73	14
Sunday	74	15
Monday	66	13
Tuesday	67	13
Wednesday	77	15
Thursday	79	16
Friday	68	14
Total	504	100
Month of year		
January	65	13
February	46	9
March	62	12
April	48	10
May	47	9
June	24	5
July	36	7
August	37	7
September	36	7
October	28	6
November	38	8
December	37	7
Total	504	100

Table 5-12 Highway traffic accidents -Summary statistics of the Environmental characteristics

Variable	Frequency	Percentage
Road surface condition		
Dry	461	91
Wet	20	4
Sandy	23	5
Total	504	100
Weather condition		
Clear	476	94
Foggy	9	2
Rain	17	3
Wind	2	1
Total	504	100

Table 5-13 Highway traffic accidents -Summary statistics of the Geographical characteristics

Variable	Frequency	Percentage
Highway		
Musafah	62	12
UmmAlnar	37	8
International Airport	62	12
Abu Dhabi-Al Ain	101	20
AbuDhabi-Tarif	120	24
Almafraq-Sawameq	46	9
Maktoom Bin Rashid	30	6
Trucks	26	5
Swihan	20	4
Total	504	100

Table 5-14 Highway traffic accidents -Summary statistics of the Accident characteristics

Variable	Frequency	Percentage
Number of Injuries		
Slight injury	88	18
Serious injury	343	68
Fatal injury	72	14
Total	503	100
Accident type		
Sideswipe	56	11
Hit pedestrian	119	24
Hit object	79	16
Rollover	84	17
Rear end	68	13
Angle accident	56	11
Other types	42	8
Total	504	100

5.4 Summary

This chapter provides a descriptive analysis of traffic accident data that were extracted from FTSS and ASCIS databases. Both urban accidents and highway accidents were described in terms of their temporal characteristics, geographical characteristics, environmental characteristics and accident characteristics. Also, interval times of the urban traffic accidents duration (reporting time, response time, clearance time) and highway traffic accident duration (response time) were described using LTM approach.

6 Analysis of Accident Duration

6.1 Introduction

This chapter starts with the preliminary analysis to decide upon the analytical method in section 6.2. Then, it progresses to the application of the proposed research methodology for analysing urban and highway traffic accidents in Abu Dhabi. The main aim of this chapter is to present the findings with an in-depth analysis of traffic accident duration using HBDMs.

The application of HBDMs to model the influence of traffic accident characteristics on urban accidents with emphasis on the intervals of reporting, responding and clearance is discussed in section 6.3. The results provide an insight into a range of accident characteristics influencing each interval time of the total accident duration.

The analysis of highway accident response time using HBDMs is explained in section 6.4. The results provide an insight into a range of accident characteristics influencing response time.

6.2 Preliminary Analysis to Decide upon the Analytical Methods

In order to select the most appropriate analytical method to model accident duration, a preliminary analysis was conducted using the dataset of highway accidents in Abu Dhabi. Section 6.2.1 presents the results of PH and AFT models that have been developed for the highway response time. Section 6.2.2 presents the most appropriate time interval that is to be used in the analysis.

6.2.1 Proportional Hazard Model (PH) Vs Accelerated Failure Time Model (AFT)

As mentioned in the Literature Review chapter, there are two approaches to investigate the effects of explanatory variables in HBDMs: the Proportional Hazard model (PH) and the Accelerated Failure Time model (AFT). The difference between these two approaches is that in a PH model the effects of covariates are to multiply the hazard rate by a factor; however, in an AFT model the effects of covariates are to multiply the mean duration by a factor.

Furthermore, not all distributions can be parameterised as either a PH model or an AFT model. As can be seen in Table 3-2, only Exponential distribution and Weibull distribution are capable of that. So, the results from fitting these two distributions can be interpreted in both ways. To further illustrate this point, a comparison of the results from the two models using Weibull distribution was applied to highway accidents response time. The following Tables 6-1 and 6-2 present the fitted AFT model and PH model respectively.

It is clear from the comparison of the results of the PH and AFT models applied to Abu Dhabi highway accident response data that there are no differences in the significant variables as well as in the level of significance of these variables. The only differences are in the magnitude of the estimated coefficients and their signs. As mentioned earlier, this is because the estimated coefficient in PH models indicates how the covariate affects the hazard rate, whereas in AFT models the estimated coefficient indicates how the covariate affects the mean duration.

Table 6-1 Highway traffic accidents -Weibull AFT model

Variable	Estimated Coefficient	t-statistics
Temporal Characteristics		
Wednesday	0.12	2.42
Tuesday	-0.11	-2.21
Morning	-0.06	-1.97
Monday	-0.10	-1.91
February	-0.43	-6.99
April	-0.22	-3.74
May	-0.22	-3.64
July	0.27	3.99
September	0.23	3.39
Geographical Characteristics		
Musafah HW	-0.23	-4.16
AbuDhabi-Tarif HW	0.25	5.54
Almafraq-Sawameq HW	-0.19	-3.09
Swihan HW	0.34	3.76
Environmental Characteristics		
Sandy	0.18	2.20
Accident Characteristics		
Hit pedestrian	-0.20	-4.63
Hit object	-0.11	-2.16
Fatal injury	0.17	3.45
Other accidents	-0.15	-2.40
Model Structure Parameters		
p (distribution shape parameter)	2.62	30.23
λ (the scale parameter)	0.05	68.52
Intercept term (_cons)	2.85	68.52
Goodness-of-fit Test		
Akaike information criterion	606.75	
Initial log-likelihood	-381.52	
Log-likelihood at convergence	-283.37	
Number of observations	504	

Table 6-2 Highway traffic accidents -Weibull PH model

Variable	Estimated Coefficient	t-statistics
Temporal Characteristics		
Wednesday	-0.32	-2.41
Tuesday	0.30	2.20
Morning	0.18	1.96
Monday	0.26	1.90
February	1.14	6.83
April	0.60	3.71
May	0.58	3.62
July	-0.72	-3.91
September	-0.61	-3.35
Geographical Characteristics		
Musafah HW	0.61	4.10
AbuDhabi-Tarif HW	-0.66	-5.38
Almafraq-Sawameq HW	0.51	3.08
Swihan HW	-0.89	-3.74
Environmental Characteristics		
Sandy	-0.49	-2.19
Accident Characteristics		
Hit pedestrian	0.52	4.56
Hit object	0.29	2.16
Fatal injury	-0.46	-3.41
Other accidents	0.41	2.39
Model Structure Parameters		
p (distribution shape parameter)	2.62	30.23
λ (the scale parameter)	0.05	68.52
Intercept term (_cons)	2.85	68.52
Goodness-of-fit Test		
Akaike information criterion	606.75	
Initial log-likelihood	-381.52	
Log-likelihood at convergence	-283.37	
Number of observations	504	

Additionally, the prediction of incident duration from the PH model can be obtained in a similar way to obtaining the prediction of incident duration from the AFT. This can be done by running the command “predict t_mean, time mean” in Stata. Figures 6-1 and 6-2 show the plots of comparison between observed and predicted durations for both models. The plots show a weak relationship between the observed and predicted duration ($R^2 = 0.15$ for both the AFT model, and the PH model). It is clear from this that there are no differences between the two models.

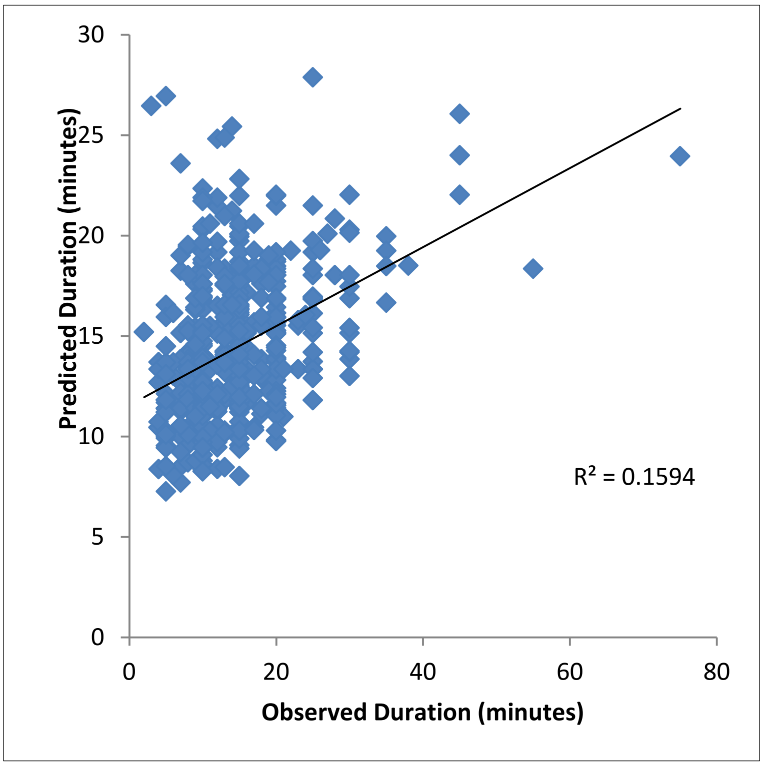


Figure 6-1 Comparison between the observed and predicted duration -Weibull PH model

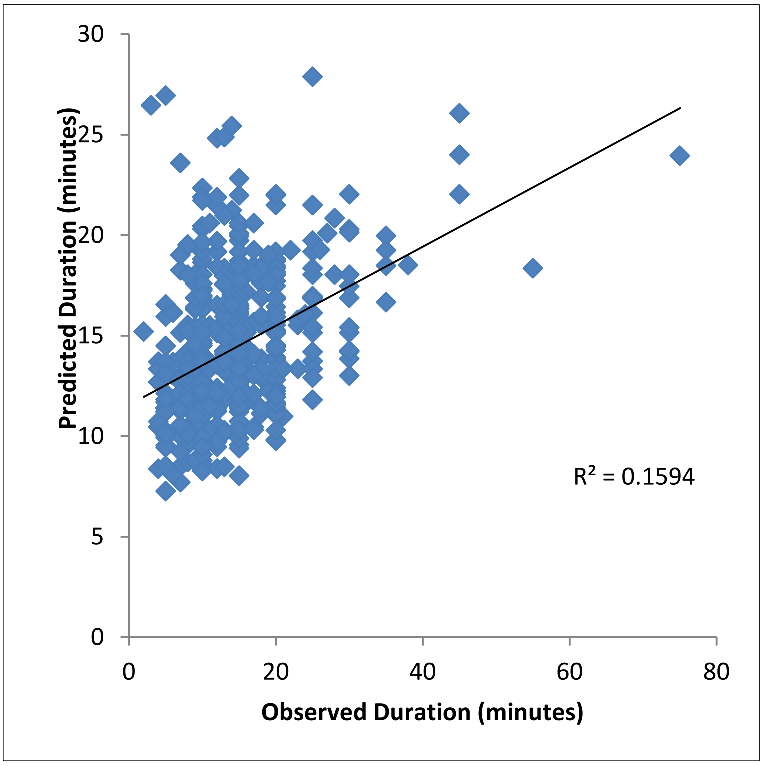


Figure 6-2 Comparison between the observed and predicted duration - Weibull AFT model

After comparing the PH and AFT models, it can be concluded that the models do not produce significantly different results, so it is not legitimate to conclude that one is better than the other.

Basically, there are two reasons which help to make clear the preference for the AFT model over the PH model:

<u>1. AFT coefficients are easier and more intuitive to interpret.</u>

AFT models measure the direct effect of the explanatory variables on the survival time instead of hazard rate, as in the PH model. This feature makes the AFT model more appropriate to achieve the objective of predicting the mean survival time than the PH model. Also, this characteristic helps researchers who are unfamiliar with HBDMs to easily understand and interpret the results because the parameters measure the effect of the correspondent covariate on the mean survival time. Furthermore, Bradburn *et al.*, (2003) suggests that “the parametric approach offers more in the way of predictions, and the AFT formulation allows the derivation of a time ratio, which is arguably more interpretable than a ratio of two hazards” (Bradburn *et al.*, 2003).

In other words, in the PH model the sign of the estimated coefficient shows whether the covariate increases or decreases the hazard rate. So, a positive coefficient means that the covariate increases the hazard rate and as a result decreases the duration. However, in the AFT model the sign of the estimated coefficient shows whether the covariate increases or decreases the duration. For instance, a positive coefficient indicates that the covariate increases the duration and a negative coefficient indicates that the covariate reduces the duration (Reid, 1994; Kay and Kinnersley, 2002; Jiezhi Qi, 2009).

<u>2. AFT models allow for more probability distribution assumptions, which serve as robustness checks.</u>

In PH models, there are relatively few probability distributions that can be used for the survival time (Exponential, Weibull, or Gompertz distributions). However, in AFT models more probability distributions can be used, including Exponential, Weibull, log-logistic, and log-normal (Table 3-2).

Considering the advantages of using AFT model over PH model, this study uses AFT model when modelling accident durations.

6.2.2 1 minute interval time vs 5 minutes interval time

As mentioned in the Data Description chapter, duration data were found to be rounded to values that are multiples of 5 minutes. To measure the effects of this rounding, both 1 minute interval time and 5 minutes interval time of highway accident response time will be investigated using three different distributions: Weibull, Log-normal, and Log-logistic.

The following Tables 6-3, 6-4, and 6-5 display the results of modelling 1 minute response time using Weibull distribution, Log-normal distribution, and Log-logistic distribution. The selection of these distributions was based on the decision that was made in the previous section 6.2.1 to apply the AFT model rather than the PH model and these are the three distributions that can be written in AFT metric.

Table 6-3 Highway traffic accidents - Weibull AFT model considering 1 minute interval time

Variable	Estimated Coefficient	t-statistics	Percentage change
Temporal Characteristics			
Wednesday	0.12	2.42	13.01
Tuesday	-0.11	-2.21	-11.04
Morning	-0.06	-1.97	-6.68
Monday	-0.10	-1.91	-9.68
February	-0.43	-6.99	-35.38
April	-0.22	-3.74	-20.49
May	-0.22	-3.64	-20.04
July	0.27	3.99	31.80
September	0.23	3.39	26.48
Geographical Characteristics			
Musafah HW	-0.23	-4.16	-20.97
AbuDhabi-Tarif HW	0.25	5.54	28.66
Almafraq-Sawameq HW	-0.19	-3.09	-17.70
Swihan HW	0.34	3.76	40.90
Environmental Characteristics			
Sandy	0.18	2.20	20.87
Accident Characteristics			
Hit pedestrian	-0.20	-4.63	-18.23
Hit object	-0.11	-2.16	-10.51
Fatal injury	0.17	3.45	19.20
Other accidents	-0.15	-2.40	-14.70
Model Structure Parameters			
p (distribution shape parameter)	2.62	30.23	
λ (the scale parameter)	0.05	68.52	
Intercept term (_cons)	2.85	68.52	
Goodness-of-fit Test			
Akaike information criterion	606.75		
Initial log-likelihood	-381.52		
Log-likelihood at convergence	-283.37		
Number of observations	504		

Table 6-4 Highway traffic accidents - Log-normal AFT model considering 1 minute interval time

Variable	Estimated Coefficient	t-statistics	Percentage change
Temporal Characteristics			
February	-0.47	-7.39	-37.87
April	-0.29	-4.63	-25.66
May	-0.24	-3.90	-22.06
Friday	0.11	2.16	12.51
Wednesday	0.11	2.22	12.20
Geographical Characteristics			
Musafah HW	-0.28	-4.82	-24.67
Almafraq –Swamiq HW	-0.33	-5.16	-28.64
Swihan HW	0.31	3.27	36.62
AbuDhabi-ALAin HW	-0.15	-3.21	-14.19
Accident Characteristics			
Hit pedestrian	-0.07	-1.65	-7.20
Overturn	0.16	3.17	17.91
Model Structure Parameters			
β (distribution shape parameter)	2.68	82.42	
σ (the scale parameter)	0.40	31.74	
Intercept term (_cons)	2.68	82.42	
Goodness-of-fit test			
Akaike information criterion	564.03		
Initial log-likelihood	-335.73		
Log-likelihood at convergence	-264.92		
Number of observations	504		

Table 6-5 Highway traffic accidents - Log-logistic AFT model considering 1 minute interval time

Variable	**Estimated Coefficient**	**t-statistics**	**Percentage change**
Temporal Characteristics			
February	-0.45	-7.18	-0.30
April	-0.28	-4.45	-0.20
May	-0.25	-4.08	-0.20
Geographical Characteristics			
Musafah HW	-0.22	-4.15	-0.20
AbuDhabi-Tarif HW	0.09	2.04	9.80
Almafraq-Sawameq HW	-0.25	-3.78	-0.20
Swihan HW	0.39	4.28	48.10
Environmental Characteristics			
Sandy	0.19	2.15	21.60
Model Structure Parameters			
P (distribution shape parameter)	4.34	39.33	
λ (the scale parameter)	0.07	100.22	
Intercept term (_cons)	2.64	100.22	
Goodness-of-fit Test			
Akaike information criterion	557.51		
Initial log-likelihood	-329.83		
Log-likelihood at convergence	-268.75		
Number of observations	504		

Figure 6-3 presents the plots of comparing the observed duration to the predicted duration for the three distributions. As can be seen, the plots show a weak relationship between the observed and predicted duration in the three distributions ($R^2 = 0.15$ for the Weibull AFT model, $R^2 = 0.17$ for the Log-normal AFT model, and $R^2 = 0.14$ for the Log-logistic AFT model). For MAPE, Weibull AFT = 21.92%, Log-normal AFT = 18.37%, and Log-logistic = 20.43%. For RMSE, Weibull AFT = 6.68, Log-normal AFT = 6.59, and Log-logistic AFT = 6.69. Finally, testing for the best fit distribution using the Akaike Information Criterion (AIC) (Table 6-6) shows that the Log-logistic AFT model is the best fit distribution for highway accident response time as the log-logistic distribution has the lowest AIC value.

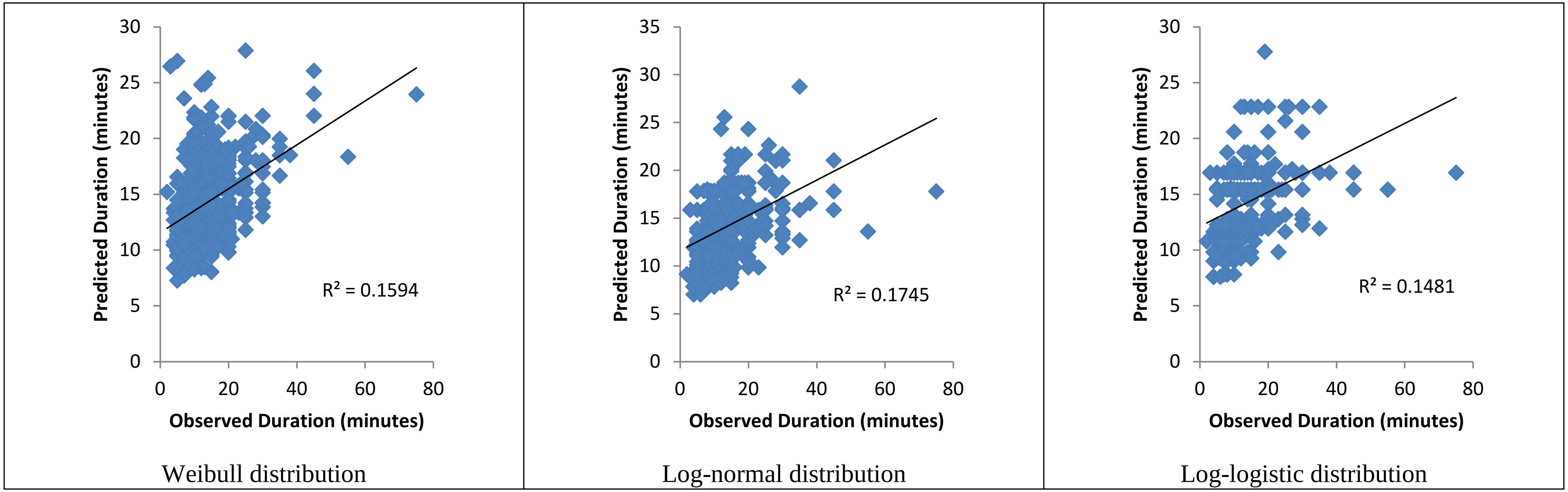


Figure 6-3 Comparison between the observed and predicted duration for Weibull, Log-normal and Log-logistic distributions considering 1 minute interval time

Table 6-6 Comparison of AIC test value for AFT response time models (1 minute interval time)

Model	-2In L	K	c	AIC
Weibull	566.74	18	2	606.75
Log-normal	529.84	11	2	564.03
Log-logistic	**537.50**	**8**	**2**	**557.51**

On the other hand, to measure the effects of rounding the duration data to values that are multiples of 5 minutes, durations are expressed in intervals of 5 minutes rather than in 1 minute and followed by fitting Weibull distribution, Log-normal distribution, and Log-logistic distribution.

Table 6-7 Highway traffic accidents - Weibull AFT model considering 5 minutes interval time

Variable	Estimated Coefficient	t-statistics	Percentage change
Temporal Characteristics			
Monday	-0.12	-2.23	-12.64
Tuesday	-0.10	-1.85	-10.78
Wednesday	0.11	2.11	11.64
April	-0.16	-2.66	-16.64
May	-0.18	-2.89	-17.96
July	0.31	4.45	37.56
September	0.23	3.32	26.83
Geographical Characteristics			
Musafah HW	-0.21	-3.53	-19.97
Umm Alnar HW	0.15	2.23	17.32
AbuDhabi-Tarif HW	0.29	6.19	34.71
Almafraq-Sawameq HW	-0.15	-2.42	-15.81
Swihan HW	0.34	3.67	41.57
Accident Characteristics			
Hit pedestrian	-0.17	-3.88	-17.16
Slight injury	-0.12	-1.86	-12.88
Serious injury	-0.17	-3.19	-16.78
Number of casualties	-0.02	-2.02	-3.03
Model Structure Parameters			
p (distribution shape parameter)	2.51	30.35	
λ (the scale parameter)	19.13	44.16	
Intercept term (_cons)	2.95	44.16	
Goodness-of-fit Test			
Akaike information criterion	645.95		
Initial log-likelihood	-382.16		
Log-likelihood at convergence	-304.97		
Number of observations	504		

Table 6-8 Highway traffic accidents - Log-normal AFT model considering 5 minutes interval time

Variable	**Estimated Coefficient**	**t-statistics**	**Percentage change**
Temporal Characteristics			
Wednesday	0.11	2.26	12.47
Friday	0.10	1.91	11.06
February	-0.44	-6.88	-37.14
April	-0.29	-4.51	-26.36
May	-0.24	-3.86	-23.1
September	0.12	1.78	13.76
Geographical Characteristics			
Musafah HW	-0.23	-3.97	-22.22
AbuDhabi- Tarif HW	0.09	2.06	10.04
Almafraq –Swamiq HW	-0.26	-4.08	-24.64
Swihan HW	0.36	3.83	44.58
Accident Characteristics			
Hit pedestrian	-0.08	-1.83	-9.15
Overturn	0.15	2.94	16.66
Model Structure Parameters			
β (distribution shape parameter)	2.61	79.71	
σ (the scale parameter)	0.41	31.74	
Intercept term (_cons)	2.61	79.71	
Goodness-of-fit test			
Akaike information criterion	561.35		
Initial log-likelihood	-337.40		
Log-likelihood at convergence	-266.67		
Number of observations	504		

Table 6-9 Highway traffic accidents - Log-logistic AFT model considering 5 minutes interval time

Variable	Estimated Coefficient	t-statistics	Percentage change
Temporal Characteristics			
Wednesday	0.10	2.07	10.79
Friday	0.09	1.81	9.99
February	-0.44	-7.03	-36.88
April	-0.31	-4.83	-26.85
May	-0.28	-4.63	-24.42
Geographical Characteristics			
Musafah HW	-0.23	-4.04	-21.63
AbuDhabi- Tarif HW	0.08	1.84	8.92
Almafraq –Swamiq HW	-0.24	-3.7	-22.77
Swihan HW	0.34	3.89	41.75
Accident Characteristics			
Hit pedestrian	-0.08	-1.93	-9.35
Overturn	0.15	3.05	16.45
Environmental Characteristics			
Sandy	0.17	2.06	19.44
Model Structure Parameters			
P (distribution shape parameter)	0.22	26.73	
λ (the scale parameter)	2.62	86.96	
Intercept term (_cons)	2.62	86.96	
Goodness-of-fit Test			
Akaike information criterion	550.72		
Initial log-likelihood	-333.63		
Log-likelihood at convergence	-261.36		
Number of observations	504		

Following that, Figure 6.4 presents the plots of comparing the observed duration to the predicted duration for the three distributions with duration expressed in 5 minute intervals. Similar to 1 minute time intervals, the plots show a weak relationship between the observed and predicted duration in all three distributions (see figure notes). Also, testing for the best fit distribution, using the Akaike Information Criterion (AIC) (Table 6-10), shows that the Log-logistic AFT model is the best fit distribution for highway accident response time as the log-logistic distribution has the lowest AIC value.

Model fit statistics are as follows: $R^2 = 0.15$ for the Weibull AFT model, $R^2 = 0.19$ for the Log-normal AFT model, and $R^2 = 0.18$ for the Log-logistic AFT mode). For MAPE, Weibull AFT model = 21.49%, Log-normal AFT model = 18.53%, and Log-logistic AFT model = 19.51%. For RMSE, Weibull AFT model = 6.73, Log-normal AFT model = 6.56, and Log-logistic AFT model = 6.59.

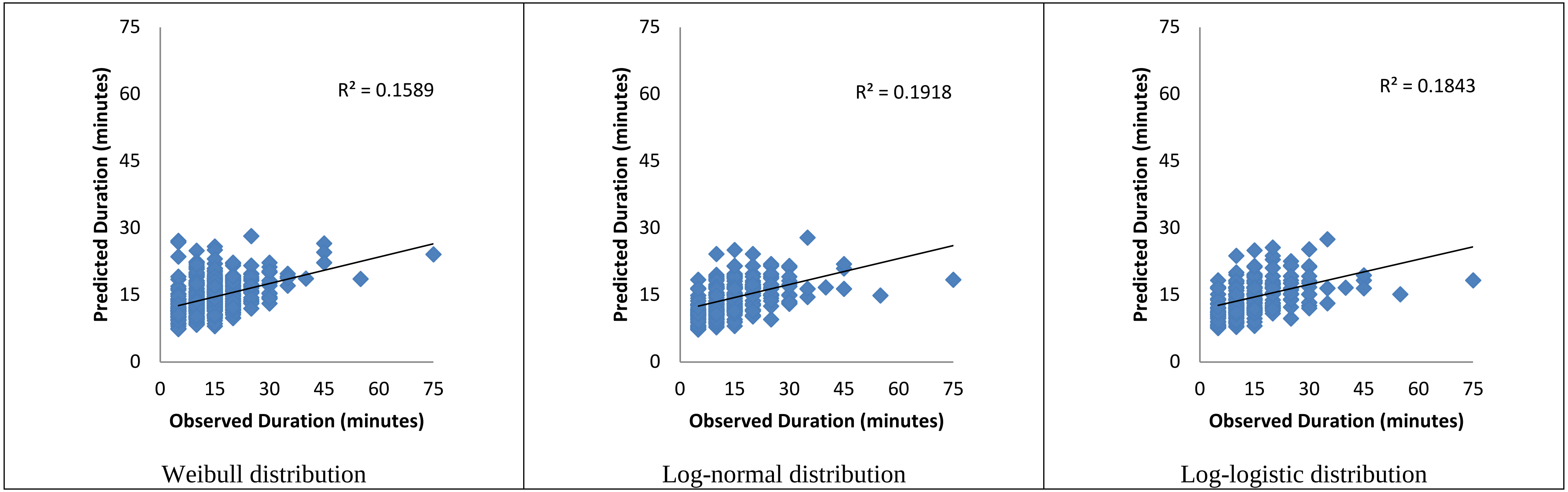


Figure 6-4 Comparison between the observed and predicted duration for Weibull, Log-normal and Log-logistic distributions considering 5 minutes interval time

Table 6-10 Comparison of AIC test value for AFT response time models (5 minutes interval time)

Model	-2In *L*	K	C	AIC
Weibull	609.94	16	2	645.95
Log-normal	533.34	12	2	561.35
Log-logistic	**522.72**	**12**	**2**	**550.72**

Investigating the effect of rounding duration to the values that are multiples of 5 minutes has shown several results. Firstly, the results obtained for two sets of fitted models with respect to 1 minute interval time and 5 minute interval time do not explain a considerable difference between them. Secondly, the results of the AIC test shows that the Log-logistic AFT model is found to be the best fit distribution regardless of the length of interval time. Finally, plotting the observed duration to predicted duration for the three distributions was found to be weak, regardless of the selection of the time interval.

As a result of the discussion in the previous paragraph, it is difficult to justify the appropriateness of grouping data into 5 minute intervals. Also the rounding off may reduce the accuracy of outputs. Even though there are some spikes in the observed duration data distribution (Table 5-1, Table 5-2, Table 5-3, Table 5-9), it is difficult to prove that there are measurement errors in the sample. In addition, grouping the data into durations of 5 minutes can be helpful for graphing and visualizing the data, but in terms of modelling this might cause a loss of valuable information in the dataset. For example, if the duration data (in minutes) is as follows: 1, 2, 3, 4, 5, 5, 10, 10, 1, 2, using a 5 minute interval will reduce the data to 8 durations of 5 minutes and 2 durations of 10 minutes. Grouping data in 5 minute intervals was considered inappropriate, as valuable information would be lost from the dataset. Therefore a 1 minute interval was considered most appropriate for the study.

6.3 Urban Accidents Duration

In this study, the data related to reporting, response and clearance time for traffic accidents were collected for accidents that occurred in the urban area of Abu Dhabi. The findings of models for each interval time are presented separately in sections 6.3.1, 6.3.2 and 6.3.3. Appendix 2 presents some of the STATA outputs.

6.3.1 Reporting Time

As mentioned earlier, reporting time is defined as the time between accident occurrence and the Abu Dhabi Urban Collision Investigation Branch (AUCIB) being informed about the accident. In general, the call operator from the police operational centre receives a phone call when an accident occurs in the system. Then, the call operator informs the AUCIB about the accident.

Descriptive analysis shows that the mean of the reporting time was 8.23 minutes, with a range span of 1 minute to 270 minutes. The standard deviation was 17.76 minutes, variance was 315.71 minutes and Kurtosis was 119.57 minutes. The mean duration was found to be comparatively smaller than the figures from previous research in other countries. For example, the mean duration of reporting time was found to be 12.2 minutes in Washington State (Nam and Mannering, 2000a).

The density distributions with respect to reporting time in this study were found to be skewed (9.54) to the right, due to the differences between the mean value and the median value (see Figure 6-5). A detailed analysis was conducted considering three distributions (Weibull, Log-normal, and Log-logistic) as discussed earlier (see sections 6.3.1.1, 6.3.1.2, 6.3.1.3). This was carried out because these distributions can be written in AFT metric.

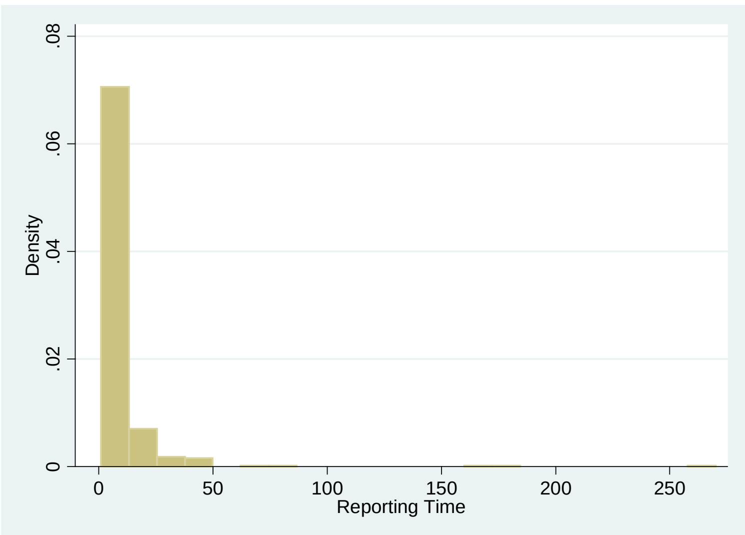


Figure 6-5 Reporting time analysis - Density distribution for accident reporting time

6.3.1.1 *Weibull Distribution*

Firstly, the reporting time data was tested with Weibull distribution. Selecting explanatory variables for the reporting time model was carried out in three steps, as explained in the Methodology chapter. The first step was to develop a base model (null model). It was found that the log-likelihood before convergence was -838.62 and AIC was 1681.24 for the base model. The baseline hazard function with respect to Weibull distribution is presented in Figure 6-6.

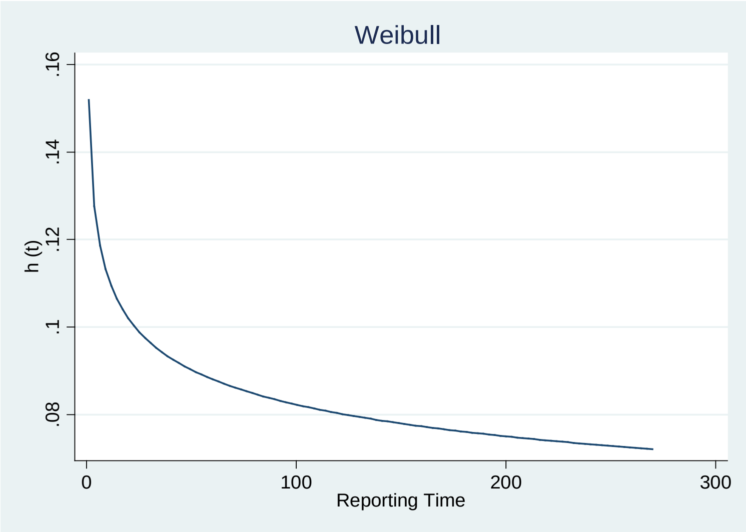


Figure 6-6 Reporting time analysis - The baseline hazard of Weibull distribution

The next step was to identify which variable, on its own, significantly decreases the statistic of $-2\log L$. Among the explanatory variables in the dataset, 41 variables show that they are significant at 85% (see Table 6-11).

Then, all of the significant variables in Table 6-11 were used in the analysis and estimated in the model. The result showed that some of these variables were insignificant in the model when integrated with other variables due to multicollinearity effects. Thus, only variables with 90% level of significance were used to estimate the final model. This step resulted in only 10 variables (Table 6-12).

Table 6-11 Urban reporting time analysis -Weibull model: the list of significant variables

Variable	-2log *L*	*LL*	AIC	t-test
Slight injury	1673.99	-836.99	1679.99	-1.90
Sunday	1670.52	-835.26	1676.52	-2.74
Monday	1666.66	-833.33	1672.66	3.16
Tuesday	1674.22	-837.11	1680.22	1.70
Friday	1673.13	-836.56	1679.13	-2.10
Hit pedestrian	1673.72	-836.86	1679.72	1.86
Angle collision	1671.95	-835.97	1677.95	-2.36
Other types of accidents	1668.91	-834.45	1674.91	2.75
Al Markaziyah	1673.80	-836.90	1679.80	-1.94
Al Meena	1673.03	-836.51	1679.03	1.99
Madinat Zayed	1673.87	-836.93	1679.87	1.77
Clear	1675.34	-837.67	1681.34	1.51
Number of vehicles involved	1665.85	-832.92	1671.85	-3.55
Dry	1675.38	-837.69	1681.38	1.48
Speeding road conditions	1672.22	-836.11	1678.22	-2.33
Failing to comply: traffic light	1674.04	-837.02	1680.04	-1.82
Exceeding statutory speed limit	1672.83	-836.41	1678.83	1.92
Failing to comply: stop sign	1673.75	-836.87	1679.75	-2.02
Other cause	1668.85	-834.43	1674.85	2.84
Rashid Bin Saeed Al Maktoum St	1672.43	-836.21	1678.43	-2.29
Eastern Ring Road	1675.08	-837.54	1681.08	-1.51
Meena St	1664.29	-832.14	1670.29	3.31
Hamdan Bin Mohammed St	1670.59	-835.29	1676.59	2.39
Sultan Bin Zayed St	1673.54	-836.77	1679.54	1.46
Residential area	1667.21	-833.60	1673.21	-3.33
Other places	1658.28	-829.14	1664.28	4.10
Junction	1674.18	-837.09	1680.18	-1.76
Car park	1673.21	-836.60	1679.21	1.89
Double road	1673.52	-836.76	1679.52	-1.96
Unknown	1668.41	-834.20	1674.41	2.80
May	1673.80	-836.90	1679.80	1.79
June	1674.24	-837.12	1680.24	1.68
September	1669.01	-834.50	1675.01	2.72
November	1671.63	-835.81	1677.63	-2.52
December	1665.29	-832.64	1671.29	-3.77
January	1673.91	-836.95	1679.91	1.75
February	1673.29	-836.64	1679.29	-2.09
Cornice Road	1667.53	-833.76	1673.53	2.84
37th Street	1675.40	-837.70	1681.40	-1.75
Delma St	1675.40	-837.70	1681.40	-1.75
Emirates Palace Roundabout	1675.40	-837.70	1681.40	-1.75

Table 6-12 Urban reporting time analysis -Weibull AFT model

Variable	Estimated Coefficient	t-statistics	Percentage Change
Temporal Characteristics			
Monday	0.31	2.51	36.69
November	-0.45	-2.74	-36.82
December	-0.78	-4.93	-54.23
February	-0.50	-3.03	-39.53
Geographical Characteristics			
Rashid Bin Saeed Al Maktoum St	-0.35	-2.48	-29.67
Eastern Ring Road	-0.28	-1.89	-24.61
Meena St	0.68	3.22	99.23
Hamdan Bin Mohammed St	0.60	2.62	83.03
Cornice road	0.92	3.95	39.53
Residential area	-0.40	-3.48	-33.59
Model Structure Parameters			
p (distribution shape parameter)	0.98	31.80	
λ (the scale parameter)	0.12	29.44	
Intercept term (_cons)	2.13	29.44	
Goodness-of-fit test			
Akaike information criterion	1611.28		
Initial Log-likelihood	-838.62		
Log-likelihood at convergence	-793.64		
Number of observations	525		

6.3.1.2 Log-normal Distribution

Secondly, Log-normal distribution was tested. Modelling the base (null) model found that the log-likelihood is -749.39 and an AIC is 1502.78. The baseline hazard function with respect to log-normal distribution is presented in Figure 6-7.

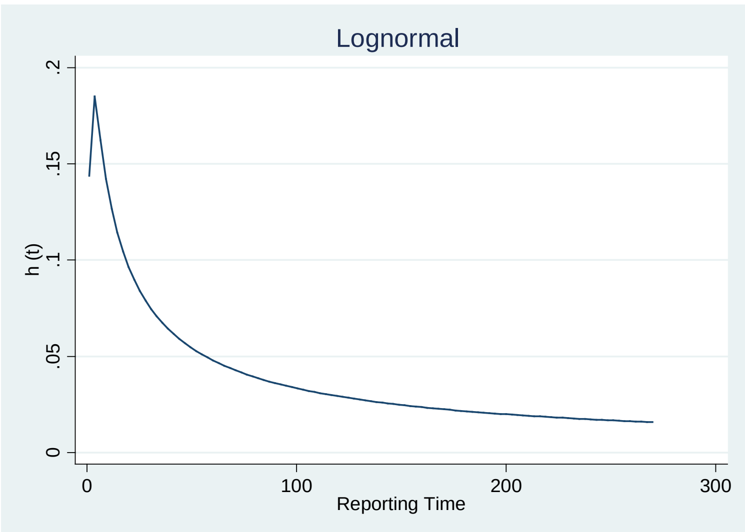


Figure 6-7 Reporting time analysis - The baseline hazard of Log-normal distribution

The second step of the variable selection procedure resulted in 17 significant variables at the level of 85% (see Table 6-13).

The final step of variable selection is to estimate the model with all the variables in Table 6-13 at once in order to find out which variable resulted in 90% level of significance in the final model. It was found that only 8 variables were significant in the model. Table 6-14 shows the resulting Log-normal AFT model with the 8 significant variables.

Table 6-13 Urban reporting time analysis -Log-normal model: the list of significant variable

Variable	**-2log L**	***LL***	**AIC**	**t-test**
Fatal injury	1493.34	-746.67	1499.34	2.34
Out of peak	1495.60	-747.80	1501.60	-1.79
Monday	1493.83	-746.92	1499.84	2.23
Friday	1495.06	-747.53	1501.06	-1.93
Angle collision	1495.16	-747.58	1501.16	-1.91
Speeding road conditions	1495.31	-747.65	1501.31	-1.87
Exceeding statutory speed limit	1496.06	-748.03	1502.06	1.65
Failing to comply: stop sign	1495.86	-747.93	1501.86	-1.71
Al Salam St	1495.52	-747.76	1502.87	1.81
Residential area	1494.16	-747.08	1500.16	-2.16
Other places	1488.74	-744.37	1494.74	3.18
November	1496.37	-748.18	1502.37	-1.55
December	1493.18	-746.59	1499.18	-2.37
March	1493.74	-746.87	1499.74	-2.25
37th St	1496.61	-748.30	1502.61	-1.48
Delma St	1496.61	-748.30	1502.61	-1.48
Emirates Palace Roundabout	1496.61	-748.30	1502.61	-1.48

Table 6-14 Urban reporting time analysis -Log-normal AFT model

Variable	Estimated Coefficient	t-statistics	Percentage Change
Temporal Characteristics			
Out of peak	-0.20	-1.98	-18.81
Friday	-0.28	-2.40	-25.15
February	-0.31	-1.95	-26.81
March	-0.38	-2.61	-32.14
May	0.41	2.89	51.76
November	-0.36	-2.25	-30.65
December	-0.41	-2.71	-33.89
Geographical Characteristics			
Meena St	0.37	1.92	45.61
Model Structure Parameters			
β (distribution shape parameter)	1.75	16.92	
σ (the scale parameter)	0.96	32.40	
Intercept term (_cons)	1.75	16.92	
Goodness-of-fit test			
Akaike information criterion	1477.62		
Initial log-likelihood	-749.39		
Log-likelihood at convergence	-728.81		
Number of observations	525		

6.3.1.3 Log-logistic Distribution

The Log-logistic distribution was tested for the data collected in Abu Dhabi. Modelling the base model (null model), it was found that the log-likelihood and AIC for the base (null) model are -751.74 and 1507.48 respectively. The baseline hazard function with respect to the Log-logistic null model is presented in Figure 6-8.

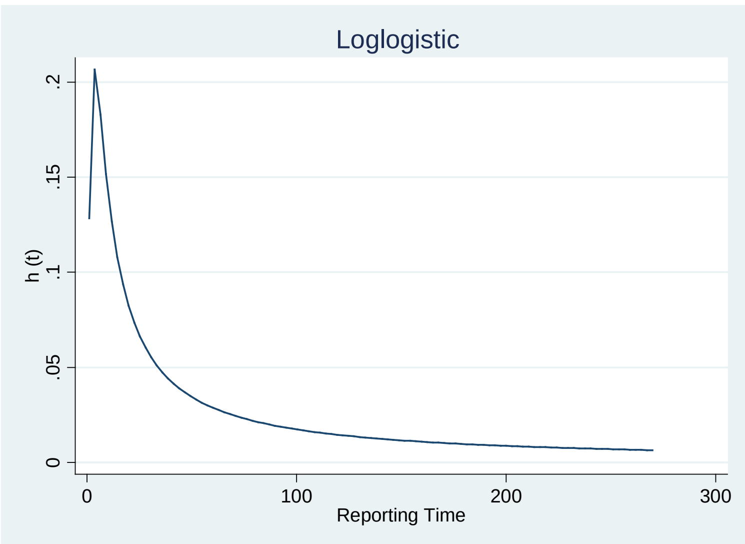


Figure 6-8 Reporting time analysis - The baseline hazard of Log-logistic distribution

The analysis was conducted as one variable at a time in order to select the significant variables in the model. Among the explanatory variables in the dataset, it was found that there were 23 significant variables at the level of 85% (see Table 6-15).

Finally, all of the significant variables in Table 6-15 were estimated in the Log-logistic AFT model. The results show that there are only 7 variables that are significant at the level of 90%. Table 6-16 presents the results of the Log-logistic AFT model.

Table 6-15 Urban reporting time analysis -Log-logistic model: the list of significant variables

Variable	**-2log *L***	***LL***	**AIC**	**t-test**
Fatal injury	1496.79	-748.39	1502.79	2.65
Out of peak	1499.45	-749.72	1505.45	-2.01
Monday	1499.85	-749.92	1505.85	1.91
Friday	1499.12	-749.56	1505.12	-2.10
Angle collision	1498.93	-749.46	1504.93	-2.14
Speeding road conditions	1500.58	-750.29	1506.58	-1.70
Exceeding statutory speed limit	1500.28	-750.14	1506.28	1.83
Failing to comply: Stop sign	1500.03	-750.01	1506.03	-1.89
Eastern Ring Rd	1501.38	-750.69	1507.38	-1.45
Al Salam St	1500.73	-750.36	1508.27	1.65
Sultan Bin Zayed St	1498.33	-749.16	1504.33	2.78
Residential area	1499.36	-749.68	1505.36	-2.03
Other places	1494.23	-747.11	1500.23	3.05
May	1488.05	-744.03	1494.05	3.96
November	1501.15	-750.57	1507.15	-1.53
December	1498.23	-749.11	1504.23	-2.30
February	1500.38	-750.19	1506.38	-1.77
March	1495.97	-747.98	1501.97	-2.78
Qasr AlAmwaj St	1500.58	-750.29	1506.59	1.91
37^{th} St	1500.87	-750.43	1506.87	-1.80
Delma St	1500.87	-750.43	1506.87	-1.80
Emirates Palace Roundabout	1500.87	-750.43	1506.87	-1.80
35^{th} Street	1501.43	-750.71	1507.43	1.56

Table 6-16 Urban reporting time analysis -Log-logistic AFT model

Variable	Estimated Coefficient	t-statistics	Percentage Change
Temporal Characteristics			
Out of peak	-0.23	-2.23	-20.65
Friday	-0.30	-2.48	-25.98
February	-0.31	-2.04	-27.04
March	-0.45	-2.95	-36.80
May	0.41	2.95	50.85
November	-0.35	-2.28	-29.71
December	-0.39	-2.72	-32.39
Model Structure Parameters			
P (distribution shape parameter)	1.82	16.59	
λ (the scale parameter)	1.76	17.08	
Intercept term (_cons)	1.76	17.08	
Goodness-of-fit test			
Akaike information criterion	1477.89		
Initial log-likelihood	-751.74		
Log-likelihood at convergence	-729.94		
Number of observations	525		

6.3.1.4 Model Selection

The results of the plots of comparing observed duration to predicted duration for the three distributions were presented in Figure 6-9. This shows that there is a weak relationship between observed and predicted durations in all distributions (see figure notes).

In terms of model fit, $R^2 = 0.07$ for Weibull AFT model, $R^2 = 0.01$ for Log-normal AFT model, and $R^2 = 0.00$ for Log-logistic AFT model. For MAPE, Weibull AFT model = 165.64%, Log-normal AFT model = 138.92, and Log-logistic AFT model = 150.05%. For RMSE, Weibull AFT model = 17.09, for Log-normal AFT model = 17.62, and Log-logistic AFT model = 17.71.

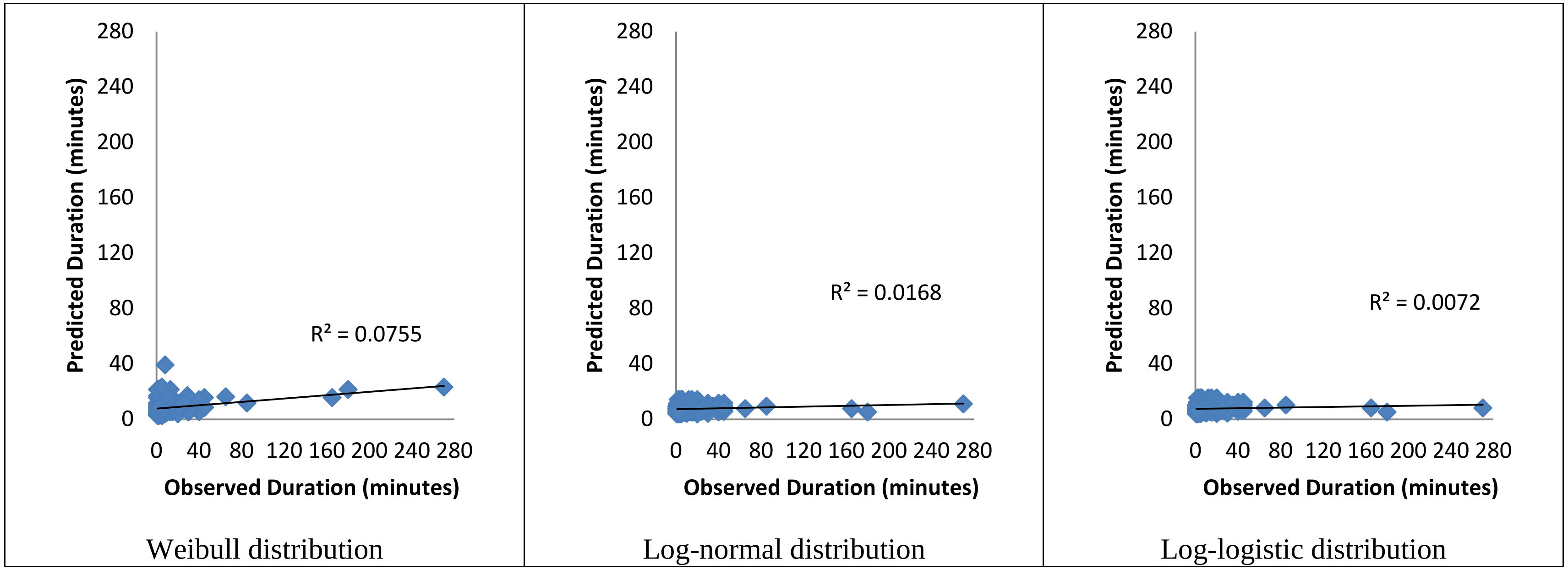


Figure 6-9 Reporting time analysis - Comparison between the observed and predicted duration for Weibull, Log-normal and Log-logistic distributions

As mentioned in the Methodology chapter, the best fit distribution is selected using AIC. Based on the results of AIC (Table 6-17), the Log-normal AFT model was selected. The following section is dedicated to interpreting the results of the selected distribution.

Table 6-17 Urban reporting time analysis -Comparison of AIC test value for AFT reporting time models

Model	-2In *L*	K	c	AIC
Weibull	1587.28	10	2	1611.28
Log-normal	1457.62	8	2	1477.62
Log-logistic	1459.88	7	2	1477.89

6.3.1.5 Interpretation of the Estimated Model

The Log-normal AFT reporting time model has 8 statistically significant explanatory variables. As shown in Table 6-14, it was found that all significant variables belonged to the category of temporal characteristics, whereas one of the variables belonged to the geographical characteristics.

Also, it is worth mentioning that accident reporting time has nothing to do with the AUCIB staff. It depends on the performance of the Abu Dhabi Police Operational Centre (APOC), which is highly dependent on the availability of the required information about an accident that can enable the call operator to assign accident investigators to the accident scene. Furthermore, APOC receives all kinds of calls regarding crimes, traffic accidents and other types of incidents.

Temporal Characteristics

Accidents that occurred out of peak period were shown to have (18.81%) lower reporting time compared to peak period accidents (Figure 6-10). As mentioned in Chapter 4, AUCIB staff will move to the accident scene if the accident resulted in serious injury or fatal injury. In this case, the call operator would wait until the first responder confirms the injury (slight, serious, fatal) status and then inform AUCIB staff if the accident had resulted in serious or fatal injuries. Since the travel time of the first responder is assumed to be faster during out of peak period, the confirmation of the injury status would be faster compared to peak period. Thus, the AUCIB staff will be reported to by the call operator more rapidly during out of peak periods compared to peak periods.

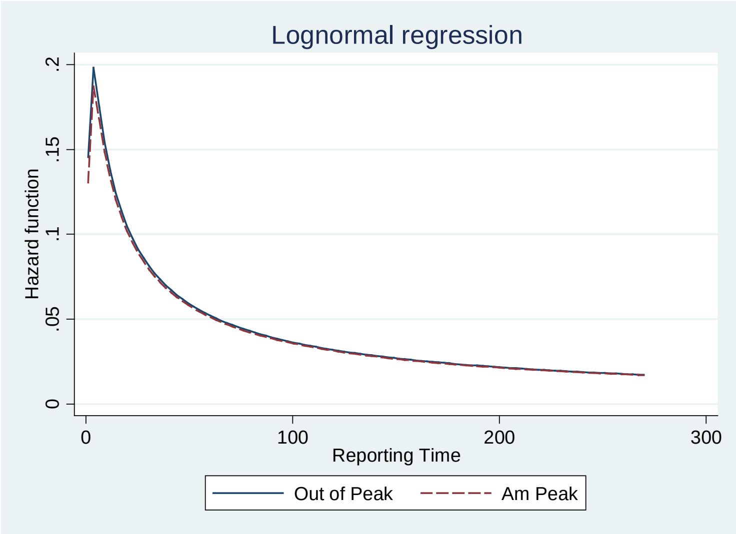


Figure 6-10 Reporting time analysis - Hazard function by peak period

Another temporal variable that was significant in the model was found to be "day of week". The results show that accidents occurring on Fridays were associated with (26.81) lower reporting time compared to other days of the week (Figure 6-11). This result was expected because the people in Abu Dhabi city travel to other emirates during the weekend (Friday and Saturday), so the number of calls received by APOC would be less during the weekend compared to weekdays. As a result, the availability of the required information regarding injury status could be the reason for this reduction in reporting time.

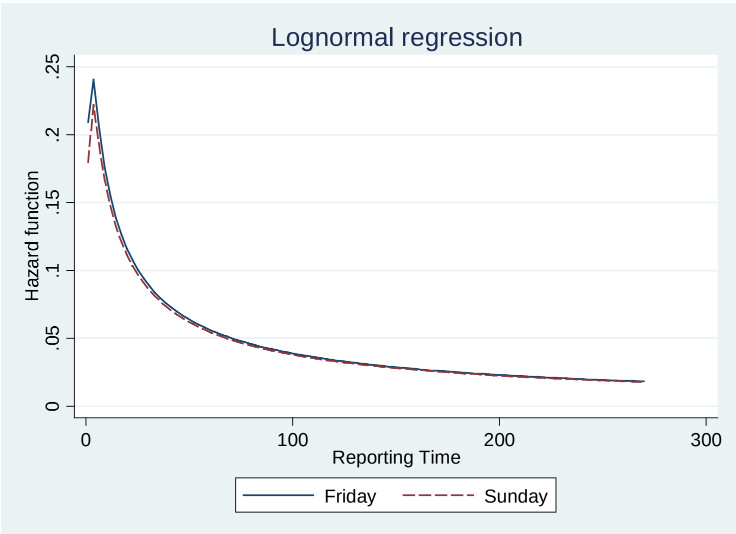


Figure 6-11 Reporting time analysis - Hazard function by day of week

Furthermore, "the month of the year" variables were found to significantly affect accident reporting time. The results show that the reporting time of accidents that occurred in February, March, November, and December were associated with (26.81%, 32.14%, 30.65%, 33.89%) lower reporting time compared to accidents in other months of the year, whereas the reporting time of accidents in May was found to have (51.76%) higher reporting time (Figure 6-12).

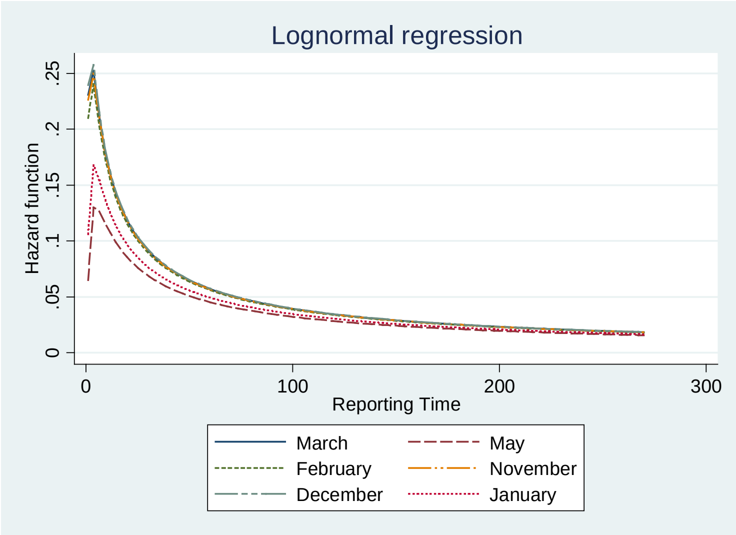


Figure 6-12 Reporting time analysis - Hazard function by month of year

Geographical Characteristics

Other variables that were found to significantly affect urban accident response time are related to accident location. Accidents that occurred on the Meena Street were found to be associated with (45.61%) higher response time compared to accidents that occurred on other roads of the city (Figure 6-13). This might be due to the failure of the CCTV system in this street.

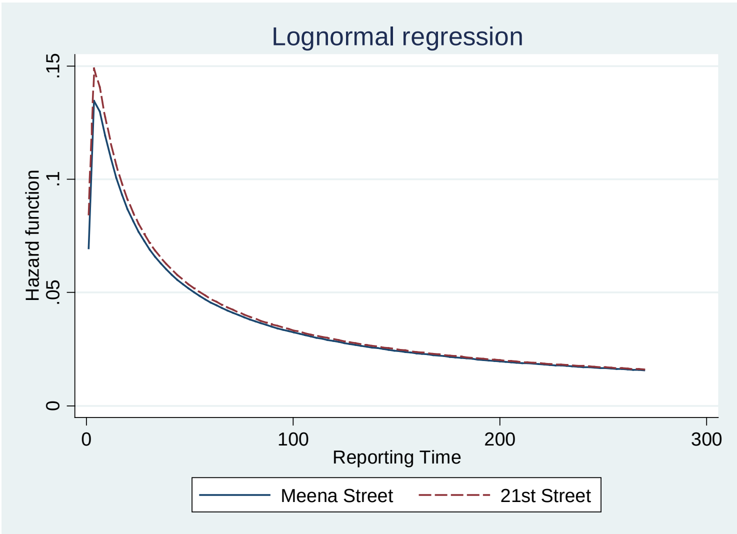


Figure 6-13 Reporting time analysis - Hazard function by street

Finally, the survival function and the hazard function are presented in Figure 6-14. Regarding the effects on reporting time, the log-normal AFT model has β = 1.75. As mentioned earlier in section 3.3.3, this distribution is similar to log-logistic distribution in terms of avoiding the assumption of monotonic hazard rate. This means that the hazard function increases first and then decreases toward zero. In this model, the hazard increased until it reached the inflection point of 3.39 minutes, after which it decreased towards zero. This implies that accidents associated with reporting times longer than 3.39 minutes are less likely to be reported soon.

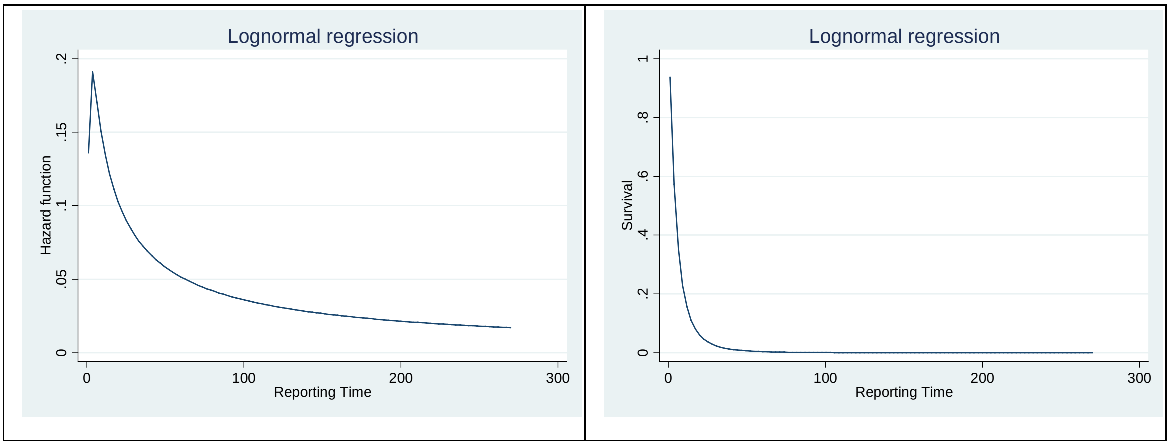


Figure 6-14 Reporting time analysis -Log-normal hazard function and survival function

Duration Prediction and Model Accuracy

For the urban reporting time, the MAPE value for the log-normal AFT model was calculated as 138.92%. It indicates that the developed model is less likely to generate an accurate predictive capability (Table 4-2). This may be due to several reasons including small sample size, inaccurate measurement of durations, or/and incomplete and incomprehensive dataset. However the development of prediction tool is considered as very important in this study as it is a very useful tool for policy makers when they devise future road safety strategies and policies in Abu Dhabi. Therefore the MAPE prediction tool has been incorporated in the analysis here regardless of the accuracy of the outcome. It is expected that the models that will be developed in the future for the same case study are will be conducted using more precise and larger datasets where the prediction accuracy will be improved to be able to use it as a useful tool for policy making.

In order to a develop decision tree for urban accident reporting time, the 8 significant variables were considered as being appropriate. Figure 6-15 and Figure 6-16 show the decision tree of reporting time in minutes, splitting first by the peak period, then day of week, following that, month of year and finally by street. If the accident has occurred during out of peak period and on a Friday, the predicted reporting time is estimated to be 5.69 minutes. If further details are available, for example the accident occurring in March, the predicted reporting time is 3.82 minutes.

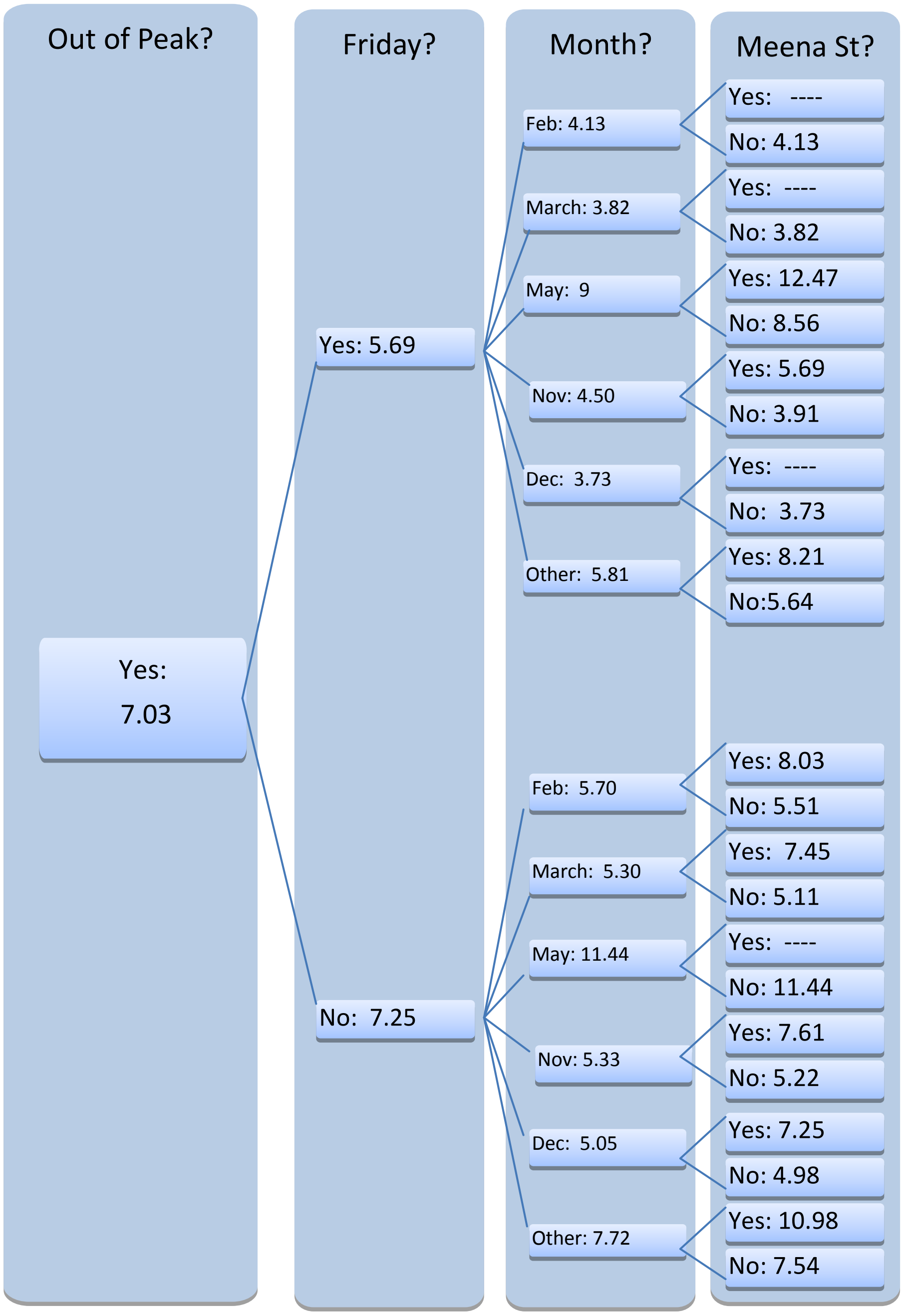


Figure 6-15 Reporting time analysis - Decision tree of out of peak periods accidents (Durations are in minutes)

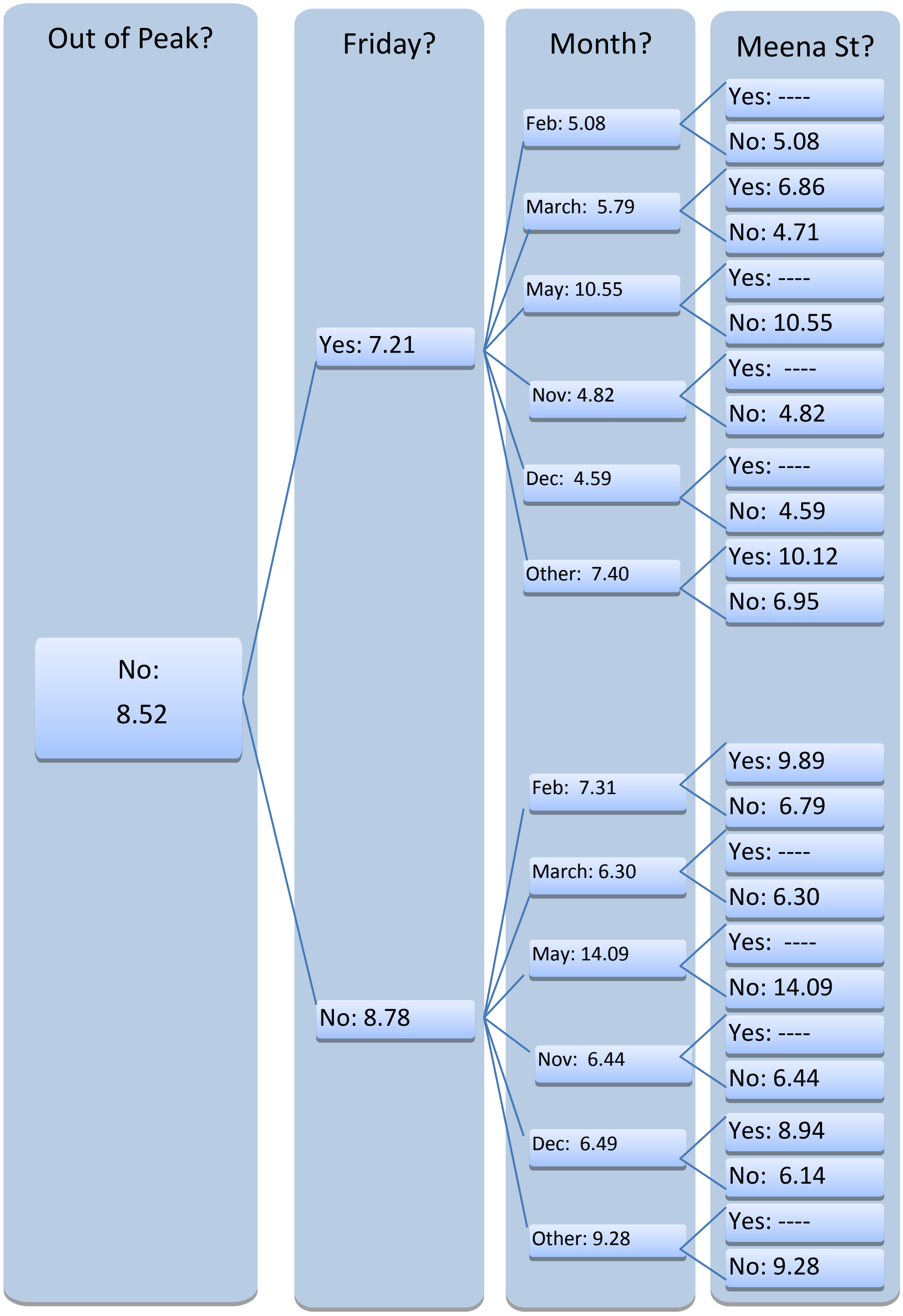


Figure 6-16 Reporting time analysis - Decision tree of peak periods accidents (Durations are in minutes)

Comparison to Previous Research

Before presenting the differences between this research and previous research, it is worth mentioning that the approach used here is not an appropriate one for comparison with the results of previous work that modelled accident duration as one piece of time. This is mainly because an empirical study (Nam and Mannering, 2000) has shown that statistically significant variables are not stable, in terms of type and effects, with respect to the interval times (reporting time, response time, clearance time). Accordingly, the comparison in this section is limited to the previous research that aimed to model each interval time of the total incident duration separately.

This research shows some differences to previous research in this context, in terms of the best fit distribution and the resultant significant explanatory variables. As found by Nam and Mannering (2000), temporal characteristics, environmental characteristics, geographical characteristics and accident characteristics were found to significantly affect reporting time. However, in this research, only variables from two categories were found to significantly affect reporting time, namely temporal characteristics and geographical characteristics.

On the other hand, the resultant best fit distribution for the reporting time in this research differs from previous research. For example, the Weibull model with gamma heterogeneity provided the best fit distribution in the study conducted by Nam and Mannering (2000). However, in this research, the Log-normal model was found to provide the best fit distribution. Finally, having stated these differences in the results, it is clear that different datasets and case study areas may yield different results.

6.3.2 Response Time

For the purpose of this study, response time is defined as the time between the AUCIB being informed about the accident and the arrival of the first accident investigator at the scene. In other words, it refers to the collision investigator's travel time to the accident scene.

Descriptive analysis shows that the mean of the response time was 6.48 minutes, with a range spanning from 1 minute to 64 minutes. The standard deviation was 5.15 minutes, variance was 26.56 minutes and Kurtosis was 34.03 minutes.

The density distributions with respect to response time in this study were found to be skewed (3.78) to the right due to the differences between the mean value and the median value (see Figure 6-17). Detailed analysis was conducted considering three distributions (Weibull, Log-normal and Log-logistic), as presented earlier (see sections 6.3.2.1, 6.3.2.2, 6.3.2.3).

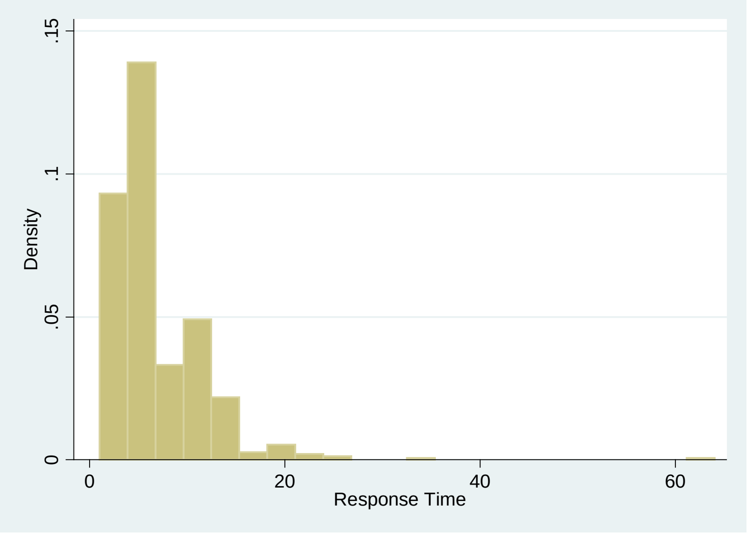


Figure 6-17 Response time analysis - Density distribution for accident response time

6.3.2.1 *Weibull Distribution*

Firstly, the response time data was tested with Weibull distribution. The first step of modelling the base (null) model found that the log-likelihood before convergence is -595.83 and AIC is 1195.66. The baseline hazard function with respect to Weibull distribution is presented in Figure 6-18.

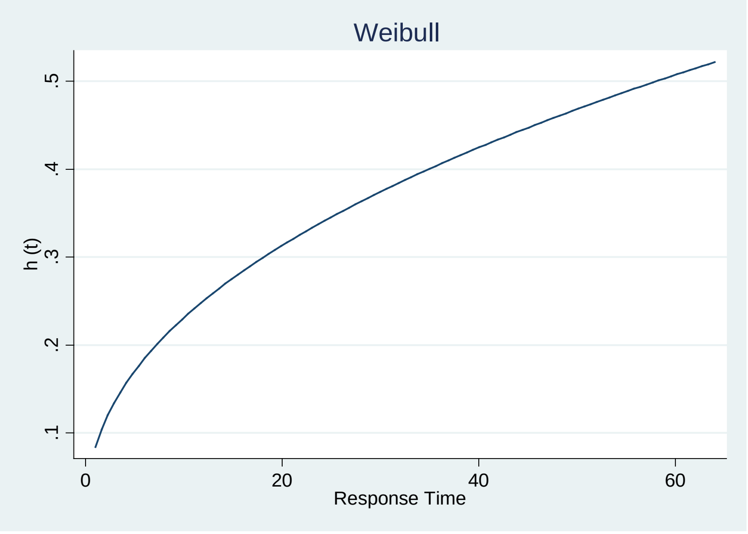


Figure 6-18 Response time analysis - The baseline hazard of Weibull distribution

The analysis was conducted as one variable at a time in order to select the significant variables in the model. Among the explanatory variables in the dataset, it was found that there were 30 significant variables at the level of 85% (see Table 6-18).

Table 6-18 Urban response time analysis - Weibull model: the list of significant variables

Variable	**-2log *L***	***LL***	**AIC**	**t-test**
Slight injury	1184.42	-592.21	1190.42	2.55
Number of casualties	1183.40	-591.70	1189.40	2.64
Afternoon	1184.92	-592.46	1190.92	-2.67
Sunday	1187.29	-593.64	1193.29	-2.18
Monday	1184.64	-592.32	1190.64	2.57
Tuesday	1185.36	-592.68	1191.36	2.45
Friday	1183.80	-591.90	1189.80	-2.95
Sideswipe collision	1185.15	-592.57	1191.15	2.48
Hit pedestrian	1189.04	-594.52	1195.04	-1.64
Rear end collision	1187.08	-593.54	1193.08	-2.23
Angle collision	1187.13	-593.56	1193.13	2.10
Other types of accidents	1188.61	-594.30	1194.61	-1.82
Al Bateen	1189.29	-594.64	1195.29	1.49
Hadbat Al Zafaranah	1184.92	-592.46	1190.92	2.48
Rain	1187.43	-593.71	1193.43	1.73
Dry	1188.67	-594.33	1194.67	-1.58
Did not leave enough space	1181.29	-590.64	1187.29	-3.49
Drink or drug driving	1185.11	-592.55	1191.11	2.42
Eastern Ring Rd	1185.20	-592.60	1191.20	-2.69
Sheikh Zayed The First St	1189.47	-594.73	1195.47	-1.54
Al Saada St	1189.74	-594.87	1195.74	-1.56
Government authority	1189.09	-594.54	1195.09	-1.64
School	1165.43	-582.71	1171.43	4.60
Junction	1186.40	-593.20	1192.40	2.30
November	1181.30	-590.65	1187.30	3.06
February	1179.99	-589.99	1185.99	-3.74
Mohammed Bin Khalifa Street	1189.69	-594.84	1195.69	-1.62
23rd St	1187.85	-593.92	1193.85	-2.85
37th St	1189.65	-594.82	1195.65	-1.85
Emirates Palace Roundabout	1187.85	-593.92	1193.85	-2.85

Subsequently, all of the significant variables in Table 6-18 were used in the analysis and estimate of the model. The results show that there were only 7 variables that significantly affected response times at the level of 90%. Table 6-19 presents the findings of the Weibull AFT model.

Table 6-19 Urban response time analysis - Weibull AFT model

Variable	Estimated Coefficient	t-statistics	Percentage change
Temporal Characteristics			
Monday	0.19	2.59	21.56
Tuesday	0.17	2.12	19.25
December	-0.21	-2.17	-19.43
February	-0.37	-3.71	-31.46
Geographical Characteristics			
Eastern Ring Road	-0.27	-3.06	-24.14
School	0.57	4.10	76.88
Environmental Characteristics			
Dry road condition	-0.55	-2.55	-42.69
Model Structure Parameters			
p (distribution shape parameter)	1.59	30.27	
λ (the scale parameter)	0.08	11.45	
Intercept term (_cons)	2.50	11.45	
Goodness-of-fit test			
Akaike information criterion	1145.16		
Initial log-likelihood	-595.83		
Log-likelihood at convergence	-563.58		
Number of observations	525		

6.3.2.2 Log-normal Distribution

Secondly, Log-normal distribution was tested. Developing the base (null) model revealed that the log-likelihood is -563.05 and AIC is 1130.11 for the base model. The baseline hazard function with respect to Log-normal distribution is presented in Figure 6-19.

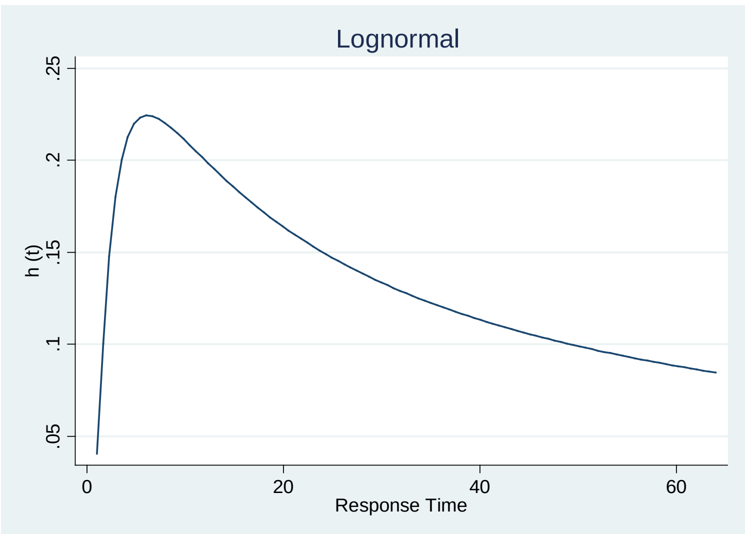


Figure 6-19 Response time analysis - The baseline hazard of Log-normal distribution

The second step of variable selection shows that 18 variables significantly affect response time at the level of 85% (see Table 6-20).

Table 6-20 Urban response time analysis - Log-normal model: list of significant variables

Variable	-2log *L*	*LL*	AIC	t-test
Slight injury	1118.33	-559.16	1124.33	2.80
Fatal injury	1118.30	-559.15	1124.30	2.80
Number of casualties	1113.13	-556.56	1119.13	3.62
Friday	1121.89	-560.94	1127.89	-2.06
Sideswipe collision	1121.94	-560.97	1127.94	2.05
Hit pedestrian	1123.57	-561.79	1129.57	-1.59
Daylight	1123.74	-561.87	1129.74	1.54
Darkness	1123.74	-561.87	1129.74	-1.54
Rain	1123.34	-561.67	1129.34	1.66
Did not leave enough space	1122.31	-561.15	1128.31	-1.95
Exceeding statutory speed limit	1123.52	-561.76	1129.52	1.61
Rashid Bin Saeed Al Maktoum St	1124.00	-562.00	1130.00	1.45
Eastern Ring Rd	1122.58	-561.29	1128.58	-1.88
Sheikh Zayed The First St	1123.70	-561.85	1129.71	-1.55
Other places	1123.19	-561.59	1129.19	1.71
February	1123.07	-561.53	1129.07	-1.75
23rd Street	1120.78	-560.39	1126.78	-2.31
Emirates Palace Roundabout	1120.78	-560.39	1126.78	-2.31

Table 6-21 Urban response time analysis - Log-normal AFT model

Variable	Estimated Coefficient	t-statistics	Percentage change
Temporal Characteristics			
February	-0.23	-2.08	-20.98
December	-0.19	-1.82	-17.97
Geographical Characteristics			
Eastern Ring Rd	-0.19	-1.97	-17.99
Environmental Characteristics			
Rain	0.83	2.03	30.35
Model Structure Parameters			
β (distribution shape parameter)	1.67	47.64	
σ (the scale parameter)	0.69	32.40	
Intercept term (_cons)	1.67	47.64	
Goodness-of-fit test			
Akaike information criterion	1125.08		
Initial log-likelihood	-563.05		
Log-likelihood at convergence	-556.54		
Number of observations	525		

The third step of variable selection is to estimate the model with the all significant variables in Table 6-20 at once in order to find which variables resulted in 90% level of significance in the final model. It was found that only 4 variables were significant in the model. Table 6-21 shows the Log-normal AFT model.

6.3.2.3 Log-logistic Distribution

The Log-logistic distribution was tested for the data collected in Abu Dhabi. Modelling the base (null) model, it was found that the log-likelihood and AIC are -562.73 and 1129.46 respectively. The baseline hazard function, with respect to the log-logistic null model, is presented in Figure 6-20.

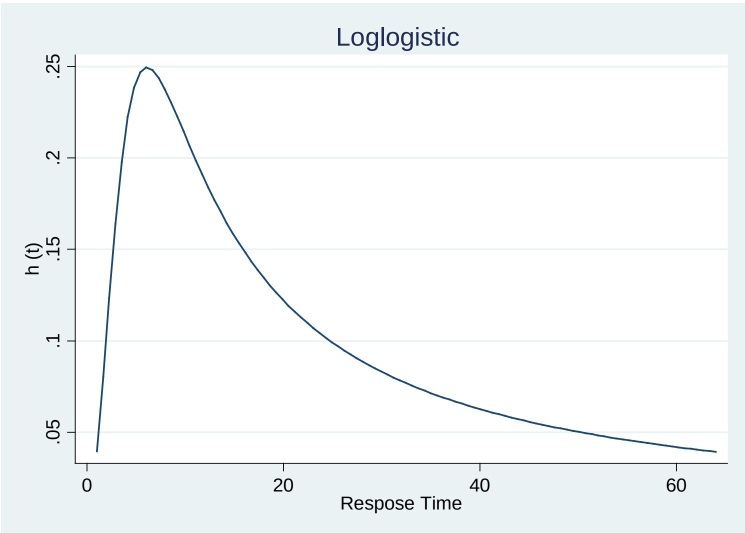


Figure 6-20 Response time analysis - The baseline hazard of Log-logistic distribution

The analysis was conducted as one variable at a time in order to select the significant variables in the model. Among the explanatory variables in the dataset, it was found that there were 21 significant variables at the level of 85% (see Table 6-22).

Table 6-22 Urban response time analysis - Log-logistic model: the list of significant variables

Variable	-2log *L*	*LL*	AIC	t-test
Slight injury	1116.93	-558.46	1122.93	2.96
Serious	1122.23	-561.11	1128.23	1.82
Fatal injury	1115.54	-557.77	1121.54	3.28
Number of casualties	1109.47	-554.73	1115.47	4.11
Monday	1123.15	-561.57	1129.15	1.52
Friday	1122.19	-561.09	1128.19	-1.81
Sideswipe collision	1117.65	-558.82	1123.65	2.82
Hit pedestrian	1121.45	-560.72	1127.45	-2.01
Daylight	1122.62	-561.31	1128.62	1.69
Darkness	1122.62	-561.31	1128.62	-1.69
Rain	1123.18	-561.59	1129.18	1.52
Did not leave enough space	1120.96	-560.48	1126.96	-2.13
Exceeding statutory speed limit	1122.83	-561.41	1128.83	1.63
Khalifa Bin Zayed St	1121.74	-560.87	1127.74	-2.02
Rashid Bin Saeed Al Maktoum St	1122.33	-561.16	1128.33	1.78
Eastern Ring Rd	1121.80	-560.90	1127.80	-1.92
Sheikh Zayed The First St	1122.22	-561.11	1128.22	-1.81
Other places	1122.13	-561.06	1128.13	1.83
February	1122.38	-561.19	1128.38	-1.75
23rd Street	1119.86	-559.93	1125.86	-2.94
Emirates Palace Roundabout	1119.86	-559.93	1125.86	-2.94

Then, all of the significant variables in Table 6-22 were used to estimate the Log-logistic AFT model. The results show that there are only 4 variables that are significant at the level of 90%. Table 6-23 presents the Log-logistic AFT model.

Table 6-23 Urban response time analysis - Log-logistic AFT model

Variable	Estimated Coefficient	t-statistics	Percentage change
Geographical Characteristics			
Khalifa Bin Zayed St	-0.48	-2.12	-38.39
Eastern Ring Rd	-0.20	-2.19	-18.89
Environmental Characteristics			
Wet Road Condition	0.44	1.27	56.24
Day light	0.11	1.84	12.09
Model Structure Parameters			
P (distribution shape parameter)	2.55	25.62	
λ (the scale parameter)	0.20	31.56	
Intercept term (_cons)	1.59	31.56	
Goodness-of-fit test			
Akaike information criterion	1125.01		
Initial log-likelihood	-562.73		
Log-likelihood at convergence	-555.51		
Number of observations	525		

6.3.2.4 Model Selection

The results of the plots of comparing the observed duration to the predicted duration for the three distributions were presented in Figure 6-21. This figure shows that the plots show a weak relationship between observed and predicted durations in all distributions (see figure notes).

In terms of model fit, $R^2 = 0.06$ for Weibull AFT model, $R^2 = 0.03$ for Log-normal AFT model, and $R^2 = 0.00$ for Log-logistic AFT model. For MAPE, Weibull AFT model = 65.92%, Log-normal AFT model = 65.33%, and Log-logistic AFT model = 71.25%. For RMSE, Weibull AFT model = 4.96, Log-normal AFT model = 5.06, and Log-logistic AFT model = 5.14.

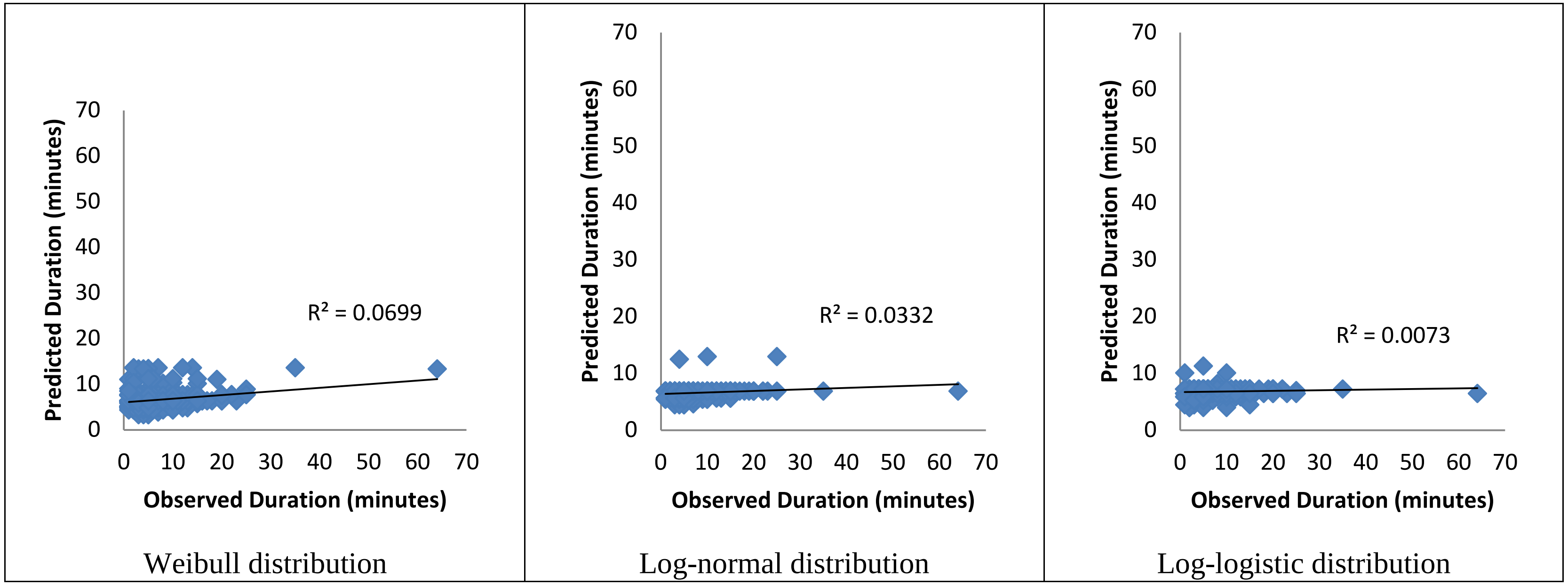


Figure 6-21 Response time analysis - Comparison between the observed and predicted duration for Weibull, Log-normal and Log-logistic distributions

According to the goodness-of-fit test (AIC), the results show that the Log-logistic AFT model is the best fit distribution for urban accident response time (Table 6-24). So, the following section will demonstrate the results of this model and the effect of the significant variables on the response time.

Table 6-24 Urban response time analysis - Comparison of AIC test value for AFT response time models

Model	-2In *L*	K	C	AIC
Weibull	1127.16	7	2	1145.16
Log-normal	1113.08	4	2	1125.08
Log-logistic	**1111.02**	**4**	**2**	**1125.01**

6.3.2.5 Interpretation of the Estimated Model

The Log-logistic AFT response time model has 4 statistically significant explanatory variables. As illustrated in Table 6-23, it was found that the significant variables belong to two categories including geographical characteristics and environmental characteristics.

Geographical Characteristics

Two variables that were found to significantly affect urban accident response time are related to accident location. Accidents that occurred on the Eastern Ring Road and Khalifa Bin Zayed Street were found to be associated with (18.89%, 38.39%) lower response time compared to accidents that occurred on other roads of the city (Figure 6-22). After carefully examining the distance between the location of the AUCIB and the roads included in this study, this finding could be interpreted as being attributable to the finding that these roads are the nearest roads to the location of the AUCIB. Thus, it was expected that the travel time of AUCIB staff to accidents on these roads would be faster compared to the response time to accidents that occurred on other roads of the city.

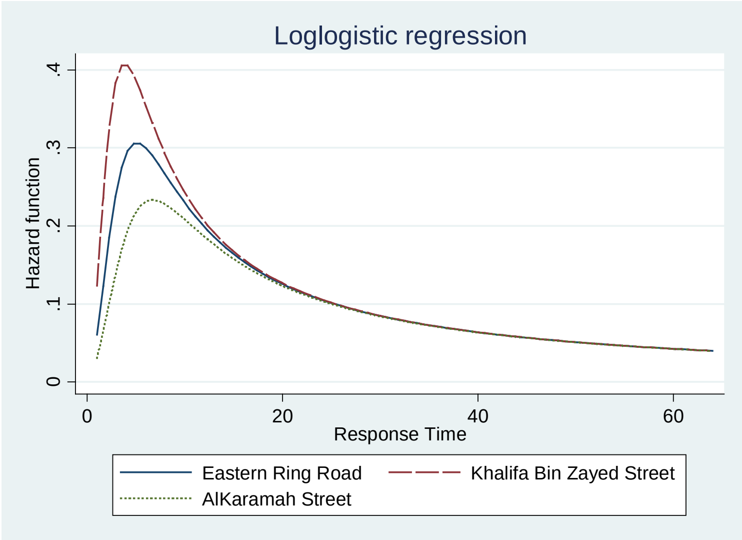


Figure 6-22 Response time analysis - Hazard function by road

Environmental Characteristics

In terms of the environmental characteristics, two variables were found to significantly affect the response time. Accidents that occurred on a wet road surface condition were associated with (56.24%) higher response time compared with accidents that happened under other road surface conditions (Figure 6-23). As expected, a wet road surface will cause a delay to the travel time to the accident scenes, but not clear conditions. Another factor is day light condition. Accidents that occurred under day light condition were associated with (12.09%) higher response time compared with accidents that happened during the night (Figure 6-24). This might be because 63% of the accidents occurred during the day compared to 37% at night.

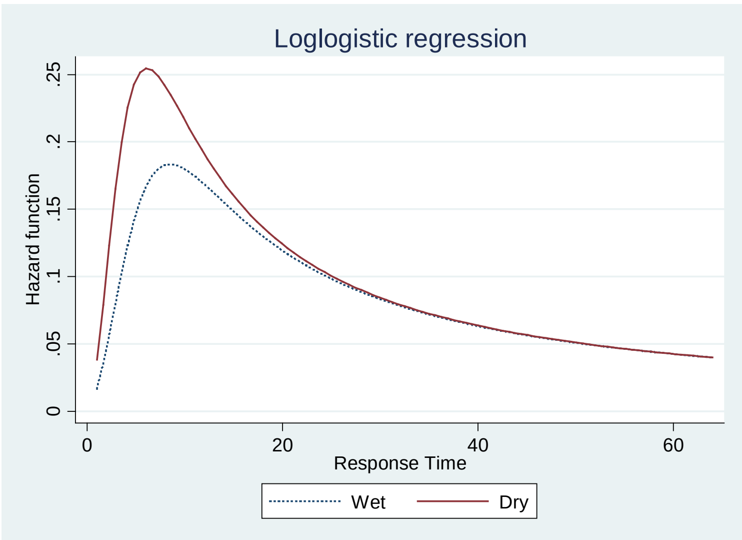


Figure 6-23 Response time analysis - Hazard function by road surface

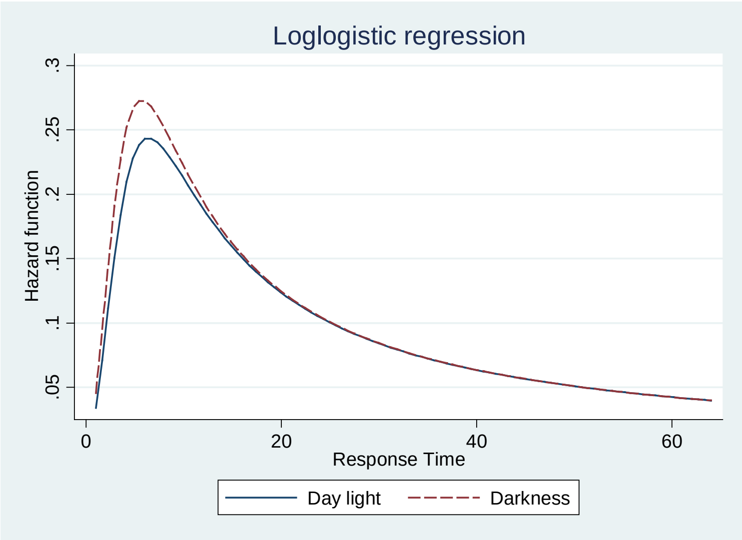


Figure 6-24 Response time analysis - Hazard function by light condition

Finally, the survival function and the hazard function are presented in Figure 6-25. Regarding the effects on response time, the log-logistic AFT model has P = 2.55. As mentioned earlier in section 3.3.3, if P > 1 this means having a non-monotonic slope and increasing from zero to a maximum at time $t = [(P - 1)^{1/P}] / \lambda$ and decreasing thereafter. In other words, there is higher chance of response for the accident before the maximum of the log-logistic distribution is reached after duration of about 5.93 minutes. However, after this time, the further the time gone without an AUCIB response, the less likely it is to have an AUCIB response soon.

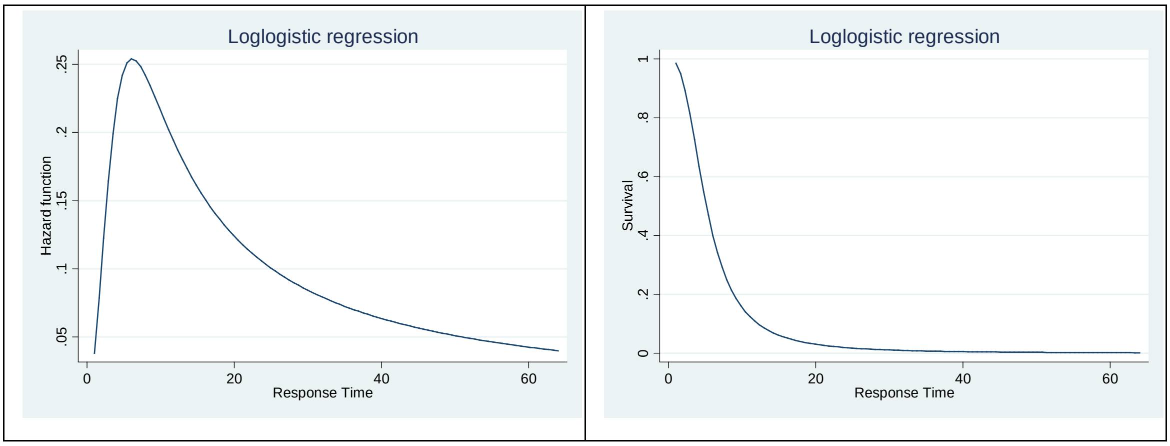


Figure 6-25 Response time analysis - Log-logistic hazard function and survival function

Duration Prediction and Model Accuracy

After fitting the log-logistic AFT model, the prediction of mean duration was performed. Afterwards, the MAPE value for the log-logistic AFT model was calculated as 71.25%, meaning that the prediction with the developed model is less likely to generate an accurate predictive capability (Table 4-2). As mentioned in the urban reporting time section, this might be because of small sample size, inaccurate measurement of durations, or/and incomplete and incomprehensive dataset. However, it is expected that the prediction accuracy will be improved in future when using more precise and larger datasets.

The significant independent variables that resulted from the urban response time model were utilised to develop a decision tree for urban accident response time. These variables are day light, wet, Khalifa Bin Zayed Street and Eastern Ring road. As shown in Figure 6-26 and Figure 6-27, the first split is made according to the day light condition, then according to road surface condition, and finally by road. This tree can be used to predict the response time in minutes for urban accidents. For example, if the accident occurred on the day light condition and was on a wet road surface, the predicted time is 9.09 minutes. However, if the accident happened on another other road surface condition, the predicted time is 6.98 minutes.

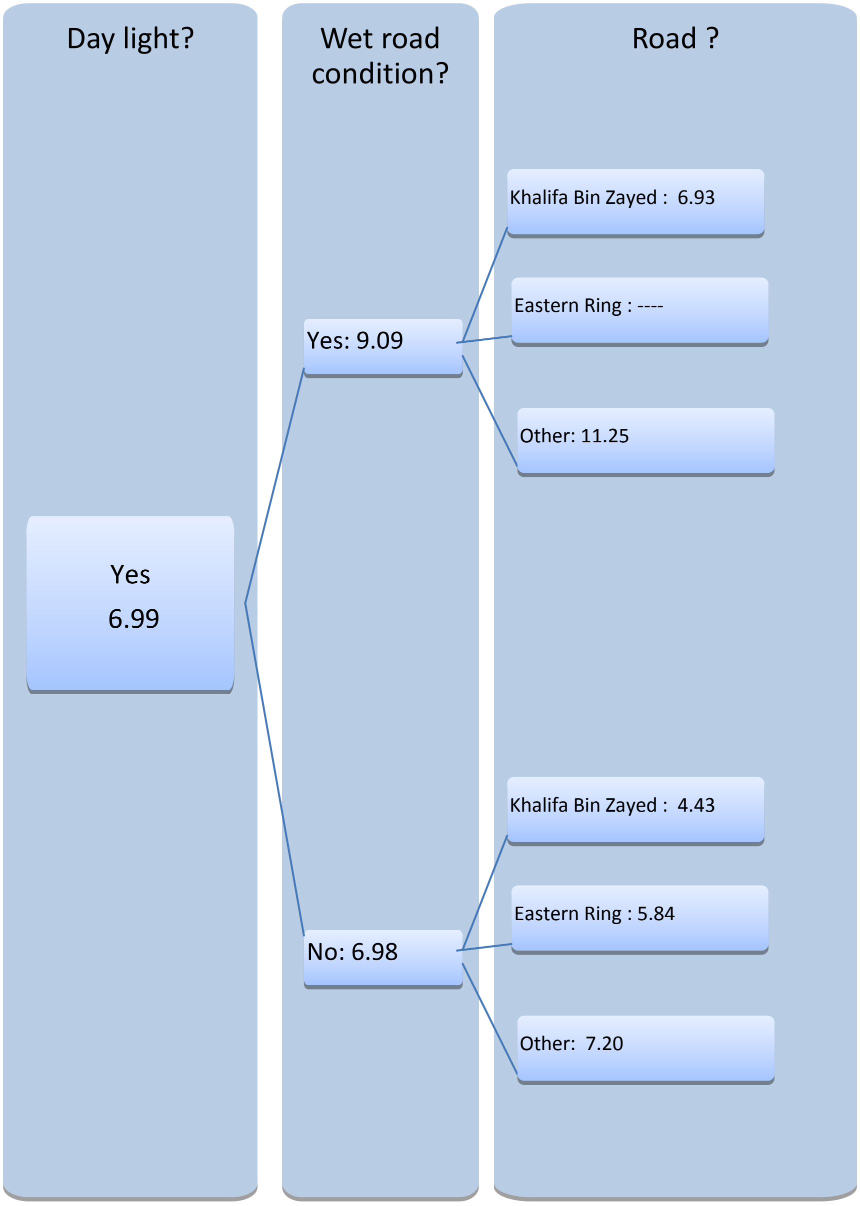


Figure 6-26 Response time analysis - Decision tree of day light accidents (Durations are in minutes)

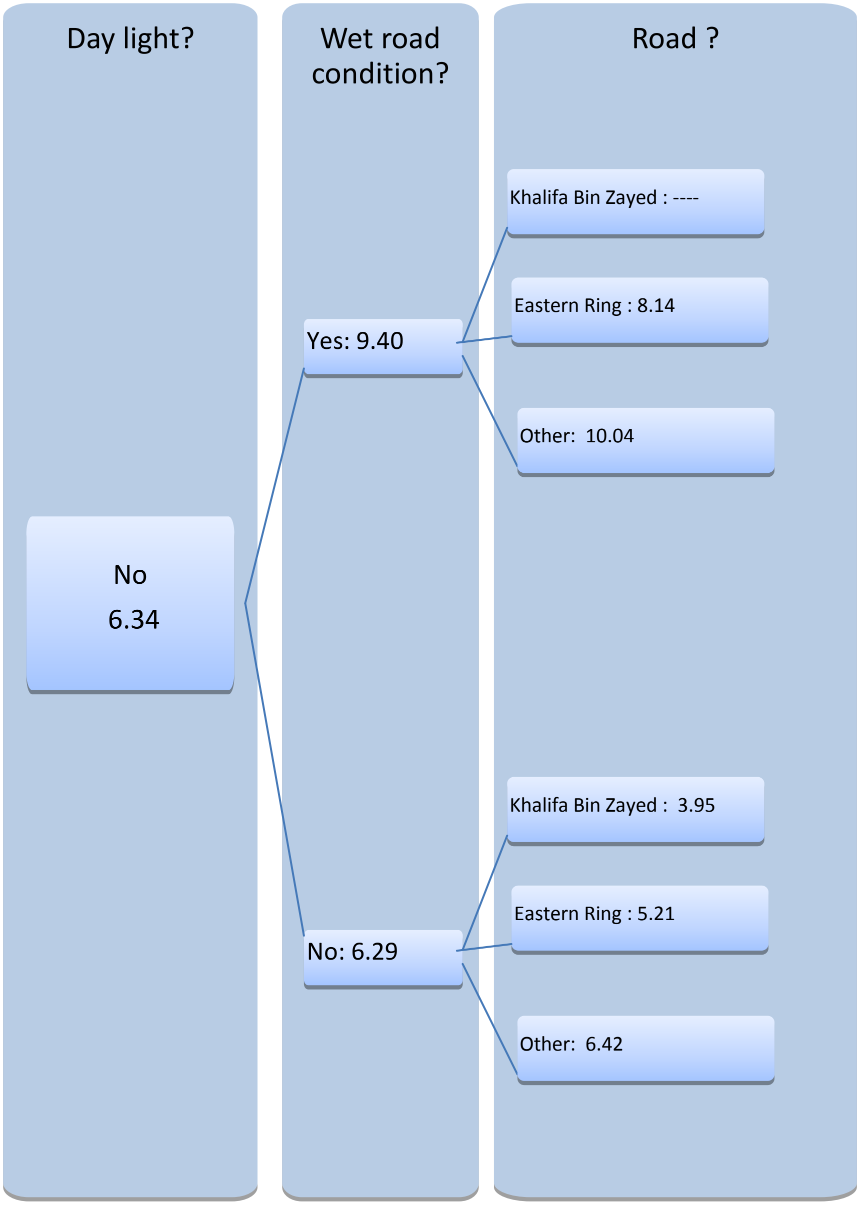


Figure 6-27 Response time analysis - Decision tree of night light accidents (Durations are in minutes)

Comparison to Previous Research

As mentioned with regard to the previous model, the comparison in this research is restricted to the previous researches that have approached modelling each interval time of the total accident duration. For the response time, two studies, by Lee and Fazio (2005), and Nam and Mannering (2000), were found.

Similarly to these previous researches, this study ascertained that some of the geographical characteristics variables, namely the accident location variable, were significantly affecting accident response time. Furthermore, some variables in the environmental characteristics, including road surface condition and light condition, were also affecting accident response time. However, in contrast to the previous researches, none of the temporal characteristics or accident characteristics were found to be significantly affecting accident response time. The resultant differences between this study and the previous research might be due to the use of different datasets and location which may yield different outcomes.

Additionally, the resultant best fit distribution for urban accident response time was found to be Log-logistic in this study, whereas it was found to be Weibull with gamma heterogeneity in Nam and Mannering (2000). However, it is not possible to compare the fitted distribution of this study with the study conducted by Lee and Fazio (2005), because Lee and Fazio applied the Cox Proportional Hazard Model, which does not require distribution assumption.

6.3.3 Clearance Time

The last interval time in urban accident duration is clearance time. In this research, clearance time has been defined as the time between the arrival of the first accident investigator at the scene and the departure time of the investigators from the scene. In other words, it refers to the time spent gathering all the available evidence from the accident scene.

Descriptive analysis shows that the mean of the clearance time was 26.26 minutes, with a range spanning from 1 minute to 130 minutes. The standard deviation was 17.83 minutes, variance was 318 minutes and Kurtosis was 8.27 minutes.

The density distribution with respect to clearance time in this study was found to be skewed (1.86) to the right due to the differences between the mean value and the median value (see Figure 6-28). A detailed analysis was conducted, considering three distributions (Weibull, Log-normal, and Log-logistic), as presented earlier (see sections 6.3.3.1, 6.3.3.2, 6.3.3.3).

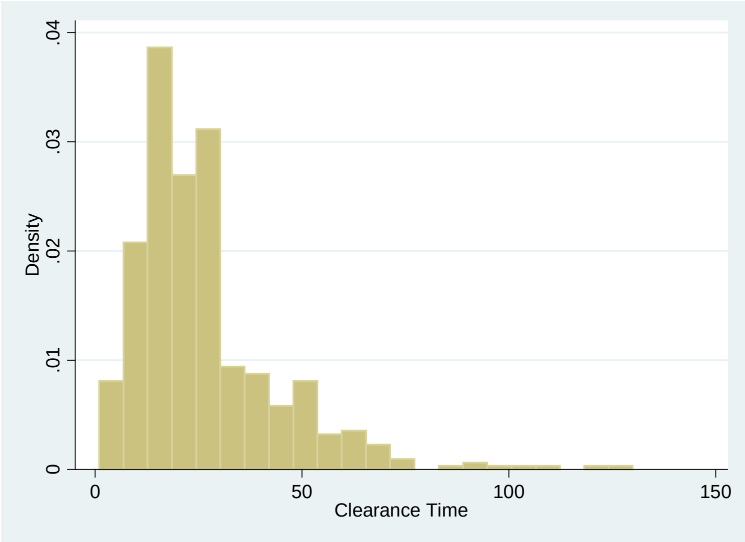


Figure 6-28 Clearance time analysis - Density distribution for accident clearance time

6.3.3.1 *Weibull Distribution*

Firstly, the clearance time data was tested with Weibull distribution. The first step was to develop a base (null) model. It was found that the log-likelihood before convergence is -551.50 and AIC is 1264.09 for the base model. The baseline hazard function with respect to Weibull distribution is presented in Figure 6-29.

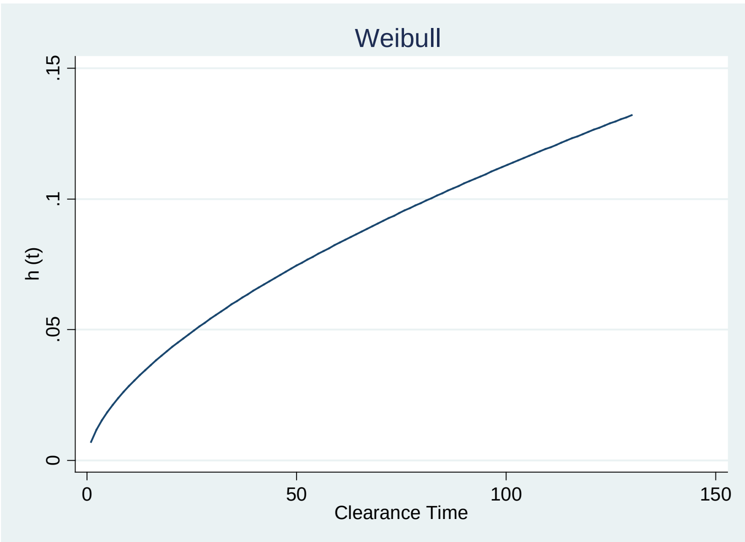


Figure 6-29 Clearance time analysis – The baseline hazard of Weibull distribution

The next step was to identify which variable, on its own, significantly decreased the statistic of -2log L. Among the explanatory variables in the dataset, 42 variables show that they are significant at 85% (see Table 6-25).

Table 6-25 Urban clearance time analysis - Weibull model: the list of significant variables

Variable	-2log *L*	*LL*	AIC	t-test
Slight injury	1255.04	-627.52	1261.04	2.14
Serious injury	1252.08	-626.04	1258.08	2.69
Fatal injury	1253.28	-626.64	1259.28	2.37
Number of casualties	1247.60	-623.80	1253.60	3.21
AM peak	1254.31	-627.15	1260.31	-2.54
Out of peak	1252.57	-626.28	1258.57	2.83
Evening	1257.68	-628.84	1263.68	1.55
Saturday	1257.32	-628.66	1263.32	1.62
Weekday	1256.45	-628.22	1262.45	-1.89
Weekend	1256.45	-628.22	1262.45	1.89
Hit pedestrian	1229.76	-614.88	1235.76	-5.83
Hit object	1250.73	-625.37	1256.73	2.86
Other types of accidents	1257.49	-628.75	1263.49	1.56
Al Mushrif	1257.08	-628.54	1263.08	1.68
Al Bateen	1255.56	-627.78	1261.56	2.03
Madinat Zayed	1255.45	-627.72	1261.45	-2.27
Al Khalidyah	1252.93	-626.46	1258.93	-2.90
Al Wahdah	1257.20	-628.60	1263.20	-1.83
Clear	1253.71	-626.85	1259.71	-2.24
Rain	1247.79	-623.90	1253.79	2.65
Number of vehicles involved	1238.47	-619.23	1244.47	4.44
Wet	1256.00	-628.00	1262.00	1.79
Did not leave enough space	1257.97	-628.98	1263.97	-1.50
Exceeding statutory speed limit	1257.03	-628.51	1263.03	1.62
Failing to comply: stop sign	1257.36	-628.68	1263.36	1.60
No road user consideration	1241.34	-620.67	1247.34	-4.68
Sheikh Zayed the First St	1250.76	-625.38	1256.76	-3.31
Meena St	1254.53	-627.26	1260.53	2.21
Hamdan Bin Mohammed St	1251.87	-625.93	1257.87	-3.18
31st St	1257.77	-628.88	1263.77	-1.86
Commercial area	1251.11	-625.55	1257.11	-3.03
School	1254.17	-627.08	1260.17	2.28
Junction	1256.45	-628.22	1262.45	1.91
Car park	1251.40	-625.70	1257.40	-3.29
July	1257.82	-628.91	1263.82	-1.56
August	1243.06	-621.53	1249.06	-4.60
October	1251.89	-625.94	1257.89	-3.05
January	1252.88	-626.44	1258.88	2.53
March	1238.31	-619.15	1244.31	4.37
23ed St	1257.41	-628.70	1263.41	-2.23
35th St	1250.48	-625.24	1256.48	2.08
29th St	1254.00	-627.00	1260.00	1.76

Table 6-26 Urban clearance time analysis - Weibull AFT model

Variable	Estimated Coefficient	t-statistics	Percentage change
Temporal Characteristics			
Out of peak	0.15	2.55	16.27
August	-0.31	-3.54	-26.73
January	0.24	2.56	28.21
March	0.40	4.94	49.49
Geographical Characteristics			
Meena St	0.29	2.58	33.66
Commercial area	-0.17	-3.50	-15.72
Car park	-0.33	-2.85	-28.68
Environmental Characteristics			
Clear	-0.56	-2.75	-43.19
Accident Characteristics			
Hit object	0.30	3.06	36.23
Number of casualties	0.03	2.10	3.93
Number of vehicles involved	0.13	4.90	14.23
Model Structure Parameters			
p (distribution shape parameter)	1.86	18.74	
λ (the scale parameter)	0.03	16.27	
Intercept term (_cons)	3.51	16.27	
Goodness-of-fit Test			
Akaike information criterion	1003.85		
Initial log-likelihood	-551.50		
Log-likelihood at convergence	-488.92		
Number of observations	525		

Subsequently, all of the significant variables in Table 6-25 were used in the analysis and the model was estimated. The result showed that some of these variables were insignificant in the model when integrated with other variables due to multicollinearity effects. Thus, only variables with 90% level of significance were used to estimate the final model. This step resulted in only 11 variables (Table 6-26).

6.3.3.2 Log-normal Distribution

Secondly, Log-normal distribution was tested. Modelling the base (null) model revealed that the log likelihood is -539.09 and AIC is 1244.30. The baseline hazard function with respect to log-normal distribution is presented in Figure 6-30.

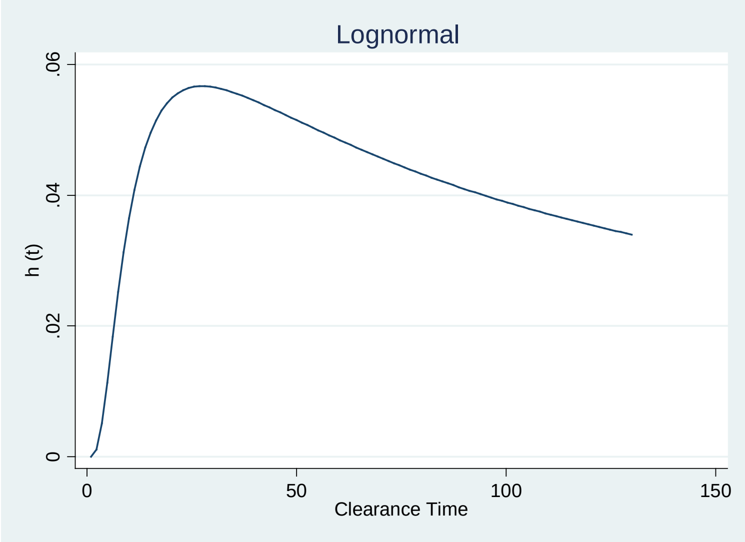


Figure 6-30 Clearance time analysis – The baseline hazard of Log-normal distribution

The second step of the variable selection procedure resulted in 42 significant variables at the level of 85% (see Table 6-27).

Table 6-27 Urban clearance time analysis - Log-normal model: the list of significant variables

Variable	-2log *L*	*LL*	AIC	t-test
Slight injury	1230.36	-615.18	1236.36	3.17
Serious injury	1234.77	-617.38	1240.77	2.36
Fatal injury	1233.11	-616.55	1239.11	2.69
Number of casualties	1226.84	-613.42	1232.84	3.69
AM peak	1235.02	-617.51	1241.02	-2.30
PM peak	1237.80	-618.90	1243.80	-1.58
Out of peak	1231.98	-615.99	1237.98	2.89
Morning	1237.02	-618.51	1243.02	-1.81
Evening	1235.82	-617.91	1241.82	2.12
Monday	1237.44	-618.72	1243.44	-1.69
Hit pedestrian	1220.15	-610.07	1226.15	-4.53
Hit object	1234.27	-617.13	1240.27	2.46
Other types of accidents	1237.27	-618.63	1243.27	1.74
Al Mushrif	1236.67	-618.33	1242.67	1.91
Al Madina Al Riyadiya	1237.44	-618.72	1243.44	-1.69
Al Bateen	1236.96	-618.48	1242.96	1.83
Madinat Zayed	1237.89	-618.94	1243.89	-1.55
Al Khalidyah	1234.67	-617.33	1240.67	-2.38
Hadbat Al Zafaranah	1237.83	-618.91	1243.83	-1.57
Rain	1233.24	-616.62	1239.24	2.66
Number of vehicles involved	1227.54	-613.77	1233.54	3.59
Wet	1237.02	-618.51	1243.02	1.81
Failing to comply: traffic light	1237.94	-618.97	1243.94	1.54
Exceeding statutory speed limit	1233.89	-616.94	1239.89	2.54
Failing to comply: stop sign	1237.24	-618.62	1243.24	1.75
No road user consideration	1229.63	-614.81	1235.63	-3.28
Sheikh Zayed the First St	1234.38	-617.19	1240.38	-2.44
Meena St	1233.16	-616.58	1239.16	2.68
Al Falah St	1236.63	-618.31	1242.63	-1.92
Hamdan Bin Mohammed St	1234.61	-617.30	1240.61	-2.39
Al Khaleej Al Arabi St	1236.52	-618.26	1242.52	1.95
Commercial area	1235.19	-617.59	1241.19	-2.26
Bridge	1238.05	-619.03	1244.05	1.50
Petrol station	1237.79	-618.90	1243.79	1.58
Car park	1234.22	-617.11	1240.22	-2.47
August	1236.11	-618.05	1242.11	-2.05
March	1232.58	-616.29	1238.58	2.79
April	1238.18	-619.09	1244.18	1.45
Mohammed Bin Khalifa St	1237.34	-618.67	1243.34	-1.72
23rd St	1237.85	-618.92	1243.85	-1.56
35th St	1234.71	-617.35	1240.71	2.37
29th St	1235.93	-617.96	1241.93	2.09

The final step of variables selection is to estimate the model with all the variables in Table 6-27 at once in order to find out which variables resulted in a 90% level of significance in the final model. It was found that only 12 variables were significant in the model. Table 6-28 shows the resulting Log-normal AFT model with the 12 significant variables.

Table 6-28 Urban clearance time analysis - Log-normal AFT model

Variable	Estimated Coefficient	t-statistics	Percentage change
Temporal Characteristics			
Evening	0.11	1.99	11.66
Monday	-0.19	-2.63	-17.50
March	0.28	3.04	33.08
Geographical Characteristics			
Meena St	0.34	2.74	41.53
Mohammed Bin Khalifa St	-0.79	-2.18	-54.65
Environmental Characteristics			
Weather Condition: Rain	0.94	2.59	156.42
Foggy	-2.13	-3.39	-88.15
Accident Characteristics			
Rollovers collision	0.33	2.17	40.32
Hit object	0.29	2.51	33.67
Other types of accidents	0.19	1.99	21.46
Number of casualties	0.04	2.24	4.42
Number of vehicles involved	0.13	4.23	14.90
Model Structure Parameters			
β (distribution shape parameter)	2.63	36.13	
σ (the scale parameter)	0.62	32.40	
Intercept term (_cons)	2.63	36.13	
Goodness-of-fit Test			
Akaike information criterion	1021.47		
Initial log-likelihood	-539.09		
Log-likelihood at convergence	-496.73		
Number of observations	525		

6.3.3.3 Log-logistic Distribution

The Log-logistic distribution was tested for the data collected in Abu Dhabi. Modelling the base (null) model, it was found that the log-likelihood and AIC for the base model was -530.48 and 1001.33 respectively. The baseline hazard function with respect to the Log-logistic null model is presented in Figure 6-31.

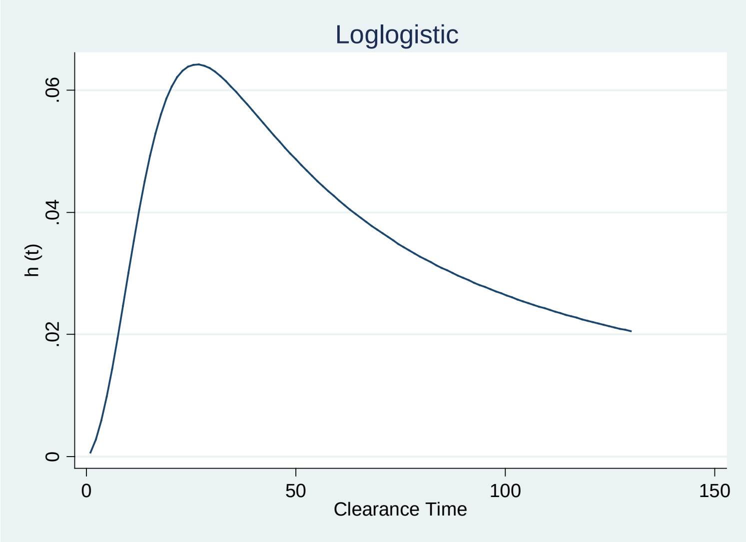


Figure 6-31 Clearance time analysis - The baseline hazard of log-logistic distribution

The analysis was conducted with one variable at a time in order to select the significant variables in the model. Among the explanatory variables in the dataset, it was found that there were 47 significant variables at the level of 85% (see Table 6-29).

Table 6-29 Urban clearance time analysis - Log-logistic model: the list of significant variables

Variable	-2log *L*	*LL*	AIC	t-test
Slight injury	1212.73	-606.36	1218.73	3.19
Serious injury	1216.20	-608.10	1222.20	2.61
Fatal injury	1215.02	-607.51	1221.02	2.83
Number of casualties	1208.51	-604.25	1214.51	3.73
AM peak	1220.10	-610.05	1226.10	-1.66
PM peak	1219.32	-609.66	1225.32	-1.89
Out of peak	1215.75	-607.87	1221.75	2.67
Morning	1219.24	-609.62	1225.24	-1.91
Evening	1217.89	-608.94	1223.89	2.23
Monday	1219.13	-609.56	1225.13	-1.94
Sideswipe collision	1220.35	-610.17	1226.35	1.59
Hit pedestrian	1200.89	-600.44	1206.89	-4.73
Hit object	1216.46	-608.23	1222.46	2.55
Other types of accidents	1220.06	-610.03	1226.06	1.68
Al Mushrif	1216.29	-608.14	1222.29	2.58
Al Madina Al Riyadiya	1220.32	-610.16	1226.32	-1.58
Al Bateen	1219.63	-609.81	1225.63	1.81
Madinat Zayed	1219.95	-609.97	1225.95	-1.71
Al Khalidyah	1218.46	-609.23	1224.46	-2.11
Rain	1216.11	-608.05	1222.11	2.88
Number of vehicles involved	1207.46	-603.73	1213.46	4.06
Wet	1219.96	-609.98	1225.96	1.70
Did not leave space	1220.63	-610.31	1226.63	-1.49
Failing to comply: traffic light	1219.28	-609.64	1225.28	1.90
Exceeding statutory speed limit	1211.74	-605.87	1217.74	3.54
Failing to comply: stop sign	1218.11	-609.05	1224.11	2.24
No road user consideration	1209.79	-604.89	1215.79	-3.63
Other cause	1220.20	-610.10	1226.20	-1.64
Sheikh Zayed the First St	1218.20	-609.10	1224.20	-2.16
Meena St	1215.82	-607.91	1221.82	2.68
Al Falah St	1219.45	-609.72	1225.45	-1.86
Hamdan Bin Mohammed St	1216.22	-608.11	1222.22	-2.60
Al Khaleej Al Arabi St	1219.61	-609.80	1225.61	1.81
Al Saada St	1220.38	-610.19	1226.38	1.61
Commercial area	1215.22	-607.61	1221.22	-2.78
Petrol station	1220.51	-610.25	1226.51	1.53
Car park	1216.95	-608.47	1222.95	-2.44
August	1216.90	-608.45	1222.90	-2.45
October	1215.95	-607.97	1221.95	-2.63
January	1219.42	-609.71	1225.42	1.88
February	1220.81	-610.40	1226.81	-1.44
March	1216.17	-608.08	1222.17	2.59
April	1220.73	-610.36	1226.73	1.47
23rd St	1219.65	-609.82	1225.65	-2.04
37th St	1220.55	-610.27	1226.55	1.67
35th St	1217.11	-608.55	1223.11	3.00
29th St	1218.07	-609.03	1224.07	2.65

Finally, all of the significant variables in Table 6-29 were estimated in the Log-logistic AFT model. The results show that there are only 12 variables that are significant at the level of 90%. Table 6-30 presents the results of the Log-logistic AFT model.

Table 6-30 Urban clearance time analysis - Log-logistical AFT model

Variable	**Estimated Coefficient**	**t-statistics**	**Percentage change**
Temporal Characteristics			
Morning	-0.14	-2.43	-13.29
Monday	-0.14	-2.12	-13.82
August	-0.20	-2.26	-18.50
March	0.22	2.35	24.84
Geographical Characteristics			
Meena St	0.39	3.32	48.95
Commercial area	-0.10	-2.02	-10.19
Environmental Characteristics			
Rain	0.94	2.63	157.37
Foggy	-2.20	-4.50	-88.97
Accident Characteristics			
Rollovers collision	0.35	2.42	42.50
Hit object	0.27	2.34	31.72
Number of casualties	0.03	1.94	3.60
Number of vehicles involved	0.14	4.91	15.88
Model Structure Parameters			
P (distribution shape parameter)	2.93	-29.44	
λ (the scale parameter)	0.05	39.44	
Intercept term (_cons)	2.82	39.44	
Goodness-of-fit Test			
Akaike information criterion	1001.33		
Initial log-likelihood	-530.48		
Log-likelihood at convergence	-486.66		
Number of observations	525		

6.3.3.4 Model Selection

The results of the plots of comparing the observed duration to the predicted duration for the three distributions were presented in Figure 6-32. This figure shows that the plots show a weak relationship between the observed and predicted duration in all distributions (see figure notes).

In terms of model fit, $R^2 = 0.15$ for Weibull AFT model, $R^2 = 0.15$ for Log-normal AFT model, and $R^2 = 0.17$ for Log-logistic AFT model. For MAPE, Weibull AFT model = 57.12%, Log-normal AFT model = 51.33%, and Log-logistic AFT model = 60.00%. For RMSE, Weibull AFT model = 16.40, Log-normal AFT model = 16.39, and Log-logistic AFT model = 16.23.

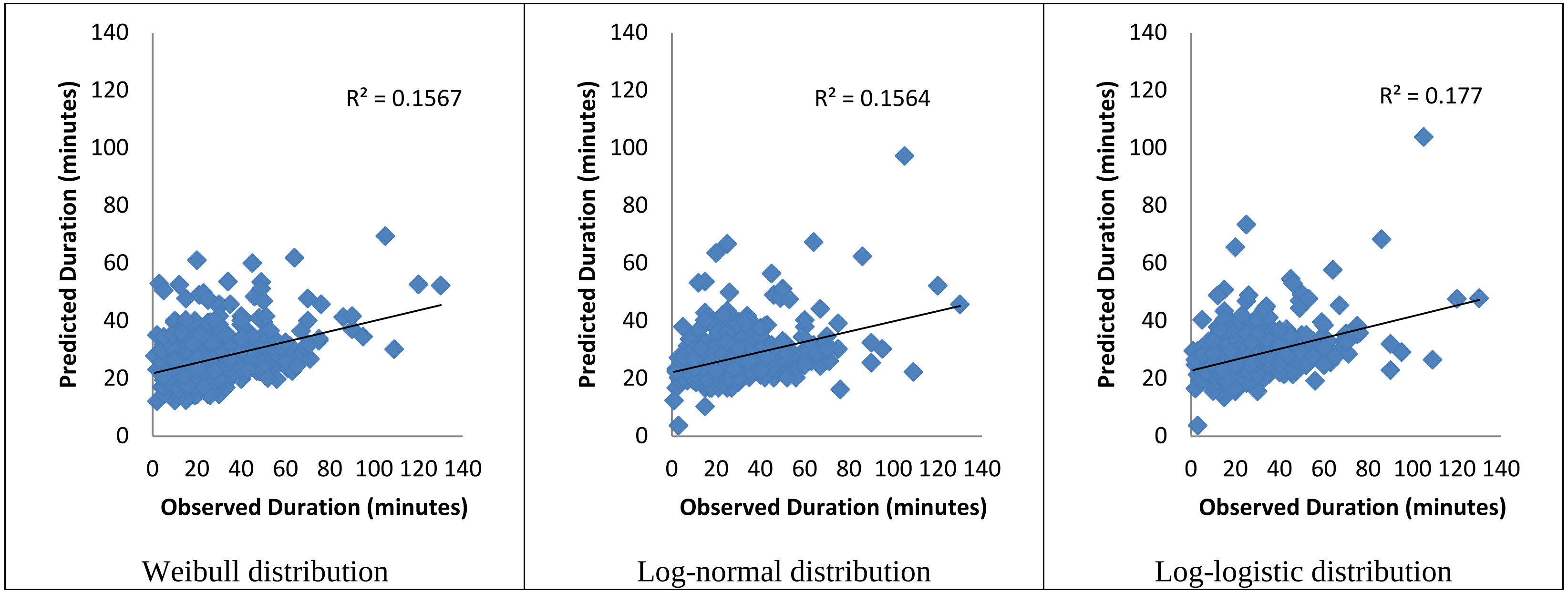


Figure 6-32 Clearance time analysis - Comparison between the observed and predicted duration for Weibull, Log-normal and Log-logistic distributions

According to the goodness-of-fit test (AIC), the results show that the Log-logistic AFT model is the best fit distribution for urban accident clearance time (Table 6-31). The following section demonstrates the results of this model and the effect of the significant variables on the clearance time.

Table 6-31 Urban clearance time analysis - Comparison of AIC values for AFT clearance time models

Model	-2In *L*	K	c	AIC
Weibull	977.84	11	2	1003.85
Log-normal	993.46	12	2	1021.47
Log-logistic	**973.32**	**12**	**2**	**1001.33**

6.3.3.5 Interpretation of the Estimated Model

The Log-logistic AFT model has 12 statistically significant explanatory variables. As shown in Table 6-30, it was found that the significant variables belong to all of the categories used in this research, including temporal characteristics, geographical characteristics, environmental characteristics and accident characteristics. This subsection presents the significant variables and their explanations.

Temporal Characteristics

The first variable in this category that has been found to significantly affect clearance time is time of day. Accidents occurring in the morning were shown to have (13.29%) lower clearance time compared to accidents that occurred during other day times (Figure 6-33). This result should be explained carefully. The morning period started from 12:01am until 11:59am. Normally the traffic flow will be less than other times of the day with the exception of the morning peak period. So, when the traffic flow is light, the investigators will be able to gather all the required information faster than during heavy traffic conditions. Furthermore, it was found that there is a common desire among collision investigators to clear an accident scene as soon as possible during the peak period to avoid more traffic congestion. Unfortunately, this is a weakness in TIM procedure that may affect collision investigators' work in gathering enough evidence from accident scenes which would help to subsequently explore the main causes of accidents and how they occurred. This finding is in agreement with previous research, which found that the attitude among accident management personnel in Washington State is to classify accidents occurring out of rush hours as less important accidents (Jones, Janssen and Mannering, 1991).

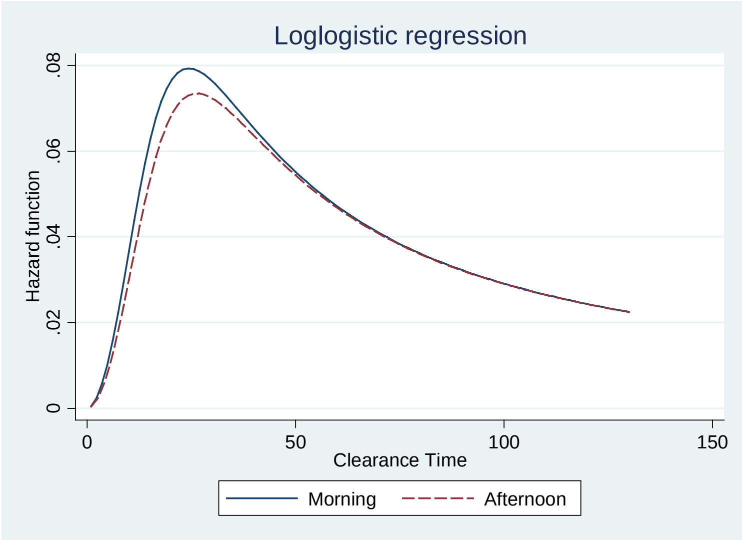


Figure 6-33 Clearance time analysis - Hazard function by time of day

Additionally, two months of the year were found to be significantly affecting clearance time. One of these months (August) was associated with (18.50%) shorter clearance time, whereas accidents that occurred in March were associated with (24.84%) longer clearance time (Figure 6-34). This is probably due to the occurrence of fatal injuries in all months of the year, with the exception of August. In general, when a fatal injury occurs, Crime Scene personnel will be involved in order to carry out forensic work. This will make collision investigators' work much longer and, as a result, the clearance time is expected to be longer.

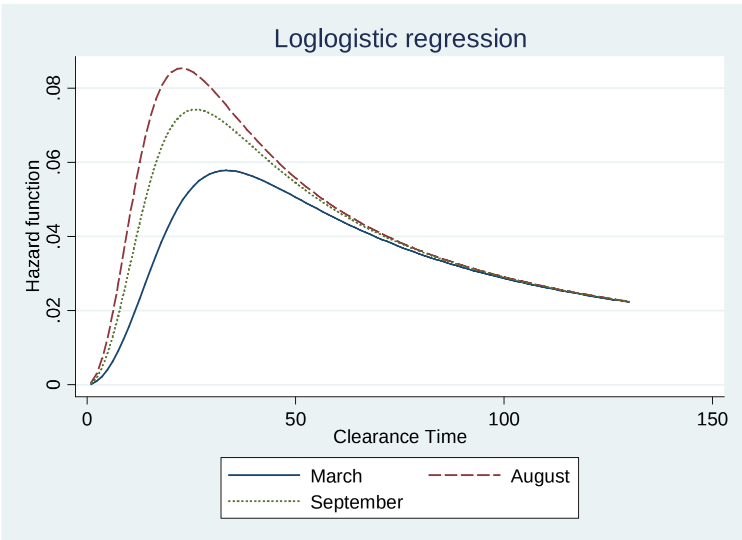


Figure 6-34 Clearance time analysis - Hazard function by month of year

Finally, accidents that happened on Mondays were associated with (13.82%) lower clearance time compared to accidents that happened on other days of the week (Figure 6-35).

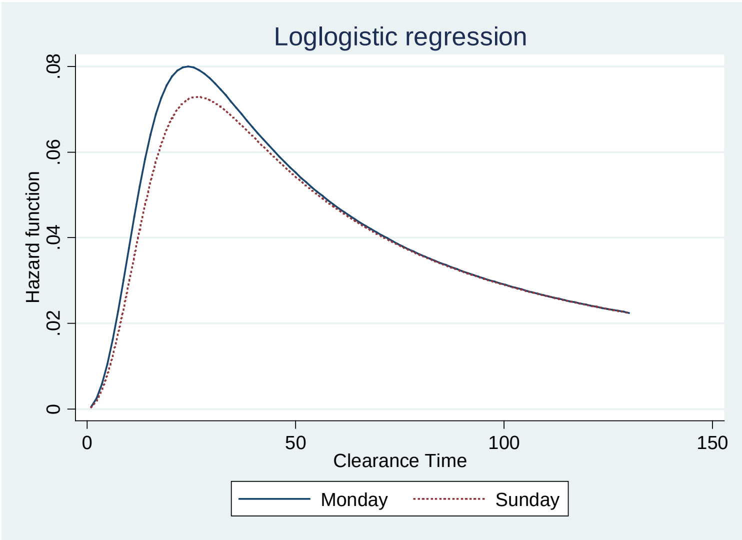


Figure 6-35 Clearance time analysis - Hazard function by day of week

Geographical Characteristics

Among the geographical variables, the "Meena Street" variable was found to significantly affect clearance time. Accidents occurring on "Meena Street" were associated with (48.95%) longer clearance time compared to accidents that happened on other roads of the city (Figure 6-36). These findings require careful interpretation due to the lower number of accidents occurring in this street and some other streets (23rd Street, 31st Street and 35th Street). However, after a thorough examination of the data of these accidents, it can be observed that there are high numbers of injuries and fatal injuries in streets associated with longer clearance time compared to other streets. This is logically accepted, because in practice clearing accidents that involve a higher number of injuries or fatal injuries requires more time. These findings are similar to those of previous studies (Chung, 2009; Jones, Janssen and Mannering; 1991; Nam and Mannering, 2000;).

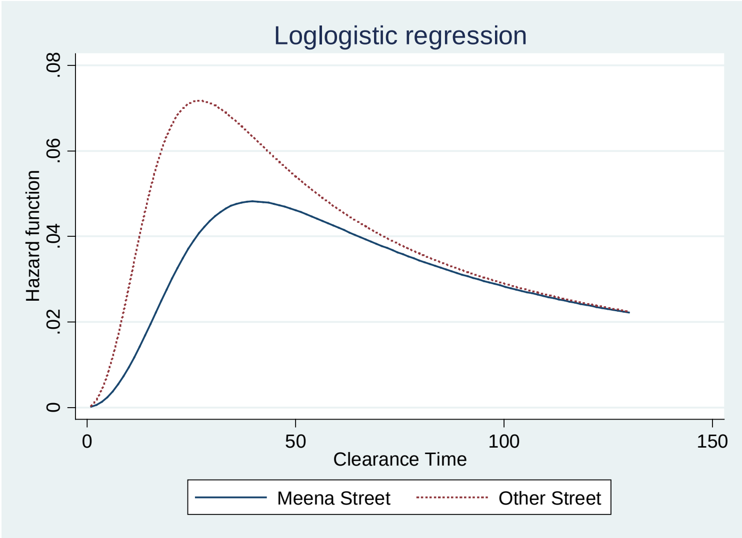


Figure 6-36 Clearance time analysis - Hazard function by street

Another location variable that was found to significantly affect clearance time is related to the nature of accident location. Accidents that occurred in commercial areas were associated with (10.19%) lower clearance time compared to accidents located in different areas (Figure 6-37). It is difficult to interpret these findings, but they could be related to the traffic conditions in these locations and the perception of the necessity to speed up the clearance time in congested areas.

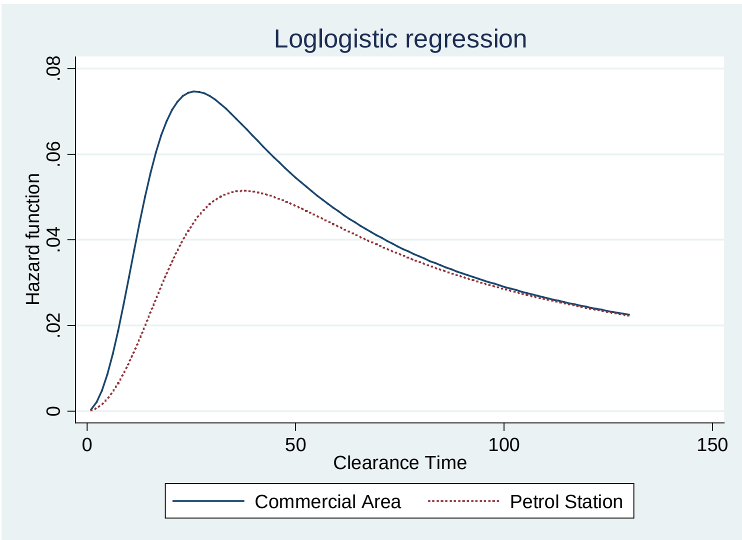


Figure 6-37 Clearance time analysis - Hazard function by location

Environmental Characteristics

Weather conditions were found to significantly affect the clearance time. Accidents that occurred in foggy weather conditions were associated with (88.97%) lower clearance time whereas accidents that occurred in rain conditions were associated with (157%) longer clearance time (Figure 6-38). As expected, rain may cause some delay in clearing accident scenes, but not clear conditions. Thus, the accident clearance process is faster during clear weather.

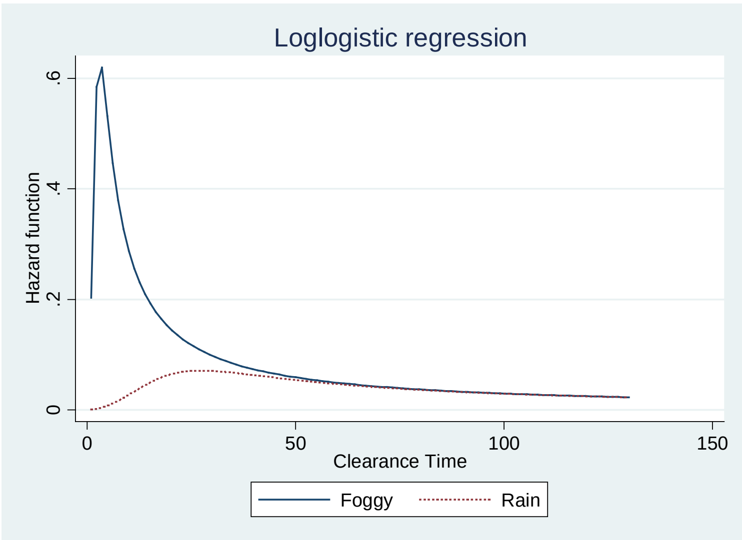


Figure 6-38 Clearance time analysis - Hazard function by weather condition

Accident Characteristics

As expected, longer accident clearance time was observed when there was an increase in the number of casualties and the number of vehicles involved (3.60%, 15.88% respectively). Furthermore, ‘hit object’ accidents and “rollovers accidents had (31.72%, 42.50%) longer clearance time compared to the clearance time of other accident types (Figure 6-39). Investigating the dataset showed that ‘hit object’ accidents resulted in 1 fatal injury only, a finding which was unexpected. Longer clearance time was expected to be associated with other accidents such as ‘hit pedestrian’ accidents, where 21 fatalities occurred. However, similar to the unexpected result of the out of peak variable, the reason was found to be related to a point of weakness in TIM procedures. The reason, which resulted from a discussion with a collision investigator in Abu Dhabi, was that collision investigation procedure varies among collision investigators: some investigators will spend a large amount of time in order to gather all the required information and witness statements from the accident scene, whereas other investigators collect the basic critical information in the accident report and leave other information and witness statements to be taken after clearing the scene. Thus, the time spent to write down the non-basic information and witnesses' statements by the latter investigators will not be considered as a part of the accident clearance time. This fact highlights the absence of clear guidelines on collision investigation and the need for a standard procedure that would be followed by all collision investigators.

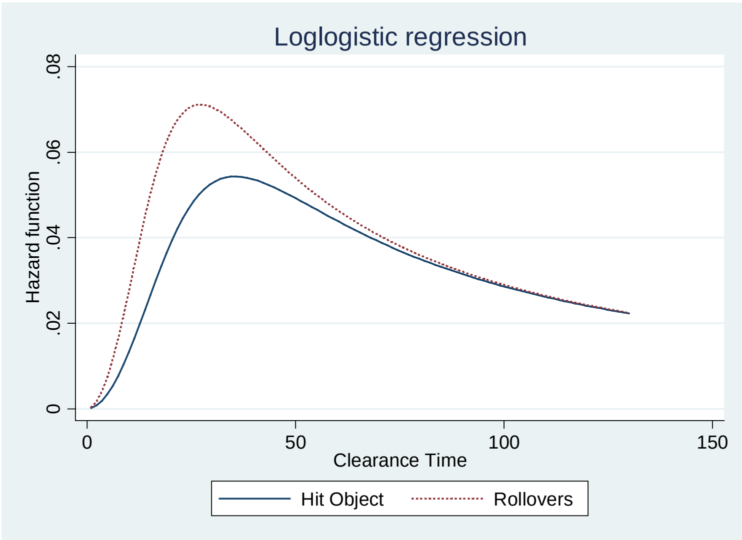


Figure 6-39 Clearance time analysis - Hazard function by accident type

Finally, the survival function and the hazard function are presented in Figure 6-40. Regarding the effects on response time, the log-logistic AFT model has P = 2.93. As mentioned earlier in section 3.3.3, if P > 1 this means having a non-monotonic slope and increasing from zero to a maximum at time $t = [(P - 1)^{1/P}] / \lambda$ and decreasing thereafter. In other words, there is higher chance of clearance of the accident before the maximum of the log-logistic distribution is reached after a duration of about 25 minutes. However, after this time, the further the time gone without clearing an accident, the less likely it is to have clearance soon.

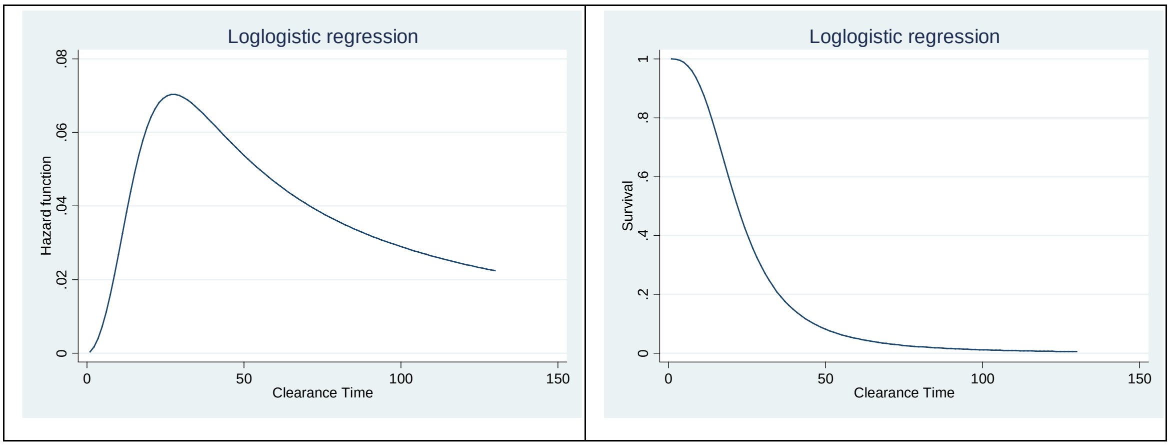


Figure 6-40 Clearance time analysis - Log-logistic hazard function and survival function

Duration Prediction and Model Accuracy

After fitting the log-logistic AFT model, the prediction of mean duration was performed. Then, the accuracy of this prediction was conducted by MAPE as illustrated in the Methodology chapter. For the urban clearance time, the MAPE was found to be 60%. According to Table 4-2, is less likely to generate an accurate predictive capability. Several reasons can cause that such as inaccurate measurement of durations, or/and incomplete and incomprehensive dataset. So, collecting accurate duration and comprehensive dataset are expected to improve the prediction accuracy in the future.

The decision tree for urban accident clearance time is large compared to reporting time and response time decision trees (Figure 6-41, Figure 6-42, Figure 6-43, Figure 6-44, Figure 6-45, Figure 6-46). This is mainly because of more variables being used to develop this tree compared to the previous decision trees. The variables are morning, Monday, August, March, Menna Street, commercial area, rain, foggy, rollovers accident, and hit object accident. The first split was made according to accident type, then morning, followed by Monday, then commercial area and, following that, Menna Street, month of year and finally weather condition. Similar to the previous trees, when more information about accident characteristics was available, a different prediction of clearance time is presented. Also, it should be mentioned that each accident type has two decision trees because of the high number of the significant variables.

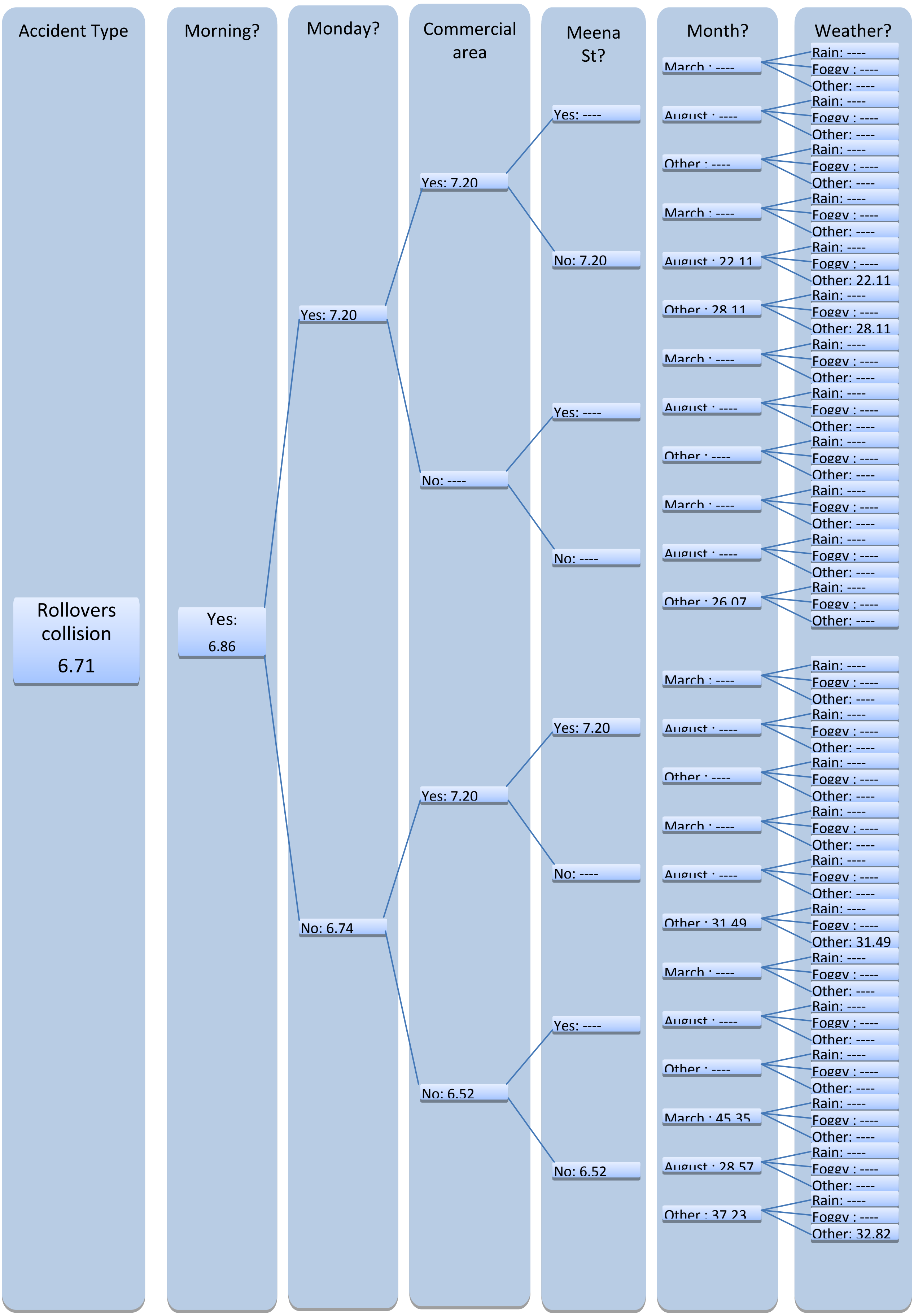


Figure 6-41 Clearance time analysis - Decision tree of Rollovers accidents that happened in the morning (Durations are in minutes)

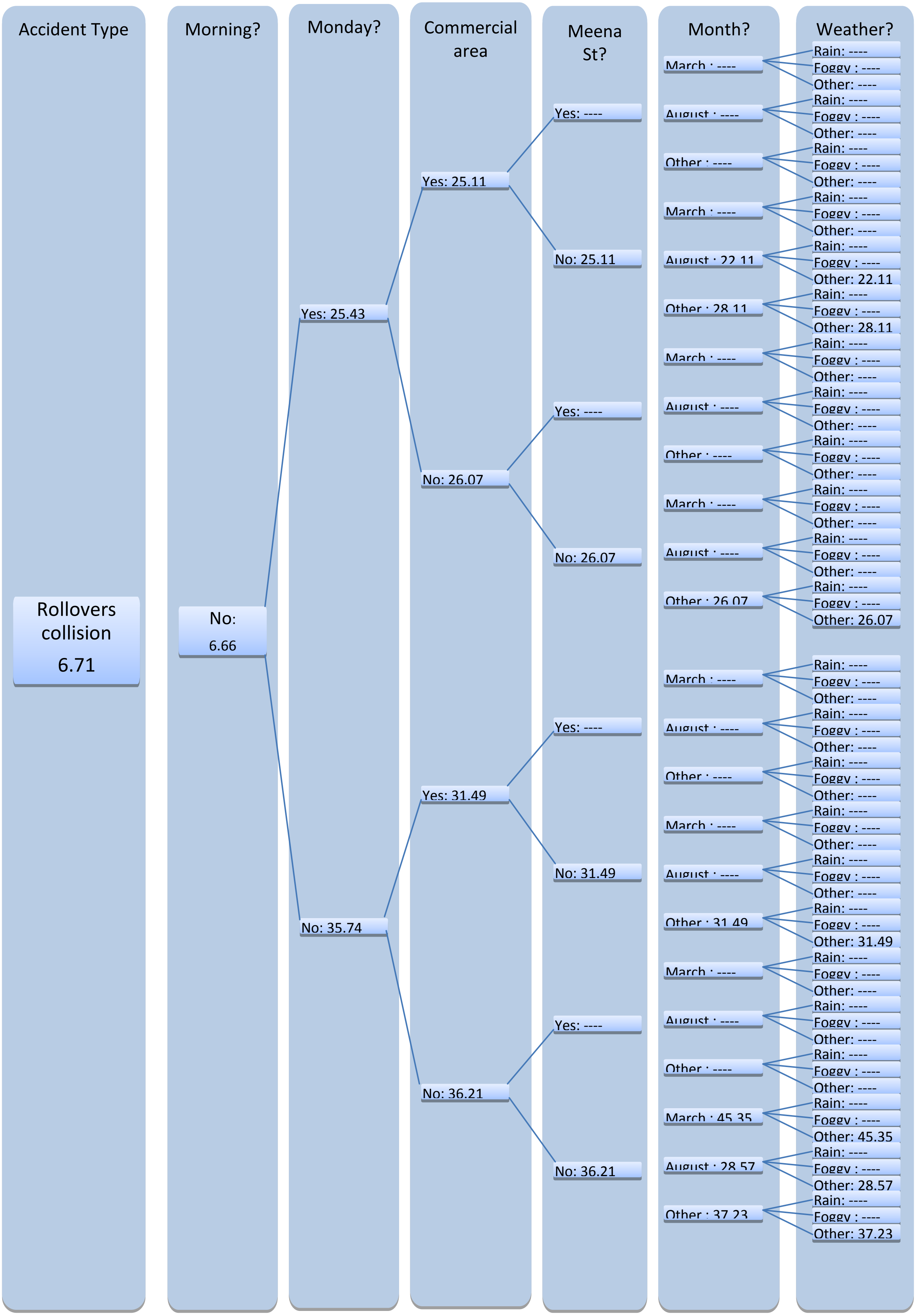


Figure 6-42 Clearance time analysis - Decision tree of Rollovers accidents that did not happen in the morning (Durations are in minutes)

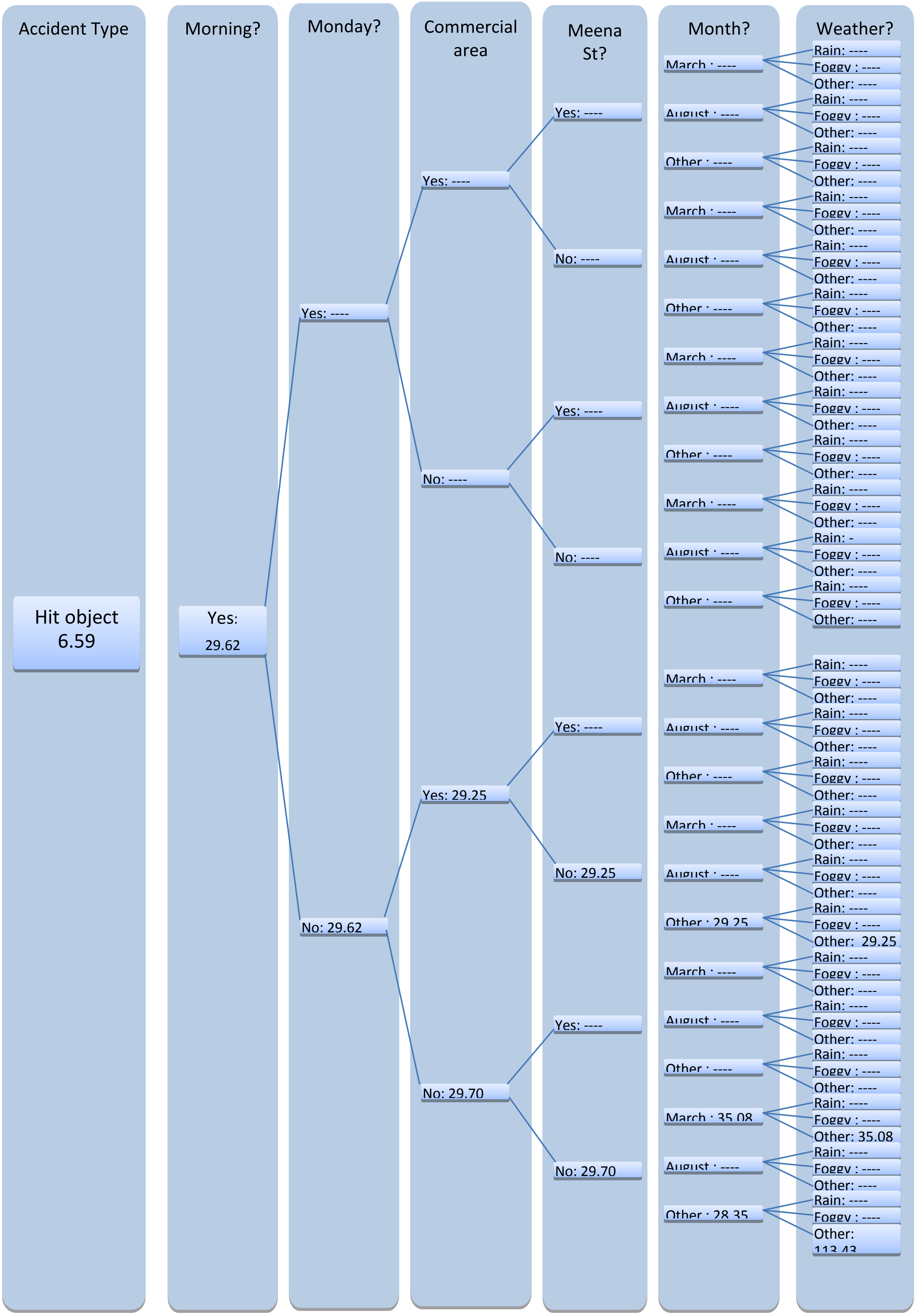


Figure 6-43 Clearance time analysis - Decision tree of Hit object accidents that happened in the morning (Durations are in minutes)

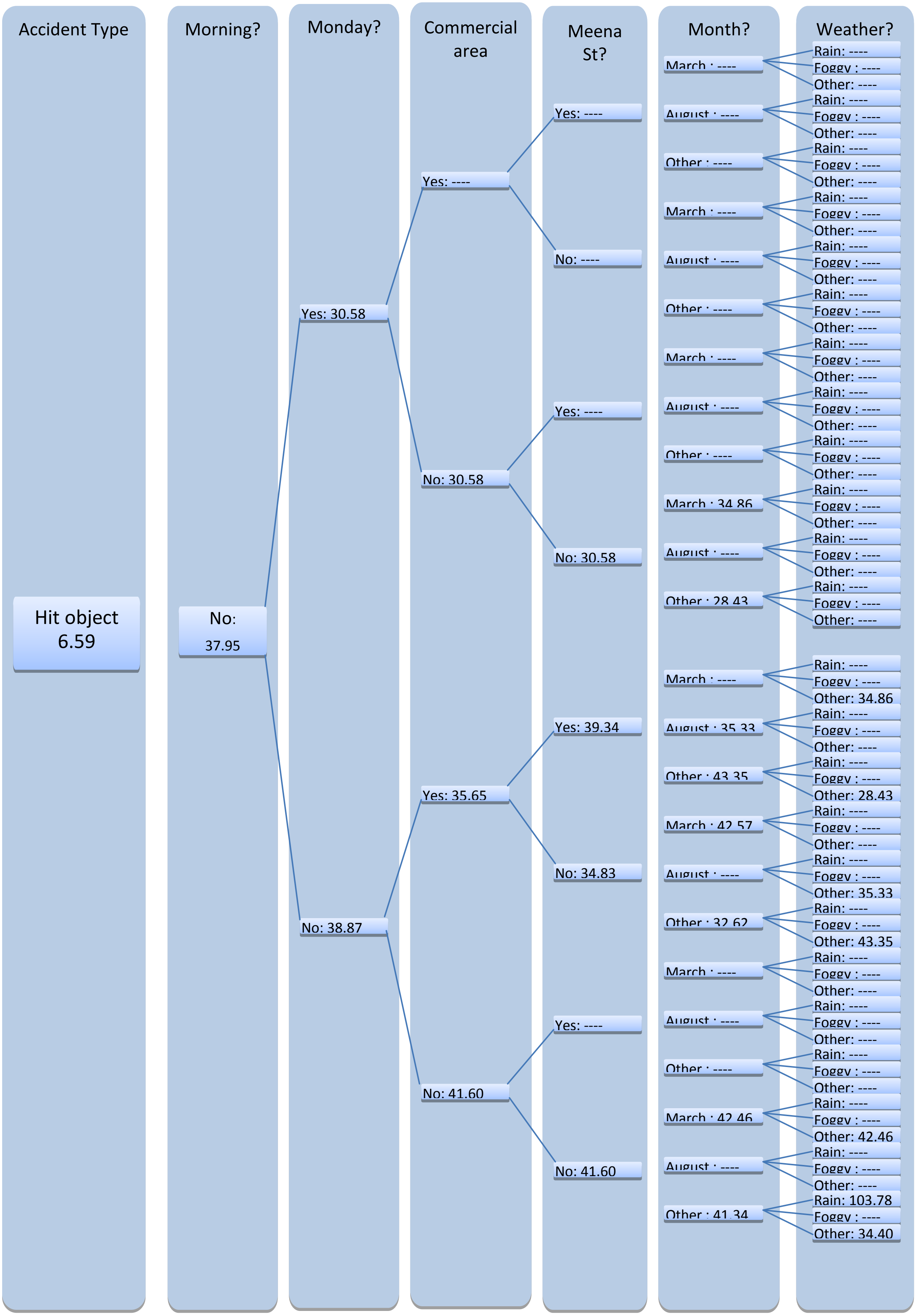


Figure 6-44 Clearance time analysis - Decision tree of Hit object accidents that did not happen in the morning (Durations are in minutes)

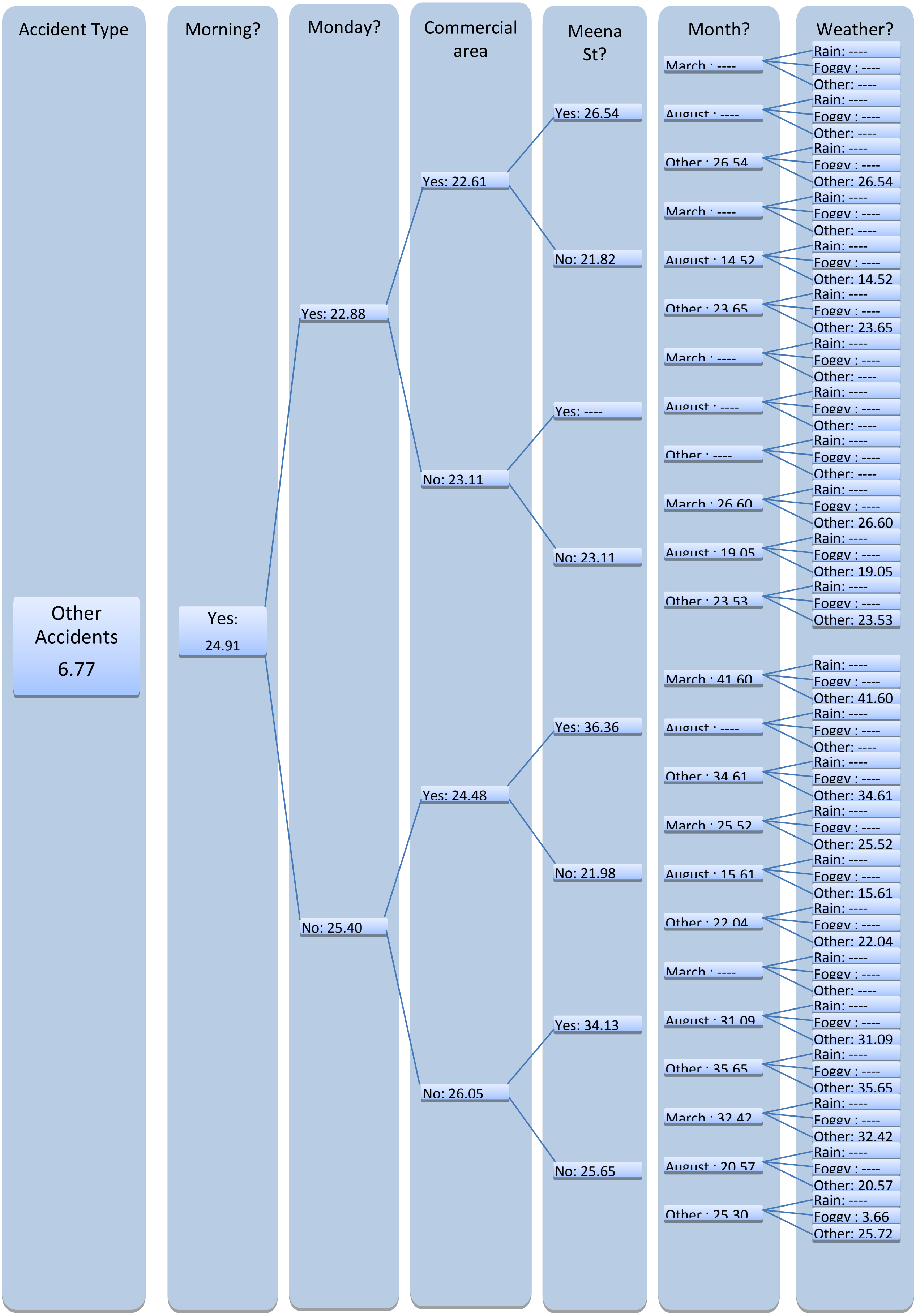


Figure 6-45 Clearance time analysis - Decision tree of other accidents that happened in the morning (Durations are in minutes)

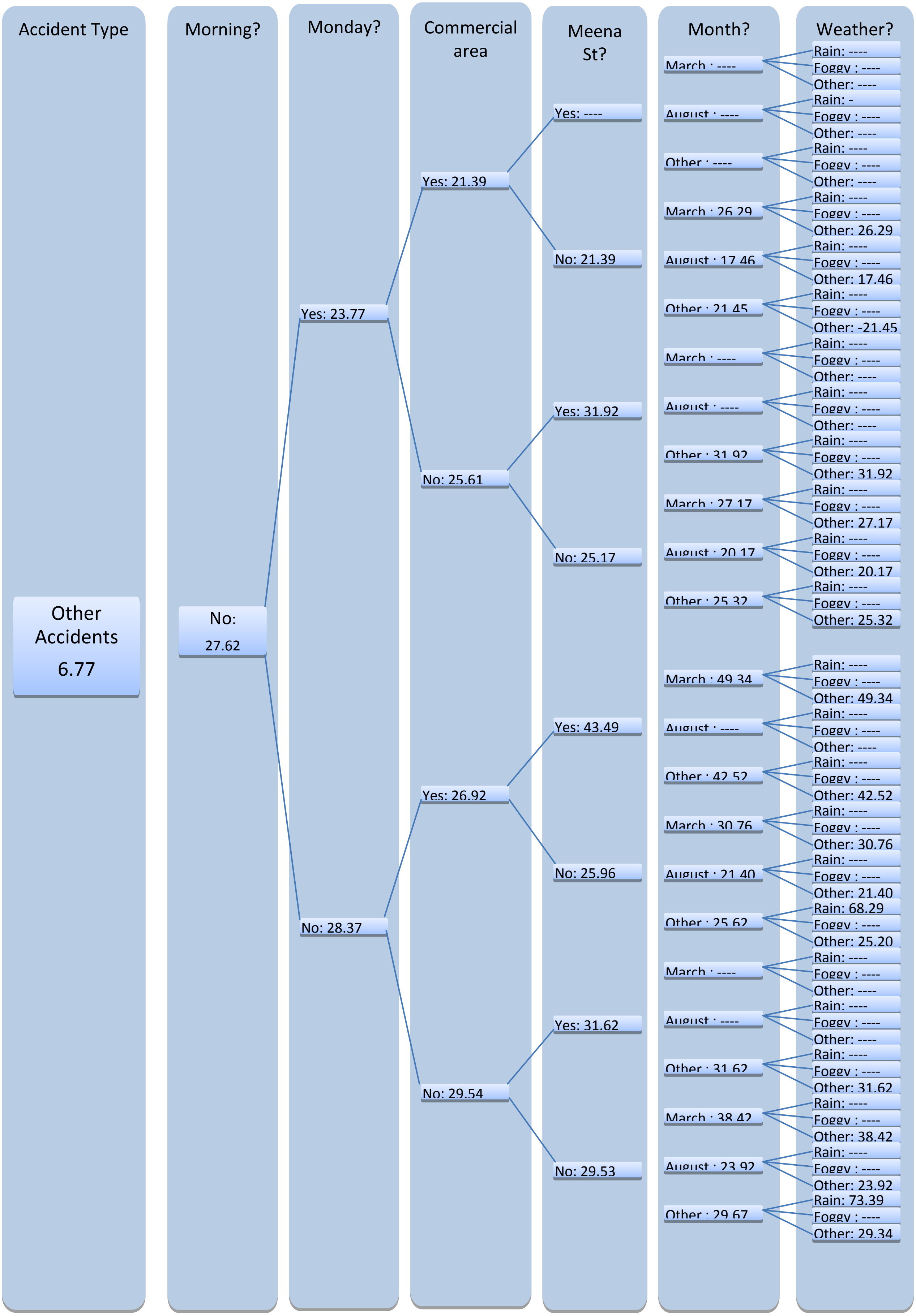


Figure 6-46 Clearance time analysis - Decision tree of other accidents that did not happen in the morning (Durations are in minutes)

Comparison to Previous Research

As mentioned with regard to the previous model, the comparison in this research is restricted to the previous researches that approached modelling each interval time of the total accident duration. For response time, two relevant studies were found: Lee and Fazio (2005) and Nam and Mannering (2000).

This research shows some similarities in terms of the best fit distribution and the resultant significant explanatory variables. Previous research by Nam and Mannering (2000) found that the significant variables belong to four categories including temporal characteristics, geographical characteristics, environmental characteristics, and accident characteristics. However, they were not significant in the study of Lee and Fazio (2005). Furthermore, some of the environmental variables in this research (e.g. fog, rain weather conditions) were found to be significant in this study and significant in the previous research by Nam and Mannering (2000). Also, the day of week variable was statistically significant in previous research, and found to be significant in this study.

In addition, the resultant best fit distribution for the clearance time shows the same result in this research. For example, the Log-logistic distribution without heterogeneity provided the best fit distribution in Nam and Mannering's study (2000). Also, it is not possible to compare the fitted distribution of this study with Lee and Fazio's study (2005) because Lee and Fazio applied the Cox Proportional Hazard Model which does not require distribution assumption. Finally, having stated these differences, it is clear that different datasets and case study areas may yield different results.

6.4 Highway Accidents Duration

In contrast to the urban accident duration, only one interval time of the total highway accident duration was collected, this being the response time. This section presents the findings of modelling response time using Weibull distribution, Log-normal distribution, and Log-logistic distribution.

6.4.1 Response Time

Similar to the urban accident duration, highway accident response time has been defined as the time between the Abu Dhabi Highway Collision Investigation Branch (AHCIB) being informed about the accident and the arrival of the first accident investigator at the scene. In other words, it refers to the collision investigators' travel time to the accident scene.

Descriptive analysis shows that the mean of the response time was 14.22 minutes with a range spanning from 2 minutes to 75 minutes. The standard deviation was 7.26 minutes, variance was 52.72 minutes and Kurtosis was 15.63 minutes. Furthermore, the density distributions, with respect to response time in this study, were found to be skewed (2.42) to the right, due to the differences between the mean value and the median value (see Figure 6-47).

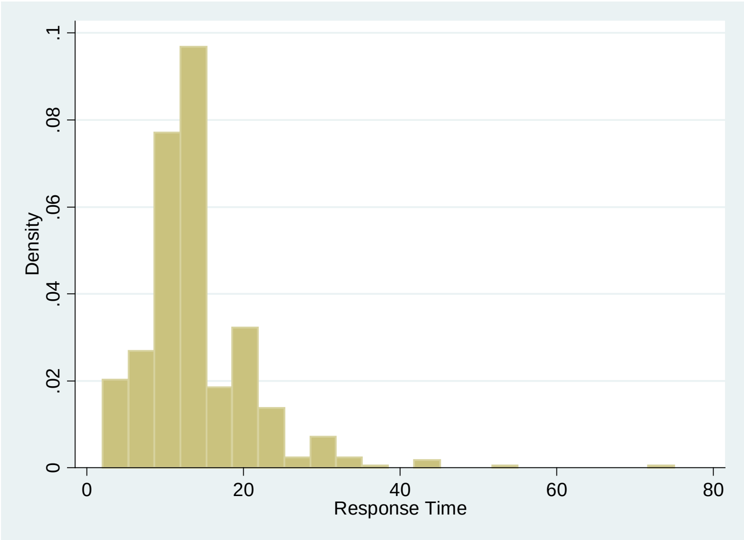


Figure 6-47 Highway response time analysis - Density distribution for accident response time

6.4.1.1 *Weibull Distribution*

Firstly, the reporting time data was tested with Weibull distribution. Selecting explanatory variables for the response time model was carried out in three steps, as explained in the Methodology chapter. The first step was to develop a base (null) model. It was found that the log-likelihood before convergence is 381.52 and AIC is 767.056 for the base model. The baseline hazard function, with respect to Weibull distribution, is presented in Figure 6-48.

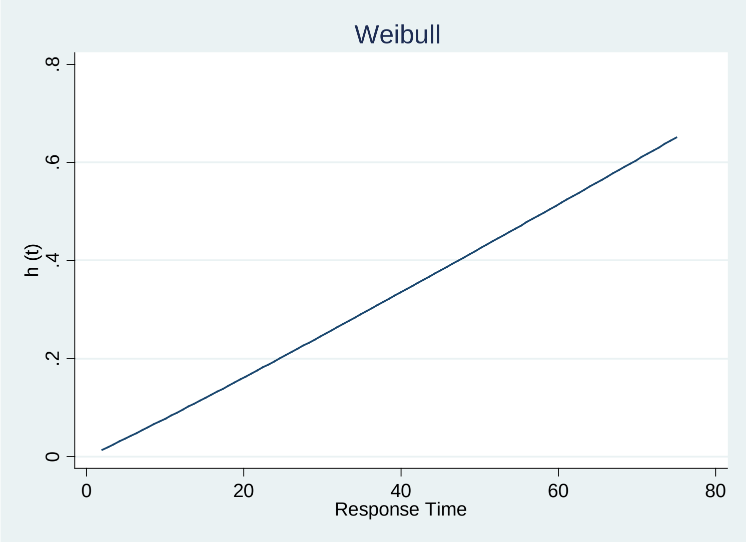


Figure 6-48 Highway response time analysis - The baseline hazard of Weibull distribution

The next step was to identify which variable, on its own, significantly decreases the statistic of -2log L. Among the explanatory variables in the dataset, 23 variables show that they are significant at 85% (see Table 6-32).

Table 6-32 Highway response time analysis - Weibull model: the list of significant variables

Variable	**-2log *L***	***LL***	**AIC**	**t-test**
Friday	760.09	-380.05	766.09	1.68
Monday	757.14	-378.57	763.14	-2.55
Tuesday	757.06	-378.53	763.06	-2.57
Wednesday	754.69	-377.34	760.69	2.82
Hit pedestrian	754.50	-377.25	760.50	-3.03
Hit object	760.44	-380.22	766.44	-1.66
Overturn	757.75	-378.88	763.75	2.23
Rear end	759.30	-379.65	765.30	1.89
Daylight	760.58	-380.29	766.58	-1.57
Darkness	760.58	-380.29	766.58	1.57
Serious	755.94	-377.97	761.94	-2.66
Fatal	741.59	-370.79	747.59	4.48
Musafah HW	746.31	-373.15	752.31	-4.46
International Airport HW	757.38	-378.69	763.38	-2.50
AbuDhabi-Tarif HW	732.92	-366.46	738.92	5.41
Almafraq-Sawameq HW	752.52	-376.26	758.52	-3.51
Swihan HW	756.43	-378.21	762.43	2.36
February	738.70	-369.35	744.70	-5.60
April	756.91	-378.45	762.91	-2.62
May	757.15	-378.58	763.15	-2.57
July	759.61	-379.80	765.61	1.78
September	754.35	-377.17	760.35	2.77
November	759.15	-379.57	765.15	1.91

Then, all of the significant variables in Table 6-32 were used in the analysis and the model was estimated. The results showed that some of these variables were insignificant in the model when integrated with other variables due to multicollinearity effects. Thus, only variables with 90% level of significance were used to estimate the final model. This step resulted in only 18 variables (Table 6-33).

Table 6-33 Highway response time analysis - Weibull AFT model

Variable	**Estimated Coefficient**	**t-statistics**	**Percentage change**
Temporal Characteristics			
Wednesday	0.12	2.42	13.01
Tuesday	-0.11	-2.21	-11.04
Morning	-0.06	-1.97	-6.68
Monday	-0.10	-1.91	-9.68
February	-0.43	-6.99	-35.38
April	-0.22	-3.74	-20.49
May	-0.22	-3.64	-20.04
July	0.27	3.99	31.80
September	0.23	3.39	26.48
Geographical Characteristics			
Musafah HW	-0.23	-4.16	-20.97
AbuDhabi-Tarif HW	0.25	5.54	28.66
Almafraq-Sawameq HW	-0.19	-3.09	-17.70
Swihan HW	0.34	3.76	40.90
Environmental Characteristics			
Sandy	0.18	2.20	20.87
Accident Characteristics			
Hit pedestrian	-0.20	-4.63	-18.23
Hit object	-0.11	-2.16	-10.51
Fatal injury	0.17	3.45	19.20
Other accidents	-0.15	-2.40	-14.70
Model Structure Parameters			
p (distribution shape parameter)	2.62	30.23	
λ (the scale parameter)	0.05	68.52	
Intercept term (_cons)	2.85	68.52	
Goodness-of-fit Test			
Akaike information criterion	606.75		
Initial log-likelihood	-381.52		
Log-likelihood at convergence	-283.37		
Number of observations	504		

6.4.1.2 Log-normal Distribution

Secondly, Log-normal distribution was tested. Modelling the base (null) model revealed that the log-likelihood was -335.73 and AIC was 675.46. The baseline hazard function with respect to log-normal distribution is presented in Figure 6-49.

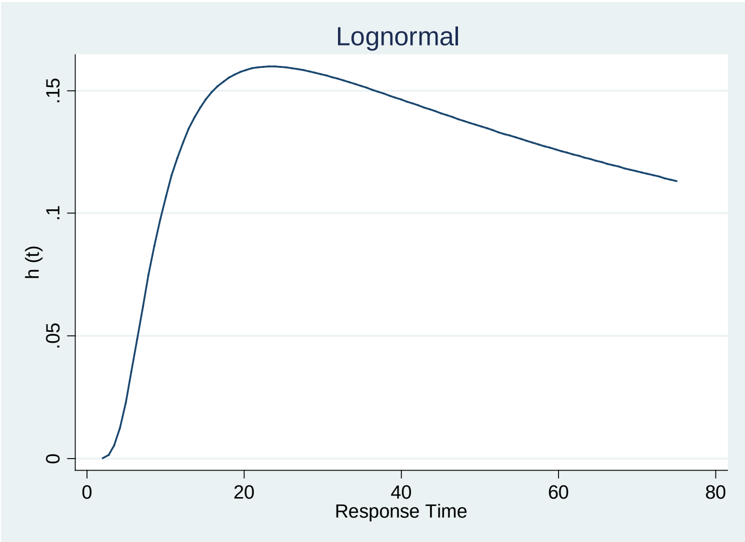


Figure 6-49 Highway response time analysis - The baseline hazard of Log-normal distribution

The second step of the variable selection procedure resulted in 18 significant variables at the level of 85% (see Table 6-34).

The final step of variable selection was to estimate the model with all the variables in Table 6-34 at once in order to find out which variables resulted in 90% level of significance in the final model. It was found that only 11 variables were significant in the model. Table 6-35 shows the resulting Log-normal AFT model with the 8 significant variables.

Table 6-34 Highway response time analysis - Log-normal model: the list of significant variables

Variable	-2log *L*	*LL*	AIC	t-test
Friday	668.13	-334.06	674.13	1.83
Monday	666.34	-333.17	672.34	-2.27
Wednesday	669.24	-334.62	675.24	1.49
Hit pedestrian	664.69	-332.34	670.69	-2.61
Overturn	662.16	-331.08	668.16	3.06
Dry	669.19	-334.60	675.19	-1.51
Sandy	668.82	-334.41	674.82	1.63
Musafah HW	658.83	-329.41	664.83	-3.58
UmmAlnar HW	666.22	-333.11	672.22	2.29
AbuDhabi-Tarif HW	661.93	-330.96	667.93	3.10
Almafraq-Sawameq HW	653.74	-326.87	659.74	-4.25
Swihan HW	658.29	-329.14	664.29	3.65
February	638.79	-319.39	644.79	-5.81
March	669.02	-334.51	675.02	1.56
April	663.61	-331.80	669.61	-2.81
May	666.99	-333.49	672.99	-2.12
August	667.29	-333.64	673.29	2.05
September	665.52	-332.76	671.52	2.44

Table 6-35 Highway response time analysis - Log-normal AFT model

Variable	Estimated Coefficient	t-statistics	Percentage change
Temporal Characteristics			
February	-0.47	-7.39	-37.87
April	-0.29	-4.63	-25.66
May	-0.24	-3.90	-22.06
Friday	0.11	2.16	12.51
Wednesday	0.11	2.22	12.20
Geographical Characteristics			
Musafah HW	-0.28	-4.82	-24.67
Almafraq –Swamiq HW	-0.33	-5.16	-28.64
Swihan HW	0.31	3.27	36.62
AbuDhabi-ALAin HW	-0.15	-3.21	-14.19
Accident Characteristics			
Hit pedestrian	-0.07	-1.65	-7.20
Overturn	0.16	3.17	17.91
Model Structure Parameters			
β (distribution shape parameter)	2.68	82.42	
σ (the scale parameter)	0.40	31.74	
Intercept term (_cons)	2.68	82.42	
Goodness-of-fit test			
Akaike information criterion	564.03		
Initial log-likelihood	-335.73		
Log-likelihood at convergence	-264.92		
Number of observations	504		

6.4.1.3 Log-logistic Distribution

The Log-logistic distribution was tested for the data collected in Abu Dhabi. Modelling the base (null) model, it was found that the log-likelihood and AIC for the base model was -329.83 and 663.67 respectively. The baseline hazard function with respect to the Log-logistic null model is presented in Figure 6-50.

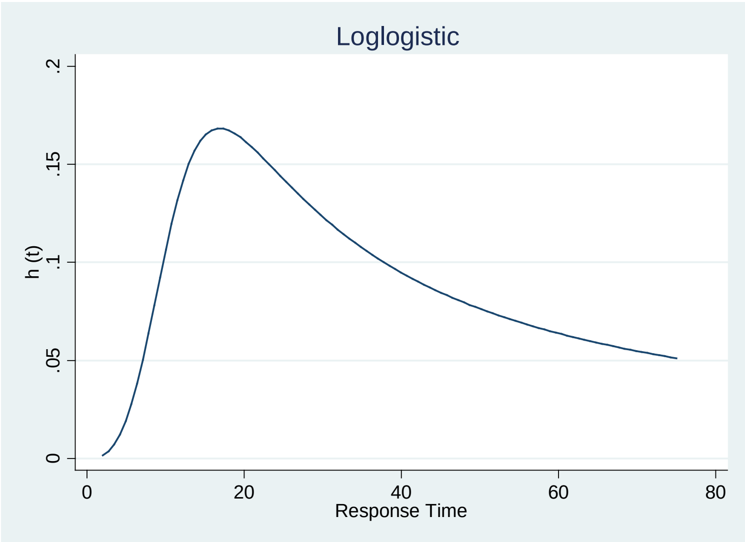


Figure 6-50 Highway response time analysis - The baseline hazard of log-logistic distribution

The analysis was conducted as one variable at a time in order to select the significant variables in the model. Among the explanatory variables in the dataset, it was found that there were 17 significant variables at the level of 85% (see Table 6-36).

Finally, all of the significant variables in Table 6-36 were estimated in the Log-logistic AFT model. The results revealed that there were only 8 variables that were significant at the level of 90%. Table 6-37 presents the results of Log-logistic AFT model.

Table 6-36 Highway response time analysis - Log-logistic model: the list of significant variables

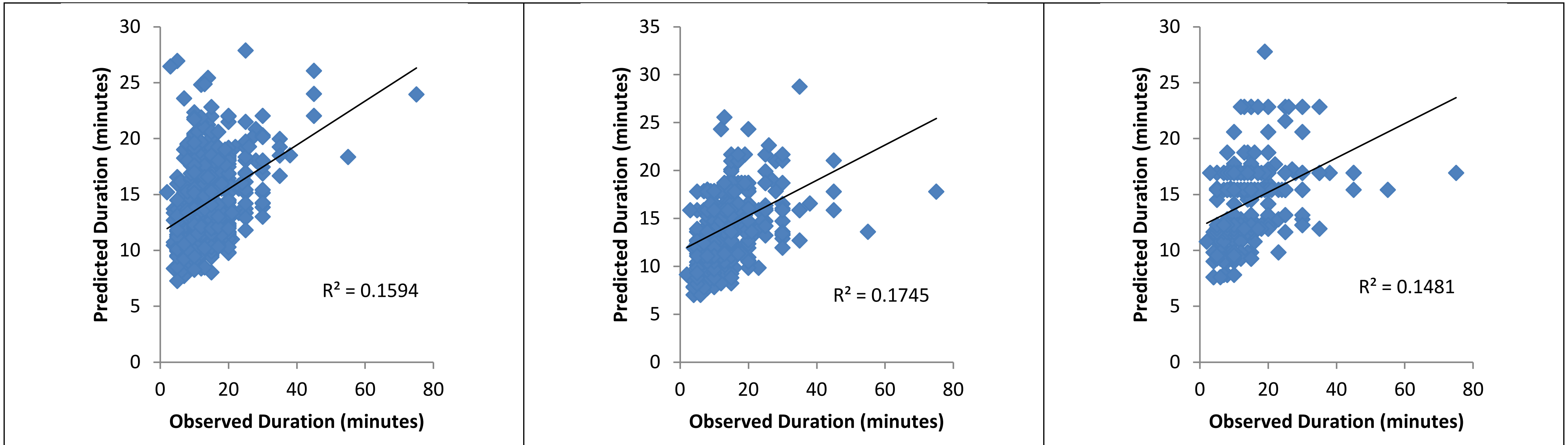
Variable	**-2log *L***	***LL***	**AIC**	**t-test**
Monday	656.35	-328.17	662.35	-1.82
Hit pedestrian	652.42	-326.21	658.42	-2.70
Overturn	650.16	-325.08	656.16	3.10
Dry	656.53	-328.26	662.53	-1.79
Sandy	654.71	-327.35	660.71	2.28
Musafah HW	641.96	-320.98	647.96	-4.28
UmmAlnar HW	654.58	-327.29	660.58	2.26
AbuDhabi-Tarif HW	650.95	-325.47	656.95	2.96
Almafraq-Sawameq HW	645.36	-322.68	651.36	-3.76
Swihan HW	645.26	-322.63	651.26	3.86
February	628.96	-314.48	634.96	-5.65
March	655.78	-327.89	661.78	1.98
April	650.01	-325.00	656.01	-3.15
May	654.38	-327.19	660.38	-2.31
June	657.11	-328.55	663.11	1.61
August	654.91	-327.45	660.91	2.19
September	656.72	-328.36	662.72	1.71

Table 6-37 Highway response time analysis - Log-logistic AFT model

Variable	**Estimated Coefficient**	**t-statistics**	**Percentage change**
Temporal Characteristics			
February	-0.45	-7.18	-26.30
April	-0.28	-4.45	-24.50
May	-0.25	-4.08	-22.29
Geographical Characteristics			
Musafah Highway	-0.22	-4.15	-20.45
AbuDhabi-Tarif Highway	0.09	2.04	9.80
Almafraq-Sawameq Highway	-0.25	-3.78	-22.55
Swihan Highway	0.39	4.28	48.10
Environmental Characteristics			
Sandy	0.19	2.15	21.60
Model Structure Parameters			
P (distribution shape parameter)	4.34	39.33	
λ (the scale parameter)	0.07	100.22	
Intercept term (_cons)	2.64	100.22	
Goodness-of-fit Test			
Akaike information criterion	557.51		
Initial log-likelihood	-329.83		
Log-likelihood at convergence	-268.75		
Number of observations	504		

6.4.1.4 Model Selection

The results of the plots of comparing the observed duration to the predicted duration for the three distributions were presented in Figure 6-51. This figure shows that the plots show a weak relationship between the observed and predicted duration in the three distributions (please see note on Figure 6-3 for model fit statistics, as this is a repeat of Figure 6-3).



Weibull distribution | Log-normal distribution | Log-logistic distribution

Figure 6-51 Highway response time analysis - Comparison between the observed and predicted duration for Weibull, Log-normal and Log-logistic distributions

The results of the AIC test show that the Log-logistic model is the best fit distribution for the response time (Table 6-38). Thus, the following section will explain in detail the process of data analysis for the Log-logistic model.

Table 6-38 Highway response time analysis - Comparison of AIC values for AFT response time models

Model	-2In *L*	K	c	AIC
Weibull	566.74	18	2	606.75
Log-normal	529.84	11	2	564.03
Log-logistic	**537.50**	**8**	**2**	**557.51**

6.4.1.5 Interpretation of the Estimated Model

The Log-logistic AFT response time model has 8 statistically significant explanatory variables. As shown in Table 6-37 it was found that all significant variables belong to three categories used in this research, including temporal characteristics, geographical characteristics, and environmental characteristics. This section presents the significant variables and their explanations.

Temporal Characteristics

Three months of the year (February, April and May) were found to be significant in the model. Accidents occurring in these months were shown to have (26.30%, 24.50%, 22.29%) shorter response time compared to accidents that happened in other months (Figure 6-52). This result was unexpected because the number of accidents per month and the severity level per month are not significantly different throughout the year. This result raised an important question about aspects of current practice of accident response that might affect these three months in particular. As mentioned in Chapter 5, AHCIB records included 356 parallel accidents (see Figure 6-53) that could not be attended immediately by collision investigators due to their involvement in investigation of other accidents. While the parallel accidents are not included in this analysis, they are likely to affect the response time. This is because the investigators deal with two or more accidents at the same time, which may significantly limit the availability of investigators in the office to respond to further accidents. Analysis of the dataset used in this research showed that February, April and May contained fewer parallel accidents. Therefore collision investigators were available more of the time for immediate response during these three months. This has been noted as a weak point in current TIM practice, as all accidents should be treated with the same priority. Recruiting more collision investigators or distributing the current staff closer to known accident hot spots would be measures that might facilitate a swifter response to any accident.

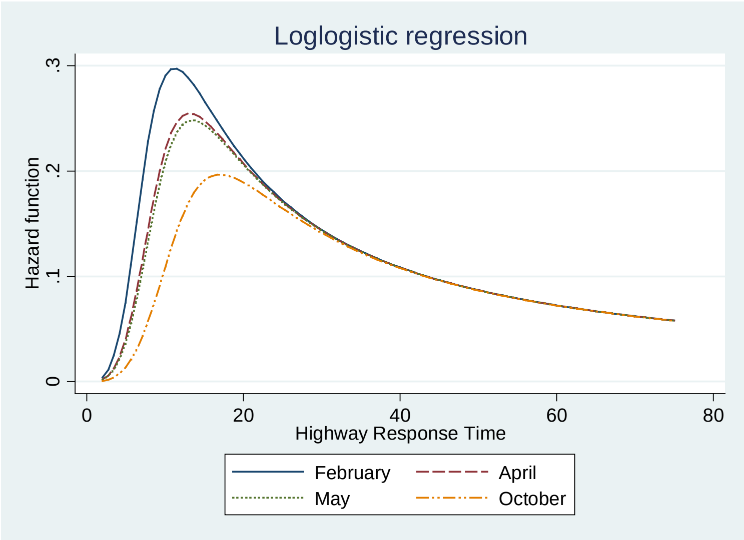


Figure 6-52 Highway response time analysis - Hazard function by month

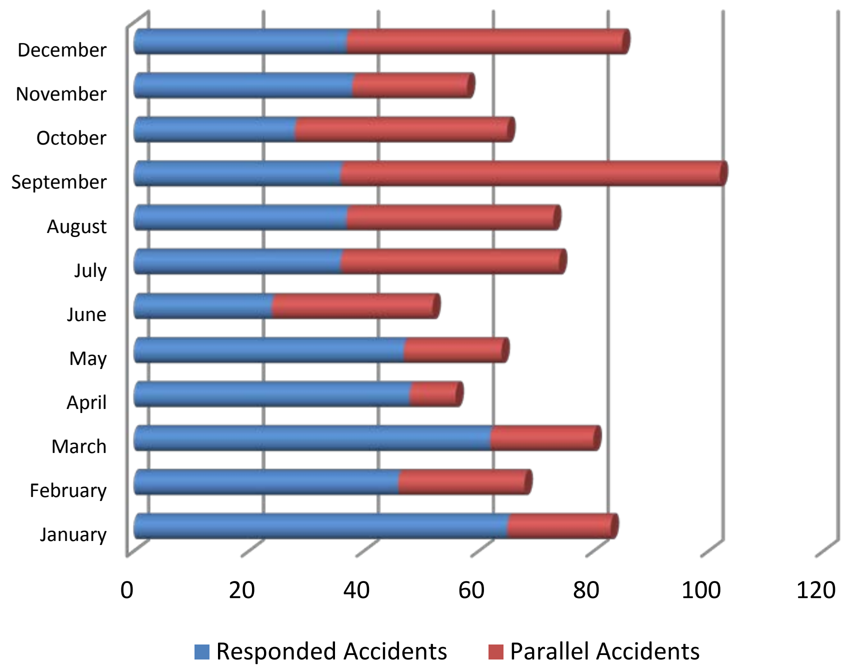


Figure 6-53 Highway response time analysis - Frequency of total accidents per month

Geographical Characteristics

The location of the accidents was found to have a significant effect on the accident response time model. Accidents that happened on the Mussafah and Al Mafraq–Sawameq highways were associated with (20.45%, 22.55%) lower response time, whereas accidents occurring on the Abu Dhabi–Tarif and Sweihan highways were associated with (9.80%, 48.10%) longer response time (Figure 6-54). These results were expected, because the AHCIB is located at one end of the Al Mafraq–Sawameq highway (which is 20 kilometres long). This also applies to Mussafah highway, which is 15 kilometres long from the location of the AHCIB. On the other hand, the Abu Dhabi–Tarif highway is over 100 kilometres long, and the Sweihan highway starts 40 kilometres away from the AHCIB, being 50 kilometres in length (Figure 6-55). As expected, accidents that happened at a further distance from the AHCIB location were shown to have longer response time.

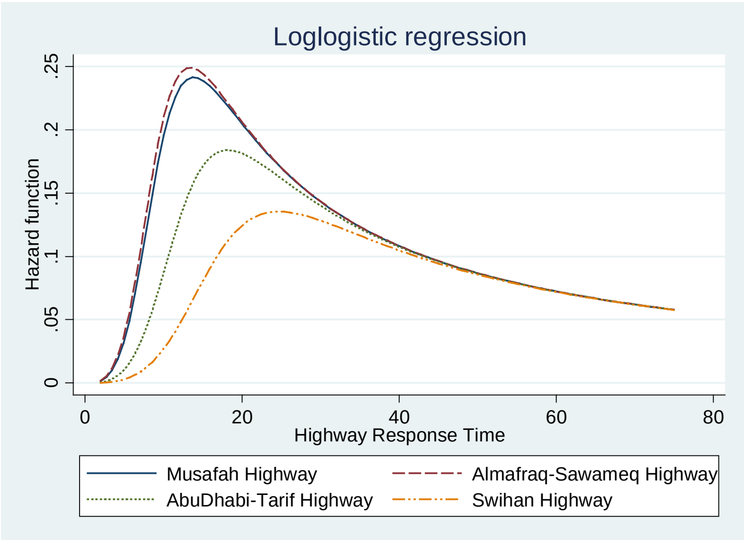


Figure 6-54 Highway response time analysis - Hazard function by highway

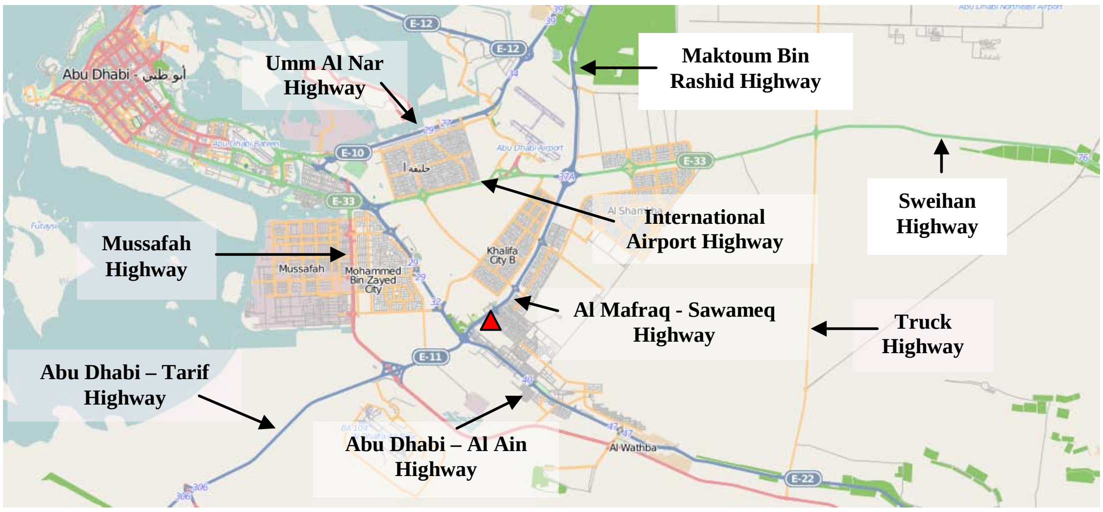


Figure 6-55 AHCIB (▲) and the highways included in this study

Environmental Characteristics

The road surface condition was found to have a significant effect on the response time. Accidents that occurred on a sandy road surface were associated with (21.60%) longer response time compared to accidents that occurred on other road surface conditions (Figure 6-56). Again, this result was expected because collision investigators cannot drive on a sandy road surface in a similar manner to when the road surface is clear and dry, so the travel time is expected to be longer in sandy conditions.

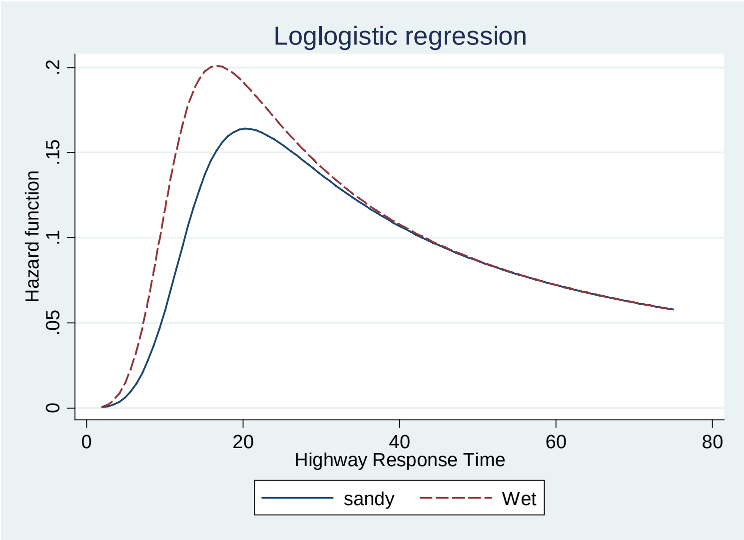


Figure 6-56 Highway response time analysis - Hazard function by road surface condition

Finally, the survival function and the hazard function are presented in Figure 6-57. Regarding the effects on response time, the log-logistic AFT model has P = 4.34. As mentioned earlier in section 3.3.3, if P > 1 this means having a non-monotonic slope and increasing from zero to a maximum at time $t = [(P - 1)^{1/P}] / \lambda$ and decreasing thereafter. In other words, there is higher chance of response for the accident before the maximum of the log-logistic distribution is reached after duration of about 18.6 minutes. However, after this time, the further the time gone without an AHCIB response, the less likely it is to have an AHCIB response soon.

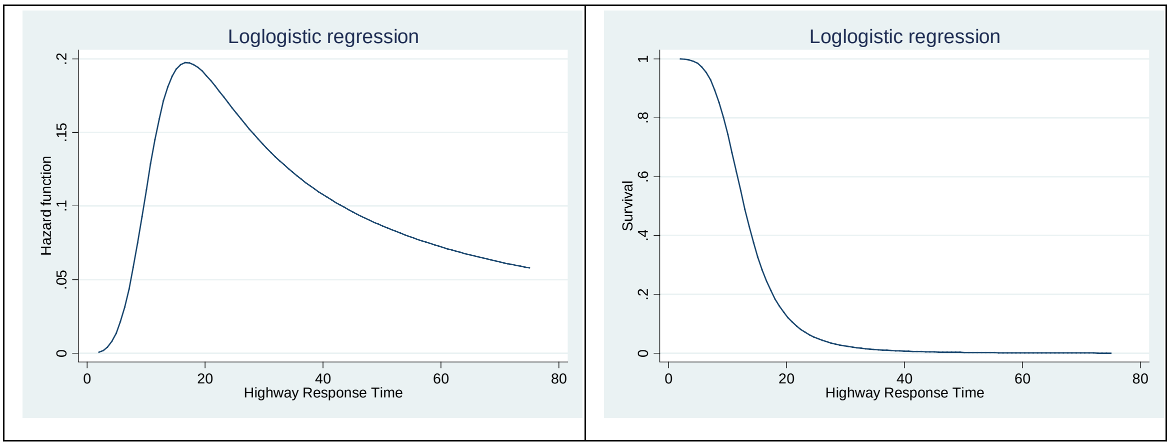


Figure 6-57 Highway response time analysis - Log-logistic hazard function and survival function

Duration Prediction and Model Accuracy

After fitting the log-logistic AFT model, the prediction of mean duration was performed. Then, the accuracy of this prediction conducted by MAPE as illustrated in the Methodology chapter. For the highway response time, the MAPE value for the log-logistic AFT model was calculated as 20.4%. This result means that the prediction with the developed model has a good forecasting accuracy according to Table 4-2.

To develop a decision tree for highway accident response time, 8 significant variables were considered: sandy weather condition, February, April, May, Musafah Highway, Abu Dhabi – Tarif Highway, Almafraq – Sawameq Highway, Swihan Highway. It starts by splitting according to weather condition, then by month of year, and finally by highway (Figure 6-58, Figure 59). If an accident occurs in sandy weather condition and in February, the predicted response time is 15.41 minutes; however if it occurs in May, the predicted response time is 17.62 minutes.

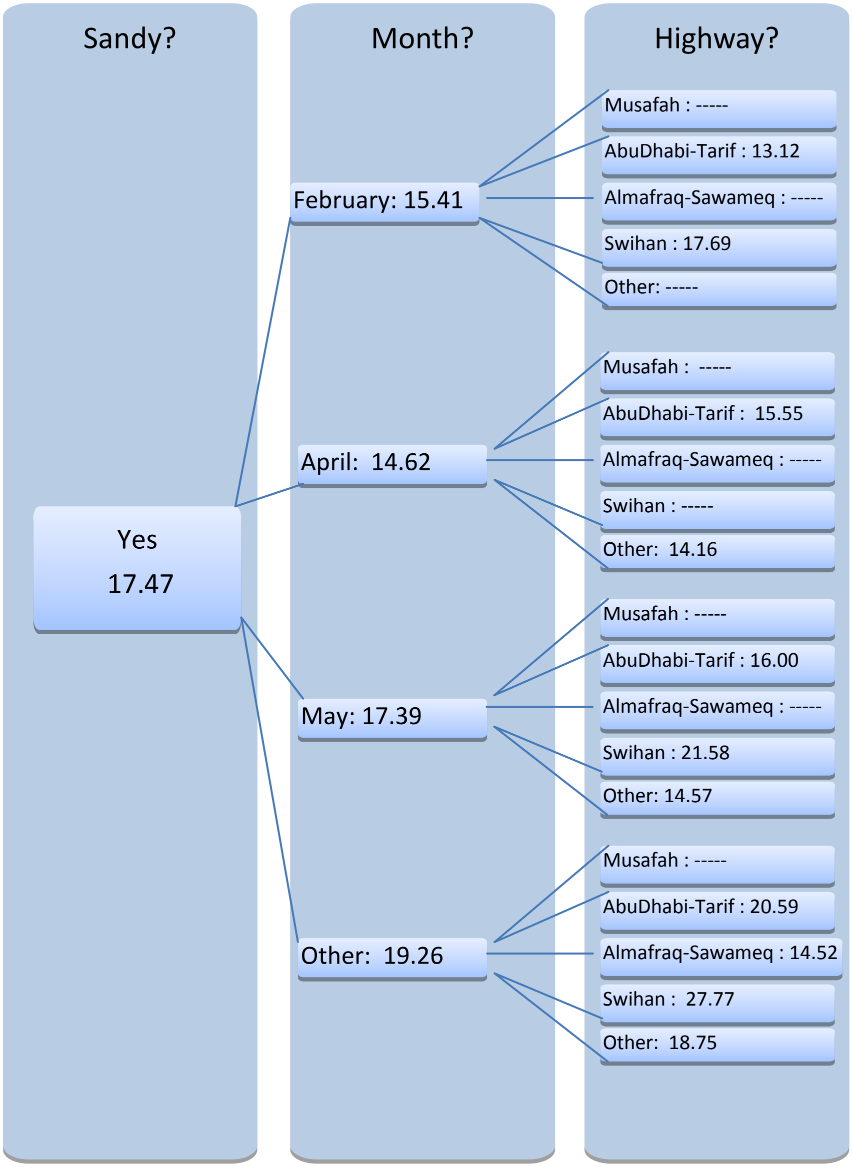


Figure 6-58 Highway response time analysis - Decision tree of sandy weather condition accidents (Durations are in minutes)

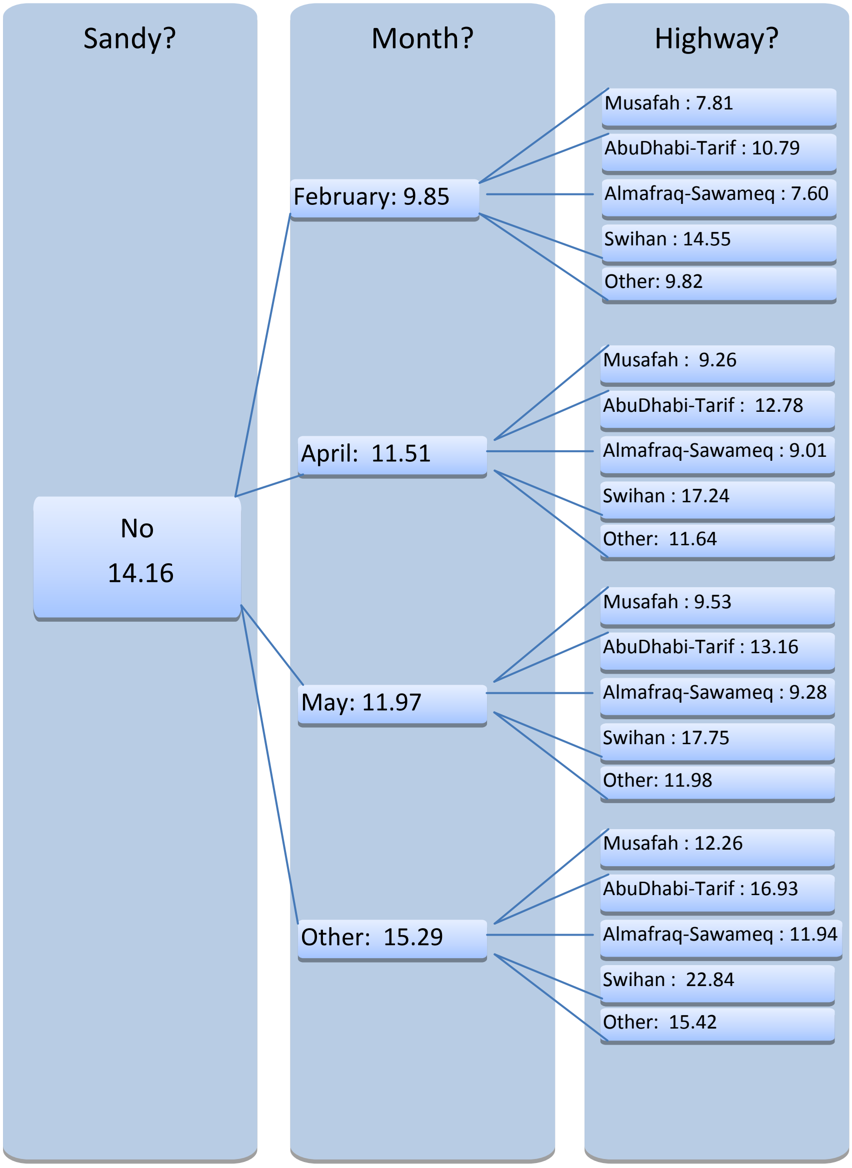


Figure 6-59 Highway response time analysis - Decision tree of other weather condition accidents (Durations are in minutes)

Comparison to Previous Research

Similar to urban accident interval time, the comparison in this section is limited to previous research that has modelled each interval time of the total incident duration separately (Lee and Fazio, 2005; Nam and Mannering, 2000).

In previous studies by Lee and Fazio (2005), and (Nam and Mannering, 2000a) temporal characteristics, environmental characteristics, geographic characteristics and accident characteristics were found to have a significant effect on response time. However, none of the accident characteristics variables were found to be significant in this study. Also, weather condition variables have been statistically significant in previous research, but were not found to be significant in this study.

On the other hand, the best fit distribution for the response time found in this research is different to previous research. For example, the Weibull with frailty distribution provided the best fit distribution in another study (Nam and Mannering, 2000a). However, it is not possible to compare the fitted distribution of this study with Lee and Fazio's study (Lee and Fazio, 2005), which applied the Cox Proportional Hazard Model, a model that does not require distribution assumption. Finally, having noted these differences, it is still possible for different datasets and different case study locations to yield different results.

6.5 Summary

This chapter presented the results of applying HBDMs with emphasis on using the AFT metric to investigate the effects of accident characteristics on each interval time of the total urban traffic accident duration (reporting, response and clearance time) and highway traffic accident duration (response time). The results show that the fitted distribution for one interval time might not fit another interval time. Also, there was no similarity in the explanatory variables that affect each interval time.

In addition to this, the results have shown that there are many points of weakness in the current practices of traffic accident management in Abu Dhabi. Furthermore, a comparison of each model's results with previous work was conducted at the end of each model.

Although the accuracy level was found to be low in most models, it was decided to develop a decision tree because they are of benefit to practitioners in terms of using a prediction tool that is quick and easy to understand. Also, based on the results of a decision tree, practitioners will be able to make decisions regarding whether to employ a traffic diversion around an accident scene or not. Furthermore, the predicted time could be disseminated to the road users which could ease incident related congestion. As a result, it is expected that the decision trees may enhance the effectiveness of the TIM process. Thus, the decision trees have been developed based on the predicted values of duration from the fitted model. As a fist attempt of visualising the outputs of the model, decision trees were designed in this study. It is expected that the decision trees proposed here will be improved in the future using the new and more accurate data coming into the system.

7 Conclusions and Recommendations

7.1 Conclusions

The aim of this research was to investigate the effects of traffic accident characteristics on each interval time (reporting time, response time and clearance time) of the urban traffic accidents and response time of highway traffic accidents duration in Abu Dhabi, the capital of the UAE. To achieve the aim of this research, fully parametric HBDMs were used, with emphasis on the Accelerated Failure-Time (AFT) metric. Although final models did not have an overall excellent fit, best-fitting models were developed, and their results can guide both current knowledge and future research. Further, although much work has been done to study incident duration, few studies have focused on urban area accidents in developing countries.

The research described in this thesis used primary data from Abu Dhabi to evaluate the impact of various factors on each interval time belong to total duration of traffic incidents such as reporting time, response time, and clearance time. Therefore, this research has made both methodological and practical contributions. This thesis contributed to the methodology of applying AFT models to each time interval of an accident rather than the whole duration. Adopting this approach proved that model estimation results have diverse factors that affect reporting time, response time, and clearance time. For example, temporal characteristics were found to affect reporting time and clearance time, but not response time. Also, environmental characteristics were found to affect response time and clearance time, but not reporting time. Also, the “daylight” variable was found to affect response time, but not reporting time or clearance time. These examples show that the methodological approach adopted in this thesis has the merit of providing more insight into factors that affect duration. Also, these findings have an important implication for accident responders whose work influence one time interval and not the whole accident duration. Also the research conducted in this thesis uses a unique set of data collected for the purpose of this research, proposes a methodology to be able to investigate the data with a useful set of analytical methods including duration models for model development and MAPE for testing the predictive accuracy of the models. Since accident data are scarce in Middle Eastern countries in general, Abu Dhabi in particular, this research provides a useful contribution both to the knowledge and practice.

It is identified that there are no previous analyses in this subject happened in Abu Dhabi to date. So the outcome of this research is not only unique but also has practical implication on how academic research can be linked in to practice. . Therefore it is expected that the results provide the Abu Dhabi authorities with a vast amount of knowledge regarding the impact of various factors on traffic incident duration leading to generate an effective incident management system for the future. The authorities will be able to use the decision trees as a basis to estimate the predicted incident duration. Also, the authorities can utilize the recommendations put forward in this thesis to improve traffic incident management and traffic safety in Abu Dhabi.

This study began by reviewing previous work on the analysis of traffic incident duration. Investigation showed that several approaches have been applied in order to investigate the factors that affect incident duration and develop an estimation model of incident duration. These approaches are Linear Regression Models, Time Sequential Models, Nonparametric Regression Method, Decision Tree and Classification Trees, Bayesian Networks (BNs), Discrete Choice Models, Fuzzy Logic (FL), Artificial Neural Network (ANN) and Hazard-Based Duration Models (HBDMs). Among these approaches, HBDMs were found to have many advantages over the previous methods in that they give more insight into duration dependence through hazard function. They also have the ability to deal with censored observations. Thus, a methodology framework was developed to estimate various models considering three distributions (Weibull, Log-logistic, and Log-normal)

An attempt was made to collect traffic accident data and accident duration data from Newcastle urban area. Several sources were considered, including the Newcastle City Council CCTV System, Northumbria Police logs, and Traffic and Accident Data Unit (TADU) records. However, it was not possible to measure the length of incident duration and collect incident characteristics because of several barriers, such as the absence of cameras recording, lack of accident location co-ordinates and missing traffic flow data. Thus, it was difficult to check the validity of police information related to traffic accidents, and an alternative approach was consequently chosen.

This alternative was to check the possibility of obtaining more data regarding traffic accidents from Northumbria Police. After several communications with the Operations Department Chief Inspector and the Chief Constable of Northumbria Police, it proved impossible to get any further information regarding accident characteristics and accident duration from Northumbria Police due to provisos in the Freedom of Information Act, whereby public authorities may refuse a request for information if extracting such information would take more than 18 hours. Thus, the focus of the study area moved to Abu Dhabi.

This study is therefore based on the metropolitan network and highway area in the city of Abu Dhabi, the capital of the United Arab Emirates (UAE). Two databases were utilised to extract the data for this study. The first one is the Federal Traffic Statistics System (FTSS), which covers all traffic accident records in the UAE. The FTSS database has comprehensive accident-related information, such as temporal characteristics (time of day, day of week and month of year), geographical characteristics (road name and location) and accident characteristics (severity level, weather condition, injury details and vehicle details). The second database is the records of Abu Dhabi Serious Collision Investigation Section (ASCIS). These records contain the details of accident duration, including reporting time, response time, clearance time and the total time.

The first model to be developed was urban accident reporting time. The results show that the average reporting time was 8.23 minutes and the Log-normal AFT model provided the best fit to the accident reporting time. Also, reporting time was found to vary based upon temporal characteristics and geographical characteristics in this research. Accidents occurring out of the peak period were associated with lower reporting time compared with peak period accidents. Since Abu Dhabi Serious Collision Investigation Section (ASCIS) staff will move to the accident scene if the accident has resulted in serious injury or fatal injury, this result could be associated with the availability of accident information regarding injury status for the accidents which occurred out of the peak period. Another significant temporal variable was day of week. The results show that accidents that occurred on Friday were associated with a lower reporting time compared to other days of week. Moreover, the results show that the reporting time of the accidents that occurred in February, March, November, and December was associated with lower reporting time compared with accidents in other months of the year, whereas the reporting time of accidents in May were associated with higher reporting time. These findings might be interpreted as a result of the occurrence of accidents during the off-peak period. Accidents in Meena Street were associated with longer reporting time compared with accidents that occurred on other streets.

The second model developed was urban accident response time. The results show that the average response time was 6.48 minutes, and the Log-logistic AFT model provided the best fit to the accident response time. Response time was found to vary based on some variables of geographical characteristics and environmental characteristics. Accidents that occurred on the Eastern Ring Road and Khalifa Bin Zayed Street were found to be associated with a lower response time compared with accidents on other roads of the city. This is probably because this road is the nearest road to the location of the AUCIB. Also, the results show that accidents in wet road surface conditions experienced higher response time compared with those in dry or sandy conditions. Finally, accidents that occurred in day light experienced higher response time compared with those that occurred at night. Examining the dataset used in this research demonstrates that this may have occurred because 63% of the accidents happened during the day light.

Urban accident clearance time was the third model. The results show that the average clearance time was 26.26 minutes, and the Log-logistic AFT model provided the best fit to the accident clearance time. Clearance time varied based on 12 statistically significant explanatory variables. Accidents that occurred in the morning were associated with lower clearance time compared with the clearance time of accidents during other times of day. Additionally, two months of the year were found to affect clearance time significantly. One of these months (August) was associated with shorter clearance time, whereas the other month (March) was found to be associated with longer clearance time. This is probably due to the occurrence of fatal injuries in all months of the year, with the exception of August. Also, accidents that occurred on Monday were found to have less clearance time compared to other days of week.

Furthermore, the street variable (Meena Street) was associated with longer clearance time. After carefully examining the data of these accidents, this finding could be interpreted as being attributable to the high number of injuries and the occurrence of fatal injuries in this street, compared with other streets. Additionally, accidents that occurred in commercial areas were associated with lower clearance time. It is difficult to interpret these findings, but they could be related to the traffic conditions in these locations and the desire to hasten the clearance time in congested areas. In terms of the environmental variables, as expected, accidents that occurred in rain weather conditions were associated with longer clearance time, whereas accidents that occurred in foggy conditions were found to be associated with longer clearance time.

‘Hit object’ and “rollover” accidents had a longer clearance time compared with the clearance time of other accident types. This may be because collision investigation procedure varies among collision investigators; some investigators will spend a long time gathering all the required information and witness statements from the accident scene, whereas other investigators collect the basic critical information in the accident report and leave other information and witness statements to be taken after clearing the scene. Thus, the time taken to write down the non-basic information and witnesses’ statements by the latter investigators will not be considered as a part of the accident clearance time. This finding highlights the absence of clear guidelines for collision investigation and the need for a standard procedure that should be followed by all collision investigators. Finally, it was found that longer accident clearance time was observed when there was an increase in the number of casualties and the number of vehicles involved.

The fourth model concerned highway accident response time. The results demonstrated that the average response time was 14.22 minutes, and the Log-logistic AFT model provided the best fit to the accident response time. 8 significant variables were found to be significantly affecting highway accident response time. Three months of the year (February, April and May) were found to have shorter response times than other months. This might be because of the high number of parallel accidents (356 accidents) that occurred in this year. Furthermore, accidents occurring on the Mussafah and Al Mafraq–Sawameq highways were associated with a lower response time, whereas accidents on the Abu Dhabi–Tarif and Sweihan highways were associated with a longer response time. These results were expected because of the distance between these highways and the location of AHCIB.

Moreover, accidents which occurred when the road surface was covered by sand were associated with a longer response time. Again, this result was expected because collision investigators cannot drive on a sandy road surface in a similar manner to when the road surface is clear and dry.

All of the results described in this research can assist traffic accident investigators and traffic operators to gain more insight into accident duration dependence. Also, an estimation of each interval time of the total accident duration can be developed based upon these findings. Although the prediction accuracy was found to be inaccurate, this research could be used as a guide for better prediction if further information and more accurate measurements of duration were collected. This prediction might lead to a better use of available resources and taking decisions that may ease the adverse impacts of traffic accidents, such as traffic diversion and traffic information dissemination.

7.2 Recommendations

The recommendations presented in this section are divided into two parts. The first part focuses on recommendations to improve the process of traffic accident management in Abu Dhabi. First of all, an investigation of the TIM process in Abu Dhabi shows that there is no traffic information dissemination to the public concerning the status of traffic conditions in Abu Dhabi. Currently, drivers are simply caught in traffic congestion. Thus, it is recommended to develop traffic information dissemination systems such as traffic radio or SMS messages. Such systems might play an important role in reducing the harmful impacts of traffic accidents on the accident scene and the surrounding area.

Another recommendation is regarding the current practice of recording accident duration in ASCIS. It was found that duration data are recorded manually and the call operator closes the accident log after informing ASCIS. Therefore, investment in a system to track the location of any traffic accident responder by the Abu Dhabi Police Operational Centre is highly recommended. Also, the log of any traffic accident should not be closed until the last responder leaves the accident scene. This might help to extract an accurate time for each interval time of the accident duration, as well as to evaluate the performance of traffic accident responders.

Also, research findings indicated the drawbacks of the current practice of accident response in Abu Dhabi, particularly the absence of the capability to respond to the accident immediately after informing AHCIB due to the absence of staff availability. These effects could be mitigated by recruiting more collision investigators or distributing the current staff closer to known accident hot spots. These measures should result in a reduction in response time to each accident, which may, in turn, lead to a reduction in the severity of injuries or the number of fatalities.

Other drawbacks of the current accident management procedure in Abu Dhabi were the perception of opening the road to traffic as soon as possible if the accident occurred during peak periods and the diversity of collision investigation procedures among the personnel. Mitigating them could be achieved by means of several measures, such as publishing a standard collision investigation guideline and strategic clearance time targets based on the accident severity. These measures could ensure the fulfilment of a high level of investigation quality for each accident which may lead to identifying the main causes of the accident.

The second part of the recommendations is dedicated to future work. Collecting further data is needed to enable analysis of the other interval times of accident duration which cannot be analysed in this study, including highway accident reporting time, highway accident clearance time, highway accident recovery time and urban accident recovery time. Examples of these data include the location of the vehicle after the accident, damage rate, vehicle type, the involvement of hazard material, the age of casualties, number of lanes blocked, daily traffic conditions, presence of road works and traffic flow data. Such an investigation would assist ASCIS to understand duration effects of all interval times of the total accident duration and, as a result, identify and facilitate application of the best measures to improve the efficiency of the TIM process.

8 References

Abdel-Rahim, A. (2004) 'Incident Management', in *Handbook of Transportation Engineering*.

Abu Dhabi Health Authority (2009) *Injury deaths 2008 - Abu Dhabi Emirate*. Available at: http://www.roadsafety.ae/index_en.html.

Acock, A.C. (2006) *A Gentle Introduction to Stata*.

Akaike, H. (1974) 'A New Look at the Statistical Model Identification', *IEEE Transactions*, 19(6), pp. 716-723.

Al kathairi, A., Mufti, R., Garib, A., Williams, B. and Karim, A. (2001) 'Traffic Characteristics in the Arab Gulf Region: Case Study in Abu Dhabi, United Arab Emirates', *Transportation Research Record: Journal of the Transportation Research Board*, Volume 1768 / 2001, pp. 51-60.

Alison, P. (1984) *Event history analysis: regression for longitudinal event data*.

Allison, P.D. (1995) *Survival analysis using the SAS system: a practice guide*. SAS company.

Altman, D.G. (2003) *Practical statistics for medical research*. London: Chapman & Hall.

Anderson, T. (2003) *Review of Current Practices in Recording Road Traffic Incident Data: With Specific Reference to Spatial Analysis and Road Policing Policy*.

Austroads (2007) *Improving Traffic Incident Management: Literature Review* (AP-R296/07). Available at: www.austroads.com.au.

Authority, A.D.T. (2009) *Facts and Figures*. Available at: http://www.visitabudhabi.ae/en/uae.facts.and.figures/population.aspx.

Balke, K. and Cooner, S. (2000) *Use of Photogrsmmetry for Investigation Of Traffic Incident Scenes* (TX-99/4907-2).

Ben-Akiva, M. and Lerman, S. (1985) *Discrete Choice Analysis: Theory and Application to Travel Demand (Transportation Studies)*. Massachusetts: MIT Press.

Bener, A. and Crundall, D. (2005) 'Road traffic accidents in the United Arab Emirates compared to Western countries', *Advances in Transportation Studies an International Journal*, A6.

Bhat, C.R. (1996) 'A generalized multiple durations proportional hazard model with an application to activity behavior during the evening work-to-home commute', *Transportation Research Part B: Methodological*, 30(6), pp. 465-480.

Bhat, C.R. (2000) 'Duration Modelling', in Hensher, D.A. and Button, K.J. (eds.) *Handbook of Transport Modelling*. Elsevier Science, London, pp. 91-112.

Bhat, C.R., Frusti, T., Zhao, H., Schönfelder, S. and Axhausen, K.W. (2004) 'Intershopping duration: an analysis using multiweek data', *Transportation Research Part B: Methodological*, 38(1), pp. 39-60.

Biggs, D., DeVille, B. and Suen, E. (1991) 'A method of choosing multiway partitions for classification and decision trees', *Journal of Applied Statistics* 18(1), pp. 49-62.

Bishop, C. (1995) *Neural networks for pattern recognition*. Clarendon, Oxford.

Blossfeld, H., Golsch, K. and Rohwer, G. (2007) *Event History Analysis with Stata*.

Box-Steffensmeier, J.M. and Jones, B.S. (2004) *Event History Modeling : A Guide for Social Scientists*.

Bradburn, M., Clark, T., Love, S. and Altman, D. (2003) 'Survival Analysis Part II: Multivariate data analysis – an introduction to concepts and methods', *British Journal of Cancer*, 89, pp. 431-436.

Bradburn, M.J., Clark, T.G., Love, S.B. and Altman, D.G. (0000) 'Survival Analysis Part II: Multivariate data analysis - an introduction to concepts and methods', *Br J Cancer*, 89(3), pp. 431-436.

Brieman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J. (1984) *Classification and regression trees*. Belmont: Wadsworth International Group.

Brücher, H., Knolmayer, G. and Mittermayer, M. (2002) 'Document classification methods for organizing explicit knowledge', *European Conference on Organizational Knowledge, Learning, and Capabilities* Athens. Available at: http://www.ie.iwi.unibe.ch/publikationen/berichte/resource/WP-140.pdf.

Carson, J.L., Legg, B., Mannering, F. L., Nam, D. and Nee, J (1999) 'Are incident management programs effective? Findings from Washington State', *Transportation Research Recoord* 1683, pp. 8-13.

Charles, P., Dunn, K., Fitzgerald, P., Peters, N. and Reardon, G. (2003) 'Establishing inter-agency regional incident management in Brisbane Australia', *Proc. 10th World Congress and Exhibition on Intelligent Transport Systems and Services*.

Chatterjee, K. and Ma, K. (2008) 'Time taken for residents to adopt a new public transport service: examining heterogeneity through duration modelling ', *Transportation*, 36(1), pp. 1-25.

Choi, K. (1996) *Predicting Freeway Traffic Incident Duration an Expert System Context Using Fuzzy Logic*. Southern California.

Chung, Y. (2009) 'Development of an accident duration prediction model on the Korean Freeway Systems', *Accident Analysis & Prevention*, 42(1), pp. 282-289.

Clark, T., Bradburn, M., Love, S. and Altman, D. (2003) 'Survival Analysis Part IV: Further concepts and methods in survival analysis', *British Journal of Cancer*, (89), pp. 781-786.

Cleves, M., Gould, W. and Gutierrez, R. (2004) *An Introduction to Survival Analysis using Stata*. Stata Press.

Collett, D. (1994) *Modelling Survival Data in Medical Research*. London: Chapman & Hall.

Cox, D.R. (1972) 'Regression Models and Life-Tables', *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(2), pp. 187-220.

Cox, D.R. and Oakes, D. (1984) *Analysis of Survival Data*. Chapman & Hall.

Cross, S.S., Harrison, R.F. and Kennedy, R.L. (1995) 'Introduction to neural networks', *Lancet*, 346(8982), pp. 1075-1079.

Derose, J. (1964) 'An analysis of reandom freeway traffic accidents and vehicle disabilities', *Highway Research Record*, 59, pp. 53-65.

Dougherty, M. (1995) 'A review of neural networks applied to transport', *Transportation Research, Part C: Emerging Technologies*, 3 C(4), pp. 247-260.

Ettema, D., Borgers, A. and Timmermans, H. (1995) 'Competing risk hazard model of activity choice, timing, sequencing, and duration', *Transportation research record*, pp. 101-109.

Farradyne, P. (2000) *Traffic Incident Management Handbook*. Available at: http://www.scopus.com/scopus/inward/record.url?eid=2-s2.0-0003417236&partnerID=40&rel=R7.0.0

FHWA (2004) *Traffic Incident Management.* Available at: http://www.ops.fhwa.dot.gov/aboutus/one_pagers/tim.htm (Accessed: February 5th).

FHWA (2005) *Traffic Congestion and Reliability: Linking Solutions to Problems*. Available at: http://ops.fhwa.dot.gov/congestion_report_04/executive_summary.htm#footer3

Florida, D.o.T. (2005) *Best Practices for Traffic Incident Management in Florida*. Available at: http://www.dot.state.fl.us/research-center/Completed_Proj/Summary_TE/FDOT_BC353_47_rpt.pdf.

Fox, J. (1997) *Applied regression analysis, linear models, and related methods*.

Fu, H. (2008) 'Identifying repeat DUI crash factors using state crash records', *Accident Analysis & Prevention*, 40(6), pp. 2037-2042.

Garib, A., Radwan, A.E. and AlDeek, H. (1997) 'Estimating magnitude and duration of incident delays', *Journal of Transportation Engineering-Asce*, 123(6), pp. 459-466.

Garson, G.D. (2009) 'Event History Analysis'. October, 2009. Available at: http://faculty.chass.ncsu.edu/garson/PA765/event.htm.

Gilbert, C. (1992) 'A duration model of automobile ownership', *Transportation Research Part B: Methodological*, 26(2), pp. 97-114.

Giuliano, G. (1989) 'Incident characteristics, frequency, and duration on a high volume urban freeway', *Transportation Research, Part A: General*, 23 A(5), pp. 387-396.

Gloder, M. (2008) *Parametric Models*.

Golder, M. (2012) Obtaining prediction of mean duration after fitting frailty mode.

Golob, T.F., Recker, W.W. and Leonard, J.D. (1987) 'An analysis of the severity and incident duration of truck-involved freeway accidents', *Accident Analysis and Prevention*, 19(5), pp. 375-395.

Greene, W. (2003) *Econometric Analysis*. 5 edn. Prentice-Hall

Hamed, M.M., Jaradat, A.S. and Easa, S.M. (1998) 'Analysis of commercial mini-bus accidents', *Accident Analysis & Prevention*, 30(5), pp. 555-567.

Hamed, M.M. and Mannering, F.L. (1993) 'Modeling travelers' postwork activity involvement: toward a new methodology', *transportation Science*, 4, pp. 381-394.

Haykin, S. (1999) *Neural Networks: a comprehensive foundation*. Prentice Hall.

Heckman, J. and Singer, B. (1984) 'A method for minimizing the impact of distributional assumptions in econometric models for duration data', *Econometrica*, 52, pp. 271-320.

Hensher, D.A. and Button, K.J. (2000) *Handbook of Transport Modelling*.

Hensher, D.A. and Mannering, F.L. (1994) 'Hazard-based duration models and their application to transport analysis', *Transport Reviews*, 14(1), pp. 63-82.

Jacobson, L., Legg, B. and O'Brien, A. (1992) *Incident Management Using Total Stations*.

Jenkins, S.P. (2005) *Survival Analysis*. Available at: http://www.iser.essex.ac.uk/files/teaching/stephenj/ec968/pdfs/ec968lnotesv6.pdf (Accessed: July, 2008).

Jensen, F. (2001) *An Introduction to Bayesian Networks*. UCL Press, London.

Jiezhi Qi (2009) *Comparison of Proportional Hazards and Accelerated Failure Time Models*.

Johnson, R. and Wichern, D. (2003) *Applied multivariate statistical analysis*. Prentice Hall

Jones, B., Janssen, L. and Mannering, F. (1991) 'Analysis of the Frequency and Duration of Freeway Accidents in Seattle', *Accident Analysis and Prevention*, 23(4), pp. 239-255.

Jong, G.D. (1996) 'A disaggregate model system of vehicle holding duration, type choice and use', *Transportation Research Part B: Methodological*, 30(4), pp. 263-276.

Jovanis, P.P. and Chang, H.-L. (1989) 'Disaggregate model of highway accident occurrence using survival theory', *Accident Analysis & Prevention*, 21(5), pp. 445-458.

Kalbfleisch, J. and Prentice, R. (2002) *The Statistical Analysis of Failure Time Data*.

Kaplan, E.L. and Meier, P. (1958) 'Nonparametric Estimation from Incomplete Observations', *Journal of the American Statistical Association*, 53, pp. 457-481.

Kass, G.V. (1980) 'An Exploratory Technique for Investigating Large Quantities of Categorical Data', *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 29(2), pp. 119-127.

Kay, R. and Kinnersley, N. (2002) 'On the use of the accelerated failure time model as an alternative to the proportional hazards model in the treatment of time to event data: a case study in influenza', *Drug Information Journal*, 36, pp. 571–579.

Keizer, S., Akker, R. and Anton, H. (2002) 'Dialogue act recognition with Bayesian Networks for Dutch dialogues', *SIGdial Workshop on Discourse and Dialogue*. Philadelphia, PA. Available at: http://delivery.acm.org/10.1145/1120000/1118134/p88-keizer.pdf?key1=1118134&key2=9901764121&coll=GUIDE&dl=GUIDE&CFID=75602040&CFTOKEN=22011264.

Kharoufeh, J.P. and Goulias, K.G. (2002) 'Nonparametric identification of daily activity durations using kernel density estimators', *Transportation Research Part B: Methodological*, 36(1), pp. 59-82.

Khattak, A.J., Schofer, J.L. and Wang, M.H. (1995) 'A Simple Time-Sequential Procedure for Predicting Freeway Incident Duration', *Ivhs Journal*, 2(2), pp. 113-138.

Kim, H.J. and Choi, H. (2001) 'A comparative analysis of incident service time on urban freeways', *Journal of International Association of Traffic and Safety Sciences*, 25(1), pp. 62-72.

Kim, W., Natarajan, S. and Chang, G. (2008) 'Empirical Analysis and Modeling of Freeway Incident Duration', *Conference on Intelligent Transportation Systems*. Beijing, China, October 12-15, 2008. Available at: http://www.asmemesa.org/itsc08/ITSC2008/papers/0306.pdf.

Kitamura, R., Chen, C. and PEendyala, R. (1997) ' Generation of Synthetic Daily Activity-Travel Patterns ', *Transportation Research Record: Journal of the Transportation Research Board*, 1607, pp. 154-162.

Kleinbaum, D. and Klein, M. (2005) *Survival Analysis: A Self-Learning Text (Statistics for Biology and Health).*

Klembaum, D.G. (1996) *Survival analysis: A self learning text*. New York: Springer.

Knibbe, W., Alkim, T., Otten, J. and Aidoo, M. (2006) 'Automatic Estimation of Incident Duration on Dutch Highways. ', *IEEE Intelligent Transportation Systems*

Conference. Toronto, Canada, September 17-20,2006. Available at: http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=01706853.

Lange, K.L., Little, R.J.A. and Taylor, J.M.G. (1989) 'Robust Statistical Modeling Using the t Distribution', *Journal of the American Statistical Association*, 84(408), pp. 881-896.

Lapham, S.C., Kapitula, L.R., C'De Baca, J. and McMillan, G.P. (2006) 'Impaired-driving recidivism among repeat offenders following an intensive court-based intervention', *Accident Analysis and Prevention*, 38(1), pp. 162-169.

Lee, J.T. and Fazio, J. (2005) 'Influential Factores in Freeway Crash Response and Clearance Time by Emergency Management Services in Peak Periods. ', *Traffic Injury Prevention*, 6(4), pp. 331-339.

Lewis, C.D. (1982) *industrial and business forecasting method*.

Lin, P., Zou, N. and Chang, G. (2003) 'Integration of a Discrete Choice Model and a Rule-Based System for Estimation of Incident Duration: a Case Study in Maryland', *Transportation Research Board*. Available at: http://attap.umd.edu/document/TRB2004_1.pdf.

Lin, T.D., Jovanis, P. and Yang, C.Z. (1993) 'Modeling the Safety of Truck Driver Service Hours Using Time-dependent Logistic Regression', *Transportation Research Record*.

Lindley, J.A. (1986) *Qualification of Urban Freeway Congestion and Analysis of Remedial Measures* (RD/87-052).

Liu, M.J., Mao, B.H., Gao, L.P., Chen, S.K. and Sun, Q.X. (2011) 'An Exploratory Hazard-based Analysis of the First Discharge Headway', *6th International Symposium on Highway Capacity and Quality of Service* Stockholm, Sweden June 28 – July 1. pp. 536–547.

Loh, W.Y. and Shin, Y.S. (1997) 'Split selection methods for classification trees', *Statistica Sinica*, 7(4), pp. 815-840.

Mannering, F. (2007) 'Course Notes: Statistical and Econometric Methods', Available at: http://cobweb.ecn.purdue.edu/~flm/CE697N_files/Stat-econ-classnotes-07.pdf.

Mannering, F., Jones, B., Garrison, D., Sebranke, B. and Janssen, L. (1990) *Generation and Assessment of Incident Management Strategies*. Available at: http://www.wsdot.wa.gov/research/reports/fullreports/204.5.pdf.

Mannering, F.L. (1993) 'Male/female driver characteristics and accident risk: Some new evidence', *Accident Analysis & Prevention*, 25(1), pp. 77-84.

Mannering, F.L. and Hamed, M.M. (1990) 'Occurence, frequency, and duration of commuters' work-to-home departure delay', *Transportation Research Part B: Methodological*, 24(2), pp. 99-109.

Margiotta, R., Voorhies, K. and Lomax, T. (2004) 'Measuring and Communicating the Effects of Traffic Incident Management Improvements', *Transportation Research Board*. Washington D.C. Available at: http://64.233.179.104/scholar?hl=en&lr=&q=cache:fKPMFHGpdmIJ:onlinepubs.trb.org/onlinepubs/nchrp/nchrp_w64.pdf+measuring+and+communicating+the+effects+of+traffic+incident+management+improvement.

Mendel, J.M. (1995) 'Fuzzy logic systems for engineering: a tutorial', *Proceedings of the IEEE*, 83(3), pp. 345-377.

Minnesota, D.o.T. (1982) *I-35 Incident Management and the Impact of Incidents on Freeway Operation*.

Nam, D. and Mannering, F. (2000a) 'An exploratory hazard-based analysis of highway incident duration', *Transportation Research Part A: Policy and Practice*, 34(2), pp. 85-102.

Niemeier, D. and Morita, J. (1996) 'Duration of trip-making activities by men and women', *Transportation* 4(23), pp. 353-371.

Ozbay, K. and Kachroo, P. (1999) *Incident Management in Intelligent Transportation Systems*.

Ozbay, K. and Noyan, N. (2006) 'Estimation of incident clearance times using Bayesian Networks approach', *Accident Analysis and Prevention*, 38(3), pp. 542-555.

Parmer, M. and Machin, D. (1995) *Survival analysis*. UK: John Wiley and Sons Ltd.

Peeta, S., Ramos, J. and Gedela, S. (2000) *Providing Real-Time Traffic Advisory and Route Guidance to Manage Borman Incidents On-Line Using the Hoosier Helper Program* (FHWA/IN/JTRP-2000/15).

Piantadosi, S. (1997) *Clinical Trials: A Methodologic Perspective*. New York: Wiley.

Qi, Y. and Teng, H. (2004) 'Integrated Modeling Approach to Total Incident Delay', *Transportation Research Recoord*, (1895), pp. 46-54.

Qi, Y. and Teng, H. (2008) 'An information-based time sequential approach to online incident duration prediction', *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, 12(1), pp. 1-12.

Raub, R. and Schofer, J. (2007) 'Managing Incident on Urban Arterial Roadways', *Transportation Research Recoord*, 1603, pp. 12-19.

Reid, N. (1994) 'A conversation with Sir David Cox', *Statistical Science*

9(3), pp. 439-455.

Rillie, I. and Byard, N. (2006) *Investigation of Incident Related Congestion: Phase 1* (Project Report UPR/T/026/06). Available at: http://www.ha-research.gov.uk/projects/index.php?id=1011.

Schönfelder, S. and Axhausen, K.W. (2000) 'Analysing the rhythms of travel using survival analysis', *Annual Meeting of the Transportation Research Board* Washington (DC) 2001

Sethi, V., Koppelman, F.S., Flannery, C.P., Bhomderi, N. and Schofer, L. (1994) *Duration and travel time of incidents – ADVANCE Project Technical Report*, Available at: http://www.easts.info/2003journal/papers/1734.pdf.

Shin, C. (2003) 'Developing of Freeway Incident Duration Prediction Models', *Journal of the Eastern Asia Society for Transportation Studies*, 5, pp. 1734-1744.

Smith, B.L., Qin, L. and Venkatanarayana, R. (2003) 'Characterization of freeway capacity reduction resulting from traffic accidents', *Journal of Transportation Engineering*, 129(4), pp. 362-368.

Smith, B.L. and Smith, K.W. (2000) *An Investigation Into Incident Duration Forecasting for Fleetforward*. Available at: http://www.gmupolicy.net/its/papers3/FleetForward.htm.

Smith, B.L., Williams, B.M. and Keith Oswald, R. (2002) 'Comparison of parametric and nonparametric models for traffic flow forecasting', *Transportation Research Part C: Emerging Technologies*, 10(4), pp. 303-321.

Smith, K. and Smith, K.L. (2001) *Forecasting the clearance time of Freeway Accidents* (STL-2001-01.). Available at: http://www.cts.virginia.edu/docs/UVACTS-15-0-35.pdf.

Songchitruksa, P., Balke, K., Zeng, X., Chu, C.L. and Zhang, Y. (2009) *A Guidbook for Effective use of Incident Data at Texas Transportation Management Centers* (FHWA/TX-09/0-5485-P2).

StataCorp (2007) *Stata Statistical Software: Release 10*.

Stathopoulos, A. and Karlaftis, M.G. (2002) 'Modeling Duration of Urban Traffic Congestion', *Journal of Transportation Engineering*, 128(6), pp. 587-590.

Steele, F. (2005) 'Event History Analysis', Available at: http://eprints.ncrm.ac.uk/88/1/MethodsReviewPaperNCRM-004.pdf.

Sullivan, E.C. (1997) 'New model for predicting freeway incidents and incident delays', *Journal of Transportation Engineering-Asce*, 123(4), pp. 267-275.

Teng, H. and Oi, Y. (2003) 'Detection-delay-based freeway incident detection algorithms ', *Transportation Research Part C: Emerging Technologies*, 11 (3-4), pp. 265-287.

Teodorovic, D. (1999) 'Fuzzy logic systems for transportation engineering: the state of the art', *Transportation Research Part A*, 33(5), pp. 337-364.

TRB, T.R.B. (1994) *Special Report 209: Highway Capacity Manual*. Washington, D.C.

UK, H.A. (2002) *Incident Management Study*. Available at: http://www.highways.gov.uk/aboutus/627.aspx

Versavel, J. (2004) 'Automatic Video Incident Detection: The Standard ITS Technology Today for Traffic and Tunnel Safety. ', *ITS in Europe Conference: Moving Towards an Integrated Europe, ERTICO, May, Budapest.*

Wang, J.J. (1996) 'Timing utility of daily activities and its impact on travel', *Transportation Research Part A: Policy and Practice*, 30(3), pp. 189-206.

Wang, M. (1991) *Modelling Freeway Incident Clearance Times*. Northwestern University.

Wang, W., Chen, H. and Bell, M.C. (2005a) 'A Review of Traffic Incident Duration Analysis', *Journal of Transportation System Engineering and Information Technology*, 5(3), pp. 127-140.

Wang, W., Chen, H. and Bell, M.C. (2005b) 'Vehicle breakdown duration modeling', *Journal of Transportation and Statistics*, 8(1), pp. 75-84.

Washington, S.P., Karlaftis, M.G. and Mannering, F.L. (2003) *Statistical and Econometric Methods for Transportation Data Analysis.* .

Wei, C.H. and Lee, Y. (2007) 'Sequential forecast of incident duration using Artificial Neural Network models', *Accident Analysis and Prevention*, 39(5), pp. 944-954.

Wei, L.J. (1992) 'The accelerated failure time model: A useful alternative to the cox regression model in survival analysis', *Statistics in Medicine*, 11(14-15), pp. 1871-1879.

Yamamoto, T. and Kitamura, R. (2000) 'An analysis of household vehicle holding durations considering intended holding durations', *Transportation Research Part A: Policy and Practice*, 34(5), pp. 339-351.

Yuan, F. and Cheu, R. (2003) 'Incident detection using support vector machines ', *Transportation Research Part C: Emerging Technologies*, 11(3-3), pp. 309-328

Zhao, X., Li, R. and Yu, X. (2009) *Intelligent Computation Technology and Automation, 2009. ICICTA '09. Second International Conference on*.

Zhong, M. and Hunt, J.D. (2005) 'Modeling Household Weekend Activity Durations in Calgary', *2005 Annual Conference of the Transportation Association of Canada* Calgary , Canada, September 19th – 21st.

8.1 Appendix 1

Accident Data Collection Form

<table>
<tr><td colspan="6">Operation Room Reference Number :</td></tr>
<tr><td colspan="6">Accident Report Reference Number :</td></tr>
<tr><td colspan="2">Date :</td><td>Day :</td><td>Month :</td><td>Sector :</td></tr>
<tr><td>Street :</td><td>Junction :</td><td>Place :</td><td colspan="2">Co-ordinates :</td></tr>
<tr><td colspan="2">Dispatching Time :</td><td>Arrival Time :</td><td>Leaving Time :</td><td>Accident type :</td></tr>
</table>

8.2 Appendix 2

```
. stset   Reporting_Time

     failure event:  (assumed to fail at time=Reporting_Time)
obs. time interval:  (0, Reporting_Time]
 exit on or before:  failure

------------------------------------------------------------------------------
      525  total obs.
        0  exclusions
------------------------------------------------------------------------------
      525  obs. remaining, representing
      525  failures in single record/single failure data
     4325  total analysis time at risk, at risk from t =         0
                             earliest observed entry t =         0
                                  last observed exit t =       270

.
end of do-file
```

Figure 8-1 Declare reporting time to be survival time data

```
. streg , dist(weibull) nohr nolog time

         failure _d:  1 (meaning all fail)
   analysis time _t:  Reporting_Time

Weibull regression -- accelerated failure-time form

No. of subjects =          525                     Number of obs   =       525
No. of failures =          525
Time at risk    =         4325
                                                   LR chi2(0)      =      0.00
Log likelihood  =   -838.62091                     Prob > chi2     =         .

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   2.009297   .0534891    37.56   0.000      1.90446    2.114133
-------------+----------------------------------------------------------------
       /ln_p |   -.142759   .0287407    -4.97   0.000    -.1990898   -.0864282
-------------+----------------------------------------------------------------
           p |    .866963   .0249171                      .8194763    .9172014
         1/p |   1.153452    .033151                      1.090273    1.220292
------------------------------------------------------------------------------

.
end of do-file
```

Figure 8-2 Fit Weibull survival distribution with no variables

```
      failure _d:  1 (meaning all fail)
   analysis time _t:  Reporting_Time

Weibull regression -- accelerated failure-time form

No. of subjects =          525                    Number of obs   =       525
No. of failures =          525
Time at risk    =         4325
                                                  LR chi2(10)     =     89.96
Log likelihood  =    -793.64255                   Prob > chi2     =    0.0000

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      Monday |   .3125752   .1244728     2.51   0.012     .0686129    .5565375
Rashid_Bin~t |  -.3519715   .1420709    -2.48   0.013    -.6304253   -.0735177
Eastern_Ri~d |  -.2825317   .1491523    -1.89   0.058    -.5748648    .0098013
    Meena_St |   .6893233   .2142073     3.22   0.001     .2694846    1.109162
Hamdan_Bin~t |   .6045116   .2308914     2.62   0.009     .1519729     1.05705
residentia~a |  -.4093151   .1175708    -3.48   0.000    -.6397496   -.1788806
    November |   -.459327   .1678944    -2.74   0.006    -.7883941     -.13026
    December |  -.7816193   .1584071    -4.93   0.000    -1.092092    -.471147
    February |  -.5031104   .1661513    -3.03   0.002    -.8287611   -.1774598
corniche_r~d |   .9231843   .2334384     3.95   0.000     .4656534    1.380715
       _cons |   2.133896   .0724799    29.44   0.000     1.991838    2.275954
-------------+----------------------------------------------------------------
       /ln_p |  -.0126364   .0314438    -0.40   0.688     -.074265    .0489923
-------------+----------------------------------------------------------------
           p |   .9874431   .0310489                      .9284256    1.050212
         1/p |   1.012717   .0318436                      .9521885    1.077092
------------------------------------------------------------------------------

.
end of do-file
```

Figure 8-3 Fit Weibull survival distribution with variables

```
      failure _d:  1 (meaning all fail)
   analysis time _t:  Reporting_Time

Weibull regression -- accelerated failure-time form
                      Gamma frailty

No. of subjects =          525                    Number of obs   =       525
No. of failures =          525
Time at risk    =         4325
                                                  LR chi2(5)      =     20.32
Log likelihood  =    -739.36498                   Prob > chi2     =    0.0011

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      Monday |   .2133085   .1118563     1.91   0.057    -.0059258    .4325428
residentia~a |  -.2519358   .1100425    -2.29   0.022    -.4676152   -.0362564
    November |  -.2966569   .1538816    -1.93   0.054    -.5982594    .0049455
    December |   -.370258   .1453488    -2.55   0.011    -.6551363   -.0853797
    February |  -.3479263   .1560623    -2.23   0.026    -.6538027   -.0420499
       _cons |   1.464846   .0835717    17.53   0.000     1.301048    1.628643
-------------+----------------------------------------------------------------
       /ln_p |   .6845849   .0754251     9.08   0.000     .5367545    .8324153
     /ln_the |   .2769268   .1633074     1.70   0.090    -.0431498    .5970033
-------------+----------------------------------------------------------------
           p |   1.982949    .149564                      1.710447    2.298865
         1/p |   .5042995   .0380368                      .4349974    .5846427
       theta |    1.31907   .2154138                      .9577679    1.816667
------------------------------------------------------------------------------
Likelihood-ratio test of theta=0: chibar2(01) =   152.62 Prob>=chibar2 = 0.000

.
end of do-file
```

Figure 8-4 Fit Weibull survival distribution with gamma heterogeneity

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
exponential	525	-852.5203	-793.7239	11	1609.448	1656.345
weibull	525	-838.6209	-793.6425	12	1611.285	1662.446
gompertz	525	-796.41	-774.9393	12	1573.879	1625.039
lognormal	525	-749.3939	-732.8695	12	1489.739	1540.9
loglogistic	525	-751.7412	-737.2013	12	1498.403	1549.563

Note: N=Obs used in calculating BIC; see [R] BIC note

Figure 8-5 Akaike information criterion test

. tab January , su (Reporting_Time)

January	Summary of Reporting Time Mean	Std. Dev.	Freq.
No	8.0264766	17.948148	491
Yes	11.294118	14.823261	34
Total	8.2380952	17.768509	525

.
end of do-file

Figure 8-6 Produces one-way tables of frequency of January

. tab January if Friday==1, su (Reporting_Time)

January	Summary of Reporting Time Mean	Std. Dev.	Freq.
No	5.6891892	6.3654153	74
Yes	18.666667	22.941956	3
Total	6.1948052	7.6914836	77

.
end of do-file

Figure 8-7 Produces one-way tables of frequency of January conditional on Friday