Crash Outcomes Data Evaluation System: An Evaluation of Medical Crash Costs

FINAL REPORT

Sponsored by the Iowa Department of Public Health

May 1999

CTRE

Center for Transportation Research and Education

IOWA STATE UNIVERSITY

The opinions, findings, and conclusions expressed in this publication are those of the authors and not necessarily those of the Iowa Department of Transportation.

CTRE's mission is to develop and implement innovative methods, materials, and technologies for improving transportation efficiency, safety, and reliability while improving the learning environment of students, faculty, and staff in transportation-related fields.

Crash Outcomes Data Evaluation System: An Evaluation of Medical Crash Costs

FINAL REPORT

Principal Investigator
Reginald R. Souleyrette
Associate Professor of Civil and Construction Engineering
Associate Director for Transportation Planning and Information Systems
Center for Transportation Research and Education

Graduate Research Assistant
Brad Estochen

Sponsored by the Iowa Department of Public Health

Center for Transportation Research and Education
Iowa State University
2901 S. Loop Drive, Suite 3100
Ames, Iowa 50011
Telephone: 515-294-5453
Fax: 515-294-0467
http://www.ctre.iastate.edu

May 1999

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Current methods of identifying high vehicular-crash locations attempt to quantify impacts. Dollar values are placed on crash-related intangibles to allow for commensurate analysis. Total costs are then used to identify deficient highway segments or intersections. However, the process used to determine cost estimates for damage, injuries, and fatalities is suspect.

By combining highway crash and medical records this study evaluates National and Iowa medical cost-component estimates. The study makes use of NHTSA's Crash Outcome Data Evaluation System (CODES). The use of CODES data results in estimates of crash cost which are significantly different from those used by FHWA and the Iowa DOT.

The sensitivity of locating high crash areas to improved estimates is explored. It is demonstrated that, the identification and ranking process currently used by the Iowa DOT is relatively insensitive to medical costs. That is, even though medical cost represent a significant and difficult-to-quantify portion of total crash cost, changing them results in only minor differences in location rank. Although mitigation prioritization is not found to receive a great benefit from the use of CODES, the system does show promise for other applications, e.g., determining appropriate societal crash costs resulting from various safety improvements.

CHAPTER 1 INTRODUCTION

In 1997 over 6.7 million motor vehicle crashes occurred in the United States. It is widely recognized that the occurrence of crashes results from the complex interaction among a driver, vehicle, and roadway (1). Crash data are critical elements of the evaluation of motor vehicle safety standards and the identification and mitigation of traffic safety problems, particularly at high crash locations (2).

Current methods of identifying high crash locations attempt to quantify the impact of highway crashes. Dollar values are placed on intangibles to allow for commensurate analysis. The total costs are then used to determine highway segments or intersections in need of improvement.

Though this approach is ingrained in practice, financial estimates of damage, injuries, and fatalities for a given location are prone to errors. These errors bias total cost estimates. Injury classifications that police officers use to code injuries on crash reports may not reflect the actual severity of injuries. For example, major injuries range from a broken arm to a severed spinal cord. Additional errors are introduced into the process as default costs are assigned to crashes fitting vague injury classifications. Clearly, incorporating actual medical costs from hospital records would improve accuracy of crash costs estimates. In turn, these new estimates may improve the identification of high crash locations.

1.1 Research Objectives

Linking hospital records with crash data presents an opportunity to improve the estimates of the crash costs used for safety analysis. This research uses Iowa's Crash

Outcome Data Evaluation System (CODES) to address the following four research objectives:

- To determine the average motor vehicle crash costs for fatalities and injuries occurring in the state of Iowa by using Iowa's CODES database.
- To determine the crash costs of specific types of crashes used in performing safety analysis on Iowa's roadways. These specific types include the kind of collision and the speed limit of the roadway that an injury crash has occurred on.
- To compare the actual medical costs with those recommended by the Federal Highway Administration (FHWA) and Iowa Department of Transportation (DOT) and determine if there is a systematic bias to the medical cost assignments.
- To evaluate the impact on Iowa's high crash location process of using new estimates of medical crash costs. This evaluation will determine what impact, if any, the new injury cost values have on the identification of unsafe roadways.

The National Highway Traffic Safety Administration initiated the CODES program in seven states to determine the effectiveness of seatbelt use nationwide. Currently 17 states have linked, or are in the process of linking, the statewide medical files with statewide crash records. This research utilizes medical crash costs derived from Iowa's CODES and traffic crash record files and includes all crashes on all Iowa roadways.

By using probabilistic linkage the medical and hospital records can be combined allowing evaluation of a new data source. The CODES data set is used to address the research objective previously presented. Using Spearman's correlation coefficient the ranking of the high crash locations can be examined with respect to new crash costs.

Findings of this research conclude that the medical estimates used by the Iowa DOT and FHWA are significantly different from the medical charges resulting from motor vehicle crashes in Iowa. However, the impact of changing the crash cost estimates does not significantly effect the overall ranking of the high crash locations in Iowa.

1.2 Report Organization

This report is presented in five chapters. Chapter 1, Introduction, outlines the objective and scope of the research. Chapter 2, Literature Review, summarizes current published research in the areas of crash analysis and economic costs. Chapter 3, Methodology, describes the approach used to determine the medical costs of motor vehicle crashes in Iowa and provides the analysis procedure based on the CODES data. Chapter 4, Analysis and Results, documents the results of the methodology. Finally, Chapter 5, Conclusion, summarizes the research and introduces future research possibilities in the subject area.

CHAPTER 2 LITERATURE REVIEW

The purpose of this research is to use information derived from joining crash records with medical databases to address the four research objectives presented in Chapter 1. This chapter describes the current practice used to identify high crash locations and determine what locations should be improved. Also included is a brief explanation on how the economic values assigned to injuries and fatalities are ascertained. This information is the basis for determining the traffic crash costs presented later in this chapter. Finally, the Crash Outcome Data Evaluation System (CODES), a national effort to combine crash and hospital records, is briefly explained.

2.1 The Determination of High Crash Locations

Identifying deficient locations on a roadway, i.e., those experiencing an above average number of motor vehicle crashes, is a goal of transportation safety engineers (3). Performing this identification allows examination of possible engineering, enforcement, and educational measures to improve safety at these locations (4). Three approaches are used in the identification of such areas: crash frequency, crash rate, and a combination of the crash frequency and crash rate approaches (5). All are discussed in detail.

2.1.1 Crash Frequency Identification

The crash frequency approach assesses the number of crashes at several locations within a system (e.g., intersections and links) to determine high crash locations. The average number of crashes per year over a specified time period is determined for the section of interest. Next, these locations are compared to the average number of crashes for similar types of roadways. An average crash frequency is determined and a confidence interval can then be calculated. If a location experiences more crashes than the maximum value in the confidence interval, it is considered to be a high crash location (5).

For example, Table 2-1 shows that Location 1 has the highest number of crashes, and therefore should be a higher priority than the other eight locations. It should be noted, however, that this approach does not address all areas equally. For example, areas that have high volumes of traffic naturally experience a higher number of motor vehicle crashes.

Table 2-1 Crash Statistics Table

Location	Number of Crashes	Crash Rate (Per million Vehicle Miles)
1	82	5.2
2	58	6.3
3	9	1.8
4	45	7.1
5	72	4.8
6	29	7.2
7	18	2.4
8	53	3.3
9	3	9.8

2.1.2 Crash Rate Identification

To avoid the bias in relation to high volume roadways crash rates can be used to compare each location on the roadway network. This calculation uses the total number of crashes, the average annual daily traffic (AADT), and the length of roadway to determine the average crash rate for each type of roadway. Crash rates are examined for sections, spots (e.g., sections that have a length of less than 0.1 miles), and intersections on the roadway network. The crash rate is determined for each location of interest and is then compared to an average crash rate. After determining the individual rates for the subject areas, the crash rate is determined for every similar roadway in the network.

Finally, a confidence interval for the crash rate can be determined using statistics. A confidence interval can be constructed at the desired confidence level. The upper limit of this confidence region is the critical crash rate. Once the critical rate has been established, each section is compared to this value. If the crash rate exceeds the critical rate the location is flagged as a high crash location.

While this approach uses statistical significance to determine high crash locations, the analyst may obtain misleading results. In Table 2-1, for example, location 9 has the highest rate; however, it also has the lowest frequency of crashes. Using this approach for low volume roadways may yield a high crash rate; however, safety improvement funds may be better spent in other locations (6). For example, if two property damage crashes (resulting in $8,000 damage) occurred on a road with an AADT of 100 the resulting crash rate would be 0.02. If a $50,000 safety improvement is 100 percent effective two crashes would be eliminated. Eliminating $8,000 in crashes with a safety improvement cost $50,000 is not an effective use of safety funds.

2.1.3 Combination Approach - Crash Rate-Frequency Method

As seen above, there are inherent problems with both the crash frequency and crash rate approach. A phone/email survey of 12 state DOTs high crash identification process determined that a two-stage combination approach was used. This is an attempt to eliminate the problems that can occur from using only one method to identify high crash locations.

The first stage of the combination approach uses the crash frequency. Next, locations meeting exceeding a frequency limit are evaluated using the crash rate analysis method. Though slight variations are used throughout the nation, the general approach is to select locations that meet a minimum threshold of crashes. In some states analysts consider the injury severity of crashes, i.e. they may choose all locations that have more than ten crashes or at least one fatality per year. This eliminates the low volume roadway that may inherently have a high crash rate unless it meets the criteria.

2.1.4 Variations and Extensions of Current Approach

The previous processes appear to be the standard for identifying high crash locations in the states surveyed. However, alternative approaches of identifying high crash locations continue to be investigated. Zeeger noted that proper crash analysis requires additional information beyond the typical traffic crash data (7). Many states use only statewide crash data, which tends to focus solely on locations with a history of a high number of crashes, thus equating the identification of hazardous locations solely with crash history. However, analysts may wish to examine characteristics of crashes, such as crash severity, and focus on predicting future crashes rather than looking at historical trends.

Although the high crash identification process is well established there have been attempts to adjust its components. For instance, concerns have been expressed about the omission of crash severity in the processes previously mentioned. Therefore, Tamburri and Smith (8) introduced the concept of a safety index. This concept addresses the issue of sites where severe crashes deserve priority compared to locations with few injury crashes. Each crash location is assigned a characteristic combination of crashes according to the percentage of fatal, injury, and property damage only crashes. Additionally, they introduced using weights for various types of crashes. This attempts to weigh injury and fatal crashes by equating them to an equivalent property damage-only crash.

Jorgensen (9) introduced two ideas to the identification of high crash locations. First, he concluded that a multivariate model should be used to determine the expected number of crashes at a location rather than using crash rates. The multivariate model takes into consideration several variables used to predict the number of crashes, e.g., AADT and roadway width. This differs from the traditional view that the expected number of crashes is proportional to the crash rate. Second, he determined that the number of crashes determined by the multivariate model should be used instead of the average crash rate. In principal, this is equivalent to using the expected value of crashes instead of a simple crash rate.

Taylor and Thompson (10) proposed that a hazard index be created for each roadway section. This index would come from a compilation of the crash frequency, crash rate, severity, volume-to-capacity ratio, sight distance, conflicts, and erratic maneuvers.

McGuigan (11) proposed that the difference between the actual and expected number of crashes for each intersection be calculated. This quantifies the potential crash reduction for all locations of interest. This approach attempts to identify promising locations for improvements instead of identifying locations with elevated crash counts.

Flak and Barbaresso (12) suggested a cumulative listing of crash types to be used as a basis for comparison for similar sections of roadways. The idea is that some locations may exceed expected values for specific crash causes compared to other locations. For instance, two similar intersections may vary widely in the number of left-turning crashes. The analysis of crash rates for specific crash types can improve the identification of problem areas and causal factors. This deviates from traditional methods of determining high crash locations based upon crash totals across all crash types.

Hauer and Persaud (8,13,14) proposed additional enhancements to the high crash location process. These recommendations use an Empirical Bayes (EB) method, which takes into account prior information to predict future events. Hauer states that the EB approach can be used to assist in the identification of high crash locations. Output from this analysis suggests the probability that a section of road exceeds the critical value, thus identifying it as a high crash location. Persaud examined the impact of a limited accident history using the EB approach and compared the results with those found using traditional methods. His findings showed similar estimates to those found when using traditional methods.

2.2 Choosing Priority Locations for Improvement

Once the high crash locations have been identified by a transportation agency, corrective measures can be recommended. There is no standardized process to sift through several possible locations to identify a small subset of locations that can benefit most from safety improvements (5). There are a variety of approaches for determining which locations have priority for safety improvements, such as value lost and crash severity (5). Regardless of the approach, the safety improvements must be justified in economic terms (15,16).

Once high crash locations are identified by a transportation agency, an in-depth review of each site is performed to determine possible deficiencies based on the contributing circumstances obtained from the crash data. The transportation agency then recommends engineering, education, or enforcement solutions (17). Various studies have investigated what effect individual changes to the roadway environment may have on the crash frequency of a roadway. These studies have resulted in crash reduction factors that express the percent reduction in the number of crashes attributed to a specific roadway improvement (18,19).

Typically, crash reduction factors are used to predict the safety effect of a proposed treatment. For example, widening the shoulder of a roadway may result in a 10 percent reduction of crashes. This approach is preferred due to the simple concept and application of the crash reduction factor (20). To get the safety improvement the number of crashes is multiplied by the crash reduction factor, resulting in the number of crashes that can be eliminated.

Public investment decisions are generally evaluated using benefit-cost analysis (21). The benefits of roadway improvements and safety enhancements are compared to the cost of such improvements. (A benefit-cost ratio can be determined by comparing the costs of the improvements with the expected dollar value of injuries saved. If the ratio is above one, every dollar amount invested returns more than one dollar in benefits.) This decision process is used to determine which locations would experience increased benefits by reducing the number of crashes compared to the cost of implementing such changes. A major component used to determine the benefit-cost ratio is the dollar value assigned to various types of crashes involving property damage, injuries, and fatalities.

2.3 Determining Societal Costs of Crashes

A key component in the allocation of highway resources is determining the cost savings that result from a specific improvement. The savings estimates largely depend on the values placed on personal injuries and fatalities. Currently there are three approaches used to determine the value placed on human life and personal injuries: the direct cost approach, the human capital cost approach, and the comprehensive cost approach (also known as willingness to pay). There are two economic classifications of costs: direct and indirect.

These two types of costs are explained below, followed by a discussion of the three approaches used to quantify the economic costs of motor vehicle injuries and fatalities.

2.3.1 Direct and Indirect Costs

Direct costs are those that can be directly measured and assigned to a specific output or work activity. For example, the labor and material costs associated with a product, service, or construction activity are direct costs.

Indirect costs are those that are difficult to attribute or allocate to a specific output or work activity. This term is used for those cost elements closely linked to the support of operations. Attempting to allocate these costs based on a cause-effect basis to individual units of output involves effort to allocate directly to a specific output; therefore these costs are incurred indirectly as a result of an activity. Common tools, general supplies, and equipment maintenance usually cannot be assigned to only one specific task or project. Therefore, the costs incurred by these items are classified as indirect costs.

2.3.2 Three Methods of Motor Vehicle Financial Crash Cost Determination

Each of the three approaches to determine the financial costs of motor vehicle injuries differs in what direct and indirect costs are incorporated into a final cost estimate. The direct cost approach associated with traffic crashes includes costs directly resulting from motor vehicle crashes. These direct costs include hospital and medical treatment, insurance administration, workplace costs, emergency service costs, travel delay, legal expenses, and property damage. These costs are relatively easy to determine because they would not have existed without the incidence of a motor vehicle crash.

The human capital cost approach attempts to incorporate some indirect costs into the price component of motor vehicle crashes. The main indirect cost incorporated into the final cost estimate is the lost household production (e.g., productive activities in the home are lost due to injury or premature death) and wages (corresponding to the length of injury or lifetime earning potential) resulting from a traffic crash. This information is determined from average values for household production (based on the average cost per hour across functional classifications for household work) and wage rates.

The final approach to determine the traffic crash costs uses a comprehensive cost approach, also known as willingness to pay. Comprehensive costs combine direct costs and human capital costs with an estimate of pain and suffering. The willingness to pay concept attempts to determine the value of life by determining what people actually pay or would be willing to pay to reduce their probability of death. Two components create this cost concept: the value individuals place on safety and health, and the cost saved by the rest of society by preventing an injury or death. Unlike the direct and human capital costs, these results tend to be inconsistent and have a wide variability.

Determining the cost of pain and suffering is a difficult task; it involves placing a value on human feelings and pain. Several studies have produced various values based on this comprehensive concept. The cost values ranged from $50,000 to $15,000,000 (23).

The three approaches used to determine crash costs each have very different values. This is expected due to the nature of each approach. The costs are often derived from different bases or for different purposes, thus resulting in significantly different estimates. While direct cost and human capital cost approaches have sufficient information available to determine their estimates, there is no uniform approach to determine the value placed on losing a limb or mental anguish incorporated into the comprehensive cost estimate. There are several estimates available for use in benefit-cost analysis of roadways, but the underlying principal of the estimate should be examined before deciding what cost estimates to use.

2.4 Case Studies of Crash Cost Determination

Estimating the economic impact of motor vehicle crashes is not an exact process (23). There are several facets involved in obtaining a complete estimate of the economic impact. Several studies have attempted to quantify the financial impact of motor vehicle crashes (24,25,26,27,28,29); the FHWA and NHTSA have continually examined this subject area and have produced several publications explaining the task of determining an estimate of the financial cost of motor vehicle crashes. These costs represent the present value of lifetime costs for both reported and unreported crashes. The most recent estimate was determined from 40,676 fatalities, 5.2 million nonfatal injuries, and 27 million property damage crashes in 1994 (26).

2.4.1 FHWA Studies

Several sources of information were used to generate the FHWA's "economic crash cost estimates." These include the Fatality Analysis Reporting System (FARS), the Crashworthiness Data System (CDS), the General Estimate System (GES), the National Automotive Sampling System (NASS), and the National Health Interview Survey (NHIS). Additionally, injury estimates are provided to FHWA from various states.

There are several components of the financial costs of motor vehicle crashes, including costs associated with market production (lost wages), household production (cleaning and maintenance), workplace costs (lost productivity), legal issues, medical treatment, property damage, insurance, emergency response, premature funeral, vocational rehabilitation, and travel delay.

Incorporating information from the various databases provides an estimate of the overall financial traffic crash costs (16). The injury information is classified using the Maximum Abbreviated Injury Scale (MAIS), with values ranging from 1 to 5. The MAIS score indicates the potential threat to life, and thus requires medical judgement (25). MAIS scores are revised roughly once every five years based on medical outcome studies (27). Table 2-2 displays the meaning of each MAIS score.

Table 2-2 MAIS Score Definitions (17)

MAIS Score	Meaning	Examples
1	Minor Injury	Whiplash, bruise, broken tooth
2	Moderate Injury	Closed leg fracture, finger crush
3	Serious Injury	Open leg fracture, amputated arm, major nerve laceration
4	Severe Injury	Partial spinal cord severance, concussion (unconscious less than 24 hours)
5	Critical Injury	Complete spinal cord severance, concussion (unconscious more than 24 hours)

In contrast to the MAIS score, which measures the potential threat to life, the KABCO scale is used by police officers when investigating a motor vehicle crash to record the injury sustained in a motor vehicle crash. The KABCO injury designation was created for injury coding by police officers at the scene. It does not require the officer to make a medical judgement (27). Instead the officer assigns a K if a fatality occurred, an A for a incapacitating injury, a B for an evident injury, and a C if someone claims to be injured but seems uninjured (30). There are drawbacks to using the KABCO injury classification, specifically inconsistent reporting of injuries. An officer may code injuries as more severe if they include women or excess blood; also, the officer's experience has an impact on how the crash records get coded (18). Officers who rarely deal with motor vehicle crashes may tend to classify injuries as more severe compared to their counterparts who deal with crashes on a daily basis.

The financial costs determined by the FHWA in 1994 are broken down by the MAIS score. Significant sources of financial loss, such as medical costs, are highly dependent on the injury outcome and not the immediate threat to life (26). A comprehensive listing of the injury and fatality costs can be found in Table 2-3.

Table 2-3 Comprehensive Fatality and Injury Costs (17)

Injury Severity	1994 Values
MAIS 1	$10,840
MAIS 2	$133,700
MAIS 3	$472,290
MAIS 4	$1,193,860
MAIS 5	$2,509,310
Fatal	$2,854,500

Additionally, the financial costs can be broken down into crash cost categories. This is done for the fatal and injury crashes in Figure 2-1 and Figure 2-2. This allows an average value for each component to be estimated, based on the findings from the several databases used to determine the total costs. Medical costs are one component of the total costs, for instance, and are estimated to be 22.5 percent of the total costs. The medical costs for a MAIS 2 injury is $30,751 (22.5 percent of $133,700).

Due to the difference in the scales used by the enforcement and medical personnel, a conversion was created by the FHWA to convert from the MAIS scale to the KABCO scale when using crash costs for analysis purposes. A conversion table, as seen in Table 2-4, was created based on the NASS data from 1982 to 1985. This table is only valid when converting crash costs from MAIS scale to the KABCO scale, not vice versa (27).

2.4.2 Other Studies of Crash Costs

The FHWA and NHTSA are not the only organizations that offers an estimate of traffic crash costs. In 1993, a survey was conducted to determine the basis and use of motor vehicle crash costs in each state (23). All fifty states responded to the survey, 45 of which use crash costs in analysis. These 45 states provided three uses of the crash cost data, including performance of benefit-cost evaluations for highway improvements and upgrades (76%), analyzing particular safety related improvements or upgrades (69%), and evaluating design exceptions (22%).

At the time of the survey, the NHTSA publication, The Economic Cost of Motor Vehicle Crashes, 1994, was not yet released. Of the 45 states that use financial crash costs in analysis there were 11 diverse sources used to obtain the value of financial crash costs, as seen in Table 2-5.

All of these studies use estimates to produce national averages. Currently there is no nationwide incidence database organized by injury outcome (26). However, in 1991 the Intermodal Surface Transportation Efficiency Act (ISTEA) provided funds to the National

Highway Traffic Safety Administration (NHTSA) to examine the possibility of creating such a database.

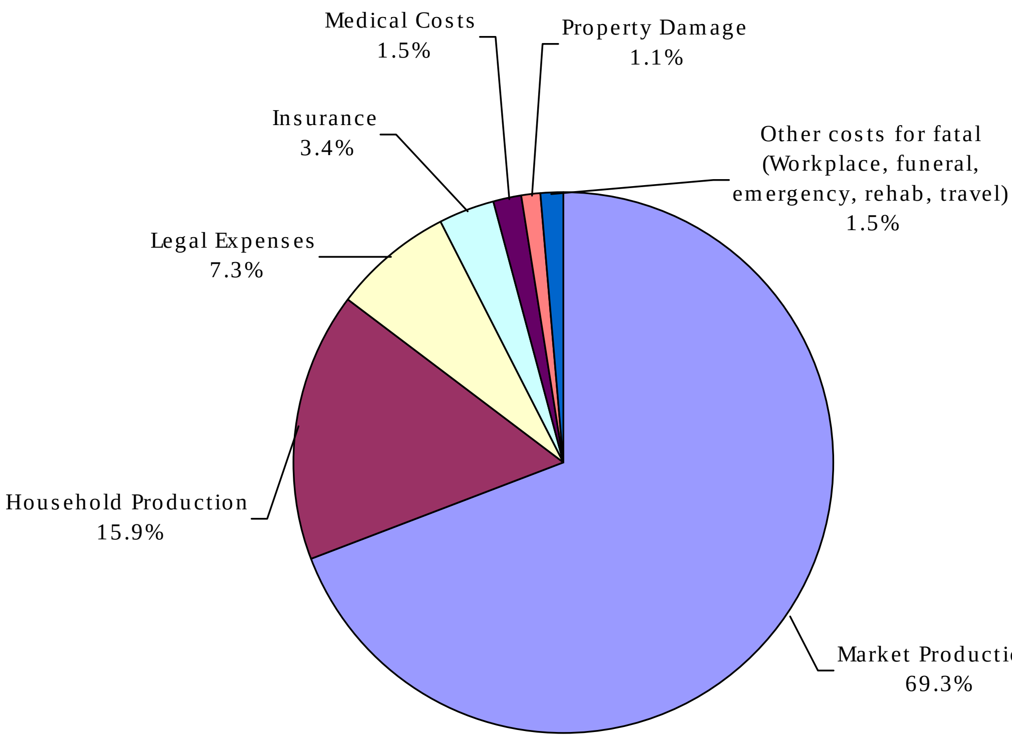

Figure 2-1 Financial Components of Fatal Traffic Crash Costs (26)

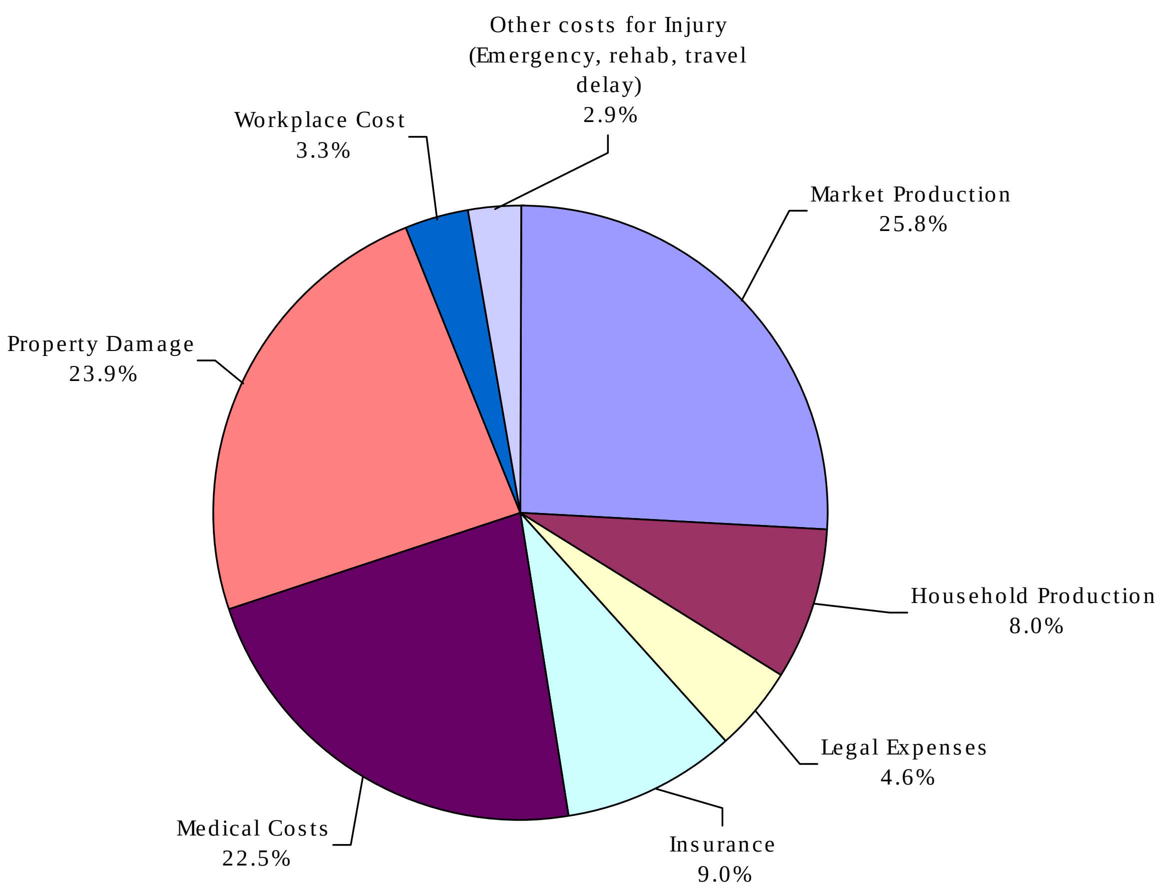

Figure 2-2 Financial Components of Injury Traffic Crash Costs (26)

Table 2-4 Conversion Table from MAIS to KABCO (27)

MAIS\KABCO	O	C	B	A	K
0	92.70%	20.50%	5.20%	1.50%	0%
1	7.00%	70.90%	78.80%	48.60%	0%
2	0.20%	7.00%	12.60%	28.00%	0%
3	0.03%	1.50%	3.10%	16.90%	0%
4	0.00%	0.06%	0.30%	2.80%	0%
5	0.00%	0.01%	0.10%	1.70%	0%
Fatal	0.0001%	0.013%	0.026%	0.50%	100%
All	100%	100%	100%	100%	100%

Table 2-5 States Source of Crash Costs (23)

Source	Number of respondents
National Safety Council Estimates	13
FHWA memorandum on “Motor Vehicle Accident Costs” T 7570.1 (June 30, 1988)	13
FHWA 1991 report “The Costs of Highway Crashes” FHWA-RD-91-055	7
Research by your agency	2
NHTSA 1983 Report “The Economic Cost to Society of Motor Vehicle Accidents” DOT HS 806 342	2
FHWA 1984 report “Alternative Approaches to Accident Cost Concepts” FHWA/RD-83/079	2
AASHTO 1988 publication “A Manual on User Benefits Analysis of Highway and Bus-Transit Improvements”	2
State insurance claims for injury costs and PDOs	1
Rollins and MacFarland, “Costs of Motor Vehicle Accidents,” Transportation Research Record 1086 (1986)	1
Kragh, Miller and Reinert (1986) “Accident Costs for Highway Safety Decision Making” Public Roads, Vol. 50 No.1, pp. 15-20.	1
Other	1

2.5 CODES – A New Approach

CODES is a coordinated approach to obtain medical and financial outcome information related to motor vehicle crashes for highway safety and injury control decision making (31). With the data integrated from several sources, answers to questions that have eluded researchers can now be answered. For instance, what do linked crash and injury data tell us about who is more at risk for serious accidents or high health care costs? One answer determined by the Wisconsin CODES team is that younger and older drivers are at greater risk for high costs when crashes occur. What do linked data tell us about the impact of different behaviors on outcome? The Missouri CODES team identified higher ejection rates for pickup truck drivers who have a lower seat belt use rate. And, more generally, what improvements in data quality result from CODES? The Utah CODES team reported that 10 percent of motor vehicle occupants, designated as uninjured on the crash report, sought medical attention (32). Likewise, this information can be used to determine medical crash costs for injuries sustained in motor vehicle collisions.

Several agencies respond to motor vehicle crashes; unfortunately, not all of the information is documented in one location to provide a complete assessment of these events occurring (33). For instance, separate data files are created documenting: 1) the crash events, 2) injuries occurring, 3) emergency response units, and 4) hospital treatment until release from a doctor's care. Therefore, it is difficult to obtain all the information needed to completely understand the entire crash event and final outcome. For this reason NHTSA funded the implementation of a crash outcome database in seven states by linking their isolated databases, such as the crash and hospital information. In 1996, the CODES states successfully linked statewide crash reports with several injury-related databases, including hospital discharge, rehabilitation, death certificates, and insurance claim information. This linking process creates statewide comprehensive information concerning all parties involved in a motor vehicle crash, from the crash scene through the health care system.

Creating a linked database relies heavily on state data resources (e.g., medical and crash data) (34). The minimum requirements for linkage include a computerized, population-based, statewide crash and injury data set. These data should contain identifiers that can be used to categorize the people involved in the crash. Discriminating variables are used to identify the same event based on common information recorded by each data source. Without these variables, which are the backbone to the process, the linkage would not be possible. The cost of the linkage process depends on the quality and availability of the state data files.

Linking several databases together was a difficult task before the process was automated using computers. Ad-hoc methods were used to determine if several records in different databases correspond to a matched record. Traditionally, the linkage involved manual inspection of several variables in both databases and corresponding records were determined based on the number of identical variables. These ad-hoc procedures often involved several passes through large data sets, with little statistical justification, that were hard to duplicate (35). Current advances in software have created processes to link data together in a statistically warranted manner.

2.5.1 CODES Applications

Currently 19 states have linked, or are in the process of linking, crash records with statewide medical files. This creates a new database for a robust analysis of new topics unattainable using traditional traffic crash records. CODES-based studies have investigated many topics (36), including the evaluation of injury occurrence by age, sex, participant type, geographic locations, time and date, type of injury, and type of treatment for injuries. Health care costs have also been evaluated using this new information. Examples of health care estimates range from additional medical charges due to the lack of wearing a seat belt to examining the burden traffic crashes impose on the Medicaid program. Traffic safety has benefited from studies resulting from the CODES data set. This research is summarized by NHTSA's publication Revised Catalog of Types of Applications Implemented Using Linked State Data (36). Several studies listed below have examined relationships between medical costs and crash, vehicle, and personal characteristics, including:

- medical impact of striking fixed objects
- injury outcome by vehicle type
- restraint use and injury severity
- child safety seat effectiveness
- head injuries associated with motorcycle use
- medical impact of alcohol-related crashes
- injury outcome for crashes involving wildlife

Additionally, applications for emergency medical services and evaluation of data quality have been examined by various states (36).

Hawaii is a leader in using the CODES information for transportation safety analysis purposes. Karl Kim, a researcher from the University of Hawaii, has developed a crash database in a geographic information system (GIS), which allows additional analysis beyond traditional tabular analysis. Spatial patterns (37) and zonal generators (38) were examined to determine the spatial relationship between trip generating activities and motor vehicle crashes. These studies attempted to: 1) describe the spatial distribution of crash locations and determine how the types of crashes can be spatially differentiated, and 2) explain the spatial patterns using a number of variables to show that activities-generating trips also indirectly predict crashes.

Furthermore, several models relating to transportation safety have been created to explain relationships between various crash characteristics. Several subjects are addressed in these models, including relationships between:

- motor vehicle crash types and injuries sustained (39);
- injury severity, vehicle type, and restraint use (40);
- driver characteristics and injury severity (41);
- driver age and gender relating to motor vehicle crash involvement (42).

2.6 Summary of Literature Review

Determination of highway safety improvements is a complex task that requires precise information to complete. First, the high crash locations must be identified by using an approach that will identify unsafe sections of road. Next, the areas are reviewed to determine the proper engineering, enforcement, and educational plans to address these deficient locations.

Each improvement has a cost associated with the implementation. These costs are weighed against the benefits that will be obtained as a result of such safety improvements. Crash reduction factors are used to approximate the effect of safety improvements. These crash reduction factors estimate the number of injuries and deaths that can be prevented if a plan is implemented.

The number of injuries and fatalities prevented are benefits of a proposed safety improvement. Several studies have attempted to place a dollar value on the prevention of injuries; however, there is no uniform process of estimating the appropriate dollar value. Several databases are used to incorporate various direct and indirect cost components associated with human life. One of the direct cost components of this estimation is the medical cost incurred by motor vehicle crash victims. Currently there is no nationwide incident database that is organized by injury outcome.

NHTSA has promoted the linkage of medical and crash records to assist in creating a statewide crash outcome database. This database contains information from the crash scene as well as information obtained as injured parties pass through the medical system. This provides information to address questions previously unanswerable, including precise medical costs incurred by motor vehicle crash victims. This approach may facilitate greater understanding into safety evaluations, allowing efforts to address these shortcomings.

Several states currently have linked databases that are used to enhance the safety analysis. Models have been developed using the new information to address several components of motor vehicle crashes. However, few studies evaluating the medical cost estimates have been used to evaluate injury crashes. This presents an opportunity to evaluate the medical crash costs associated with motor vehicle collisions. Once the approach is determined it could be transferred to other states, providing them with a process to evaluate the medical crash costs associated with safety analysis in their respective state.

CHAPTER 3 METHODOLOGY

This research examines medical crash costs based on the CODES project developed using the state of Iowa's traffic crash records file, maintained by the Iowa DOT, and the statewide medical information records collected by the Iowa Department of Public Health.

The methodology presented in this chapter is specific to the data available in the state of Iowa after 1996. This constraint exists because several conditions must be met to proceed with the probabilistic linkage technique. One condition not met in Iowa before 1996 was the existence of a statewide medical database. However, upon completion of such requirements this process is readily transferable to Iowa, as well as other states, to evaluate the crash costs used, if any, for identifying high crash locations. Furthermore, other states' approaches to determining the high crash locations must be adapted as Iowa possesses a unique approach to determine the high crash locations within its jurisdiction.

First, this chapter explains a probabilistic matching procedure used to link crash and medical records. This process is responsible for linking the crash and medical records. Second, the geographic information system-based crash analysis tool currently being developed for the Iowa DOT is briefly discussed. Third, the current high crash location process used in Iowa will be explained. Fourth, the process for determining the medical crash costs is presented, as well as how these values will be used in the analysis section. Fifth, the approach used to analyze the high-crash location process is illustrated. Finally, an all-encompassing overview is presented to clearly depict the approach to satisfying the four research objectives outlined in Chapter 1: 1) determine the medical crash costs for Iowa crashes based on a probabilistic linkage approach, 2) determine the medical crash costs for specific types of crashes of interest to safety analysts, 3) compare the medical costs from the Iowa data to the national average estimates published by the Federal Highway Administration, and 4) evaluate the impact of Iowa's high crash location process with respect to these new estimates.

3.1 Probabilistic Linkage

The use of new computer applications makes it easier to link the two data sets together. It can be a straightforward process as long as two stipulations are satisfied: the data sources must have comparable units of analysis, and each data source must contain unique identifiers to discriminate among an event or person. However, several issues must be addressed before attempting probabilistic linkage on the databases.

Background and technical information pertaining to the probabilistic linkage technique came from two sources: Jaro (35) and NHTSA/National Association of Governors' Highway Safety Representatives (34) publications. The following is a brief summary presented from these publications.

3.1.1 Characterizing Data Sources

The first step of the probabilistic linkage process is to identify the data resources available for linking. When a traffic crash results in an injury, several different organizations are involved until the final release from medical care. Each of these resources should be examined (34). A typical motor vehicle crash resulting in an injury may involve emergency medical service (EMS), emergency department resources, and hospital personnel. Figure 3-1 creates an exhaustive list of possible medical care providers for victims of a motor vehicle crash.

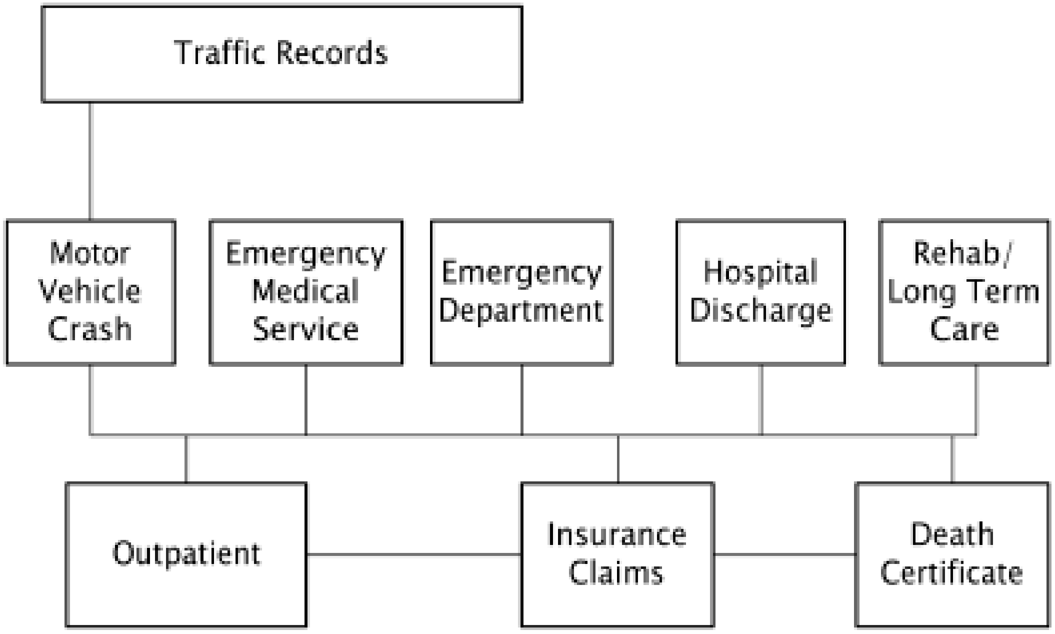

Figure 3-1 Possible Medical Care Documentation for Traffic Crash Victims (34)

The traffic record contains non-medical information only; however, other data sources contain medical and financial information. While the traffic crash records contain little medical information, linking the crash records with the medical records allows financial outcome information and other medical measures to be analyzed.

Non-medical data sources are obtained from the police crash report. However, due to a variety of reporting thresholds used throughout the nation, not all crashes are recorded (22,43). A measure of injury is usually available through the KABCO scale. As mentioned earlier, law enforcement officials document injuries according to severity rather than potential threat to life.

Other non-medical information is available to enhance the analysis of traffic crashes. One such example is the driver's license file, which contains driver-specific data including the driver's license number, date of birth, social security number, and a history of convictions (if any).

Medical information can come from a variety of sources, including EMS reports, trauma units, hospital discharge and rehabilitation records, long-term health care (nursing home) information, and outpatient data. These sources are briefly discussed below.

3.1.1.1 EMS Records

EMS records include information about individuals who are treated and transported to a hospital by emergency medical service staff. If victims are transported via ambulance, the EMS record is the first medical documentation completed for persons with injuries sustained in a motor vehicle crash. There may be several EMS reports for each individual due to multiple methods of transportation (e.g., ambulance to local hospital, then on to regional trauma center).

3.1.1.2 Emergency Department Data

Emergency department data are first recorded in the emergency department and then transferred to the patient's medical record when services are complete. Usually emergency department data does not contain information on service costs, as the patients do not receive medical bills from each entity of the hospital. The emergency department may have more than one record for a single person due to multiple treatments of the same injury. For individuals not transported by EMS, the emergency department is the source of treatments and diagnoses.

3.1.1.3 Hospital Discharge Data

Hospital discharge data represents a rich source of information that is not available from crash records. Discharge records contain medical information on the entire hospital stay. These data have been standardized for reporting to the Health Care Financing Administration for Medicare and Medicaid payments. Information provided includes: patient, hospital, and provider identifiers, procedures and diagnoses, disposition, injury severity score (ISS), and medical charges. Inpatient data are usually collected by licensed medical records technicians and are the basis for payment information. This results in a highly reliable data source. However, access to this information may be restricted by legislation and/or be expensive to obtain.

3.1.1.4 Long Term Care Data

Long-term care data are collected by each individual agency for Medicare and Medicaid payments. Computerized and standardized records usually do not exist at the statewide level and contain little information to assess the impact of traffic crashes. Since it focuses largely on the patient’s level of impairment and vital information, they provide little information useful to assess traffic crash injuries.

3.1.1.5 Outpatient Data and Other Sources of Information

Outpatient data are usually collected and maintained by each individual organization and are not considered part of the statewide primary medical care information. Additional information may be available from hospitals and government agencies, including the Fatality

Analysis Reporting System and trauma registries. This information is considered a subset of a larger data set and provides little additional data on injuries.

3.1.1.6 Death Certificates

The death certificate contains information on medical cause, time, location, and contributing injuries for all deaths. Standardized codes are used to document causes of death. However, they do not address medical condition before death. Death certificates are computerized according to a uniform national standard.

3.1.1.7 Insurance Claim Data

Although medical insurance data seem to promise a robust source of information, only limited medical information is created through the claims process. Treatment and payment information may be available to those who use Medicare and Medicaid for their primary insurance, while others may have private insurance companies to pay any medical expenses incurred. The major advantage of claims data is that the outpatient and inpatient medical information may appear in the same file. However, no detailed medical information is provided, only the costs associated with the care provided. Also, these costs can be suspect because the amount paid by an insurance company may not reflect the charges applied by the medical facility. Additionally, an individual may have several insurance providers and each company will have its own record of this event. Due to the nature of the health care insurance industry, the data sets are usually incomplete or lagging due to the time required to reconcile the various medical charges assessed by the providers.

3.1.2 Summary of Medical Data

As presented in the previous sections, there are a variety of places to obtain medical information. Medical data can be tied to one or more sources of information; therefore, care must be used when determining which elements of the database to use for analysis. Additionally, sufficient documentation concerning the data must be available for all parties involved.

A large amount of information in these data sets is confidential, and may have legislation governing their use. For this reason, an advisory committee is required by NHTSA for each state attempting to link this information and includes representatives responsible for each data set being linked together. For example, individuals involved with crash reporting and EMS data reporting should be members of this advisory committee. New technologies allow data access while protecting patient confidentiality. Specifically, several patient identifiers are released to facilitate the data linkage process and are subsequently removed from the final data set.

To ensure the highest success rate for the data linkage process, the data should be edited and standardized. This process includes a manual inspection of various fields within each data set to determine if there are any obvious data entry errors. Data element definitions must agree in both data sets. For example, in the gender option for the files, if 1 represents "male" then all other files should use the same convention. This approach increases the number of records linked by ensuring a clean data set.

3.2 Requirements for Data Linkage

The objective of this process is to identify two records, in separate files, that correspond to the same individual. Both files must contain corresponding fields to initiate the linkage process. For example, in order to match on Social Security Number and age, both files must contain this information.

The process of linking two files can be a time consuming task. All of the possible record pairs could be created representing every possible combination of the data values in each table. If table 1 contained *n* number of records and table 2 contained *m* number of records the potential resulting pairs would be the product of the two number of records (*n* x *m*) number of possible pairs.

The objective of the record linkage process is to classify each pair as belonging to one of two sets (35). The first of these two sets is the set of matched record pairs (M). The second set is the set of unmatched pairs (U). Any combination of two records from the respective database must be assigned to the U or the M group.

3.2.1 The "Blocking" Concept

Attempting to combine two small databases containing 1,000 records can create 1,000,000 possible matched pairs. However, the maximum number of pairs in the M group can only be 1,000. In order to make this process feasible it must be possible for the analyst to examine the possible matches and discard any based on manual inspection.

This is the concept of blocking. By dividing both files into mutually exclusive and exhaustive subsets and only searching for matches within a subset, the process of linkage becomes manageable (e.g., instead of examining all possible combinations of the variables within each data set, at least one variable must match).

Blocking eliminates an excess amount of unnecessary attempts to link two databases by linking all records that have an exact match on the variable of interest. However, problems tend to occur if the information is miscoded or left blank in one of the data sets. This process automatically assigns a miscoded record to the U group, as well as any non-matching variables (residual group). To avoid assigning a miscoded record to the U group, multiple passes are made on the records in each file. Any records that are not matched in one blocking scheme can be rematched in a different blocking arrangement. So the residuals are the input for the second attempt at matching.

Smaller blocks are more efficient than larger blocks. Restrictive blocking schemes are best used on the first pass, and the majority of the records will match on the first pass. This means that additional passes will have smaller number of records and will be able to process quickly, even with a very non-restrictive blocking scheme.

Fields that are better for blocking include those with a high number of potential values. However, the reliability of such data is also a key factor directly resulting in the reliability of the link. Data sources that have a high potential to be misreported or omitted will not benefit the data linkage process. Therefore, an evaluation of the potential for a variable as part of a blocking scheme is needed. This evaluation comes in the form of a weight on the linking variables.

3.2.2 Variable Weighting

To discern if a match of variables should constitute a linkage of two records, a weight is assigned to each field. This measure is also known as the discriminating power for each variable, which is based on the probabilities associated with the variable.

There are two probabilities associated with each variable. The first probability is the m probability. The m probability represents the likelihood that the information is recorded correctly. This can be determined by subtracting the error rate for the variable of interest from 1. For example, if gender disagrees 10 percent of the time due to transcription and reporting errors, then the m probability will be 0.90 (1-0.1). The larger the m probability, the better the chance of a reliable field.

The second probability (u) used for probabilistic linkage is the likelihood that a field will agree based on the incorrect value. For example, the probability of gender agreeing at random is 0.5. This results in 4 possible combinations of variables to match, as seen in Table 3-1. Two of the four matches would yield a correct match, thus resulting in a 0.5 probability that the variables matched by chance.

The weights that correspond to the m and u probabilities are calculated as the logarithm to the base 2 of the quotient of m and u, if the match is positive. However, if a

Table 3-1 Possible Combinations of Gender Variables

File A	File B
Male	Female
Male	Male
Female	Male
Female	Female

match does not occur, the weighting can be calculated by taking the logarithm to the base 2 of the quotient of 1-m and 1-u, resulting in a negative weight.

The following, based on Jaro (35), shows an application of the weighting process. In this example the weight assigned for gender and social security number is examined. It should be intuitive that a correct match on gender is not as strong as a match on social security number.

In the example presented in Table 3-2, we have a 10 percent error rate for the gender variable and a 40 percent error rate with the social security number. As displayed in Table 3-1, the u probability of the gender variable is 50 percent. For this example we will assume that there is a one in one million chance of obtaining a match based on entering an incorrect social security number.

As shown in Table 3-2, the social security number receives a larger positive weight compared to the gender-matching variable (e.g., 0.85 vs. 22.51). It is important to also note that this approach does not apply a stiff negative weight for the social security number if a match is not obtained; instead a larger negative weight is assigned to gender (e.g., -2.32 vs. -1.32). The distribution of the composite weights is generally bimodal with the non-matching observations receiving negative values and the matches receiving positive values.

Table 3-2 Parameters for Example Weighting

	Gender	Social Security Number
Error Rate	10%	40%
m probability	0.90	0.60
u probability	0.50	0.0000001
Positive Weight = Log_2 (m/u)	0.85	22.51
Negative Weight = Log_2 [(1-m)/(1-u)]	-2.32	-1.32

3.3 Selection of Blocking and Linkage Variables

Blocking and linkage variables must provide discrimination among the events and individuals in the databases to be linked. Direct and indirect identifiers can be used to assist in the linkage process by identifying specific individuals and events. These direct identifiers, as shown in Table 3-3, have the ability to link records directly, based on their unique nature.

Table 3-3 Direct Identifiers

Examples of Direct Identifiers
Name
Social Security Number
Crash Record Number
Vehicle Identification Number
Driver's License Number
Patient ID

Indirect variables can promote linkage as well. However, they cannot facilitate linkage independently; they must be used together with other indirect variables. Indirect variables, displayed in Table 3-4, are available for events as well as individuals. Both direct and indirect variables are used in the linkage process, due to the potential errors that may occur in each respective database.

Table 3-4 Indirect Identifiers

Examples of Indirect Identifiers	
Age	Date
Birth Date	Times
Gender	EMS Region
County	Injured Area
City	Residence Zip Code

3.4 Anticipated Linkage Results

The process of linking two files through probabilistic linkage creates matched pairs, unmatched pairs, and pairs that require review to determine if they are indeed matches. The initial assignment to one of these three categories is based on the composite weight score. Though low thresholds can increase the rate of positive matches, this increases the potential to falsely assume a record is a match when it is actually a non-matching pair. The opposite is true for increasing composite weight threshold. However, the general approach is to provide a wide range of scores identifying the cases for review to minimize the number of records unjustly classified as a non-match (34).

The review allows the analyst to determine if any potential matches not meeting the match threshold should be included in the matched status. Linkage results can also be reviewed by examining histograms of the weights assigned to potential matched pairs as well as other graphical representations of the linkage outcome. Figure 3-2 shows an example of a graphical representation of the linkage results. As expected, there are a large number of unmatched pairs (U) with low weights. However, as the weight increases the number of incorrectly matched pairs decrease and the matched pairs (M) increase. Each of the line presented in Figure 3-2 represents the cumulative positive and negative weights from each variable used to link the two databases. As expected, the unmatched pairs have a large negative weight, while the matched pairs have smaller negative weights This graph also provides information on the overall linkage process. For example, a large gap between the matched and non-matched records represents the amount of discrimination power of the

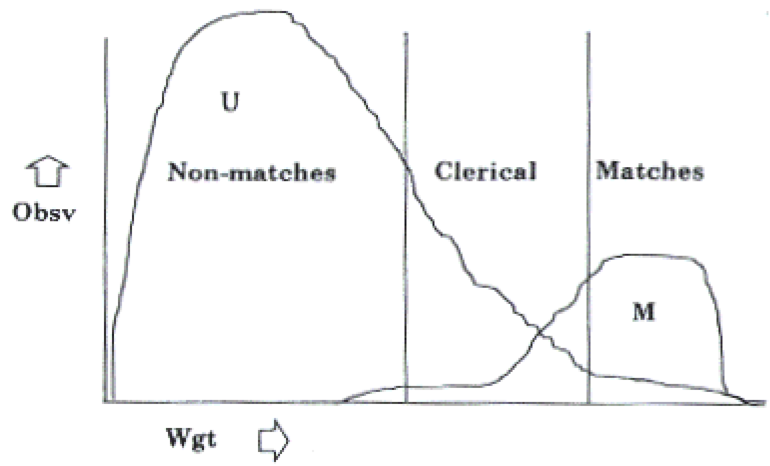

Figure 3-2 Graphical Representation of Linkage Results (35)

linkage process. Once the review is complete the data set is finalized, resulting in a linked data set ready for analysis.

This validation process provides analysts with an understanding of what was successful and what was ineffective in the linkage process. Also exposed is the quality of the data used for linking, which may provide answers to why records were and were not linked correctly. The graphical representation can also note any systematic bias in the linked data caused by missing data, reporting thresholds, and several other factors. Precise documentation of every component used to link the data eliminates future problems when attempting to link new data. The linkage of this information should be viewed as an annual event, not just a one-time occurrence.

3.5 Enhanced Analysis with GIS

Geographic analysis has been shown to enhance decision-making. A geographic information system (GIS) is a computerized database management system that allows the capture, storage, management, analysis, and presentation of spatial information. A GIS contains two types of information, spatial and attribute data. Spatial data are objects that have an orientation in two or three-dimensional space, such as lines, points, or polygons (44) whose locations are depicted using a geographic coordinate system. Information about these objects is found in the attribute table. These attribute tables contain tabular information that can normally be found in current databases. In some cases the tables used in the GIS environment are identical to traditional databases (45).

Creating an application with a graphical interface can eliminate confusion for analysts (46). Various applications have been developed using GIS (47,48,49). For instance, Iowa's system of analyzing crash data in GIS, the GIS-Accident Location Analysis System (GIS-ALAS), has spatial data (700,000 point locations and 290,000 roadway segments) and attribute data (crash characteristics obtained from the officer reports) (50). GIS-ALAS provides a graphical interface to Iowa's ALAS system.

3.6 Building on Iowa's Crash Analysis System

Currently desktop analysis of crashes is performed in Iowa using the Personal Computer Accident Location Analysis System (PC-ALAS). PC-ALAS does not utilize recent developments in computer graphics and spatial analysis methods such as geographic information systems. Currently the Iowa DOT is in the process of creating a GIS-based crash analysis tool to enhance the analysis capabilities.

Crash analysis requires complete information to identify high crash locations, especially information on the characteristics of crashes and potential contributing circumstances. This information includes data on the crashes (number of crashes, severity, number of vehicles, and fixed objects struck) as well as roadway features (geographic locations, roadway classification, surface type, and traffic control devices), vehicle information (type, size and age), human factors (age, gender, license status, and physical impairment) and environmental conditions (weather and lighting). PC-ALAS addressed all of these areas; however, one area that was weak was the geographic location.

The graphical representation of GIS-ALAS addresses this concern. The geographic information tied to each crash creates an opportunity for spatial analysis of crash records. By incorporating the geographic information into a new crash analysis tool, Iowa has created a state-of-the-art desk top crash analysis tool to analyze crashes.

3.7 Analysis of Iowa's Location Identification and Prioritization Process

Unlike other state DOT's, the Iowa DOT maintains a database that covers all reported crashes occurring within the state. The statewide data are the backbone to the analysis of high crash locations within the state. Iowa has a unique approach to determine the high crash sections of roadway in the state, as shown in Figure 3-3. Based on five years of crash data, all possible locations are analyzed to determine if they meet a minimum threshold to become an initial candidate location. To become an initial candidate location, a site needs to meet one of the following three criteria: 1 fatal crash, 4 injury crashes, or a total of 8 crashes. Once the candidate locations have been determined, a three-phased ranking scheme is used as the basis to determine the high crash locations.

The first ranking is based on the number of crashes, or frequency, occurring at each location. Each site is given a ranking based on the number of crashes: the site that has the highest frequency of crashes receives a one. This is done for each location until every site

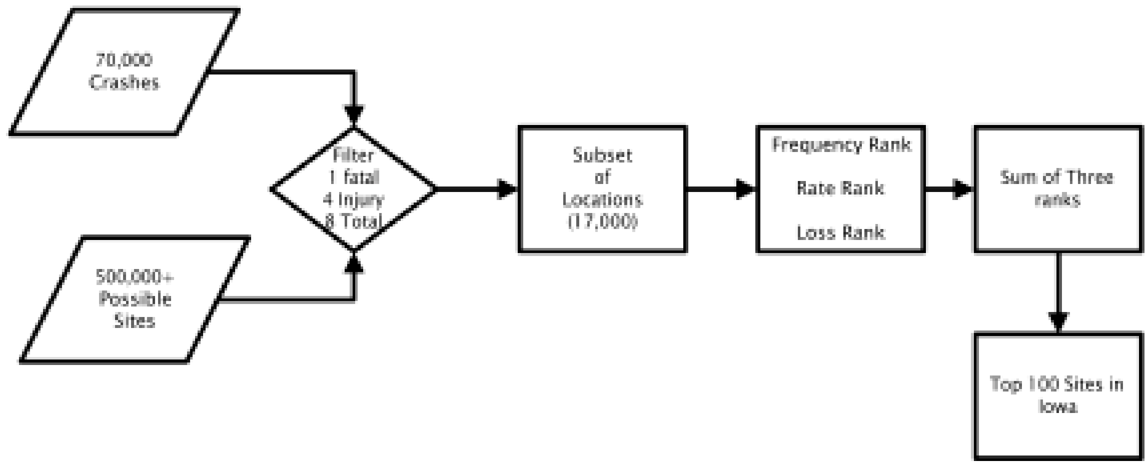

Figure 3-3 Current Approach to Identify High Crash Locations in Iowa

has a ranking; in the case of a tie each location receives the same rank and the subsequent ranking is skipped.

Second, each site is ranked according to the crash rate. The crash rate takes into consideration the length of the segment as well as the volume (intersection crash rates are determined using only the volume). The number of crashes is then divided by the exposure at each location. Again these sites are ranked, with the site that has the largest crash rate receiving the top ranking; the same implication of a tie applies to this ranking as well. In some instances the volume of traffic is unknown. Assigning these locations the top ranking (rank = 0) and then proceeding with the ranking procedure as usual avoids this problem.

Finally, each site is ranked according to the financial loss from the crashes. (This is determined by using values based on the injuries sustained in each type of crash as seen in Table 3-5. These values are then multiplied by the number of people that fall into each category.) For example, if two fatalities, four major injuries, 12 minor injuries, and 15 possible injuries occur at a location, the value loss due to injuries is $2,206,000 (2 X

$800,000 + 4 X $120,000 + 12 X $8,000 + 15 X $2,000). Property damage is incorporated as well. Officers report estimates on the crash report form. In some instances there is no estimate of property damage; when this occurs a default value of $2,000 is used. All of these values are summed up and result in a ranking based on the value lost at each location.

Table 3-5 Crash Costs by Injury Type

Type	Dollar Value
Fatal	$800,000
Major Injury	$120,000
Minor Injury	$8,000
Possible Injury	$2,000

To determine the top 100 high crash locations within the state, each of the three ranks are added together and a final ranking is performed with the lowest cumulative ranking receiving the highest ranking of a 1. Those falling within the top 100 ranking are deemed high crash locations within Iowa. In Table 3-6, an example of this process is shown, using fictitious data, for the top 13 locations throughout the state. In this instance the location represented by reference node 11111111 receives the top ranking for the statewide analysis, while reference node 21212121 receives the 13th ranking. This process is performed for approximately 17,000 locations that meet the initial threshold.

3.8 Using CODES to Revise Crash Cost Estimates

In 1996, 25,729 motor vehicle crashes in Iowa involved an injury to at least one person. Overall, 39,052 individuals were reported to have sustained injuries during this calendar year. Many of the people injured in these motor vehicle crashes received medical

Table 3-6 Example of High Crash Location Ranking Process

Reference Node	# of Crashes	Rank	Crash Rate	Rank	Dollar Loss	Rank	Total Rank	Statewide Rank
11111111	47	5	2.63	23	2,327,237	15	5+23+15 = 43	1
33333333	29	31	Unknown	0	1,909,420	20	31+0+20 = 57	2
44444444	25	35	2.76	15	2,734,603	9	35+15+9 = 59	3
22222222	24	37	2.71	19	3,150,760	4	37+19+4 = 60	4
55555555	53	1	2.46	29	1,373,300	35	1+29+35 = 65	5
77777777	40	10	2.92	8	1,120,949	47	10+8+47 = 65	5
00000000	34	21	2.40	33	2,000,850	18	21+33+18= 72	7
10101010	49	2	2.65	21	1,117,965	50	2+21+50 = 73	8
32323232	28	32	2.41	32	2,684,259	10	32+32+10= 74	9
88888888	19	51	3.15	3	1,824,587	22	51+3+22 = 76	10
99999999	18	53	2.47	28	3,501,985	1	53+28+1 = 82	11
66666666	36	18	2.28	41	1,740,548	27	18+41+27= 86	12
21212121	32	24	1.98	61	1,357,951	39	24+61+39=124	13

treatment in some form. Very little information is available from the traffic crash reports concerning the extent of the injuries sustained. However, NHTSA has provided incentives for states to link medical files with traffic crash records, thus creating a crash outcome database.

This process attempts to link the 39,052 injuries sustained from motor vehicle crashes with their corresponding hospital record. Using customized computer software, a probabilistic linkage approach is initiated to determine if two records, one in the crash file and one in the medical file, represent the same incident, or person. Upon completion of the probabilistic approach several records are successfully linked, as shown in Figure 3-4. This creates a CODES data set, which provides the foundation for the objectives in this report.

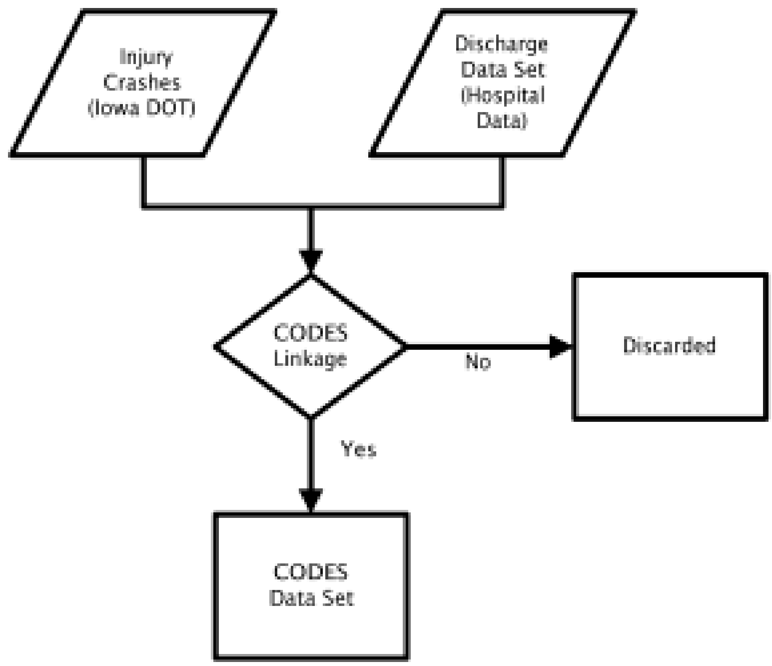

Figure 3-4 Process of Linking Crash and Medical Files

3.9 Application of Procedure to Address Research Objectives

With the creation of this new data set, by way of a probabilistic approach, several issues can be examined with respect to medical costs. This information directly leads to three of the four research objectives and is presented in Chapter 4, Analysis and Results.

The first research objective, determining the medical crash costs for Iowa crashes, can be addressed by simply taking the mean values from the new data set produced. Information from the crash file, specifically the injury classification, is used in conjunction with the medical costs charged by the medical institution. Until now, Iowa has not had a data set that contains these two pieces of information. The new estimates are reported in various forms in the analysis section, Chapter 4.

Determining the medical crash costs for specific types of crashes (e.g., head-on collisions, rear-end crashes, etc.) is the second objective of this research. Again, this will require basic statistical measures to determine these new estimates. In particular, two categories of crashes will be examined, including the type of crash and the speed limit of the roadway where the crash occurred. As before, the type of crash and the speed limit of the roadway are available from the crash report. The medical charge information is available from the statewide medical records. Medical crash cost estimates are determined for broadside, rear end, sideswipe, and single vehicle crash types. Additionally, medical crash cost estimates will be determined for the speed limit of the roadway where the crash occurred.

The third research objective is to compare these medical cost estimates of Iowa motor vehicle crashes with those obtained from the FHWA. Minor adjustments to the FHWA values must occur due to the cost of living adjustments for Iowa compared to the national average. After this adjustment, statistical tests are performed to determine if differences exist, and to what extent, between the national estimates and the actual values determined from Iowa data.

The final objective, evaluating the medical crash cost impact on Iowa's high crash location process, can be performed with the creation of the CODES database through additional analysis. As outlined earlier, Iowa uses the value loss as one basis to determine the high crash locations. These costs are associated with default values based on the severity of the motor vehicle crashes. Theoretically, with the creation of the CODES database the exact medical costs for each crash can be determined. However, not all of the injuries may appear in both the medical and crash records, thereby omitting some individuals from the CODES data set. Therefore, new estimates for the reported injury crashes will be determined from the observations in Iowa's CODES database. Changing these values provides an opportunity to determine what impact various dollar amounts have on the analysis of Iowa's high crash location process as displayed in Figure 3-5.

Statistical testing allows analysis into the sensitivity of Iowa's ranking scheme to improve estimates of injury costs to determine if there is a significant difference between the two rankings. The process outlined in this chapter allows an evaluation into the estimates

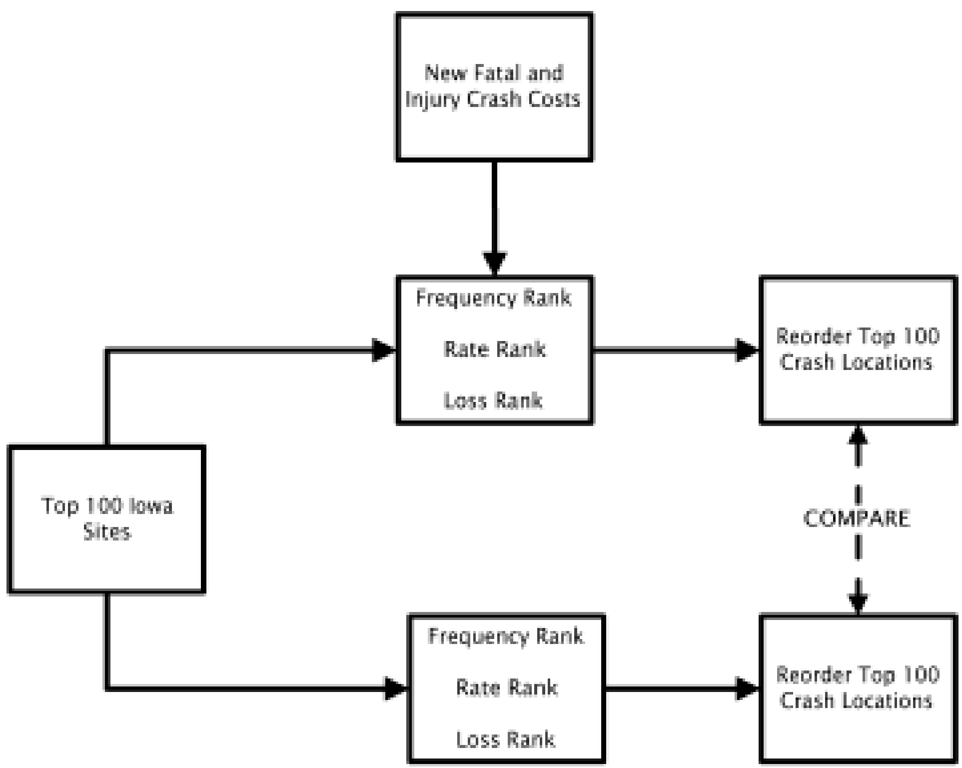

Figure 3-5 Comparison of Rankings

obtained by the FHWA and the Iowa DOT. This approach will provide medical crash cost estimates for the type of collision and speed limit of the roadways, thereby creating additional information to assist in identifying the consequences of traffic crashes. The results obtained from using this methodology are presented in the next chapter, Analysis and Results.

CHAPTER 4 ANALYSIS AND RESULTS

The goal of this analysis is to assess the estimates of crash costs as determined by the Iowa DOT and the FHWA in the new light of injury costs derived from medical records. The results from applying the methodology presented in the previous chapter are reported. First, this chapter outlines the results of several procedures conducted to make this assessment. The entire analysis included several steps that began with linking the Iowa crash records with statewide hospital discharge records for the state. Second, using GIS-ALAS, the spatial distribution of linked and unlinked medical injuries is examined. Third, new medical crash cost estimates are derived based on the new data set and compared to current medical costs used by the FHWA and Iowa DOT. Fourth, the medical crash costs are determined for two crash characteristics, speed limit and type of collision. Finally, the new medical crash costs are evaluated to determine what impact, if any, these changes have on the current high crash location process in Iowa.

4.1 Probabilistic Linkage Results

Using the medical files from the statewide database, 769 injuries were identified in the crash records and in the medical database. These records were linked using the social security number as the blocking variable. In Iowa the social security number usually serves as the driver's license number. Therefore, if the license number of an individual in the crash database is identical to the social security number in the medical database, these records are joined.

Next, the age, date of incident, sex, and time of incident were used to determine if the records linked by the blocking variable represent the same incident. Using the weighting criteria presented in Section 3.2.2 eliminated 284 records from the initial linkage based on the social security number, resulting in 485 records linked to a motor vehicle crash injury.

4.1.1 Spatial Bias

Before using any of the information obtained from the probabilistic linkage of the medical and crash records, the spatial pattern of the linked data will be examined to convey information concerning locations with high and low linkage rates. Three geographic scales, county, city, and census tract, are examined to compare the linkage rates. The results can be graphically displayed using Iowa's GIS-ALAS system, as shown in Figure 4-1.

Inspection of the spatial distribution allows an initial assessment of the quality of the linkage results. Using the geographical element of the crash records, created by GIS-ALAS,

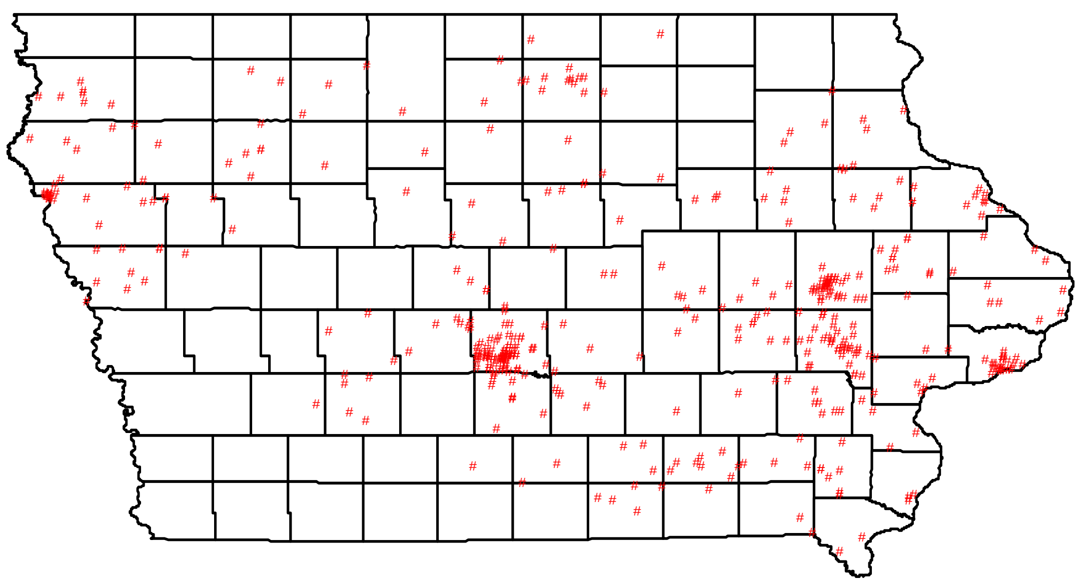

Figure 4-1 Location of Linked Crashes in Iowa

areas reporting high and low numbers and rates of successful links can be identified. This information provides an initial assessment into the quality of the linked data.

First, the county scale is examined to determine the linkage ratio throughout the state, and the results show the linkage varies. The linkage ratio, displayed in Figure 4-2, represents the number of linked injury records compared to the number of injury crashes occurring in each county. Initial inspection of the map shows that at least one record was successfully linked in several counties. However, inspection of the southwest quadrant of the state shows several counties did not have any successful linkage between the crash and medical records. This can be explained by the location of the ten trauma centers throughout the state, the source of the medical records for this study. Currently there is not a trauma center located in southwest Iowa; however, there is one located in Omaha, Nebraska, just across the border.

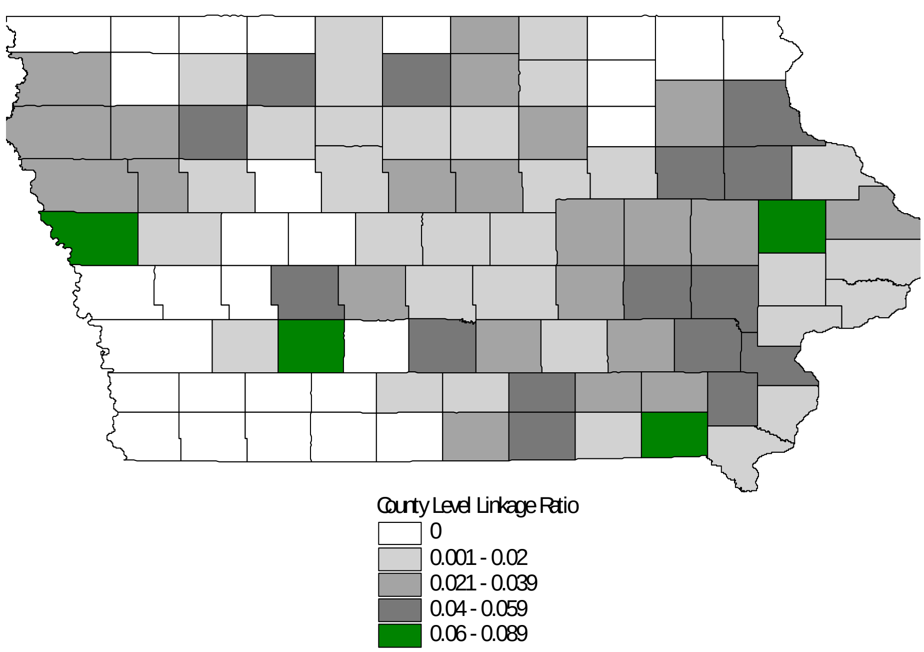

Figure 4-2 Spatial Distribution of Linked Crashes by County

Next, the linkage ratios for several cities are examined to determine if there is any bias toward one particular city or geographic area. Table 4-1 demonstrates the linkage ratio for the top 18 cities according to the number of injury crashes (those with at least 200 injury crashes in 1996). For a majority of the cities, about one percent of the crash records linked to the hospital records for the injury crashes. The low linkage ratio can be explained by the small sample size of linked records compared to the number of injury crashes. However, Council Bluffs did not have any crash records linked to the medical records. This is expected due to the trauma center location in Nebraska.

Finally, the linkage ratio at the census tract level for Polk County is displayed in Figure 4-3. The dark areas on the map represent the census tracts that have the highest

Table 4-1 Linkage Ratio for Top 18 Cities in Iowa

City	Number of Injury Crashes	Number of Linked Records	Ratio
Des Moines	3043	38	0.0125
Davenport	1363	18	0.0132
Cedar Rapids	1074	21	0.0196
Sioux City	859	15	0.0175
Council Bluffs	809	0	0.0000
Waterloo	616	2	0.0032
Dubuque	460	4	0.0087
Iowa City	445	6	0.0135
West Des Moines	399	5	0.0125
Clinton	336	2	0.0060
Ames	303	0	0.0000
Bettendorf	276	4	0.0145
Cedar Falls	276	1	0.0036
Fort Dodge	258	1	0.0039
Mason City	254	4	0.0157
Muscatine	212	3	0.0142
Urbandale	211	4	0.0190
Burlington	208	2	0.0096

linkage ratio (6-10%) in Polk County, compared to the white sections that contain the lowest linkage ratio (0%). A somewhat surprising result is the high linked ratio in rural locations compared to the lower rates experienced in the urban sections of the county. However, there are two possible reasons that the crash rates are distributed in this fashion. The first deals with the source of the data. In Des Moines there is only one trauma center. Severe injuries requiring treatment at a trauma center may be more likely to occur in rural areas. In these instances the patients are brought directly to the trauma center instead of opting for the local hospital. Second, a high number of crashes occurring in the city result in relatively minor injuries, or no injuries at all; therefore, those with minor injuries use the local hospital for treatment rather than the trauma center.

As shown in the previous table and figures, analyzing spatial patterns introduces new opportunities. Spatial analysis is not the primary objective of this research; however, displaying the information through a GIS provides an additional data quality check before the information is used in statistical testing. From the results presented the linked records are dispersed throughout the state. This ensures that data from a limited set of geographic locations do not drive the analysis for a statewide application. This supports the view that the statewide results, presented later in this chapter, are not unduly biased by the spatial pattern of linked crashes.

4.2 Crash Costs: Dealing with Non-Normality Issues

The new data set contains medical information concerning 475 injuries sustained in motor vehicle crashes (ten records from the initial linkage were eliminated because the injury classification was coded as unknown). The most important part of this data set, for purposes of this analysis, is the medical costs incurred by the crash victims. Before the information is used, an inspection of the data is performed to determine what statistical procedures can be used to address the four research objectives:

- Determine the average motor vehicle injury crash costs for the state of Iowa.
- Determine the crash costs of specific types of crashes (type of collision and speed limit).
- Compare the actual medical costs to the FHWA and Iowa DOT estimates.
- Evaluate the impact on Iowa's high crash location process using the new estimates.

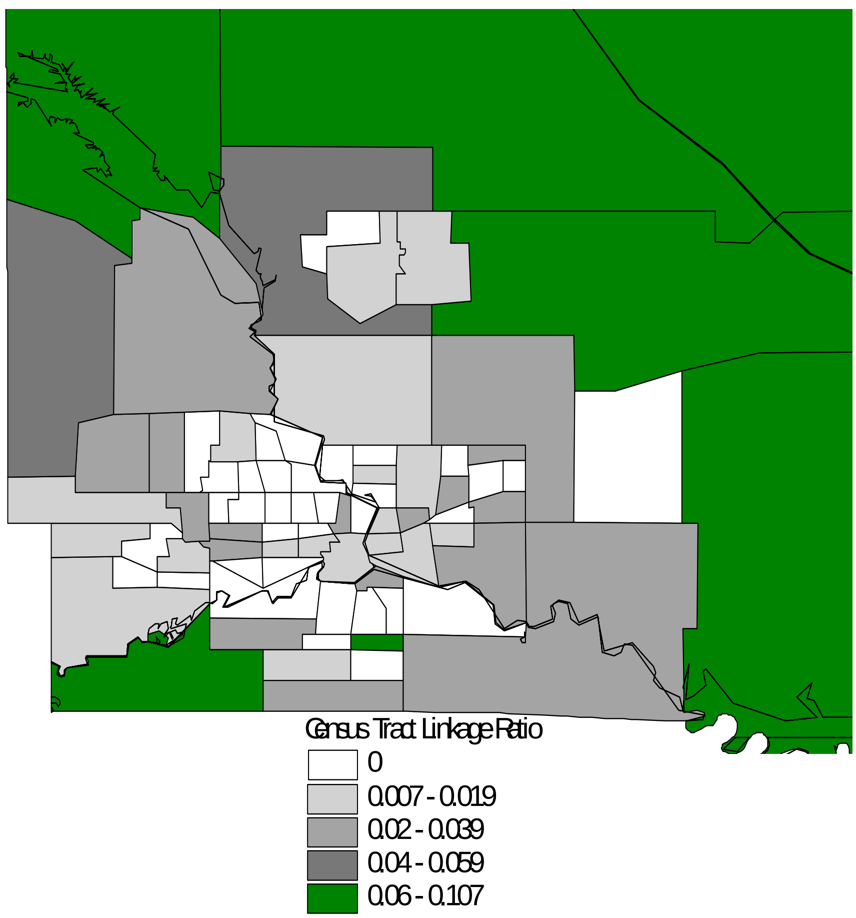

Figure 4-3 Linkage Ratio for Polk County Census Tracts

In order to determine what statistical measures are appropriate for analytical purposes, the distribution of the data must be determined. The normal distribution allows basic statistical procedures to be used for analysis; otherwise, statistical analysis becomes a more complex task. Normality can be verified by examining a histogram of the medical values, as shown in Figure 4-4. Normal distributions follow the traditional bell-shaped curve presented in elementary statistical publications.

Though the data does not take on a normal distribution, transforming the data may allow the distribution to become approximately normal. Four transformations are investigated and presented in Figure 4-5: the square root of the medical costs; the square of the medical costs; inverse of the medical costs; and the log of the medical cost. The log transformation appears to distribute the data normally and will be used in the analysis of the data.

4.2.1 Average Crash Costs

Using the normally distributed transformed data, the central tendency and dispersion of crash costs can easily be determined. These costs will be used as the basis to compare the crash costs of the FHWA and Iowa DOT. However, these costs must also be transformed to the log scale to create a common basis of analysis. This calculation uses the KABCO crash injury classification found in the traffic crash records, described in Section 2.4. Using the CODES data set, the medical crash costs (log scale) are determined for injury severity, speed limit, and collision type, as presented in Table 4-2 through Table 4-4.

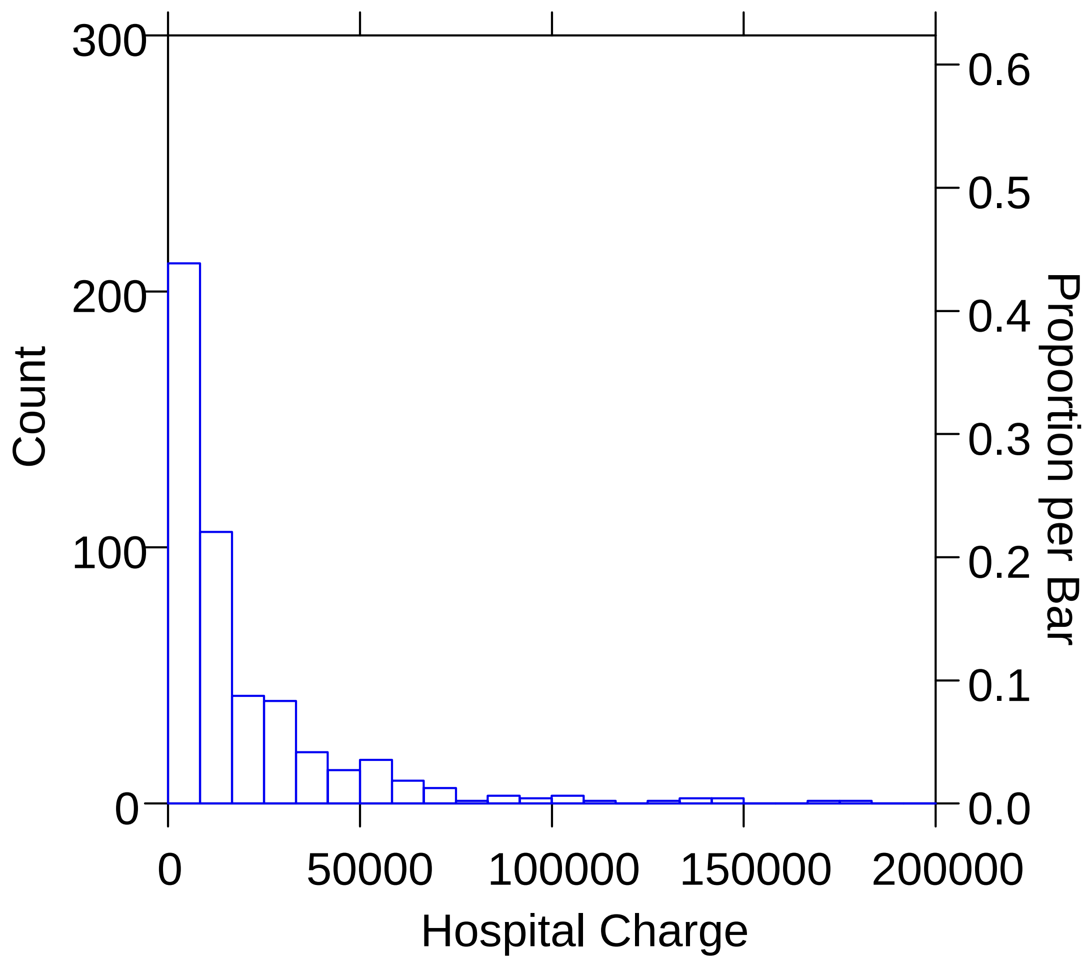

Figure 4-4 Distribution of Hospital Costs

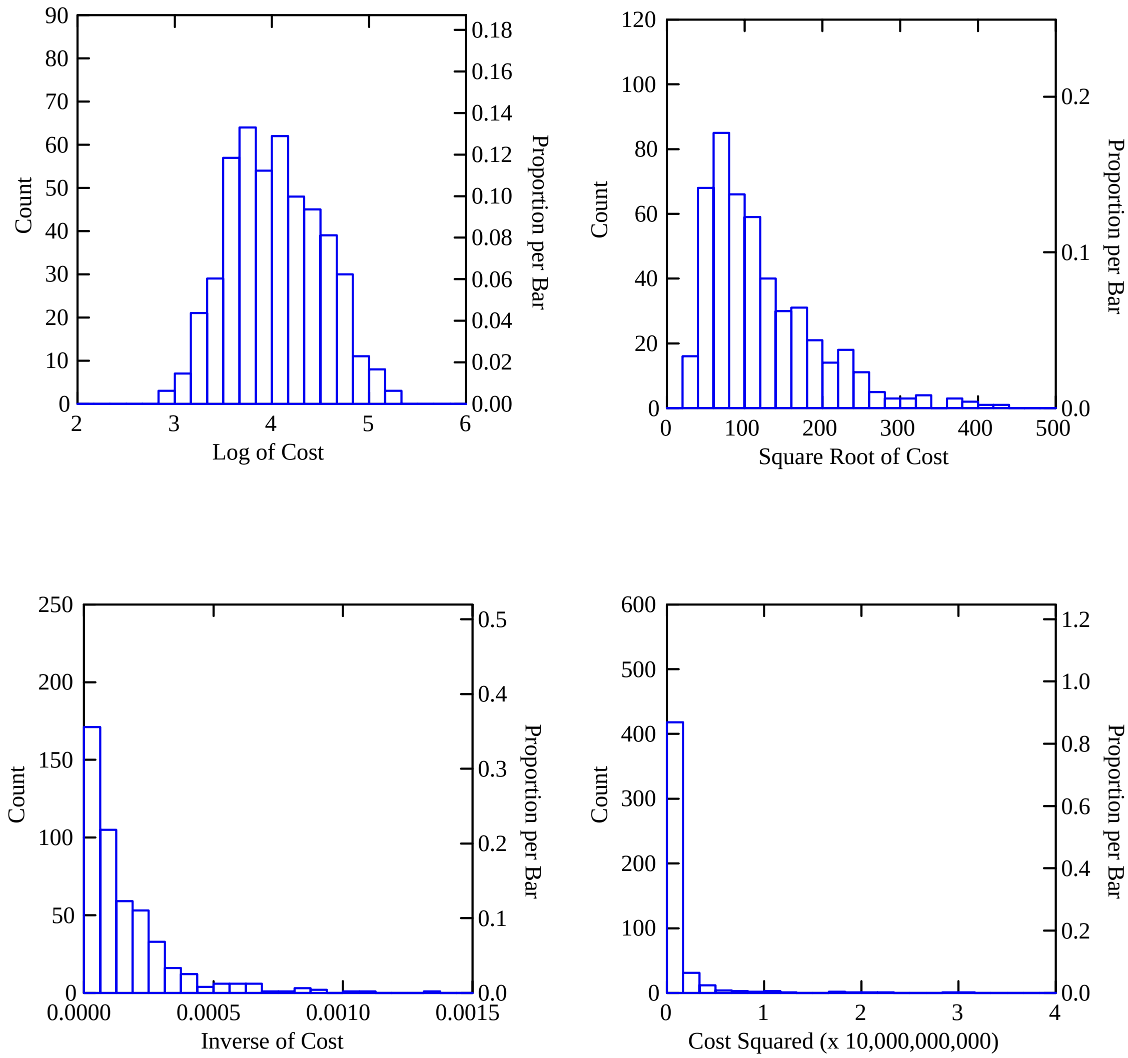

Figure 4-5 Graphical Display of Transformed Cost Data

Table 4-2 Log Medical Crash Cost by KABCO Injury Severity Scale

Crash Severity	Mean	Standard Deviation
Fatal	4.075	0.480
Major Injury	4.194	0.449
Minor Injury	3.785	0.405
Possible	3.715	0.353
Unknown	3.661	0.606

Table 4-3 Log Medical Crash Cost by Collision Type

Collision Type	Mean	Standard Deviation
Broadside	4.006	0.461
Head On	4.182	0.542
Parked Vehicle	3.772	0.236
Rear End	3.844	0.386
Side Swipe	3.886	0.386
Single	4.053	0.475
Other	4.057	0.490
Unknown	4.009	0.514

Table 4-4 Log Medical Crash Cost by Speed Limit

Speed Limit	Mean	Standard Deviation
25	3.872	0.449
35	3.875	0.441
45	3.892	0.414
55	4.132	0.481
65	3.929	0.404

4.2.2 Initial Inspection of CODES Mean Costs

Before attempting to perform a statistical analysis of the data set, the data should be examined to determine if the initial findings are intuitive. First, the crash severity is examined below, followed by the collision type and speed limit.

The pattern of mean cost values presented in Table 4-2 is reasonable. The maximum cost is associated with major injuries, and the cost decreases as the injury severity decreases. Furthermore, the expected medical costs for a fatality do not exceed the major injury estimates. Initially, this may not seem correct; however, when considering the limited medical treatment applied to some fatal traffic crash victims the expected value can often be less than that of an individual receiving major injuries requiring extensive medical attention.

Table 4-3 presents the average crash costs based on collision types. Intuitively, head-on crashes have a higher medical cost than rear-end collisions. The mean cost values of the two collision types provides support for this assumption. Additionally, crashes usually resulting in little or no injury (sideswipes and collisions with parked vehicles) also have lower medical crash costs compared to broadside and single vehicle crashes (usually resulting in higher medical costs).

Obviously, higher speeds usually result in increased severity of injuries. This parallels an increase in the medical costs as the speed limit increases. Table 4-4 clearly displays this phenomenon. However, higher speed facilities are designed to higher standards. Designing to the higher standards may reduce the injury severity (and cost) associated with crashes on high speed roadways. This is also represented in the injury crash cost estimate presented in Table 4-3, in which the injury costs for 65-mph roadways are less than those at 55 mph.

Overall, the pattern of mean values presented in the previous three tables do not indicate any unreasonable pattern of values pertaining to the medical costs resulting from motor vehicle crashes.

4.3 Comparison of CODES Results with FHWA Estimates

The Iowa DOT uses the KABCO scale on the crash reporting form and the FHWA studies on injuries use the MAIS scale to determine the crash cost estimates. Therefore, the conversion from the MAIS scale to KABCO scale, presented in Table 2-4, must be used to determine the average crash cost in the KABCO convention. However, before converting the MAIS crash costs to the KABCO scale the costs must be adjusted for geographical and historical economic differences. This conversion takes the national averages at each category and adjusts them to 1996 Iowa dollars based on the Consumer Price Index. To convert 1994 dollars, the year of the FHWA estimates, to 1996 dollars, the year of the CODES data, a conversion factor of 1.058 is used (54). Additionally, the cost must be adjusted from a national average to Iowa dollars, using a conversion factor of 0.965 (51). Therefore, the 1996 FHWA cost estimates for Iowa crashes can be estimated in Table 4-5.

Using the information obtained in Table 4-5, along with the cost component breakdown presented in Figure 2-1 and Figure 2-2, the medical estimates of the Iowa crash costs can be determined based on the FHWA crash cost estimates. According to NHTSA, the medical component consists of 1.5 percent of the total fatal crash costs and 22.5 percent

Table 4-5 FHWA Costs Estimates for Iowa Crashes

	1994 FHWA Costs	1996 FHWA Costs (1.058 Conversion factor)	1996 FHWA Costs for Iowa (0.965 Conversion Factor)
(K) Killed	$2,854,500	$3,020,061	$2,914,358
(A) Major Injury	$212,880	$225,227	$217,344
(B) Minor Injury	$46,862	$49,580	$47,845
(C) Possible Injury	$25,467	$26,944	$26,001

Table 4-6 FHWA Medical Cost Estimates for Iowa Crashes

	1996 FHWA Total Cost Estimates for Iowa Crashes	1996 FHWA Medical Cost Estimate for Iowa Crashes
(K) Killed	$2,914,358	$43,715
(A) Major Injury	$217,344	$48,902
(B) Minor Injury	$47,845	$10,765
(C) Possible Injury	$26,001	$5,850

of the injury crash costs. The resulting FHWA estimate of the medical crash costs can be found in Table 4-6.

Next, these values are converted to the log scale. This allows comparison with the medical information obtained from the Iowa CODES data set, which has a normal distribution in the log scale. Therefore, the values obtained in Table 4-6 are converted to the log scale and presented in Table 4-7. Finally, these values are now compared with the estimates obtained from the Iowa CODES database in Table 4-8.

4.3.1 Significance Testing of FHWA Medical Costs

Once the medical values of the FHWA estimates have been adjusted based on historical and geographical considerations, statistical testing will determine if there is a

Table 4-7 Log Scale of Iowa FHWA Medical Costs by Injury Type

	1996 FHWA Medical Cost Estimate for Iowa Crashes	Log 1996 FHWA Medical Cost Estimate for Iowa Crashes
(K) Killed	$43,715	4.640
(A) Major Injury	$48,902	4.689
(B) Minor Injury	$10,765	4.032
(C) Possible Injury	$5,850	3.767

Table 4-8 Comparison of FHWA and Iowa Medical Crash Costs

	FHWA Medical Costs	Iowa Medical Costs CODES Data	Sample Size
(K) Killed	4.640	4.075	25
(A) Major Injury	4.689	4.194	286
(B) Minor Injury	4.032	3.785	114
(C) Possible Injury	3.767	3.715	50

difference in the estimated medical costs presented by the FHWA and the observed medical costs obtained from the Iowa CODES database. Transforming the data into a normal distribution simplifies the statistical testing to determine if the values are indeed significantly different.

Using the t-test, the probability that the FHWA medical estimates are equivalent to the Iowa CODES medical cost estimates can be determined. A computed probability of 0.002, for example, indicates a 99.8 percent likelihood that the two values are not the same; conversely, a value of 0.997 indicates a 99.7 percent likelihood that the two values are the same.

Presented in Table 4-9 are the test statistics for the four injury classifications using the KABCO scale. Additionally, the probability (reported to three significant digits) that the two values (FHWA and CODES) are equivalent is presented.

The interpretation of the results is very straightforward. To reject the hypothesis that the FHWA and Iowa CODES medical crash costs are equivalent a low probability is needed. This clearly is the case for the estimated medical costs determined by the FHWA. Therefore, the use of statistical measures (t-test) determines that the FHWA estimates of the medical crash costs are not equal to the information obtained from the Iowa CODES data set. Therefore, we can conclude that the medical crash cost estimates of the FHWA are too high compared to the costs obtained from the Iowa CODES database.

Table 4-9 Summary of Statistical Information for Medical Crash Costs

Extent of Injury	FHWA Value	CODES Value	Sample Size	T – Statistic	Probability
Fatal	4.646	4.075	25	-5.882	0.000
Major Injury	4.695	4.194	286	-18.641	0.000
Minor Injury	4.037	3.785	114	-6.498	0.000
Possible Injury	3.733	3.715	50	-1.039	0.304

4.4 Comparison of CODES Results with Iowa Estimates

Currently the Iowa DOT uses standard values, as displayed in Table 4-2, to determine the value loss for a crash. Each fatality is assigned an $800,000 loss value; likewise the major, minor, and possible injuries are assigned loss values of $120,000, $8,000, and $2,000 respectively. Obviously, these values are quite different from those determined by the

FHWA. Therefore these values are also inspected to determine how representative they are compared to actual values observed from the Iowa CODES database. Unfortunately, the medical cost components for the Iowa DOT figures are unknown. Therefore, the NHTSA distribution of the medical costs are applied to the current Iowa injury crash costs, providing an estimate of the medical crash costs using current Iowa values. The NHTSA breakdown of the total crash costs was the only study attempting to determine each cost component. This, then, is the justification for choosing the FHWA distribution and applying it to the Iowa crash costs.

To determine the medical values for the current Iowa crash costs, the fatal crash cost is multiplied by 1.5 percent and the injuries are multiplied by 22.5 percent. This leads to the estimated Iowa medical values presented in Table 4-9. Again, the values must be converted to the log scale to have equal units of comparison; these values are also presented in Table 4-10.

Once the common basis of the log scale has been determined, comparisons to the Iowa CODES data set can be performed. This analysis will examine the medical costs of the Iowa crashes assuming the same distribution as the NHTSA estimates.

Table 4-10 Log of Iowa Crash Costs

	Iowa Crash Costs	Estimated Medical Costs	Log Medical Cost
(K) Killed	$800,000	$12,000	4.079
(A) Major Injury	$120,000	$27,000	4.431
(B) Minor Injury	$8,000	$1,800	3.255
(C) Possible Injury	$2,000	$450	2.653

4.4.1 Significance Testing of Iowa Medical Crash Costs

Applying the NHTSA distribution to the current crash costs used by the Iowa DOT allows statistical analysis of the current medical crash cost in Iowa. The testing will determine if the current medical crash costs are similar to those found in the Iowa CODES database. The log-transformed data simplifies the statistical testing by creating a normally distributed data set, thereby allowing the t-test to be used to determine if the current medical crash costs are equivalent to the CODES findings.

Table 4-10 shows the test statistics for the four injury classifications found in the Iowa crash record and CODES database. Also presented is the probability that the CODES and Iowa medical estimates are equivalent.

The results from the statistical analysis clearly indicate the major, minor, and possible injury medical crash costs currently used by the Iowa DOT are not equal to the findings from the Iowa CODES database. Therefore, from the information presented in Table 4-10 it can be said that the major injury medical crash costs currently used by the Iowa DOT is overestimated when compared to the findings in the Iowa CODES data. Furthermore, the

Table 4-11 Summary of Derived Medical Crash Statistics

Extent of Injury	Log Iowa DOT Medical Value	Iowa DOT Medical Value	Log Iowa CODES Value	Iowa CODES Value	Sample Size	T-Stat	Prob
Fatal	4.079	12,000	4.075	12,000	25	-0.042	0.967
Major	4.431	27,000	4.194	16,000	286	-8.916	0.000
Minor	3.255	1,800	3.785	6,100	114	13.972	0.000
Possible	2.653	450	3.715	5,200	50	21.275	0.000

minor and possible injuries are underestimated compared to the values obtained from the Iowa CODES database. However, the medical crash costs estimates for fatal crashes are equal to the Iowa CODES value. This is evident by examining the probability that the fatal crash costs are equal. The t-statistic of –0.042 corresponds to a 97.6 percent likelihood that the two costs are equal.

Clearly, the majority of the crash cost estimates used by the Iowa DOT are not representative of the actual costs occurring in the state. However, this assumption relied on the NHTSA distribution of the crash costs. In the next section the impact of using the incorrect values to determine high crash locations for the state of Iowa is considered.

4.5 Implications for Identification and Prioritization of High Crash Locations

The injury costs derived from the CODES database are different from those used by the Iowa DOT estimates. To examine the effect of this difference on the ranking of high crash locations in Iowa, the CODES-derived injury costs were converted to total crash costs in the following manner. An increase or decrease in medical costs was assumed to result in a corresponding change in five of the seven other crash cost components: insurance, market production, legal fees, household production, and workplace costs. This assumption is based on the idea that an increase in medical costs usually results from an increase in injury severity and, therefore, will result in additional time away from work and as well as a corresponding increase in several of the other cost components. For instance, the CODES-derived estimate of medical costs for major injuries is only 60 percent of the value estimate by the Iowa DOT ($16,000 vs. $27,000). The values of these five components were thus adjusted by multiplying by 60 percent. The other two crash cost components, property damage and “other” (e.g., travel delay, emergency medical services) were assumed to be unrelated to changes in medical charges. Total crash costs for minor and possible injuries were calculated in the same manner. The crash costs assigned to the fatal crashes will not change because the current estimates were determined to be a proper estimate. The new crash cost estimates are presented in Table 4-12.

Table 4-12 Total Crash Costs Used on Evaluation

Injury Severity	Current Value	Medical Estimate	New Estimate
Fatal	$800,000	$12,000	$800,000
Major	$120,000	$16,000	$85,000
Minor	$8,000	$6,100	$22,000
Possible	$2,000	$5,200	$18,000

The next section will examine the impact on the high crash location process and determine if using alternative values, based on the injury costs, result in different outcomes. Spearman’s rank order correlation coefficient allows the changes of a ranked list to be quantified in statistical measures. As with the correlation coefficient, Spearman’s rank order correlation coefficient will have a value between -1 and 1. Larger values mean less difference between two orderings, or a positive correlation.

4.5.1 Evaluation of High Crash Location Process Using New Estimates

The evaluation of the high crash location process examines the impact of using the new estimates derived from the CODES database. Using the values presented in Table 4-12 the ranking procedure is performed using the top 100 locations in Iowa, as presented in Section 3.9. The analysis finds the Spearman’s rank order correlation coefficient of 0.94 indicating little difference between the two rankings. However, a selection of the old versus new rankings, as shown in Table 4-13, clearly shows that the ranking of specific locations do change. Additionally, the potential for a location to move up or down the list is minimized due to the limitation of the top 100 locations. If this process is replicated for the entire set of 17,000 locations the potential ranking movement is much greater. However, with a high

Table 4-13 Sample Ranking Using New Estimates

Initial Ranking	New Ranking
1	2
2	3
3	1
4	6
5	4
6	9
7	7
8	5
9	12
10	13
11	14
12	21

Spearman's rank order correlation coefficient, it is unlikely that a location would see a substantial change in the rankings (e.g., change from 100th to 1st on the overall scale).

4.5.4 Sensitivity Analysis of High Crash Location Process

As shown in the previous section of this chapter there is a discrepancy in the values being used in analysis and the values used by the FHWA. Therefore, this section will examine the sensitivity of the high crash location identification process and determine if using alternative values based on the injury costs results in different outcomes. Spearman's rank order correlation coefficient can be used to determine what impact changing the cost values has on the ranking process.

The sensitivity of the ranking of locations to injury costs will be examined for each injury classification. To determine the impact of changing these values, a chart of the Spearman's rank order correlation coefficient is developed for several values and plotted to determine the impact on the ranking by changing the values assigned for the injury classification.

4.5.4.1 Sensitivity Analysis of Fatal Crash Costs

First, the crash cost assigned to the fatalities is examined, holding all other values constant at the default values used by the Iowa DOT ($120,000; $8,000; $2,000; for major, minor and possible injuries). As shown in Figure 4-6, a critical point is clearly evident at $1,400,000; above this point the impact on the high crash location process does not change. The ranking of locations is more sensitive to values below the current default cost than above.

4.5.4.2 Sensitivity Analysis of Major Injury Crash Costs

Next, the impact of changing the crash costs based on the default value assigned to major injuries is explored. Again all other values are held constant when changing the value assigned to the major injuries. As shown in Figure 4-8, changing the values assigned to major injuries results in a very insensitive output. Unlike the fatal costs, no clear inflection point is present. Upon further inspection an inflection point does not occur until the major injuries are assigned a cost at $1,300,000, over ten times the current value being used by the

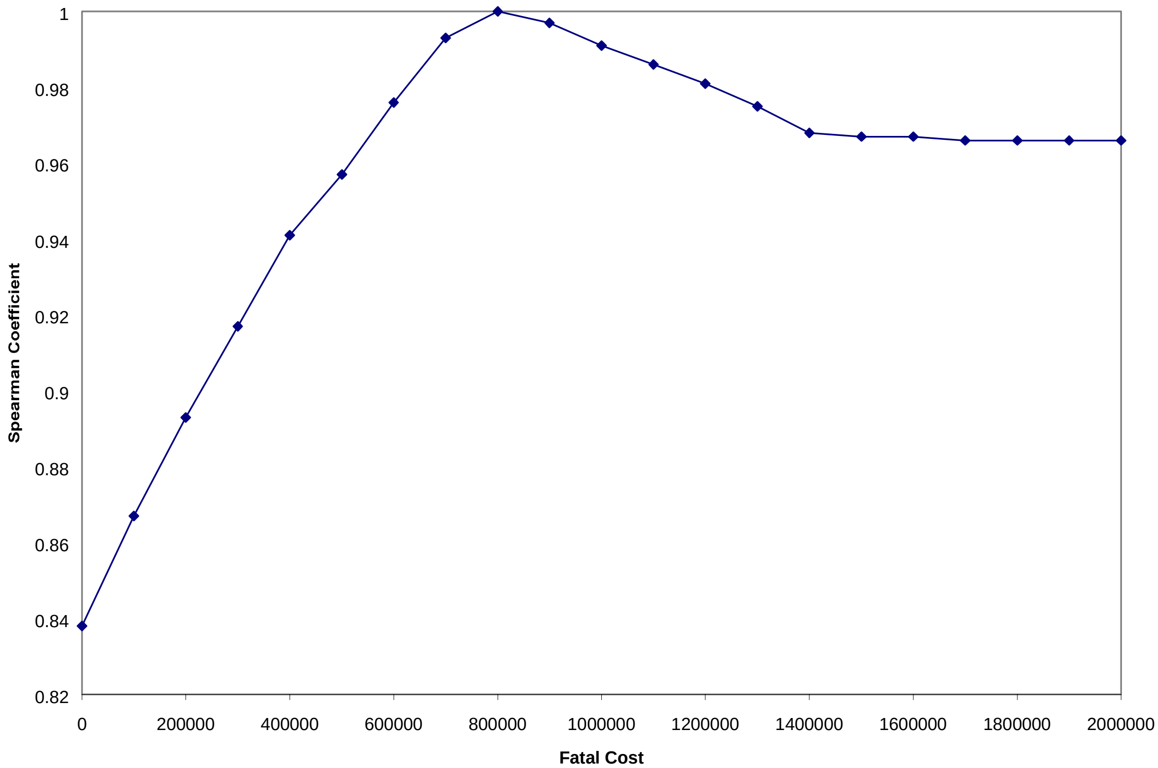

Figure 4-6 Sensitivity of Fatal Costs in Iowa High Crash Location Process

Iowa DOT. The $1.3 million value will not enter the analysis of high crash locations because it was determined that the current crash costs used by the Iowa DOT ($120,000) are over-estimated according to the CODES database. Finally, holding all other crash costs constant and only changing the major injury costs to the recommended $85,000 results in a Spearman's correlation coefficient of 0.987. Obviously changing only the value of the major injuries has little effect on the overall ranking of high crash locations.

4.5.4.3 Sensitivity Analysis of Minor Injury Crash Costs

Next, the impact of changing the minor injury crash cost is evaluated while holding all other values constant. The results are presented in Figure 4-8 and are similar to the major injury results. Thus changing the dollar amounts does not significantly change the overall ranking. As shown in Figure 4-8, the Spearman correlation coefficient only changes by a few one hundredths, signifying very little change to the overall ranking scheme. The Spearman correlation coefficient does not change more than one tenth (i.e. 1.0 to 0.9) until the value is increased to $79,000, close to 1,000 percent of the current default value of $8,000 and more than 300 percent above the recommended value of $22,000 based on the CODES data set. Clearly, changing only the value placed on minor injuries does not drastically change the overall ranking of high crash locations.

4.5.4.4 Sensitivity Analysis of Possible Injury Costs

Finally, the possible injuries are investigated holding all other crash cost values to the constant default values used by the Iowa DOT. As with the minor injuries, the possible injuries are also very insensitive to the dollar values assigned for analysis. Presented in Figure 4-9 is the distribution of the Spearman correlation coefficient for the sensitivity analysis, resulting in very little change. A change of only 0.05 (i.e. from 1.0 to 0.95) occurs at ten times the current value of $2,000 used by the Iowa DOT. Inspection of the new recommended value of $18,000 for a possible injury results in a Spearman's correlation coefficient of 0.95. Like the other values assigned to injuries and fatalities the impact of changing only the crash costs of the possible injuries has very little effect on the overall ranking of high crash locations.

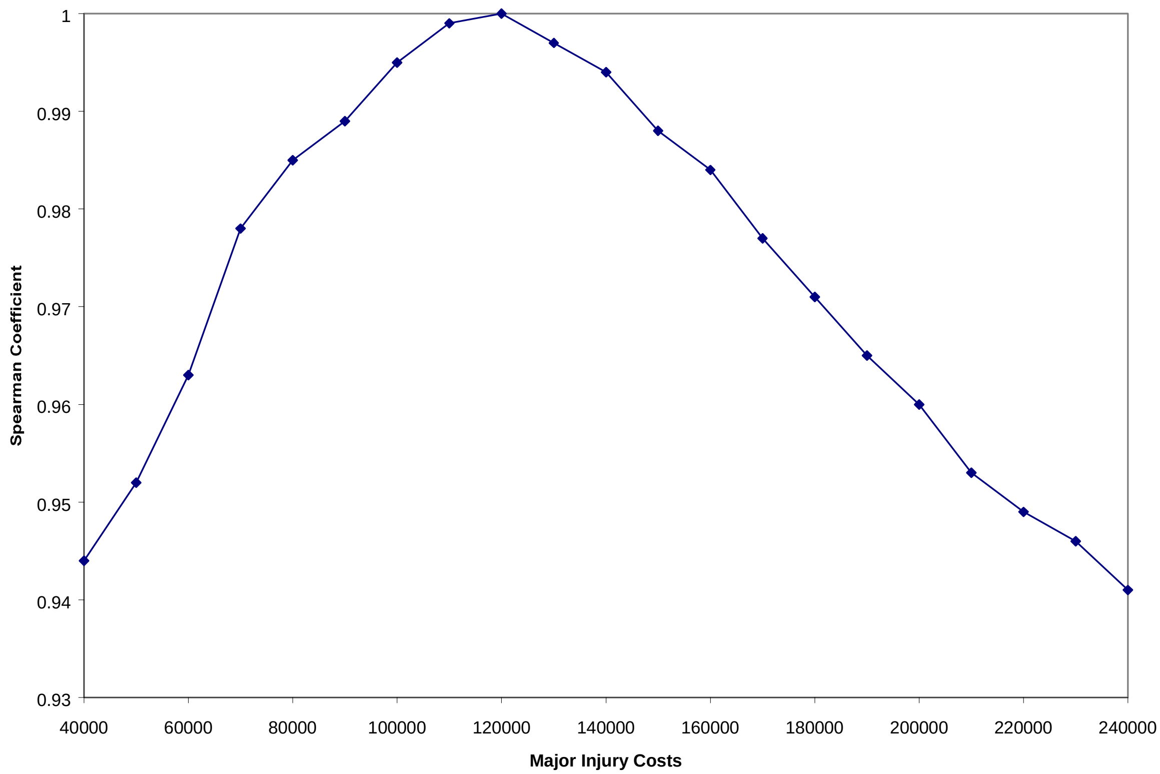

Figure 4-7 Sensitivity of Major Injury Costs in Iowa's High Crash Locations Process

4.5.4.5 Discussion of Sensitivity Analysis Findings

Overall the impact of changing each individual injury crash cost while holding the remaining values constant did not have a great impact on the overall ranking of the high crash locations. However, the underlying concept of Iowa's high crash location ranking approach does not allow one specific crash characteristic to dominate the ranking of locations. This is clearly evident from examining each individual impact of changing the costs associated with injuries, as presented in Sections 4.5.4.1 through 4.5.4.4.

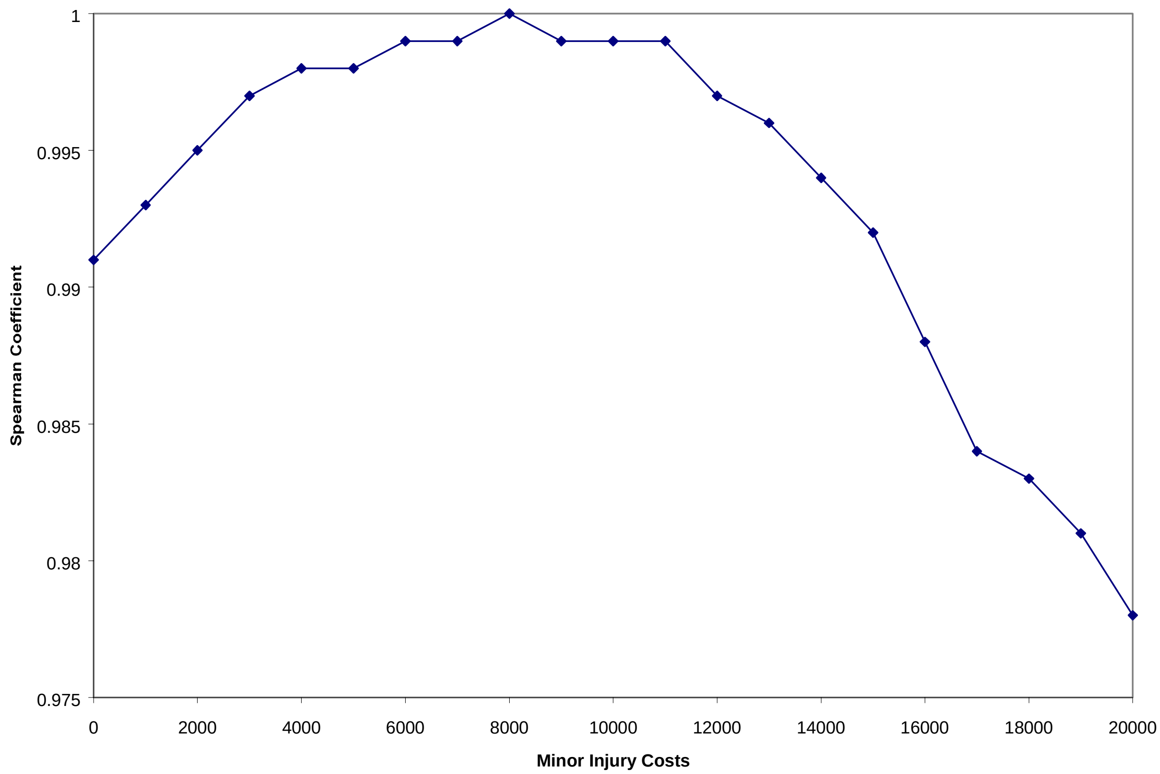

Figure 4-8 Sensitivity of Minor Injury Costs in Iowa's High Crash Locations Process

Iowa's high crash location process utilizes three measures to determine the high crash locations: crash frequency, crash rate, and value lost. Assuming all significant crashes are properly recorded, and that traffic levels can be accurately estimated, the first two components are relatively straightforward. However, the third measure relies on values which vary by source.

As presented in Section 3.7, the overall ranking placed on a deficient site is obtained through the sums of the respective ranking of the crash rate, frequency, and value lost. If fatal and injury crashes are not assigned a value, this parameter is eliminated from the ranking procedure. The impact of not using the value loss in the analysis can be examined

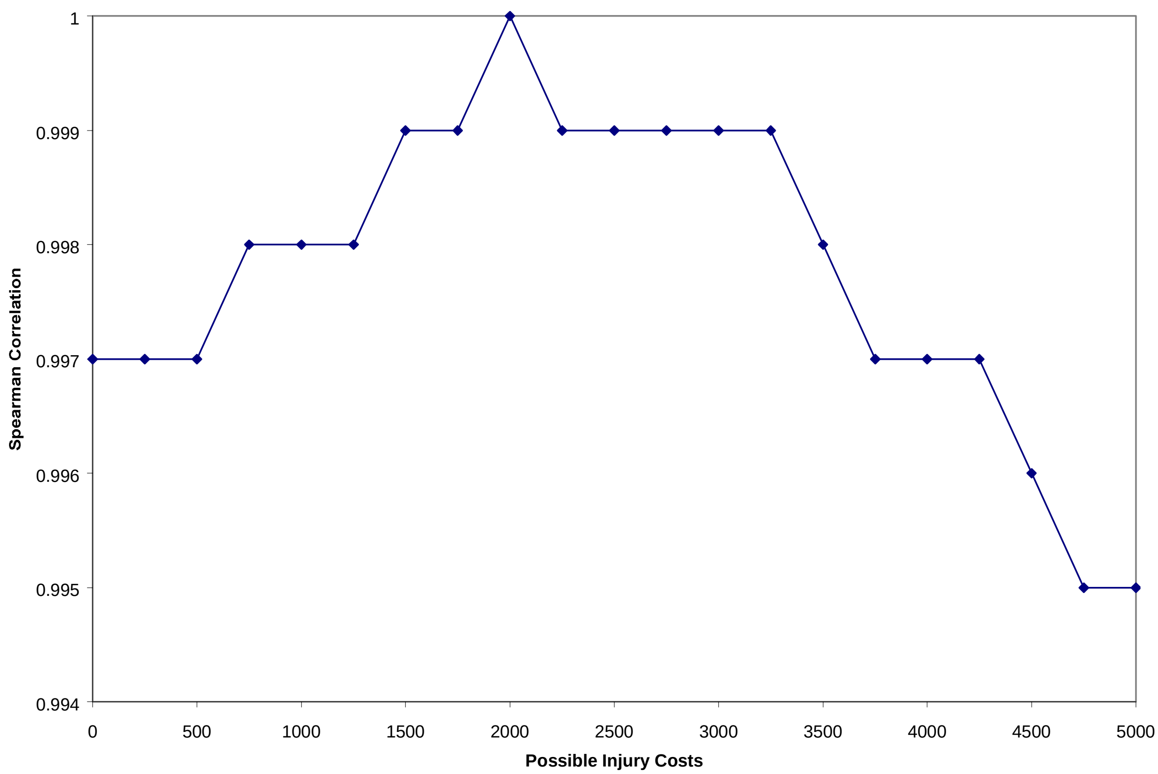

Figure 4-9 Sensitivity of Minor Injury Costs in Iowa's High Crash Locations Process

using the Spearman's rank order correlation coefficient. When no value was assigned to the injuries the Spearman's rank order correlation coefficient was 0.78.

This finding, however, does not represent a flaw in the approach to identify high crash locations. Instead it minimizes the impact of a variable that is difficult to measure, thereby relying more on information that does not require any judgment in determining the correct values to use for analysis. The Iowa DOT's approach incorporates the crash rate, crash frequency, and value loss into a comprehensive analysis approach used to identify high crash locations within the state of Iowa.

CHAPTER 5 CONCLUSION

The stated objective of this report was to evaluate medical crash costs in use by the Iowa DOT and FHWA, and to investigate the impact changing these costs would have on safety project prioritization. This report provides an analyst with an approach that better estimates crash costs associated with injuries sustained in motor vehicle crashes. The approach is readily transferable to other states, though the benefit of improved crash cost estimates on the identification of unsafe locations depends on the state's ranking process.

The final chapter of this report is presented in three sections. First, the results from this research are presented. Suggestions for additional applications of the CODES database follow. Finally, recommendations are presented to facilitate the probabilistic linking process for future studies.

5.1 Summary of Findings

To recapitulate, this research used Iowa's CODES database and crash records to address four objectives.

- To determine the average motor vehicle crash costs for fatalities and injuries occurring in the state of Iowa.
- To determine the medical costs based on speed limit and collision type
- To compare the actual medical costs with those recommended by the Federal Highway Administration (FHWA) and Iowa Department of Transportation (DOT) .
- To evaluate the impact on Iowa's high crash location process of using new estimates of medical crash costs.

Four hundred seventy-five files containing medical charges were probabilistically linked to crash records. Using the linked records, average medical charges were determined based on injury classification, type of collision, and speed limit. These values and their comparison to FHWA and Iowa DOT values are presented in Tables 5-1 through 5-3. Results demonstrate significant differences amongst FHWA, Iowa DOT and the CODES data estimates.

Table 5-1 Average Medical Charges by Speed Limit

Speed Limit	Mean
25	$7,500
35	$7,500
45	$8,000
55	$13,500
65	$8,500

Table 5-2 Average Medical Charges by Collision Type

Collision Type	Mean
Broadside	$10,000
Head On	$15,000
Parked Vehicle	$6,000
Rear End	$7,000
Side Swipe	$8,000
Single	$11,000
Other	$11,000
Unknown	$10,000

Table 5-3 Average Medical Charges by Crash Severity

Crash Severity	FHWA	Iowa DOT	CODES
Fatal	$43,715	$12,000	$12,000
Major Injury	$48,902	$27,000	$16,000
Minor Injury	$10,765	$1,800	$6,000
Possible	$5,850	$450	$5,000

While the costs are different, the benefit of improving/revising crash cost estimates is marginal as show in (Spearman's correlation coefficient of 0.94 showed the rankings before and after improvement to vary little.) Even when FHWA fatality estimates of $2.8 million are used in place of the $800,000 figure used by the Iowa DOT, resulting rank orders vary only slightly. However, while rank order changes little, stated public opposition towards the "low value" previously assigned to fatality costs may not be warranted. Alternatively, should the DOT wish to bolter public confidence in analysis predicated on value of life, this analysis has shown that analysts may indeed increase these values without undue impact on the final decision-making process.

5.2 Future Research

The probabilistic linkage methodology used in this report, and the resulting database, can be used to support further research. Specifically, CODES data can be linked to other information to study causal relationships. The data can also be used in other contexts, such as emergency response analysis. Finally, other types of non-medical crash costs can be studied through additional linkages. These are outlined briefly below.

5.2.1 Extension of Analysis

One approach to using the CODES database was presented in this report; however, several additional analyses can be performed. For example, evaluation of the hazards associated with roadside objects are possible using spatial processing and statistical techniques.

Figure 5-1 Ranking Based on a 0.94 Spearman Correlation Coefficient

Initial Ranking	New Ranking
1	2
2	3
3	1
4	6
5	4
6	9
7	7
8	5
9	12
10	13
11	14
12	21

Additional information can be obtained by using video-logs of roadway sections to enhance the analysis.

Additional information presents opportunities to develop crash prediction and cost models. Using medical and crash records, the expected number of injuries, crashes, and resulting costs can be determined based on several variables (such as restraint use, speed limit, age, gender, roadway geometry, vehicle types, and fixed object locations). Using the expected number of crashes in conjunction with economic impact of crashes presents a new approach for analyzing crashes in Iowa.

Finally, regional trauma centers are the only source of the hospital charges used in this analysis. Clearly, expanding the data set to be inclusive of all hospitals statewide will increase the linkage rate. Once the database expands the results of this research should be validated.

5.2.2 Emergency Medical Service Analysis

The CODES database is not restricted to motor vehicle crashes; emergency medical service can also be analyzed. Relating the injury outcome as a function of time is a difficult task. However, with the CODES database the time the incident occurred is now linked to the times the medical service arrived at the scene, departed the scene, and arrived at the hospital. This new information presents an opportunity to analyze the relationship between the response time, injury severity, and final outcome.

5.2.3 Other Crash Cost Determinations

As shown previously in Figures 2-1 and 2-2, there are several components responsible for the total crash costs. Providing a good estimate for a small portion (e.g., medical charges) of the total crash costs is the initial step in obtaining an accurate estimate. Expanding the probabilistic linkage technique allows other sources of information to be joined with the crash record. Estimates of lost earnings, market production, and insurance costs can all be improved once the data are linked.

5.3 Recommendations

This section recommends a change to enhance the linkage process. This change will result in an increase of linked files. Iowa's crash report form has not been changed for 20 years, though some revisions will occur in the future. Addressing the information collected at the crash scene will facilitate the linkage of hospital and crash records

The Model Minimum Uniform Crash Criteria (MMUCC) is a publication presented by NHTSA that recommends 113 items that can be collected or derived from other collected data elements. The MMUCC recommendations particularly address variables that provide discriminating and linking potential. Currently, Iowa only collects personal information about the drivers involved in a motor vehicle crash. However, collecting additional information, such as the name and social security numbers of the passengers involved in the motor vehicle crash would assist in linking the crash records to hospital records.

Therefore, when Iowa revises the crash report special attention should be given to adding discriminating variables. If Iowa follows the recommendations the number of records linked should increase.

REFERENCES

1 Andrzej P. Tarko, Shyam Eranky, Kumares C. Sinha, Rodian Scinteie, *An Attempt to Develop Crash Reduction Factors Using Regression Technique*, presented at the 78th Annual Meeting of Transportation Research Board Washington, D.C, January 1999.

2 *Traffic Safety Facts 1997*, National Highway Traffic Safety Administration's National Center for Statistics and Analysis, Washington D.C.

3 Khisty, C. Jotin and B. Kent Lall. *Transportation Engineering An Introduction*. Second Edition, Prentice Hall, Upper Saddle River, New Jersey, 1990 pp. 664-665.

4 Souleyrette, Reg, Tim Strauss, Michael Pawlovich, and Brad Estochen. *Safety Data in a GIS Environment: New Tools for the Four Es*. 24th International Forum on Traffic Records and Highway Information Systems, Minneapolis, MN, July 26-28, 1998.

5 Ezra Hauer, Identification of Sites with Promise, *Transportation Research Record 1542, Statistical Methods and Accident Analysis for Highway and Traffic Safety*, Transportation Research Board/National Research Council, Washington D.C., 1996, pp. 54-60.

6 Barbaresso James C., Brent O. Bair, Christopher R. Mann, and Gary Smith, Selection Process for Local Highway safety Projects, *Transportation Research Record 847, Analysis of Highway Accidents, Pedestrian Behavior, and Bicycle Program Implementation*, Transportation Research Board/National Research Council, Washington D.C., 1982, pp. 24-29.

7 Zegeer, C.V. *Method for Identifying Hazardous Highway Elements*, National Cooperative Highway Research Program Synthesis 128, NCHRP, 1986.

8 Tamburri, T.N. and R.N. Smith. The Safety Index: Method of Evaluating and Rating Safety Benefits. *Highway Research Record 332*. National Research Council, Washington D.C., 1970, pp. 28-29.

9 Jorgensen, N.O. *Statistical Detection of Accident Blackspots*, Proceedings from OTA-PIARC 11th International Study Week in Transportation and Safety, Brussels, Belgium. 1972.

10 Taylor, J.I. and H.T. Thompson. *Identification of Hazardous Locations*. Report FHWA-RD-77-81. Federal Highway Administration, December 1977.

11 McGuigan D.R.D. The Use of Relationships Between Road Accidents and Traffic Flow in ‘Black Sport’ Identification. *Traffic Engineering and Control*. February 1982, pp. 60-65.

12 Flak, M.A. and J.C. Barbaresso. Use of Computerized Roadway Information System in Safety Analysis. *Transportation Research Record 975, Analysis and Management of Traffic Accident Records.* Transportation Research Board/National Research Council, Washington D.C., 1982, pp. 50-55.

13 Hauer, E and B.N. Persaud. Problem of Identifying Hazardous Locations Using Accident Data. *Transportation Research Records 975, Analysis and Management of Traffic Accident Records.* Transportation Research Board/National Research Council, Washington D.C., 1984 pp. 36-43.

14 Hauer E. Empirical Bayes Approach to the Estimation of Unsafety: The Multivariate Regression Approach. *Accident Analysis and Prevention* Vol. 24 No. 5, October 1992, pp. 456-478.

15 DeGarmo, E. Paul, William G. Sullivan, and James A. Bontadelli *Engineering Economy,* Ninth Edition, MacMillan Publishing Company, New York, New York 1993, pp. 510-525.

16 Pfefer, Ronald C., Richard A. Raub, and Roy E. Lucke. *Highway Safety Data: Costs, Quality, and Strategies for Improvement, Final Report*, FHWA RD-96-192, United States Department of Transportation/Federal Highway Administration, January 1998.

17 *Identification, Analysis and Correction of High Accident Locations*, Missouri Highway and Transportation Department, Rolla, Missouri, Second Edition 1990.

18 Tarko, Andrzej, Eranky Shyam, and Sinha Kumares. *Methodological Considerations in the Development and Use of Crash Reduction Factors*. presented at the 77th Annual Traffic Research Board Annual Meeting, Washington D.C., January 1998.

19 Ivan, John N., Christian F. Davis, Fei Yuan, and Norman W Garrick. *Estimating Benefits from Specific Highway Safety Improvements: Phase 1 -Feasibility Study* Proceedings from the 78th Annual Traffic Research Board Annual Meeting January 1999.

20 Tarko, Andrzej P., Eranky Shyam, Sinha Kumares, and Rodian Scinteie. *An Attempt to Develop Crash Reduction Factors Using Regression Technique*. presented at the 78th Annual Meeting of Transportation Research Board Washington, D.C, January 1999.

21 Wright Paul H., and Norman J. Ashford. *Transportation Engineering Planning and Design*, Third Edition, John Wiley & Sons, New York, New York, 1989 pp. 344-366.

22 Blincoe, Lawrence J., and Barbara M. Fagin. *The Economic Cost of Motor Vehicle Crashes 1990*. Report DOT HS 807 876. Federal Highway Administration, National Highway Traffic Safety Administration, September 1992.

23 Forkenbrock, David, Norman S. Foster, and Thomas F. Pogue. *Safety and Highway Investment*. University of Iowa Public Policy Center for the Midwest Transportation Center, June 1994.

24 Rollins, John B., and William F. McFarland. Costs of Motor Vehicle Accidents and Injuries. *Transportation Research Record 1068, Traffic Accident Analysis, Alcohol Involvement, and Evaluation of Highway Improvements*, Transportation Research Board, National Research Council, Washington D.C., 1986, pp.1-7.

25 Rice, DP and MacKenizie, E.J. *Cost of Injury in the United States, A Report to Congress 1989*. San Francisco, CA: Institute for Health and Aging, University of California and Injury Prevention Centers, The John Hopkins University, 1989.

26 Blincoe, Lawrence, *The Economic Cost of Motor Vehicle Crashes 1994*. Report DOT HS 808 425 National Highway Traffic Safety Administration, Washington DC, 1996.

27 Miller Ted R, Charles Calhoun, W. Brian Kraugh, and Charles Zegeer. *The Cost of Highway Crashes*, Federal Highway Association, Publication Number FHWA - RD-91-055.

28 Miller, Ted, Brooke Whiting, Brenda Kragly, and Charles Zeeger, Sensitivity of a Highway Safety Resource Allocation Model to Variations in Benefit Computation Parameters, *Transportation Research Record 1124, Transportation Needs, Priorities, and Financing.* Transportation Research Board, National Research Council, Washington D.C., 1987, pp. 58-65.

29 Miller, Ted, David Levy, Rebecca Spicer, and Diance Lestina. Allocating the Costs of Motor Vehicle Crashes Between Vehicle Types, *Transportation Research Record 1635, Highway Safety Modeling, Analysis, and Design.* Transportation Research Board, National Research Council, Washington D.C., 1998, pp. 81-87.

30 *Final Model Minimum Uniform Crash Criteria*, Report DOT HS 808 745. National Highway Traffic Safety Administration/Federal Highway Administration/National Association of Governors' Highway Safety Representatives, August 1998.

31 National Highway Traffic Safety Administration Crash Outcome Data Evaluation System homepage http://www.nhtsa.dot.gov/people/ncsa/Codes/codes2.html. March 9, 1999.

32 National Highway Traffic Safety Administration. *Why Data Linkage?* DOT HS 808 461. Washington, DC: Department of Transportation, October, 1996.

33 Johnson, Sandra W. and Walker, Jonathan. *NHTSA Technical Report: The Crash Outcome Evaluation System (CODES)*. DOT HS 808 338. Washington, DC:

Department of Transportation, National Highway Traffic Safety Administration, January, 1996.

34 Johnson, Sandra W. *So You Want To Link Your Data*. DOT HS 808 426. Washington, DC: Department of Transportation, National Highway Traffic Safety Administration, July, 1996.

35 Jaro Matthew, Probabilistic Linkage of Large Public Health Data Files. *Statistics In Medicine*, Vol. 14, 1995 pp. 491-498.

36 National Highway Traffic Safety Administration. *Revised Catalog of Types of Applilcations Implemented Using Linked State Data*. Report DOT HS 908 793. United States Department of Transportation, National Highway Traffic Safety Administration, October 1998

37 Ned Levine, Karl Kim, and Lawrence Nitz, Spatial Analysis of Honolulu Motor Vehicle Crashes: I Spatial Patterns, *Accident Analysis and Prevention*, Vol. 27. No. 5 pp. 663-674, 1995

38 Ned Levine, Karl Kim, and Lawrence Nitz, Spatial Analysis of Honolulu Motor Vehicle Crashes: II Zonal Generators, *Accident Analysis and Prevention,* Vol. 27. No. 5 pp. 675-685, 1995

39 Karl Kim, Lawrence Nitz, James Richardson, and Lei Li, Analyzing the Relationship between Crash Types and injuries in Motor Vehicle Collisions in Hawaii, *Transportation Research Record 1467, Traffic and Roadway Accident Analysis and Traffic Records Research*, Transportation Research Board, National Research Council, Washington D.C., pp. 9

40 Karl Kim and Lawrence Nitz, Understanding Causes of Injuries in Motor Vehicle Crashes, *Transportation Quarterly*, Vol. 50(1), 1996, pp. 105-113.

41 Kim, Karl, Lawrence Nitz, James Richardson, and Lei Li. Personal and Behavioral Predictors of Automobile Crash and Injury Severity. *Accident Analysis and Prevention* Vol. 27 No. 4, 1995, pp. 469-481.

42 Richardson, James, Karl, Kim, Lei Li, and Lawrence Nitz. Patterns of Motor Vehicle Crash Involvement by Driver Age and Sex in Hawaii. *Journal of Safety Research* Vol. 27, No. 2, 1996, pp. 117-125.

43 Zegeer Charles V., Herman F. Huang, J. Richard Stewart, Ron Pfefer, and Jung Wang. Effects of a Towaway Reporting Threshold on Crash Analysis Results, *Transportation Research Record 1635, Highway Safety Modeling, Analysis, and Design.* Transportation Research Board, National Research Council, Washington D.C., 1998, pp. 49-57.

44 Hans, Zachary N., *A Geographic Information System for Transportation Modeling, Alternatives and Policy Analysis*, Master's Thesis. Iowa State University, Ames, Iowa, 1994.

45 Simkowitz, Howard J. 1989, Geographic Information Systems: An Important Technology for Transportation Planning and Operations. *Transportation Research Record 1236, Transportation Data 1989*, Transportation Research Board, National Research Council, Washington D.C., 1989, pp. 14-22.

46 Miller, John S. Using Technology to Help Overcome Institutional Obstacles to Improved Crash Records Processing, *Transportation Research Record 1581*, *Safety and Human Performance/Traffic Records. Accident Prediction and Analysis, and Statistical Methods*. Transportation Research Board, National Research Council, Washington D.C., 1997, pg. 16.

47 Saccomanno, F.F., K.C. Chong, and S.A. Nassar. Geographic Information System Platform for Road Accident Risk Modeling. *Transportation Research Record 1581, Safety and Human Performance/Traffic Records. Accident Prediction and Analysis, and Statistical Methods*. Transportation Research Board, National Research Council, Washington D.C., 1997, pp. 18-26.

48 Harky, David L. *Evaluation of Truck Crashes Using a GIS-Based Crash Referencing and Analysis System*, presented at the 78th Annual Transportation Research Board Annual Meeting, Washington D.C., January 1999.

49 Quiroga, Cesar A. and Darcy Bullock. Geographic Database For Traffic Operations Data. *Journal of Transporation Engineering,* May/June 1996, pp. 226-234.

50 Souleyrette, Reg, Tim Strauss, Brad Estochen Michael Pawlovich. *GIS-Based Accident Location and Analysis System (GIS-ALAS) Project Report: Phase 1*, CTRE Management Project 97-6, April 1998, pg.. 1.

51 U.S. Bureau of the Census, *Statistical Abstract of the United States: 1998* (118th edition.) Washington DC, 1998.