Beyond Crash Data: The Application of Multiple Datasets for Highway Safety Analyses

Heather Rothenberg, Michael A. Knodler, Jr. PhD

ABSTRACT

Traditional crash and fatality data have provided valuable information for use in examining highway safety. More specifically, these data have helped engineers understand crash characteristics in an effort to identify, implement, and evaluate countermeasures aimed at reducing crash frequency and severity. However, the application of these data is not without challenges as officers in the field are often asked to provide information on characteristics they are not fully able to assess at the roadside, including injury severity and roadway elements. Additional datasets including roadway inventory, crash location, hospital, and citation data can provide added information towards a more comprehensive understanding of the true relationship between drivers, vehicles, the roadway, and the resulting crashes. This study focuses on three critical elements in the application of expanded data for safety analysis: 1) data collection, 2) data compilation, and 3) data distribution.

Using the University of Massachusetts Highway Safety Data Warehouse as an example, the process of collecting 13 datasets used in highway safety data are outlined. Additionally, advanced methods for linking and employing multiple datasets are detailed and case study examples of the successful application of these advanced highway safety datasets are provided. These examples include successfully completed analyses of lane departure crashes, an examination of older driver issues, and an assessment of speeding behaviors.

INTRODUCTION

Motor vehicle crashes are a major public health problem in the United States (US). Each year over 40,000 fatalities occur along U.S. roadways resulting in costs exceeding $230 billion (National Highway Traffic Safety Administration, hereafter referred to as NHTSA). In 2005, 116,254 motor vehicle crashes occurred along Massachusetts roadways. These crashes resulted in 442 fatalities and 5,100 incapacitating injuries, totaling over 6.4 million dollars in damage. In addition to the lives lost and individuals injured in crashes, these events have financial impacts. In 2005, the costs associated with Massachusetts crashes were more than $6.4 million (1). To learn more about how and why these crashes occur, highway safety practitioners turn to crash data, and other related data sets, to gain insight that will guide decision making and program planning. Considerable resources have been allocated to improving crash data collection, transfer, access, and analysis.

In 2003, the National Highway Traffic Safety Administration (NHTSA) formed a multidisciplinary team consisting of members from NHTSA, the Federal Highway Administration (FHWA), Federal Motor Carrier Safety Administration (FMCSA), and the Bureau of Transportation Statistics (BTS) to examine how data can be used to achieve the Department of Transportation (DOT) safety goals. This team developed a list of recommendations to be implemented at both the federal and state levels towards the improvement of safety data. These recommendations were classified into five areas:

coordination and leadership; data quality and availability; electronic technologies and methods; uniform and integrated data; and facilitated data use.

Traditional crash and fatality data have provided valuable information for use in examining highway safety. More specifically, these data, collected by police officers using crash report forms, have helped engineers understand crash characteristics in an effort identify, implement, and evaluate countermeasures aimed at reducing crash frequency and severity. However, the application of these data is not without challenges as police officers in the field are often asked to provide information on crash characteristics they are not fully able to assess at the roadside, including injury severity and roadway elements. Additional datasets including roadway inventory, crash location, hospital, and citation can provide added information towards a more comprehensive understanding of the true relationship between drivers, vehicles, the roadway, and the resulting crashes. This paper focuses on three critical elements in the application of expanded data for safety analysis: 1) data collection, 2) data compilation and transformation, and 3) data access and distribution.

Using the Univeristy of Massachusetts Highway Safety Data Warehouse as an example, the process of collecting 13 datasets used in highway safety data are outlined. Additionally, advanced methods for linking and employing multiple datasets are detailed and case study examples of successful applications of these advanced highway safety datasets are provided. The examples provided in this paper include analyses of lane departure crashes, an examination of older driver issues, and an assessment of speeding behaviors.

COLLECTING AND COMPILING DATA IN A DATA WAREHOUSE

The process of collecting and compiling data in a central repository for use in data analyses is a multi-task effort. The first element in this process is gathering the data and providing it in a central site. This data warehouse serves several purposes which are outlined below.

- **Serve as a Central Repository for Data from Many Sources**
  Housing crash-related data in one location expedites the processes associated with collecting, understanding and analyzing multiple datasets. An interested traffic safety stakeholder no longer has to request each set from its respective owner and then undergo the process of preparing it for analysis. Instead, that process is done once and the final data product is provided to all traffic safety stakeholders for use. By creating a central location, duplication of efforts is eliminated and a level of standardization for data access is added.

- **Provide a Location for Validating and Interpreting Data to Obtain Consistency**
  Users from a variety of agencies are accessing the same data sets in the same formats. Data quality issues are handled uniformly and definitions are standardized.

- **Create a Location to Provide Historical View of Events**
  Storing the historical data allows analysts to examine data in the historical context in which the event occurred. Updated roadway inventory files or drivers license information may overwrite previous files with the same data at the collecting agency; in the data warehouse updated files are added to the system rather than replacing existing

files. Historical data also provide the basis for important analyses such as trend analyses or "before and after" studies. In addition to providing historical data, storing previous versions of datasets can be used to understand the data framework in which the data was collected. This can be valuable when determining how changes in data collection systems have impacted the type or amount of data collected.

- **Serve as a Decision Support System**
  A centralized data storage system allows decision makers to consider a problem area or potential solutions across a diverse set of resources. This is a powerful method for overcoming data quality issues that may exist in one data source and not in another.

- **Offload for Query Processing From the Source System**
  One problem encountered when trying to analyze data directly from a source system involves major differences in resource requirements. If the data are not moved from the source system to the warehouse, it is common that the source system is bogged down attempting to produce reports because the source system was not designed to handle the load that large scale analysis requires. The warehouse solves this problem by providing separate computing resources for the specific task of analyses.

DATA WAREHOUSE DEVELOPMENT

Data warehouses are not finite projects with a start date and end date; instead, they are built on a series of smaller projects. To that end, they should not be designed as a long-term project with one final completion goal; they should be managed as individual projects contributing towards the larger system. Developing a series of short term goals and smaller-scale projects provides the opportunity for successes early in the development of the warehouse. This increases the likelihood that both involved agencies will remain committed to the data warehouse.

For each new data set, a data schema is designed based on the elements in that specific dataset and the needs of the end users. This design process is undertaken to maximize the usability of the information in the data set while standardizing the structure to fit in the overall data warehouse system. Once the schema has been designed for a data set, the process of adding the data from the source data set to the data warehouse begins.

Data Acquisition

The University of Massachusetts Highway Safety Data Warehouse currently consists of 13 different datasets provided by a variety of data owner agencies. These are outlined in Table 1. This process is outlined in some detail as it is critical to the success of a data warehouse and can often be the most difficult process to successfully navigate.

TABLE 1 Data Included in the Massachusetts Data Warehouse

Owner	Data Set
Registry of Motor Vehicles	ALARS (Old Crash Data System)
	CDS Crash Data
	Licensed Drivers
	Registered Vehicles
Merit Rating Board	Citation Data
Massachusetts State Police	CMV Crash Data
	CMV Inspection Data
Division of Health Care Finance and Policy	Inpatient Discharge
	Emergency Department
	Outpatient Discharge
US Census Bureau	Population
Massachusetts Highway Department	Vehicle Miles Traveled
Department of Public Health, Bureau of Vital Statistics	Death Certificate

The success in acquiring such a rich collection of data – both in terms of breadth and depth – can be traced to four key concepts.

1. Data owners served as active participants in the data warehouse process. Data owners are more comfortable sharing data with a project they know and understand. Ongoing dialogue during transfer and loading also provides a level of comfort in the data sharing as the owner knows and understands how the data are being treated.
2. Data sharing was made easy for the data owners. Because resources at the collecting agency are often limited, the data warehouse should accept data in any format the data owner is willing to share it. These formats may include file transfer protocol (FTP) sharing, internal/intranet transfer, compact disk, zip disk, or backup tape. While this requires the data warehouse to have the ability to read the data format and transform the data to the appropriate file type, this greatly encourages the data owner to allocate the minimal resources necessary to prepare the data for transfer.
3. Data warehouse managers respected the needs of data owners. This includes being mindful of how data quality problems identified during the transfer and loading of data into the warehouse should be addressed. In addition, the data warehouse should strictly adhere to any privacy guidelines established for the storage and use of data. These guidelines may cover what hardware/software should be used to protect the data as well as who has access to the data, how they access data, and how the data are used.
4. Data owners were made aware of the tangible benefits associated with sharing data. For each data owner, the reward associated with sharing data may vary, but there is some benefit for all data owners.
    - Housing data in another location provides an additional set of skilled workers who are familiar with the data set and can aide in the process of finding data quality problems as well as identifying and implementing potential solutions. In addition, the second location can be used as a way to "double check" numbers and validate queries run by the data owner on their own data set.
    - Providing data to an organization that has trained analysts increases the data owner's analysis resources.

- Entering data into the data warehouse and providing web-based access means that analysts or interested parties within the data owner's agency now have easy access to the data for their own use.
- The data warehouse serves to answer outside data questions, alleviating the demand on data owners to respond to requests for information.

In Massachusetts, data owners seemed more willing to share data with the University of Massachusetts Traffic Safety Research Program (UMassSafe) for two institutional reasons. The first is that the University of Massachusetts is a public state university. Bound by many of the same rules applied to other state agencies, there was a sense that UMassSafe understood and would abide by state policy. Second, as part of a major research university, UMassSafe understands the importance of protecting data, reasonable use, and maintaining the privacy of identifying information in all of the data records. Data owners felt comfortable that their data would be handled appropriately.

Crash data, owned by the RMV, was the first data set acquired for the data warehouse. Once this data set was loaded in the data warehouse and made available through the web-based access system, it was used to leverage the sharing of data by other data owners. Citation data, which had been previously unattainable, was willingly shared once the Merit Rating Board who owns citation data could see how the crash data had been used. This illustrates the importance begin the design and development of a system with even one data set as a way to leverage the support of other data owners.

Extraction, Transformation, and Loading of Data

The process of loading, refreshing, and maintaining data from the source system in the data warehouse is the extraction, transformation, and loading (ETL) phase of an operational data warehouse. ETL of a source system into the data warehouse is part science and part art. The source systems encountered across the state government organizations differ greatly and do not lend themselves to a generalized approach to ETL. Data may be received from data owners in multiple formats, varying degrees of quality and completeness, and may or may not be documented. Unlike data that are collected for business purposes, the data in the data warehouse are generally databases collected by state agencies whose resources are limited and who therefore cannot run the data quality checks associated with business/enterprise databases. Upon receipt, the data warehouse staff may have to invest significant effort in the area of data quality checks as well as communication with data owners in the process of identifying and reconciling anomalies.

DATA ACCESS

Merely moving data into a warehouse does not fully address the need for data. Access is an integral aspect of any operational warehouse. There are two mechanisms that can be used to extract information from the data warehouse: 1) web-based Interactive Data Retrieval System (IDRS) and 2) systems-based access.

Interactive Data Retrieval System

To provide access to the data, the Interactive Data Retrieval System (IDRS) was developed as a web-based user interface that requires no "special" software and no knowledge of computer

programming or special query methods. Functionality associated with the IDRS includes the following:

- Query a variety of data sets collected and stored in different formats via a standardized user interface;
- Query multiple years of data within a dataset in the same query;
- Reset and alter portions of a query without having to begin a new query;
- Build, save, and reuse custom filters based on any of the fields available in the dataset;
- Save queries to be rerun or altered at future dates;
- View results in table, customizable chart, or customizable map format, or export results as a CSV file.

The IDRS is based on a system of web modules which allows for expansion of the system to meet user needs. Existing modules currently included in the data warehouse are the following:

- Online Query Tool which provides access to aggregate level data for users at both a "basic" and "advanced" level.
- Reports Module which provides access to a library of reports which can be searched or browsed by keyword, topic area, year produced, etc.
- Technical Assistance Module which provides a web-based form for submitting requests for assistance.

Online Query Tools

The Online Query Tools were designed to provide users with access to data in a format that allowed them to query data at different levels depending on their data analysis skill level. Management of data in the query tool allowed administrators to limit what users were offered as available data depending on what level of access they had been granted and which data sets they were approved to query. The Ad Hoc Query Tool, which provides advanced levels of access to data, is shown in Figure 1.

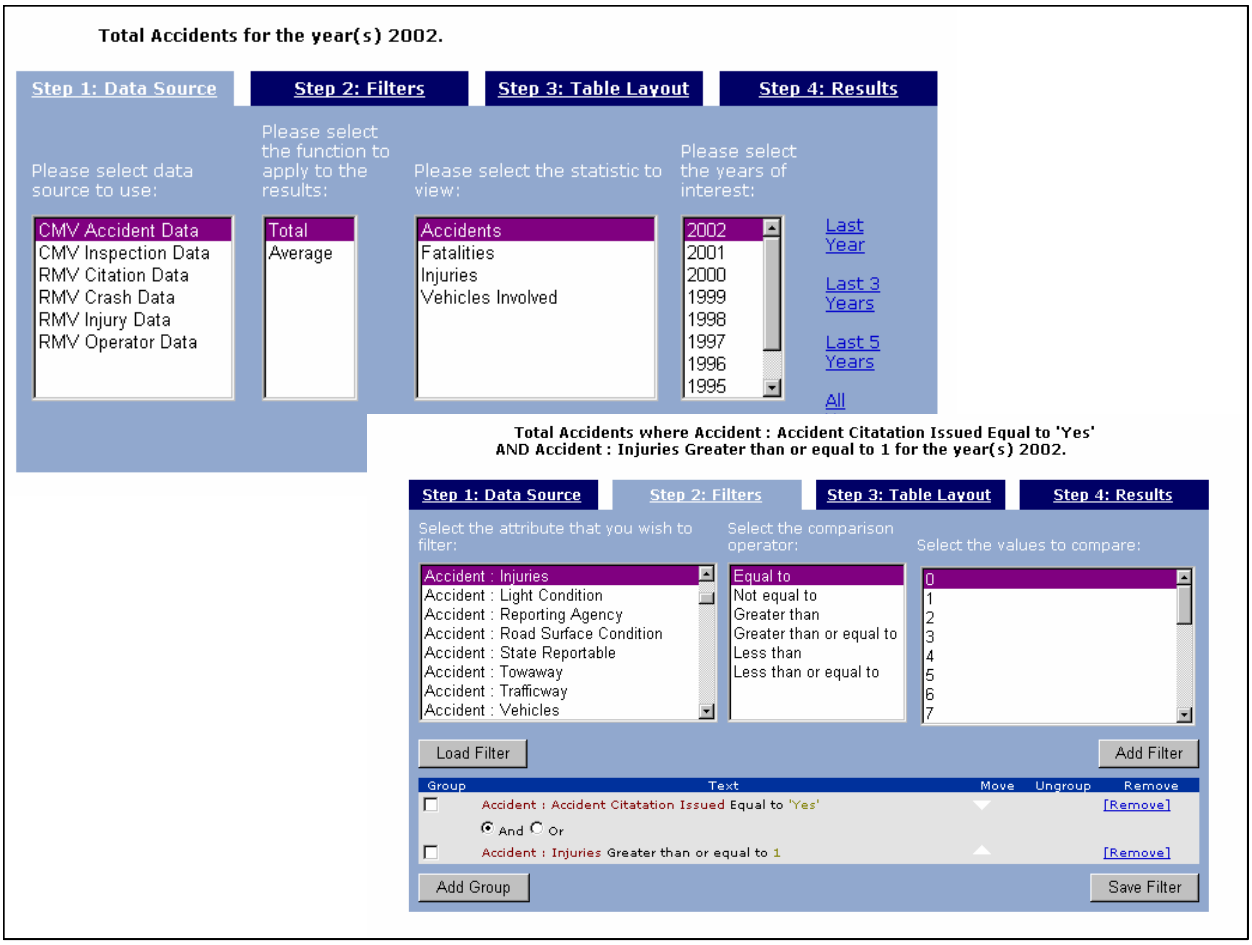


Figure 1 IDRS Ad Hoc query tool.

Reports Module

Feedback provided by users indicated interest in the development of a new module that would provide access to full documents or reports in addition to the data that was accessible through the Online Query Tools. To meet that need, a Reports module was developed and added to the system.

The Reports module provides online access to reports that have been completed as the result of an organization's data analyses and evaluation. Report managers have access to the report management features allowing the manager to publish reports in a digital format to all users of the system. This report management process includes documenting meta-data for each report and classifying the report in a standard thesaurus of crash topics. Once published, reports can be browsed by topic or searched by keywords and results can be viewed in PDF format.

There are two sections to the Reports module. The first provides Basic Stats by Topic. This section allows an organization to publish frequently used documents. This provides one or two click access to the most valuable and most used reports. In Massachusetts, this section has been used to publish basic stats allowing users to get quick answers to commonly asked questions. Figure 2 shows a screen capture of the Basic Stats element of the Reports module.

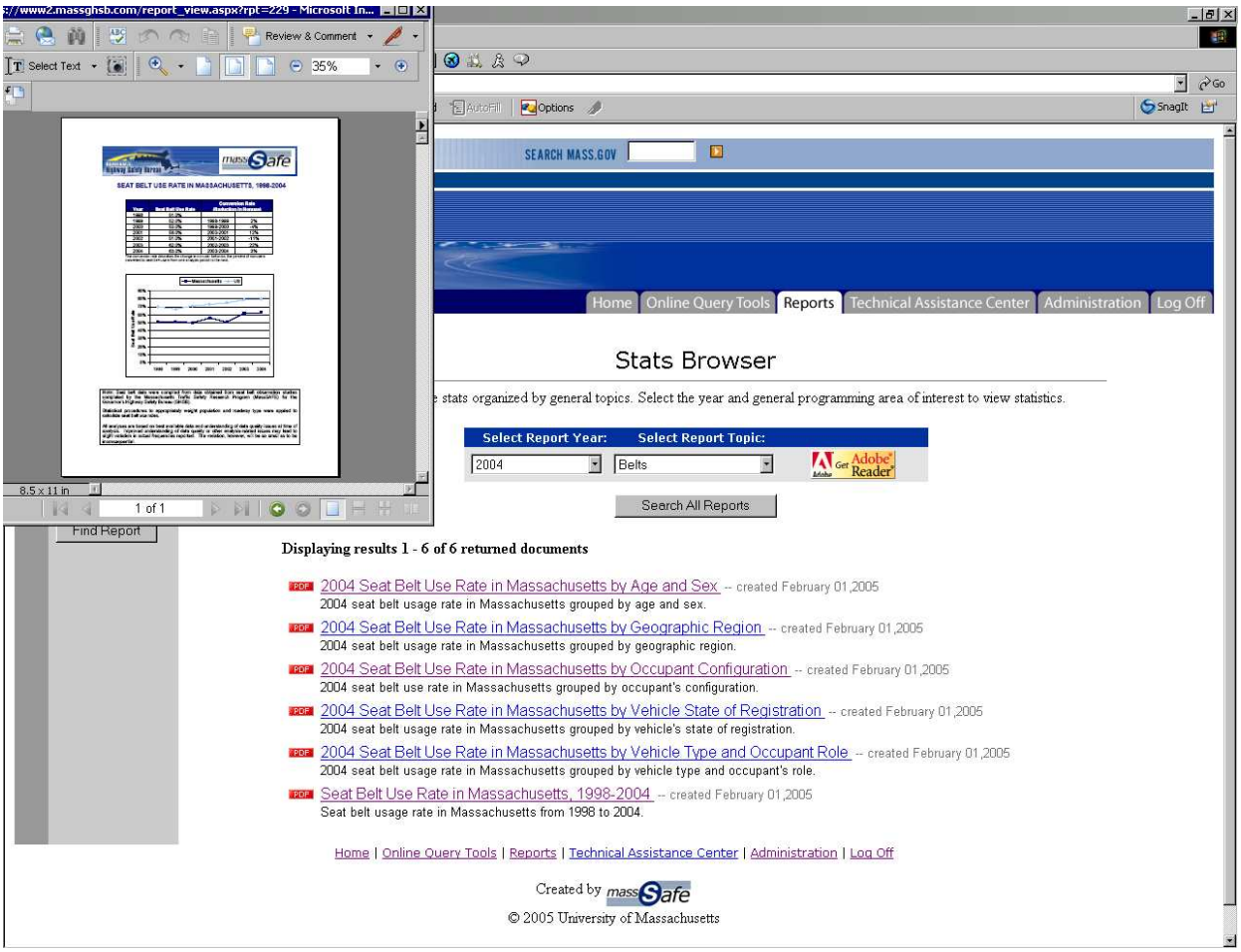


Figure 2 Basic stats browser section of reports module with selection in PDF window.

The second section of the reports module allows users to browse the entire organizations library of documents by topic using the included thesaurus of terms, or to perform full-text searches for documents based on user-selected keywords. The browse format is a familiar format for web-users enhancing the ease of the search. Reports are categorized using multiple terms contained in the thesaurus. Results from the browse capability appear as a list with an abstract and a link to view the entire document in PDF format.

Technical Assistance Center

In addition to accessing data through the Online Query Tools, users were given access to the Technical Assistance Center through the internet-based request form. This form provides a way for users to send an email to the Technical Assistance Center outlining who they are, what their question is, an ideal timeframe for response, response format, and end-use for data responses. Gathering all of this information provides the analyst fielding the request with the background necessary to determine how best to query the data to gain the desired results. Results are then returned to the user via email.

Systems-Based Access
In some instances, the query tools may not provide the level of detail required to identify problem areas and evaluate countermeasures. In these instances, data can be accessed directly through the system using SQL. All databases included in the warehouse can be queried in this manner. In addition, this method allows access to identifying information that may be critical for analysis but cannot be distributed through an aggregate level tool such as the IDRS. This identifying information might include driver information (name, license number, vehicle registration), location information (XY coordinates for crash location), citation information (violator name), etc.

DATA LINKAGE

Traditionally, analysis of crashes has been centered upon the use of police-reported information collected on state-specific crash report forms. However, the series of events surrounding a crash are more complex than the data on a crash report form can accurately record. Ideally, data should cover the events immediately preceding a crash, the characteristics of the crash itself, and the outcomes associated with the crash. The primary purpose of the Crash Outcome Data Evaluation System (CODES) is to link multiple datasets in an effort to create a more robust dataset that more effectively captures crash events. Specifically, CODES focuses on linking crash data to emergency medical services (EMS) and hospital data, to understand injuries and charges associated with crashes; other safety datasets including citation, roadway inventory, and insurance data may also be linked to provide a more comprehensive overall picture. Recognizing that this type of data collection and linkage was best possible at the state level, CODES was established by the National Highway Traffic Safety Administration (NHTSA) in 1992. Since then, some states, including Massachusetts, have been funded to develop CODES programs, conduct data linkages, and report findings to NHTSA and other traffic safety stakeholders (2).

CODES employs probabilistic linkage to link datasets with common information but no common unique identifier. Crash characteristics (i.e. time, location, object struck), person characteristics (i.e. age, sex), and vehicle characteristics (i.e. type of vehicle) that are common across data sets can be used to link person level records. In Massachusetts, the CODES linkage currently includes crash, emergency department, hospital inpatient, and death certificate data, through 2004. This type of linkage allows engineers to consider, for example, the costs associated with treating injuries sustained in crashes associated with roadway segments that have certain characteristics. This type of cost information can add significantly to dialogue, especially with decisions makers who so frequently work in terms of benefit-cost analyses, around resource allocation and countermeasure effectiveness.

The linkage of Massachusetts data is based on three match passes. Within each match pass, fields from each data set are used for two sets of linkage specifications. Match specifications are the same for each pass and are used to calculate match weights and probabilities; these specifications define the fields for comparison. Join specifications, which are different for each pass, identify candidate pairs. Each pass is run independently, and passes are merged to obtain a resulting set of linked data. Figure 3 outlines the fields used for the Massachusetts linkage. These linkages approximately 45,000 matched pairs each year.

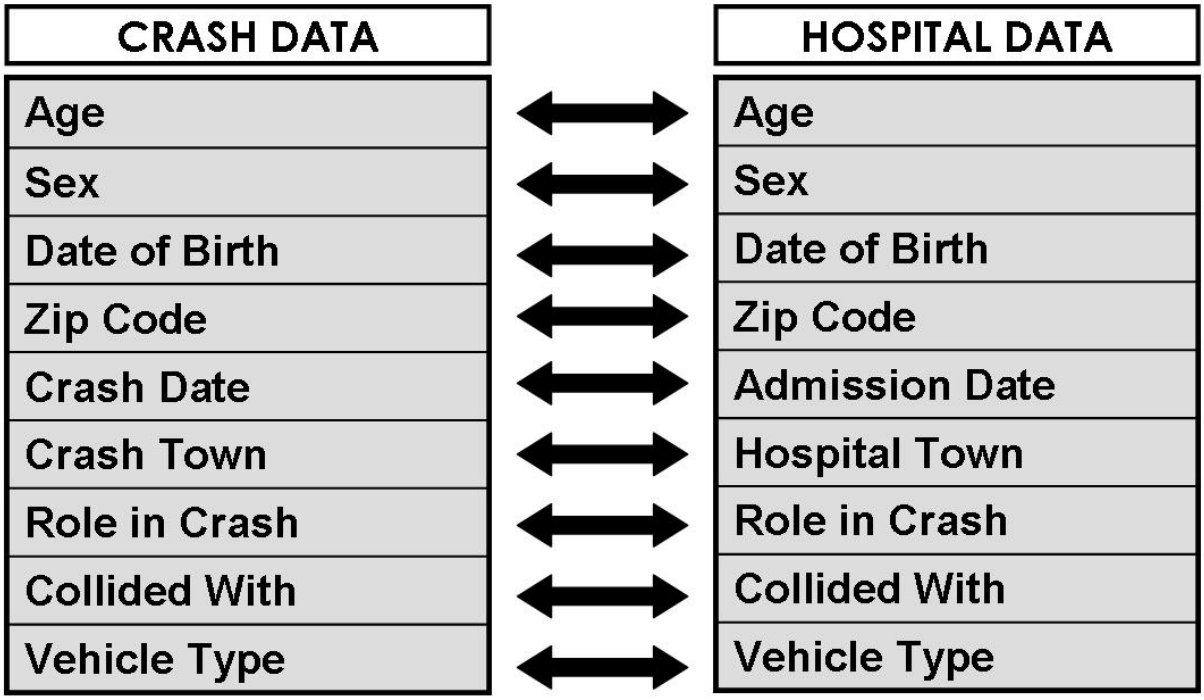


Figure 3 Fields used for Massachusetts CODES linkage.

LOCATION ANALYSES

Specific crash location information (XY coordinate) became a standard element in the Massachusetts crash data set after the initial development of the Massachusetts data warehouse and web-based tools. As such, there is no location-based information that is currently included as a fully developed module. However, through work with the Massachusetts State Police Commercial Vehicle Enforcement Section, UMassSafe has been working on the development of a web-based system for viewing crashes and crash information that is based on Google user interface.

The Massachusetts Commercial Motor Vehicle (CMV) Crash Map is an interactive tool that allows users to identify trends and pinpoint crash information involving CMVs across the state. A colored marker representing each CMV crash can be clicked to display detailed information on the crash, the vehicle, and data quality information for that particular case. Figure 4 shows an area level map of CMV crash locations and Figure 5 shows what the user sees when they select a particular location.

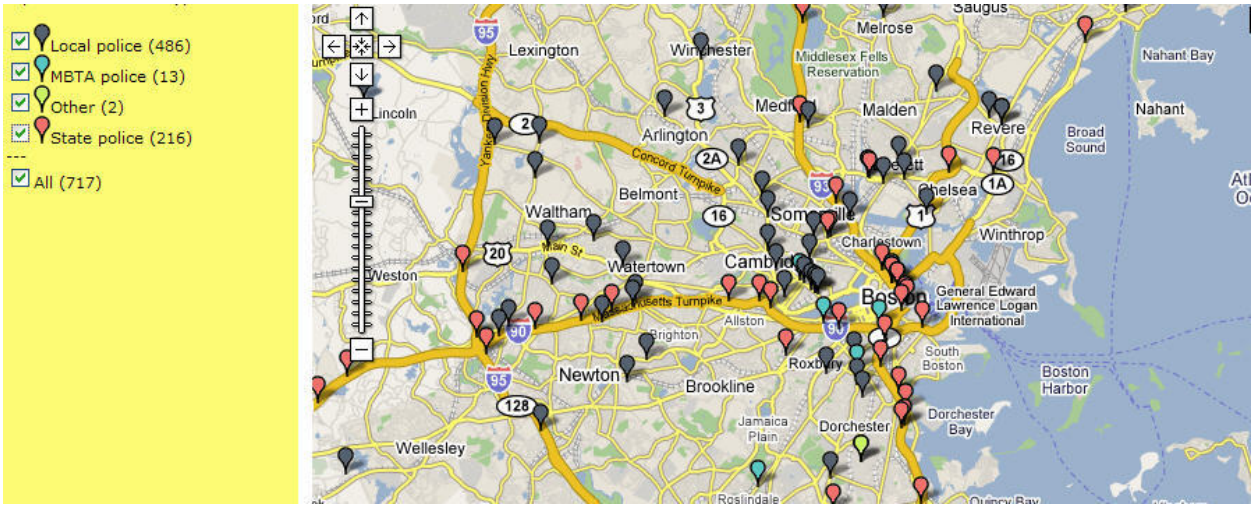


Figure 4 Area-level map of commercial vehicle crashes in eastern Massachusetts by reporting agency type.

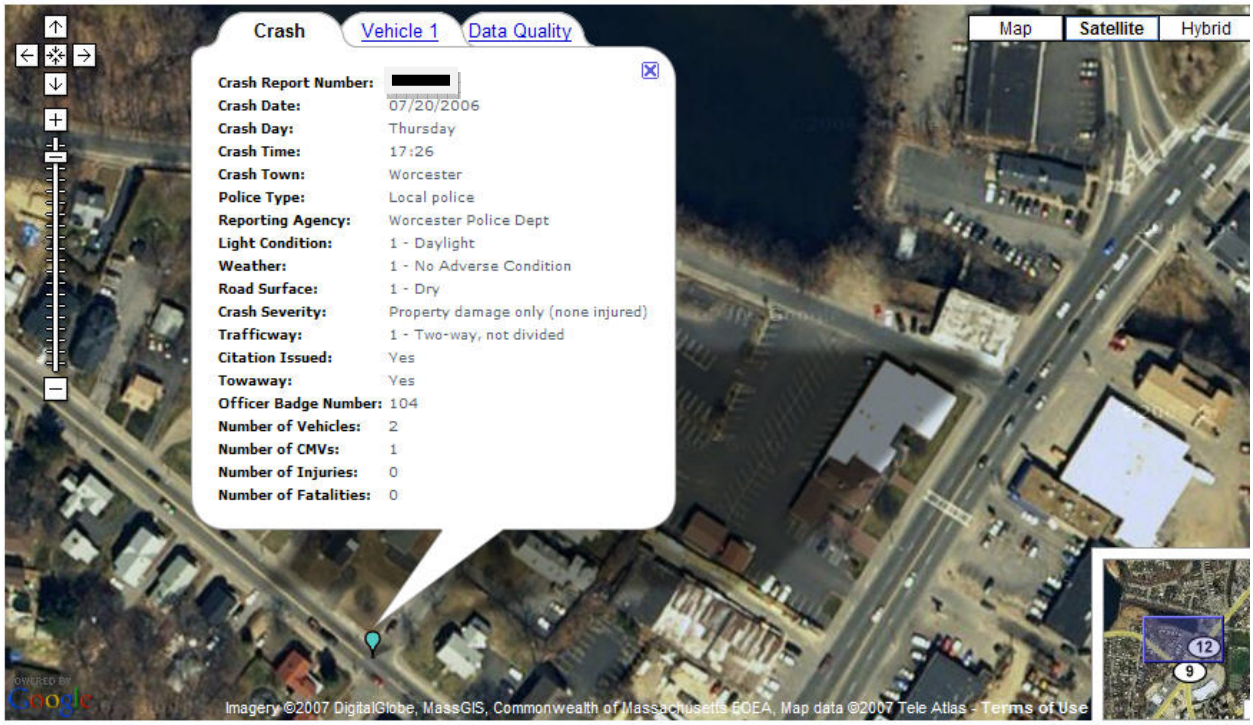


Figure 5 Crash-level map of commercial vehicle crash with additional information (crash report number blocked for privacy reasons).

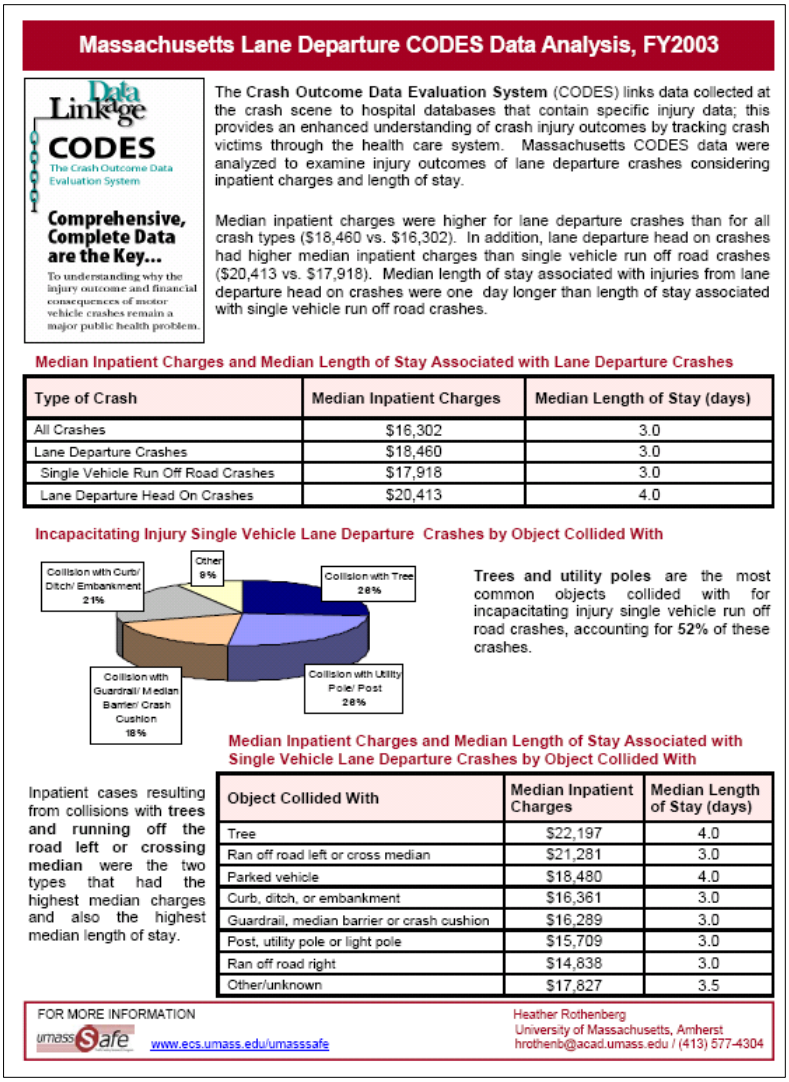
APPLICATIONS OF THE DATA WAREHOUSE IN MASSACHUSETTS HIGHWAY SAFETY

Arguably the best method for presenting the functionality of such a tool is to consider several applications in which the data have been employed. Below are the descriptions of three applications where a variety of datasets were used to examine highway safety problem areas of special interest to Massachusetts program and policy makers. In each of these examples, the University of Massachusetts Highway Safety Data Warehouse and data linkage methods (specifically CODES), were used to identify appropriate datasets, analyze them in relation to each other, and provide data-driven information that could be used in the development, implementation, and evaluation of highway safety programs and countermeasures.

Lane Departure Crashes

Lane departure crashes account for approximately 19 percent of all crashes in Massachusetts but almost 46 percent of crashes involving fatal injuries. Lane departures have been identified as a key problem area as part of the Massachusetts Strategic Highway Safety Plan and efforts in this area have been multi-agency in nature with leadership provided by the Federal Highway Administration and Massachusetts Highway Department (MassHighway). Several reports and fact sheets have been prepared for use with the Massachusetts highway safety community, including some used by MassHighway in their work with Regional Planning Agencies to address lane departure crashes at the local and regional levels.

Figure 6 shows a fact sheet prepared using linked crash and hospital data. This fact sheet provides information on charges associated with different types of lane departure crashes. Results of the analysis can provide a basis for determining where to implement countermeasures based on crash cost and severity rather than frequency alone.



Massachusetts Lane Departure CODES Data Analysis, FY2003

Data Linkage CODES
The Crash Outcome Data Evaluation System

Comprehensive, Complete Data are the Key...
To understanding why the injury outcome and financial consequences of motor vehicle crashes remain a major public health problem.

The **Crash Outcome Data Evaluation System** (CODES) links data collected at the crash scene to hospital databases that contain specific injury data; this provides an enhanced understanding of crash injury outcomes by tracking crash victims through the health care system. Massachusetts CODES data were analyzed to examine injury outcomes of lane departure crashes considering inpatient charges and length of stay.

Median inpatient charges were higher for lane departure crashes than for all crash types ($18,460 vs. $16,302). In addition, lane departure head on crashes had higher median inpatient charges than single vehicle run off road crashes ($20,413 vs. $17,918). Median length of stay associated with injuries from lane departure head on crashes were one day longer than length of stay associated with single vehicle run off road crashes.

Median Inpatient Charges and Median Length of Stay Associated with Lane Departure Crashes

Type of Crash	Median Inpatient Charges	Median Length of Stay (days)
All Crashes	$16,302	3.0
Lane Departure Crashes	$18,460	3.0
Single Vehicle Run Off Road Crashes	$17,918	3.0
Lane Departure Head On Crashes	$20,413	4.0

Incapacitating Injury Single Vehicle Lane Departure Crashes by Object Collided With

Trees and utility poles are the most common objects collided with for incapacitating injury single vehicle run off road crashes, accounting for **52%** of these crashes.

Median Inpatient Charges and Median Length of Stay Associated with Single Vehicle Lane Departure Crashes by Object Collided With

Inpatient cases resulting from collisions with **trees and running off the road left or crossing median** were the two types that had the highest median charges and also the highest median length of stay.

Object Collided With	Median Inpatient Charges	Median Length of Stay (days)
Tree	$22,197	4.0
Ran off road left or cross median	$21,281	3.0
Parked vehicle	$18,480	4.0
Curb, ditch, or embankment	$16,361	3.0
Guardrail, median barrier or crash cushion	$16,289	3.0
Post, utility pole or light pole	$15,709	3.0
Ran off road right	$14,838	3.0
Other/unknown	$17,827	3.5

FOR MORE INFORMATION
umassSafe www.ecs.umass.edu/umasssafe

Heather Rothenberg
University of Massachusetts, Amherst
hrothenb@acad.umass.edu / (413) 577-4304

Figure 6 Portion of Massachusetts CODES lane departure fact sheet.

Older Drivers

Greater fragility has been shown to contribute to high fatality rates among older road users. Older drivers, specifically, have been identified as a key problem area in the Massachusetts Strategic Highway Safety Plan. A variety of data were used to examine older road users. In addition to an analysis of linked crash and hospital data conducted through CODES, analyses were also based on citation and licensed driver data.

Results of these varied analyses have been presented in several formats including fact sheets (as shown in Figure 7), report, and presentation to the Massachusetts Traffic Records Coordinating Committee. The findings of these analyses have provided information for the ongoing discussion around older vehicle occupants involved in crashes, specifically around older drivers. Previously, similar analyses were coupled with a review of policy and legislation around teen drivers to support decision-making around Massachusetts Graduated Driver Licensing laws.

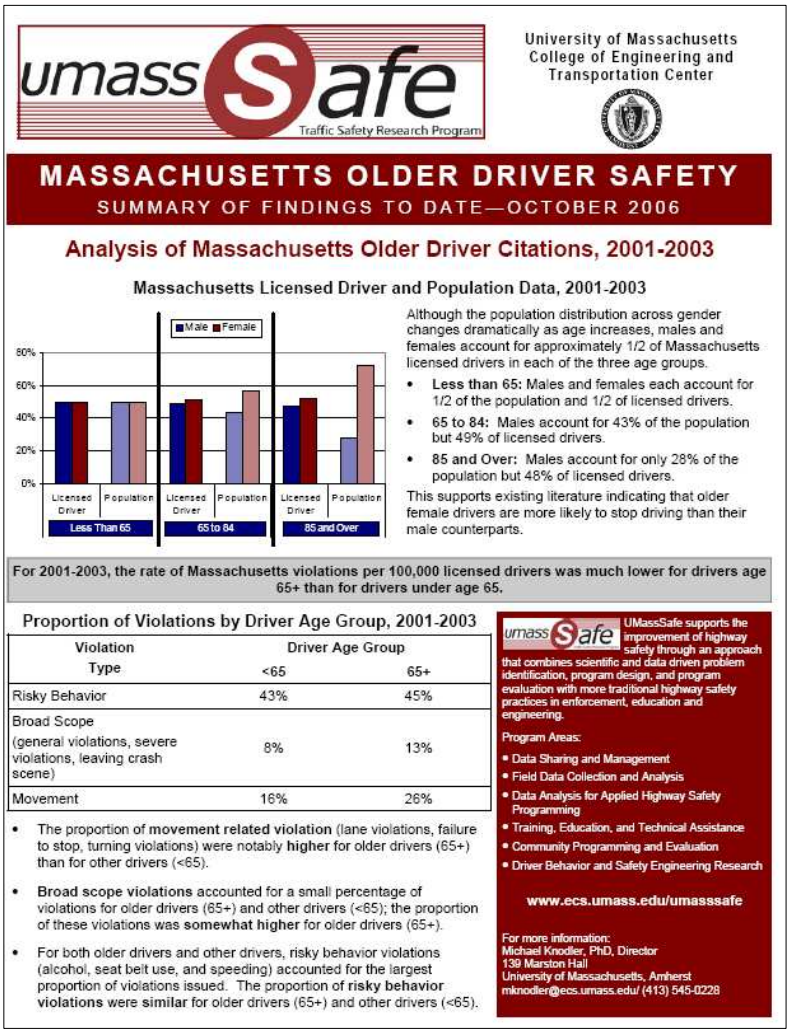


Figure 7 Portion of Massachusetts older driver fact sheet.

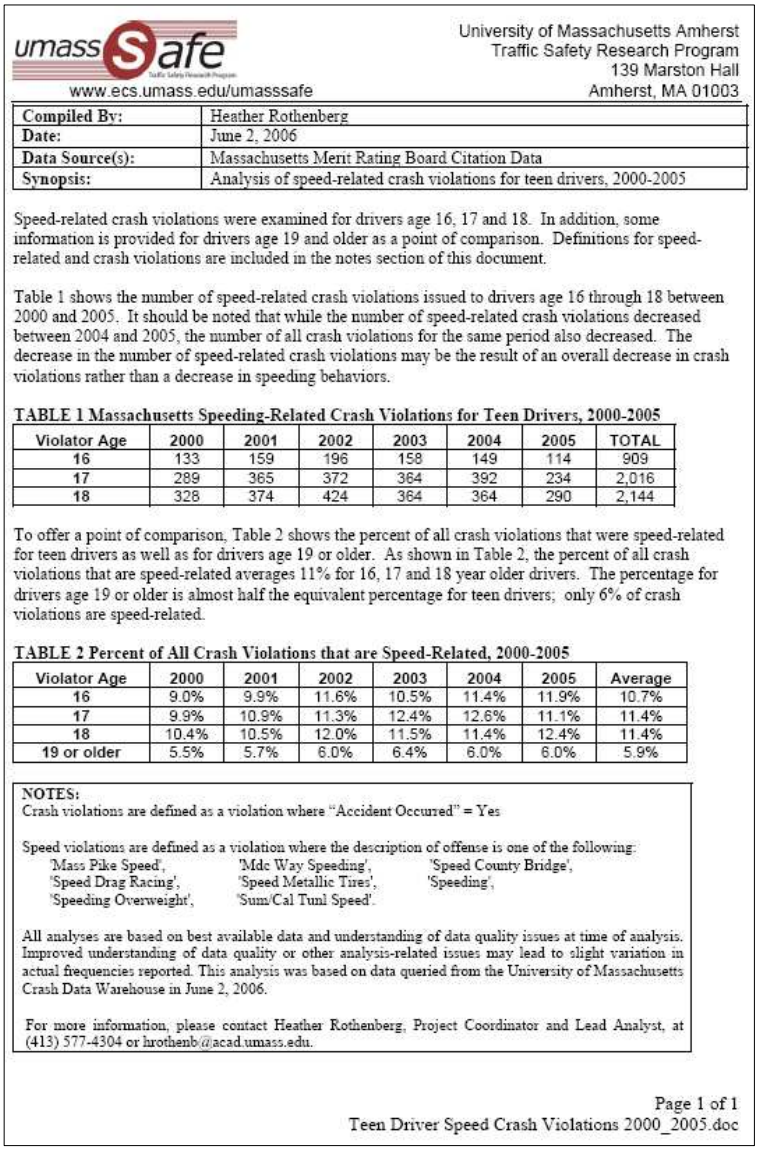
Based on the continued interest in older drivers, and data-driven policy making, the Massachusetts Registry of Motor Vehicles has submitted a project problem statement to the Executive Office of Transportation's Cooperative Research Program to undertake a data-driven policy and legislative review. This review will integrate these results as part of the work aimed at identifying options for improved licensing practices for older drivers and assessing alternative transportation methods for older residents of Massachusetts (such as improved transit, etc).

Speeding and Highway Safety

The relationship between speed and highway safety has been well documented. In 2006, nearly one-third of crash fatalities in the United States were speed-related. Many highway safety stakeholders have engaged in efforts to reduce the frequency and severity of speeding-related crashes. These efforts include enforcement initiatives, engineering based countermeasures, and public education campaigns.

Massachusetts approach to addressing speed-related crashes has always included a heavily data-driven component. In addition to evaluations of speeding crashes, citation data has been heavily relied on for information on speeding crashes, especially since the Massachusetts crash report form collects limited information on speeding. In addition to administrative datasets stored in the data warehouse (crash, citation, etc), the University of Massachusetts participated in work funded by the Federal Highway Administration to examine rational speed limit setting.

Figure 8 shows the response to a request for information on teen drivers and speed-related violations issued in a crash.



umass Safe
www.ecs.umass.edu/umasssafe

University of Massachusetts Amherst
Traffic Safety Research Program
139 Marston Hall
Amherst, MA 01003

Compiled By:	Heather Rothenberg
Date:	June 2, 2006
Data Source(s):	Massachusetts Merit Rating Board Citation Data
Synopsis:	Analysis of speed-related crash violations for teen drivers, 2000-2005

Speed-related crash violations were examined for drivers age 16, 17 and 18. In addition, some information is provided for drivers age 19 and older as a point of comparison. Definitions for speed-related and crash violations are included in the notes section of this document.

Table 1 shows the number of speed-related crash violations issued to drivers age 16 through 18 between 2000 and 2005. It should be noted that while the number of speed-related crash violations decreased between 2004 and 2005, the number of all crash violations for the same period also decreased. The decrease in the number of speed-related crash violations may be the result of an overall decrease in crash violations rather than a decrease in speeding behaviors.

TABLE 1 Massachusetts Speeding-Related Crash Violations for Teen Drivers, 2000-2005

Violator Age	2000	2001	2002	2003	2004	2005	TOTAL
16	133	159	196	158	149	114	909
17	289	365	372	364	392	234	2,016
18	328	374	424	364	364	290	2,144

To offer a point of comparison, Table 2 shows the percent of all crash violations that were speed-related for teen drivers as well as for drivers age 19 or older. As shown in Table 2, the percent of all crash violations that are speed-related averages 11% for 16, 17 and 18 year older drivers. The percentage for drivers age 19 or older is almost half the equivalent percentage for teen drivers; only 6% of crash violations are speed-related.

TABLE 2 Percent of All Crash Violations that are Speed-Related, 2000-2005

Violator Age	2000	2001	2002	2003	2004	2005	Average
16	9.0%	9.9%	11.6%	10.5%	11.4%	11.9%	10.7%
17	9.9%	10.9%	11.3%	12.4%	12.6%	11.1%	11.4%
18	10.4%	10.5%	12.0%	11.5%	11.4%	12.4%	11.4%
19 or older	5.5%	5.7%	6.0%	6.4%	6.0%	6.0%	5.9%

NOTES:
Crash violations are defined as a violation where "Accident Occured" = Yes

Speed violations are defined as a violation where the description of offense is one of the following:
'Mass Pike Speed', 'Mdc Way Speeding', 'Speed County Bridge',
'Speed Drag Racing', 'Speed Metallic Tires', 'Speeding',
'Speeding Overweight', 'Sum/Cal Tunl Speed'.

All analyses are based on best available data and understanding of data quality issues at time of analysis. Improved understanding of data quality or other analysis-related issues may lead to slight variation in actual frequencies reported. This analysis was based on data queried from the University of Massachusetts Crash Data Warehouse in June 2, 2006.

For more information, please contact Heather Rothenberg, Project Coordinator and Lead Analyst, at (413) 577-4304 or hrothenb@acad.umass.edu.

Page 1 of 1
Teen Driver Speed Crash Violations 2000_2005.doc

Figure 8 Response to data request on speed-related citations for teen drivers.

CONCLUSION

Data are a critical element to any successful highway safety effort. As such, it is critical that access be provided to data in a way that allows not only for analyses of individual datasets, but also of those datasets in relation to each other. This type of central repository that allows access for analyses meets several of the needs identified by the national multidisciplinary team established to identify how to use data to meet Department of Transportation goals. The contributions – in terms of funding, data, and guidance – works towards coordination and leadership; this was especially true in Massachusetts since the state's Traffic Records Coordinating Committee (TRCC) played an integral role in the process. Providing a central repostitory with multiple levels of access directly address the need for improved data quality and

accessibility. Web-based interfaces and the use of existing technologies, coupled with advanced software and methods used for data linkage, combine to meet the needs of users at all levels of access and analysis. The ability to incorporate work completed through other funded programs such as CODES and the commercial vehicle safety project improved opportunities to integrate data and facilitate use. Using other program to supplement the warehouse and access as it was initially developed ensures that the data stays current and that access methods are exapanded to address changing user needs.

Creating a central repository, and training analysts who can effectively use those datasets in conjunction with each other, is key to the success of highway safety efforts at the planning, implementation, and evaluation stages.

ACKNOWLEDGEMENTS

The authors of this paper would like to acknowledge the state and federal agencies that supported and participated in differing stages of the development of the Highway Safety Data Warehouse and data linkage programs. These agencies are the Massachusetts Executive Office of Public Safety and Security, Massachusetts Executive Office of Transportation, Massachusetts State Police, National Highway Traffic Safety Administration, and Federal Motor Carrier Safety Administration.

REFERENCES

1. National Center for Statistical Analysis, National Highway Traffic Safety Administration, Fact Sheets. http://www-nrd.nhtsa.dot.gov/CMSWeb/index.aspx. Accessed January 18, 2008.
2. Benavente, M., M. Knodler, H. Rothenberg. Case Study Assessment of Crash Data Challenges: Linking Databases for Analysis of Injury Specifics and Crash Compatibility Issues. *Transportation Research Record: Journal of the Transportation Research Board* Number 1953, Pages 180-186, 2006.

AUTHORS

Heather Rothenberg
University of Massachusetts Traffic Safety Research Program
139 Marston Hall
Amherst, MA 01003
Ph. (413) 577-4304
Fax (413) 577-1036
hrothenb@acad.umass.edu

Michael A. Knodler, Jr. PhD
Assistant Professor
Department of Civil and Environmental Engineering
University of Massachusetts Amherst
216 Marston Hall
Amherst, MA 01003
Ph. (413) 545-0228
Fax (413) 577-1036
mknodler@ecs.umass.edu