CHAPTER 2
LITERATURE REVIEW

2.1 Introduction

This chapter presents some background information of the available literature related to the development of accident rates, accident models and accident statistic databases. The first section deals with the major national accident databases and the second with other existing accident databases. The third section focuses on a brief description of the highway statistics available while the fourth section deals with some recent research on the development and comparison of various accident rates. Finally, the fifth section deals with studies related to the development of accident models.

The accident databases needed to be investigated further in an effort to choose the most complete database that will provide the crash frequencies. As mentioned earlier, the safety model will utilize accident rates in order to produce the accident risk of a facility based on its free-speed. Therefore, some recent papers on the development of accident rates are presented in this chapter. Accident models are also important in estimating the impact of ITS on safety despite the fact that this is a relatively new topic. Some of the accident models developed by various research groups are presented below to demonstrate the fact that most of them refer to a limited amount of variables.

2.2 Accident Statistic Databases

In this section, the major accident statistic databases are reviewed. The level of exposure makes these databases very important when compared with local and State level databases. The largest and most complete accident database is the General Estimates System (GES) which was extensively used for the development of this thesis. Other major databases include the Fatality Analysis Reporting System (FARS), and the Highway Statistics Database. These databases will be described in further detail in this chapter. Other databases include the Highway Safety Information System (HSIS) and the Crashworthiness Data System (CDS).

Figure 2-1 below illustrates the various national accident databases that will be described below.
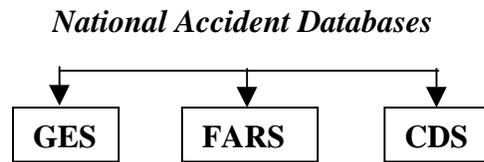
National Accident Databases

| GES | FARS | CDS |

Figure 2-1 National Accident Databases

2.2.1 General Estimates System

The General Estimates System (GES) database was developed in 1988 by the National Center for Statistics and Analysis and is operated by the National Highway Traffic Safety Administration (NHTSA). GES data are obtained from a nationally representative probability sample selected from all police-reported crashes. Primary objectives for the development of this system were the identification of traffic safety problem areas and for the database to be used as a basis for benefit/cost analyses of traffic safety initiatives.

The level of exposure makes GES, the largest accident database available in the United States. Due to its level of exposure, DOT agencies, lawyers, doctors, researchers and insurance companies use it extensively. By using the database one can estimate different accident frequencies (number of vehicle crashes). For a crash to be eligible for the GES sample, it has to meet a number of some criteria including: (a) a Police Accident Report (PAR) must be completed, (b) it must involve at least one motor vehicle traveling on a traffic way and (c) the result must be property damage, injury or death.

GES data collectors perform weekly visits to approximately 400 police jurisdictions in 60 sites across the United States. The GES 1996 file that was used for the purposes of this thesis includes approximately 56,000 Police Accident Reports (PAR's). More detailed information as to how the data are grouped and how the national estimates are produced, will be discussed in Chapter 3.

The crashes in the database are classified in a variety of ways, for example by a typical speed limit, accident severity, time-of-day and vehicle type. A statistical model that the database utilizes enables the user to extract national statistics on accident frequencies.

There are three main files (Statistical Analysis Data Sets) in the GES database that include all the variables. These files contain the following:

Accident file: contains information describing environmental conditions and roadway characteristics at the time of the crash. It includes information such as the time the crash occurred, the manner of collision and speed limit of the facility that the crash occurred.

Vehicle/Driver file: contains information describing the vehicles involved in the crash and their drivers. It includes information such as the model/make of the vehicle, model year of the vehicle, driver maneuvered to avoid, and driver distracted by.

Person file: contains general information describing all persons involved in the crash: drivers, passengers, pedestrians, pedalcyclists and non-motorists. It includes information such as age, sex and injury severity.

2.2.2 Fatality Analysis Reporting System

The Fatality Analysis Reporting System (FARS) was conceived and developed by the National Center of Statistics and Analysis (NCSA) in 1975. It contains data on all fatal traffic crashes within the 50 U.S States, the District of Columbia and Puerto Rico. FARS was developed in order to assist traffic safety professionals in identifying traffic safety problems and evaluating motor vehicle safety standards and highway safety initiatives. The so-called FARS forms are completed by each State. Such forms include the following information:

- Police Accident Reports (PARS)
- State driver licensing files
- State vehicle registration files
- State Highway Department data
- Vital Statistics

- Death Certificates

- Hospital/Coroner/Medical reports

In order for a crash to be included in FARS, it must involve a motor vehicle travelling on a traffic way and result in the death of a person (either an occupant of a vehicle or a non-motorist) within 30 days of the crash. Detailed descriptions of each fatal crash reported are included in FARS files. For each case more than 100 coded data elements that characterize the crash are included. These include elements that characterize the crash, the people involved and the vehicles. In greater detail, the data elements are reported on four forms as follows:

- Accident form: contains the time and location of the crash, first harmful event and the number of vehicles and people involved.

- Vehicle and driver forms: contains data for the vehicle type, most harmful event, drivers' license status and initial and principal impact points.

- Person form: contains data on each person involved in the crash, such as age, gender, role in the crash, injury severity and restraint use.

Finally, FARS incorporates a variety of information on alcohol related crashes with data such as overall crash alcohol estimates and driver and non-occupant Blood Alcohol Estimates (BAC). All FARS files and data are available through the so-called FARS Query System, which is fully available on the World Wide Web. Data starting from the year 1994 are available through this system where the user can choose a combination of variables that are required for the development of potential tables or the production of results with fatality injury related crashes.

2.2.3 Highway Safety Information System

The Highway Safety Information System (HSIS) was developed by the Federal Highway Administration (FHWA). The FHWA proceeded in the development of this database due to the need for a database that would serve as a tool to assist highway engineers and administrators in the decision-making process. The need for an understanding of how safety is affected by the geometric design of the roadway, the use of traffic control measures and the size and performance capabilities of the vehicles led to the development

of the HSIS. As of today, eight states are included in the HSIS including: Illinois, Maine, Michigan, Minnesota, Utah, California, North Carolina and Washington. The last three states were added in 1995 to provide better geographic coverage and increase the amount of data whereas the first five were chosen in 1987 when the HSIS was initially developed. All the above-mentioned states collect roadway-based data that provide quality data on a large number of accident, roadway and traffic variables. The selected states maintain basic crash files, roadway inventory and traffic files. The analyst can decide which of these three files are more useful to study a particular problem. These three files include a variety of subset files such as crash, roadway inventory, traffic volume, intersection, curve and grade and interchange/ramp. In brief, some of the data files include the following information:

- Crash: contains type of accident, vehicle types, sex and age of occupants, accident severity and weather conditions.
- Roadway Inventory: contains information for types of roadway, number of lanes, lane width, rural urban designation and functional classification.
- Traffic Volume: contains Annual Average Daily Traffic (AADT) data.
- Intersection: contains traffic control type, intersection type, signal phasing and turn lanes.

The above data can be used to analyze many safety problems. Modeling efforts to attempt to predict future accidents through roadway and traffic factors can be one application where the HSIS can be very beneficial. In terms of the extraction of data, the Statistical Analysis System (SAS) format is used, as is the case in the GES Database.

Only police reported accident data on the state-maintained highway system are included in the HSIS. The policies of each state and not its size determine the size of the state-maintained system. Table 2-1 below provides an indication of the quantity of data available.

Table 2-1 HSIS quantity of data available

State	Average Crashes/Year	Roadway Mileage
California	45,000	7,000
Illinois	150,000	16,000
Maine	39,000	22,000
Michigan	140,000	9,600
Minnesota	85,000	49,600
North Carolina	118,000	31,000
Utah	50,000	12,900
Washington	34,000	8,600

The HSIS was extensively used in order to study a variety of problems in the traffic safety area. Recently published research referred to the Investigation of Highway System Roadways in the HSIS States (Zegeer et al.) and the Application of an Improved Accident Analysis Method for Highway Safety Evaluations (Wang). Also, HSIS data were used for a study for the Magnitude and Severity of Passing Accidents on Two-Lane Rural Roads (HSIS, 1994). The above studies showed that while the HSIS database contains only eight of the total States in the Nation, the data are of high quality and the accident and roadway files can be linked. The Investigation of Crash Rates in the Eight Participating States, by Roadway Class (Zegeer et al.) will be described in more detail in the fourth section regarding the literature for research on crash rates using accident databases.

2.2.4 Crashworthiness Data System

The Crashworthiness Data System (CDS) was developed by the National Automotive Sampling System (NASS) in 1979 through NHTSA, an agency of the U.S. Department of Transportation. A random sample of thousands of minor, serious and fatal crashes is selected for study. Twenty-four field research teams study nearly 5,000 crashes a year involving passenger cars, light trucks, utility vehicles and vans each year. Research teams further examine each particular case in order to obtain data and evidence of skid marks, fluid spills, bent guardrails and broken glass. The crash damage is then measured and the teams photograph the vehicles. Interviews with crash victims follow and the nature and severity of injuries is determined.

All the collected data by the research teams become permanent NASS reports. CDS is then used by the NHTSA for a variety of purposes such as the identification of potential preexisting traffic safety problems. The data are also used for the evaluation of vehicle safety systems and designs. Another important area where the CDS is very useful is to examine the relationship between the type and seriousness of a crash and the resultant injuries.

2.2.5 State Data System

The State Data System was developed by the National Highway Traffic Safety Administration in the early 1980's. The data system includes accident data from seventeen States. Each State has its own police crash report format, incorporating different data elements. The State Data System incorporates data files similar in format to the General Estimates System file format. The following three files are included:

- Crash file: includes general crash characteristics describing the environmental and roadway conditions at the time of the crash.
- Vehicle file: contains information describing the vehicles involved in the crash.
- Person file: contains information describing the characteristics of the people involved in the crash: drivers, passengers, pedestrians and pedalcyclists.

The State Data System has been used in the investigation of the relative risk of subject and peer vehicles that support NHTSA's defect investigation program. Also, the data system was used in comparisons with other databases such as the GES and FARS.

2.3 Highway Statistics

Due to the fact that accident rates need to be developed in the context of this thesis, some literature review was performed. By extracting crash frequencies from the databases, one needs the denominator in order to develop crash rates (i.e. crashes per vehicle miles traveled). The highway statistics comprise a powerful tool for developing accident rates since they provide valuable measures in terms of highway mileage and the number of

vehicle miles traveled (vmt). The Federal Highway Administration (FHWA) and the Office of Highway Management publish the highway statistics publication once a year. Selected statistical tabulations are included in this publication, such as information on highway use and finance and the extent, characteristics and performance of the public highways, roads and streets in the Nation. Highway mileage and performance is extensively described in this publication. Most of the data are divided into urban and rural tables according to population and Federal-aid legislation definition and they are presented primarily on a State-by-State basis.

2.4 Development of accident rates

Many researchers have attempted to find the variables most highly associated with crashes. Bernardo and Ivan (1997) studied the prediction of the number of crashes versus the crash rate using Poisson regression. This type of regression was used to model both the crash and crash rate. Small data sets for several intersections were used for this study. Several ways of modeling highway safety were investigated, including different representations of traffic exposure and intersection effects as independent variables. Bernardo and Ivan (1997) suggested that the Poisson distribution allows for the relationship between exposure and crashes to be more accurately modeled as opposed to the linear relationship assumed in crash rate prediction. However, this study was only focused on intersections thus being site specific.

Zegeer et al. developed motor vehicle crash rates by crash type and roadway class in eight states: California, Illinois, Maine, Michigan, Minnesota, North Carolina, Utah and Washington. The accident data were extracted from the Highway Safety Information System (HSIS). The most important variables for calculating the crash rates were the urban or rural code, the functional roadway class, number of lanes and divided versus undivided roads. Eight different roadway classes were used as follows:
- Urban freeways
- Urban two-lane highways
- Urban multi-lane divided non-freeways

- Urban multi-lane undivided non-freeways

- Rural freeways

- Rural two-lane highways

- Rural multi-lane divided non-freeways

- Rural multi-lane undivided non-freeways

The results of this study showed that the most common accident type in most states was the rear end/same direction sideswipe, with angle and turning accidents as second. In terms of crash rates produced, these were lower on freeways than any other roadway class. The study showed significant variation in the crash rates, even within common roadway classes. This was justified by the differences in reporting procedures, the nature of the highway system, driving populations and other factors that varied from state-to-state. Severity, light and surface conditions and collision types were also considered, yielding very similar results for all eight states. Zegeer et al. concluded that the results from their study were considered reasonable with the crash rates to be used as a baseline in order to better understand crash trends. Knowing this kind of information can be used to improve safety for passengers and drivers by and upgrading existing highways and changing the design of new highways.

In another study, Mohamedshah and Kohls (1994) developed accident rates using the Highway Safety Information System database in order to use them for the development of the so-called Interactive Highway Safety Design Model (IHSDM). An accident prediction model was developed to produce average accident rates for different highway accident types. The objective of the study was to determine if the HSIS data could be used to develop the accident prediction model. Two states referred to as state A and B from the HSIS, were selected for the analysis. Three years of accident data were used and only accidents that occurred on interstates, state routes and U.S. routes were considered. The results of the study showed that data from a database such as the HSIS could be used to develop the accident prediction model for different roadway types. Mohamedshah and Kohls suggested that the development of average accident rates require judicious manipulation of the data and sound engineering judgement.

Zhou and Sisiopiku (1997) developed accident rates and examined their relationship with volume-to-capacity ratios. Data from Interstate I-94 in Detroit, Michigan were used to examine this relationship. Particular emphasis was given on the development of models to explain the differences between accident rates during weekends and weekdays, rear-end accidents and fixed-object collisions and property damage only accidents versus accidents involving injury and fatality. Zhou and Sissiopiku concluded that the accident rates were highest in the very low hourly volume-to-capacity (v/c) range and decreased rapidly when the v/c ratio increased. Then the rates gradually increased as the v/c ratio continued to increase. This very important as in the following chapters it will be shown that this is consistent with the results obtained in this thesis.

The final outcome of this study was that the correlation between accident rates and the volume-to-capacity values followed a general U-shaped pattern. Regarding the property damage only accident rates, these had a general U-shaped relationship with v/c values. On the contrary the rates that involved injury and fatality decreased as the v/c ratio decreased.

2.5 Accident prediction models

Vogt and Bared (1997) developed accident models for two-lane rural segments and intersections. Advanced statistical techniques were used in this study, as well as extensive accident and roadway data. Despite the fact that these models had as a primary objective the effect of design elements such models refer to a roadway class, two-lane rural roads, that has a high accident rate. The Highway Safety Information System (HSIS) was used in this study as the main source of data. Models were focused on segments and intersections. Negative binomial and Poisson models were considered for both cases. Vogt and Bared (1997) concluded that the data used for these models offered reasonable representations of the effects of highway variables on accidents. The State accident statistics for a given highway or intersection class can be used to obtain proportions of accidents by severity types. However, the Poisson, negative binomial and logistic models were used to model severities without significant results.

Kulmala (1995) studied the safety of main road junctions with the help of accident prediction models. Separate models were produced for 915 T-intersections and 847 four-way intersections. The goal of this study was to estimate the expected value of the number of accidents using Poisson or negative binomial distributions were assumed. Also, several models were made for all injury accidents involving motor vehicles only, single accidents and rear-end accidents. Results of the study showed that the accident risk increased as the traffic share on the minor road increased. Ten accident categories were used in the study as shown in Table 2-2 below. These categories included several subdivisions that made the total number of accident types studied almost 100.

Table 2-2 Accident Categories (Kulmala, 1995)

Accident Categories
Same driving direction (no turning vehicles)
Same driving direction (one turning vehicle)
Opposite driving directions (meeting accident)
Opposite driving directions (turning vehicle)
Intersecting driving directions (turning vehicle)
Pedestrian accident (on ped. crossing)
Pedestrian Accident (away from ped. crossing)
Off Road accident
Other Accident

Lack of good exposure data and the large quantity of data needed were the major disadvantages in completing this study. However, the accident prediction models provided reasonable results according to the author. A final conclusion was that the minor road at a T-intersection or a four-way intersection had a crucial effect on the safety of a junction due to the fact that the drivers approach the intersection at a higher speed, hence increasing the accident risk too.

Persaud and Musci (1995) used hourly traffic volumes in regression models for estimating the accident potential on two-lane rural roads. They used data from Ontario, Canada and used different combinations of time periods and geometric characteristics. Single vehicle accidents were particularly studied and the models showed that the accident potential was higher during the night. On the other hand, for multi-vehicle

accidents the accident potential was higher during the day. The study also emphasized the importance in differentiating between single or multi-vehicle accidents and day or night conditions.

2.6 Summary of Literature Review

The literature review presented in this chapter provided some basic background information for the topic of traffic safety and issues related with accident databases, models and rates.

Most of the research papers on the development of the accident rates concentrate on the development of rates using field data or a State accident database, i.e. the exposure level is minimized and refers to a specific area only. Another drawback is the fact that not all accident types or facility types are examined thus providing a variety of conclusions and results for those particular studies.

It must be also emphasized that the models that have been developed are not sensitive to the level of detail required to evaluate Intelligent Transportation Systems. Details such as level of congestion and travel efficiency are not considered. In addition to that, other studies are limited in scope and are not sensitive to factors impacted by ITS.

Accident prediction models examined in the review mostly relate to a specific site, either an intersection or sections of two-lane rural roads. This fact enabled the researchers in developing the models using field data and controllable variables. The issue of traffic volume as a key factor of developing accident rates was particularly examined by Persaud and Musci (1995).

The variety of accident databases, both in local and national level, were also presented and described in brief. These databases serve as a benchmark for the development of many studies related to accident prediction models and the development of accident rates. Due to the fact that most of the databases do not provide information regarding the

annual vehicle miles traveled on each facility type, the Highway Statistics are mostly used for the development of the accident rates.

Finally, after extensive research, there was not any available research regarding the impact of traffic signal coordination on traffic safety. As this is a relatively new topic, only some summarized information was presented in this literature review.