A GIS Accident System to Accompany CARE

By

Sampson E. Gholston, Ph.D.
Industrial and Systems Engineering and Engineering Management Department
The University of Alabama in Huntsville
Huntsville, Alabama

and

Michael D. Anderson, Ph.D.
Civil and Environmental Engineering Department
The University of Alabama in Huntsville
Huntsville, Alabama

Prepared by

UTCA

University Transportation Center for Alabama
The University of Alabama, The University of Alabama in Birmingham,
and The University of Alabama at Huntsville

UTCA Report 03304
June 30, 2005

Technical Report Documentation Page

1.Report No FHWA/CA/OR-	2.Government Accession No.	3.Recipient Catalog No.

4.Title and Subtitle A GIS Accident System to Accompany CARE	**5.Report Date** June 2005
7.Authors Sampson E. Gholston and Michael D. Anderson	**8. Performing Organization Report No.** UTCA Report 03304
9.Performing Organization Name and Address Industrial System Engineering and Engineering Management Department, and Civil and Environmental Engineering Department University of Alabama in Huntsville Huntsville, AL 35899	**10.Work Unit No.** **11. Contract or Grant No.** DTSR0023424
12.Sponsoring Agency Name and Address University Transportation Center for Alabama The University of Alabama Box 870205, 207 Shelby Hall Tuscaloosa, AL 35487-0205	**13.Type of Report and Period Covered** Final Report/ 01/01/2003 – 12/31/2003 **14.Sponsoring Agency Code**

15.Supplementary Notes

16.Abstract

This project developed a geographic information system (GIS) interface to access the crash data exported using the CARE software package. With advanced spatial query and display capabilities, the GIS-based system will enhance crash data query and display features. The system includes instant graphical access, enabling viewing and selecting of desired network locations (nodes and links). In addition to the interface development, this project examined the quality of incorporating data extracted using CARE on to a roadway network. The quality of data within the system was evaluated to see if the system was providing accurate tools for analysis to ensure proper decisions regarding application of safety funds for roadway improvements. The final focus of this project was examination of tools and statistical procedures to identify high accident locations within the case study counties.

17.Key Words CARE, Accident Analysis, Geographic Information Systems	**18.Distribution Statement**

19.Security Class (of this report)	20.Security Class (of this page)	21.No. of pages	22.Price

Contents

List of Tables

List of Figures

Executive Summary

Accident analysis and infrastructure investment decisions require accurate, reflective data from which to draw conclusions. This project focused on the ability to develop, verify and use a new integrated system for accident analysis based on two existing technologies. The inputs to the new system were the Critical Accident Reporting Environment system (CARE) software, which has been used for accident data manipulation and analysis, and geographic information systems (GIS), which provide spatial data display and analysis.

The combined CARE-GIS system developed as part of this project can display accidents at specific locations down to the street segment level. A statistical analysis of the CARE-GIS system was performed to demonstrate the accuracy of the system, which was shown to correctly locate a very high percentage of accidents. The CARE-GIS system was used to develop a visual map of high accident locations for Huntsville, AL using street segment specific data. Overall, this project demonstrated the development of tools to better visualize accident locations and methodologies to make better roadway infrastructure investment decisions based on the additional information.

Section 1
Introduction

The ability to analyze accident data and historical records and make recommendations to improve transportation infrastructure is important. These analyses are used to identify problem areas and allocate limited safety improvement dollars to make the most significant impact on roadway infrastructure.

Currently, Alabama uses the Critical Accident Reporting Environment system (CARE system) for accident analysis (http://care.cs.ua.edu/). CARE uses advanced analytical and statistical techniques to generate valuable information directly from the data. By following the step-by-step menus outlined on the screen, CARE is easy to use. The CARE system is capable of developing trends and over/under-representations of traffic accidents on a county level. An example output from the CARE system is shown in Figure 1-1. This image contains accident data for a test county for the year 2003. The analysis of the accidents has been further broken down to identify accident severity: the number of property damage, injuries, and fatality accidents.

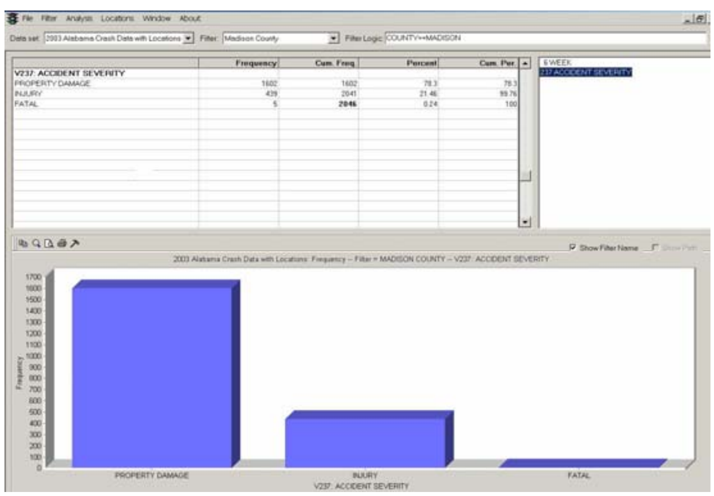

Figure 1-1. Example report generated from CARE

CARE provides a variety of accident details, however, there is no information about local roads, i.e., the spatial coordinates or specific locations where the accident occurred on local roads. This information is provided for accidents that occurred on state roads or the Interstate Highway System. Unfortunately, the existing CARE system is not capable of displaying accident locations on individual roadways for crashes located off of this system.

The goal of this project was the development of a spatial interface for the accident data exported by the CARE system for enhanced visual display and analysis. The interface utilizes geographic information system (GIS) technology to add graphical visualization for the accident data exported from the CARE system onto individual roadways. GIS is a computer system for capturing, storing, querying, analyzing, and displaying geographic data (Chang, 2002). In recent years GIS has been used in emergency planning, market analysis, transportation planning, and military applications. The ability to handle and process geographically referenced data (which describes both the location and characteristics of spatial features on the Earth's surface) distinguishes GIS from other information systems.

The data extracted using CARE and combined with GIS technology allow the identification of locations where crashes are overrepresented within a community, and offer the ability to graphically identify accidents and query the accidents by type and cause. The incorporation of the data extracted using CARE into a GIS environment, and analysis of the system developed in this project was undertaken through a series of tasks. First, a system was designed to allow the exported data to be displayed on a roadway network. This process required the addition of spatial coordinates to the accident records. In addition to developing a graphical interface for the accident data, this project included a statistical test to measure the accuracy of such a system. Several case study locations were examined to determine the accuracy of the system. This accuracy analysis was an important step to determine the merit of the system and to allow recommendations regarding the use of the system within Alabama. The final aspect of this project is the examination of a methodology that will work in conjunction with the CARE-GIS integrated system for the identification of overrepresented locations. This identification methodology will allow individual counties to best determine their sites for safety improvement and to provide support data on which to base transportation infrastructure safety improvement decisions.

The report contains five sections. The first section introduces the technology and outlines the tasks involved in the research project. The second section documents the system developed to integrate the data extracted using CARE into a GIS environment for visualization and analysis. The third section examines the accuracy of the model developed in the second section through statistical testing. This section also extrapolates the quality of data to be expected if the system developed in the second section is applied to the entire state. The fourth section examines a methodology to identify high accident locations from the CARE-GIS system. The final section presents the conclusions of this research effort.

Section 2
Design of the System

The first task of this research focused on the design of a system to incorporate accident data from the CARE system and display the accidents graphically in a GIS.

Traffic accidents in Alabama are investigated by law enforcement officers, who are responsible for completing an accident analysis form to record information about the accident cause, individuals affected, roadway and weather conditions, and most importantly for the system, the location of the accident (Department of Public Safety). For accidents that occur on the city or county roads, the accident location is defined through a node-link system. The officer collecting the accident information records the location of the accident through this node-link identification methodology. If an accident occurs at an intersection, the officer merely records the node number of the intersection on the accident reporting form. If an accident occurs along a roadway, the officer reports the link number and provided the distance to the closest node. The node numbers used by officers to locate accidents in metropolitan areas across the state have been previously defined and printed on county and city maps. However, the existing mapping system for the node maps was developed in a CAD environment; therefore, the maps do not contain locations for intersection nodes relative to any real-world coordinate system (Doolin interview, 2003). This is a limitation to using the maps as a direct method for developing a graphically based system that needed to be overcome.

The design of the interface used the accident node location information provided by the data extracted using CARE with spatial locations in the GIS environment. To accomplish this, it was necessary to develop a GIS table with the node locations for the test cases identified and correctly placed within a real-world coordinate system. This process was performed manually in the GIS environment by obtaining CAD node maps for selected cities and counties within Alabama. Spatially accurate roadway data, containing graphics and attributes by route name, were obtained from the United States Census Department as TIGER files. These data are available free of charge, which allows the system to be economical, and in ArcIEW format, commonly used GIS software within Alabama. An example of the TIGER data is shown in Figure 2-1.

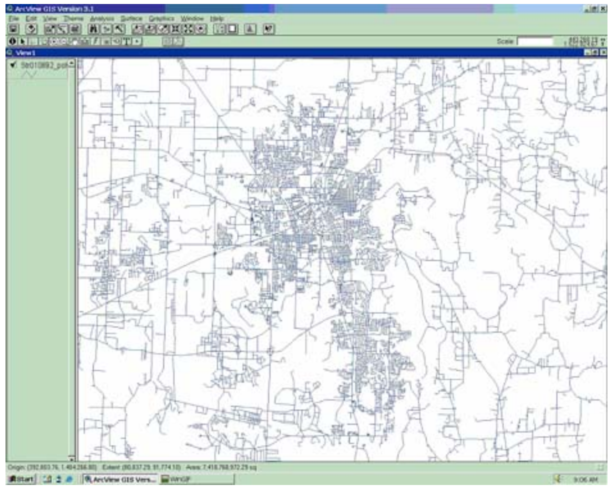

Figure 2-1. TIGER roadway file

After obtaining the GIS file with the roadway network from the Census Department, a process of heads-up digitizing was performed to transfer the node locations from the paper maps to the GIS. This process involved visual identification of the node from the paper map, and then transfer of the node to the corresponding location in the GIS environment. The individual performing the digitizing effort would then place a point in the appropriate location and attribute the node with the node number from the paper map. This heads-up digitizing was performed to generate spatially accurate node locations from which to incorporate the accident from the CARE system. Unfortunately, there are spatial errors in the paper maps and GIS data layers used to develop the node locations. However, as these items represented convenient and inexpensive tools, they were used in the project. Figure 2-2 shows the nodes digitized into GIS.

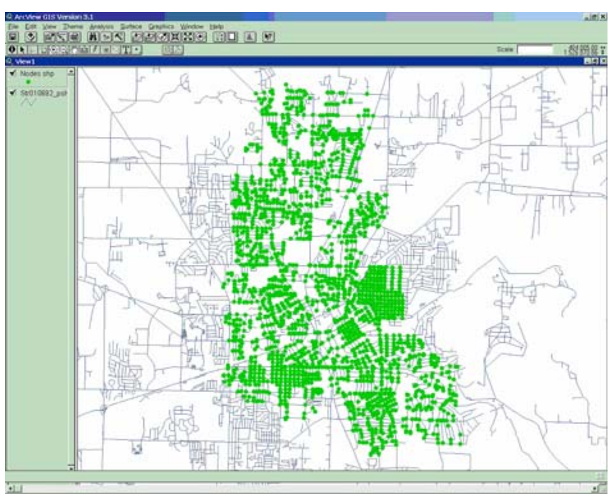

Figure 2-2. Nodes digitized to the roadway file

Once the nodes were located and attributed in the GIS, the next step was to add additional attribution that contained the X and Y coordinates for the node points. This step was performed through a simple process using standard GIS functions. After the nodes were attributed with the spatial coordinates, the data were exported into a text file containing the node number and coordinates. This text file is known as the node location file.

A separate text file of accident data was also developed from the CARE system. The text file contained the case number for the accident (important for linking the accident record to the entire collection of data recorded for the accident) and the two node numbers used to identify the accident location (as stated previously, if the accident occurred at an intersection, only one node number was reported).

The two separate text files form the entry information into a FORTRAN program written to link the two files and to add spatially accurate coordinate information to the case numbers for the accidents, essentially providing a location for each specific accident. The program was written to read the accident case number and node numbers and determine if the accident occurred at an intersection or along a roadway segment. If an intersection accident is determined through the listing of only a single node number, the program reads the node location file to identify the coordinates for the accident and writes the information into an intersection accident file. If a roadway segment accident is determined, the program reads the coordinates for the beginning and ending node from the node location file, interpolates between the two coordinates, and writes the information into a link accident file. A drawback to the system is that the program does not use more advanced information than the node numbers between which the accident occurred, and is therefore unable to place the accident with any more precision than a point taken to be mid-way between the nodes. This methodology was selected because the officer’s coding the distance between accident and the reference node was assumed to be suspect, and perhaps contain large spatial inaccuracies.

Once the intersection accident file and link accident file are developed, they can be imported into the GIS environment and the locations of the accidents can be displayed graphically through standard GIS commands. A screen shot of the intersection accident file and link accident file displayed in GIS is provided in Figure 2-3 for accidents occurring between 1993 and 2002.

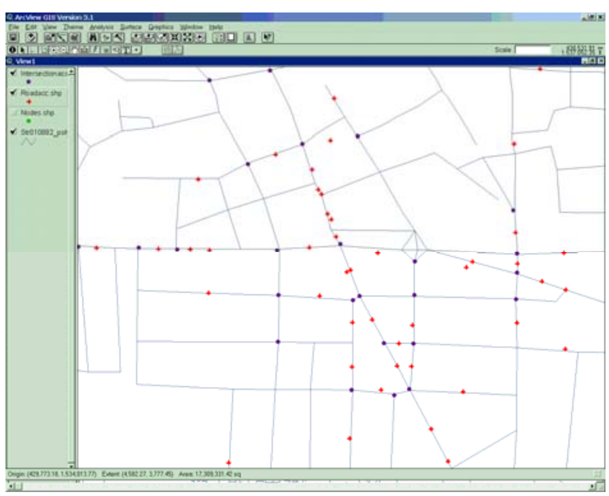

Figure 2-3. Accident locations in GIS

With the accidents correctly displayed and incorporated into the GIS through the two files, the details of the accidents can be added to the information using the case number of the accident as a key field. This implies that the case number in the graphic files can be matched with the same case number from an export of the data extracted using CARE. This allows the user to begin making queries on the data to determine the locations for specific accidents and to begin to analyze the causes behind the accidents with knowledge of the accident location. An example query using the system is shown in Figure 2-4.

Figure 2-4. Accidents where the cause was related to driver condition

Section 3
Quality of the Data

The second theme of this project was the evaluation of the system (described in the previous section) to integrate crash data into the GIS environment. As presented, the CARE-GIS system was based on the use of free data from the Census Department for the roadway graphics and the FORTRAN program written to combine the accident case number and coordinates for use in the GIS software. The statistical analysis of the system was based on a subset of accidents from three case study counties selected to as representative of rural Alabama. The system was developed for each of the three counties and the node locations were attributed in the GIS system through the heads up digitizing described in the previous section.

Node accidents took place at intersections. Since these nodes are attributed with spatial coordinates that are specifically obtained from the GIS system during the heads-up digitizing process, these accidents will be placed correctly by the system, provided they were correctly identified by the officer completing the accident report.

Link accidents took place on roadways between two nodes. The accuracy of these accidents was studied to determine if the CARE-GIS system could correctly place the location. The roadway accidents were segmented into four potential categories: on the roadway graphic/correctly placed between nodes, off the roadway graphic/correctly placed between nodes, on the roadway graphics/incorrectly placed between nodes, and off the roadway graphics/incorrectly placed between nodes. Specifics for the four categories are defined as follows:

- <u>On the roadway graphic/correctly placed between nodes.</u> These accidents represent those that were accurately attributed by the program and would plot correctly in the GIS environment.

- <u>Off the roadway graphic/correctly placed between nodes.</u> These accidents represent those that were accurately attributed by the program, but not plotted correctly in the GIS environment because of roadway graphics issues that were independent of the program. Essentially, this was caused by roadway alignment issues and linear interpolation between node numbers used by the program.

- <u>On the roadway graphic/incorrectly placed between nodes.</u> These accidents represent those that were not accurately attributed by the officer recording the information, but were correctly placed on the roadway according to the node locations.

- <u>Off the roadway graphic/incorrectly placed between nodes.</u> These accidents represent those that were not accurately attributed by the officer recording the information and were also not correctly placed in the roadway.

The accidents were analyzed using the four potential categories as a measure of accuracy of the system. If the accidents fell in the "on roadway graphic/correctly placed between nodes" or "off roadway graphic/correctly placed between nodes", then the data were deemed to be accurate and the program and CARE-GIS was deemed to be working correctly. If the accidents fell in the "on the roadway graphic/incorrectly placed between nodes" or "off the roadway graphic/incorrectly placed between nodes", then the data quality was deemed as poor and errors were logged affecting the system's ability.

The accidents for the three case study counties were analyzed as discussed previously. The results from the analysis are shown in Table 3-1.

Table 3-1. Analysis of data for the case study counties (proportions).

	County A	County B	County C
On the roadway graphic/correctly placed between nodes	0.43	0.5	0.41
Off the roadway graphic/correctly placed between nodes	0.55	0.38	0.48
On the roadway graphic/incorrect placed between nodes	0.01	0.03	0.09
Off the roadway graphic/incorrect placed between nodes	0.01	0.09	0.02

A confidence interval on the percentage of the accidents that would have been correctly placed was done for the 1,368,939 rural accidents recorded statewide during 1993 – 2002 and exported by CARE from Alabama traffic crash data.

A success was defined as either an accident point being placed on the roadway graphic correctly or off the roadway graphic correctly. Taking P as the proportion of accidents correctly placed in a random sample of size n, and q = 1 - p, an approximate 95% confidence interval on P is given by:

$$\hat{p} - Z_{\alpha/2}\sqrt{\hat{p}\hat{q}/n} < P < \hat{p} + Z_{\alpha/2}\sqrt{\hat{p}\hat{q}/n}$$

Where: The number of accidents in the three counties over the study period was n = 4481
x = 4131 (the number correctly placed)
$\hat{p}$ = x/n = 4131/4481 = 0.9218
$\hat{q}$ = 1 - $\hat{p}$ = 1 – 0.9218 = 0.0781
Level of Confidence = α = 0.05

$\Rightarrow 0.9139 < P < 0.9296$

Using the normal approximation and a confidence level of 0.05, we conclude that at the 95% confidence level the proportion of correctly placed accidents statewide would be between 0.9139 and 0.9296. The percentage of correctly located accidents would fit within the following interval:

(91.39, 92.96)

The application of the CARE-GIS system statewide would require an extensive manual digitizing effort to convert the accident node maps into a GIS-based coverage that could be used as input to the FORTRAN program. This manual digitizing effort would require approximately 3,000 person-hours to complete the GIS coverage for all 67 rural counties in Alabama and approximately 5,000 person-hours to complete the GIS coverage for the urban areas within Alabama. With considerable manual effort, it would be possible to develop a spatial visualization system that would incorporate all the accidents from Alabama traffic crash data using CARE. The decision to extend this study to the entire state will be left to personnel from the Alabama Department of Transportation.

Section 4
Analysis of Data System

The final task of this project was the use of the CARE-GIS package to develop a system to analyze the accident data to identify locations where traffic crashes are overrepresented. The identification of such locations is important as roadway accidents are often concentrated at particular roadway stretches or intersections and only after identifying these locations can appropriate counter-measures be undertaken to improve the conditions and reduce the number of accidents, therefore improving the safety of the roadway (Sarkar, 2001).

According to the 1997 Alabama Traffic Accident Facts, an Alabama driver has a 37.8% chance of getting injured or killed in a road accident (Alabama Traffic Accident Facts, 1997). Hence there is a need to analyze road accidents, find the overrepresented locations and determine the contributing factors. In this project, each roadway segment will have a unique severity value based in the total number of accidents and the resulting impact of the accidents. If one roadway segment records one hundred accidents and all the accidents involve minor injuries or property damage, then this roadway would potentially be safer than another roadway segment where only thirty accidents were recorded but they all involved fatalities or major injuries.

Using the CARE system, it is possible to compile many important accident statistics. For example, it is possible to identify the distribution of selected accident types and examine trends for accidents such as time of day, cause, and driver age. Using the CARE system and analyzing data for a sample County, it is possible to determine the distribution of accidents based on severity, as shown in Table 4-1.

Table 4-1. Accident severity distribution for a sample county

Year	Property Damage	Injuries	Fatalities	Total
1993	6,222	1,688	39	7,949
1994	6,406	1,940	37	8,383
1995	6,377	1,945	52	8,374
1996	6,273	1,871	26	8,170
1997	6,489	1,897	43	8,429
1998	6,830	1,994	39	8,863
1999	6,775	2,003	46	8,824
2000	6,703	1,936	41	8,680
2001	6,983	2,076	45	9,104
2002	7,521	2,174	39	9,734

However, identifying specific roadway segments with features that are over represented in accidents requires additional spatial information. These analyses have always been time-consuming and tedious, and they highly subjective with much room for human error (Collins, 2001). Some of the limitations in the analysis can be overcome by the use of a GIS system using the ability to capture, store, query, analyze, and display geographic data. From this understanding of the importance of GIS for accident analysis, the CARE-GIS system developed in section two of this report will be used in the analysis of high accident locations. Figure 4-1 shows a segment of roadways from the CARE-GIS system with accident locations displayed.

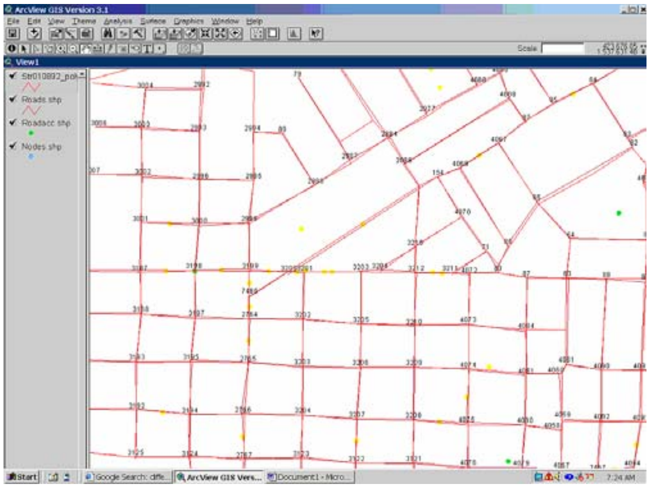

Figure 4-1. Display from CARE-GIS system

Using the CARE-GIS system, it was possible to determine the number and severity of the accidents along specific roadways. Each roadway segment will have a severity index. The methodology for computing the severity index will be based on four different types of crash severity.

1. Fatalities
2. Major injuries
3. Minor injuries
4. Property damage only.

To develop a roadway severity value for the individual roadway segments, it is necessary to combine the accidents into an aggregated total using the potential monetary value of the accident as a unifying factor. According to the Federal Highway Administration, the four different types of road accident severity levels and the associated costs are shown in Table 4-2 (Federal Highway Administration, 2003).

Table 4-2. Severity and equivalent costs

Severity Index	Equivalent Cost in Dollars
K Fatality	$3,000,000
A Major Injury	$500,000
B Minor Injury	$50,000
C Property Damage Only	$10,000

Since the cost involved in an accident depends on the severity of the accident, a severity index is given to each segment. A monetary crash cost was used to calculate the coefficients in the equivalent property damage only (EPDO) and severity index (SI) equations. To determine the coefficients for the EPDO/SI equations, the weighted average cost for the K (fatal) and A injury (major injury) and B and C injury (minor injury) and property damage only (PDO) must be computed (W. Benifield, ALDOT, 2003). The weighted average cost for the fatal and A injury crashes is computed by:

$$K\&A\,cost = \frac{(Cost_{fatalcrash})(\#\,fatal_{crashes}) + (Cost_{Acrash})(\#\,Ainjury_{crash})}{\#\,fatal_{crashes} + \#\,Ainjury_{crash}}$$

The coefficient of K and A cost is calculated by:

K& A coefficient = (K & A crash cost)/ PDO crash cost

The coefficient of B and C cost is calculated in the same manner. The EPDO is calculated by the following formula:

EPDO = K&A coefficient (K+A)+B&C coefficient (B+C) + PDO

The Severity Index (SI) is obtained by dividing the EPDO by the total number of crashes (N).

SI = [K&A coefficient (K+A)+B&C coefficient (B+C) + PDO]

The severity indices developed for all roadway segments from the case study County were calculated and displayed in the GIS system, as shown in Figure 4-2. The higher the severity index the greater the accident impact on the people of the community, based on the severity index. Hence, the severity indices were ranked from highest to the lowest. The segments having the top ranks (high indices) were considered as the hot spots. While this method uses the quantitative data currently recorded at accident scenes, future methodologies should aim for a wider analysis of the accident scene, which will take into account qualitative data. Some suggestions of qualitative data include the nature of the incident, the types of road users involved, or the existence of road safety awareness programs in the area.

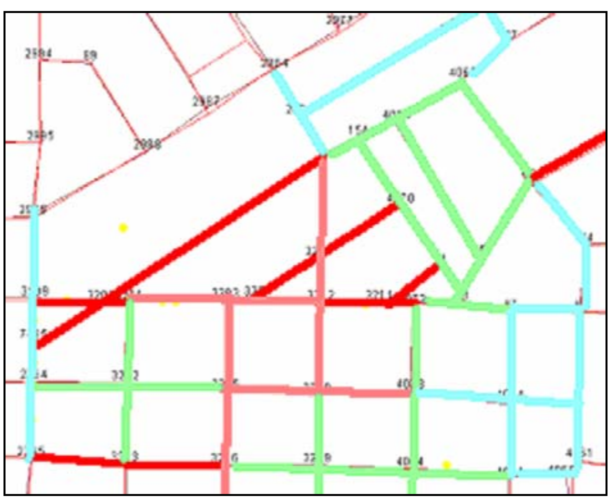

Figure 4-2. Display of severity indices

Section 5
Conclusions and Recommendations

This project focused on three main objectives, the development of a program to display data extracted using CARE in a GIS environment, a review of statistical quality of the accidents displayed in such a system, and a methodology to identify locations where accidents were overrepresented.

The program and procedure developed used CARE to extract and display the traffic accident data in GIS was a manually intensive procedure. The accident node numbers were stored in a CAD format without any real world spatial coordinates, so a great deal of time was required to perform the heads-up digitizing process to develop a coverage of accident nodes in GIS. This process also incorporated a level of potential human error, as the digitizing process is not guaranteed to be 100 percent successful. However, the digitizing process was performed with high success and the program developed to use the node locations recorded in the field was able to define accident locations spatially, at least to the extent of placing the accident between the appropriate nodes. Additionally, with the use of the accident number tagged to each accident, it was possibly to develop displays and queries for future use to better analyze the accidents within cities, and possibly develop enhanced countermeasures or educational tools to reduce accidents.

The statistical evaluation of the CARE-GIS system was performed on three case study counties. The results demonstrated that the combination of the FORTRAN program and readily available digital roadway maps can be successfully used to develop accident displays in the rural areas of the state. The accuracy correctly locating accidents and having reasonable roadway geometry from the data was shown to be approximately 92 percent for a group of case study accidents. This knowledge was beneficial as the CARE-GIS might prove to be an effective tool to map many years of previous accident records stored in the CARE system on a statewide basis.

The use of the CARE-GIS system for the development of the high accident location methodology was demonstrated using a case study city. The CARE-GIS was used to obtain the accident information necessary to support the analysis, as the CARE system alone is not able to disaggregate the accidents to the level desired for specific roadway analyses, for sites not on the Interstate or state highway system. After the accident information was obtained, the development of a severity index was performed using widely accepted statistical procedures and information provided by the Federal Highway Administration and Alabama Department of Transportation. The severity index calculation provided the ability to identify specific roadway segments that would be candidates for infrastructure improvements.

Overall, this project focused on the goal of developing tools and methodologies to potentially reduce accidents, and to make to our roadway safer, through the ability to better interpret accident records and to provide more information for individuals to evaluate accidents. The case studies performed in this research should demonstrate that these systems, though manual and time-consuming to develop, have the ability to assist in the ability to improve data analysis and investment on our transportation infrastructure.

Section 6.0
References

Alabama Traffic Accident Facts, Alabama Department of Transportation, Montgomery, AL, 1997.

Care Research and Development Laboratory, 2003, The University of Alabama, Tuscaloosa, AL, http://care.cs.ua.edu/

Chang, K. Introduction to Geographic Information Systems. McGraw-Hill, 2002

Collins, Andrew. "Using GIS to Identify crash Black Spots," Geographic Information Systems, Issue 50, Vol. 16, Number 2. September 2001.

Alabama's Uniform Traffic Accident Report, Instructional Manual for Investigating Officers. Department of Public Safety, State of Alabama. Revised 4/86.

Alabama Crash Investigation Node Maps, and other pertinent information, provided by James Doolin, Alabama Department of Transportation, 2003.

Information provided by safety section representatives from the Alabama Division of the Federal Highway Administration, Montgomery, Alabama, 2003.

Information Provided by Benifield Waymon, Multimodal Safety Section, Alabama Department of Transportation, Montgomery, AL, 2003.

Sarkar, A, K. International Forum for Rural Transportation and Development. Vol . 8, Issue 4, May 2001.