U.S. Department
of Transportation
National Highway
Traffic Safety
Administration

People Saving People
http://www.nhtsa.dot.gov

DOT HS 809 200
NHTSA Report

February 2001

Problems, Solutions and Recommendations for Implementing CODES

(Crash Outcome Data Evaluation System)

Technical Report Documentation Page

1. Report No. DOT HS 809 200	2. Government Accession No.	3. Recipients's Catalog No.	
4. Title and Subtitle Problems, Solutions and Recommendations for Implementing CODES (Crash Outcome Data Evaluation System)		5. Report Date December 2000	
		6. Performing Organization Code	
7. Author(s) (See attached)		8. Performing Organization Report No.	
9. Performing Organization Name and Address		10. Work Unit No. (TRAIS)n code	
		11. Contract of Grant No. (See Attached)	
12. Sponsoring Agency Name and Address National Highway Traffic Safety Administration 400 Seventh Street, S.W. Washington, DC 20590		13. Type of Report and Period Covered NHTSA REPORT	
		14. Sponsoring Agency Code	
15. Supplementary Notes Dennis Utter served as the NHTSA Contracting Officer's Technical Representative for the study			
16. Abstract Problems, solutions and recommendations for implementation have been contributed by 16 of the 27 CODES states and organized as appropriate under the administrative, linkage and application requirements for a Crash Outcome Data Evaluation System (CODES). The purpose of this report is to provide feedback to states interested in implementing CODES. A background section describes the funding of the CODES states and the CODES Model. Administrative problems focus on maintaining communication, supporting a collaborative source of authority, developing policies for confidentiality and release of CODES linked data, managing CODES and Institutionalization. Linkage problems focus on data access, data quality/preparation, linkage, and validation. Application problems focus on statistical issues, personnel issues, confidentiality issues, limitations for case selection, production issues, decision making and Web site development. Recommendations for successful implementation also are organized under the administrative, linkage and application categories.			
17. Key Words CODES, linked data, probabilistic linkage, administrative, application, crash, outcome measures,		18. Distribution Statement Document is available to the public through the National Technical Information Service, Springfield, VA 22161	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No of Pages 65	22. Price

Form DOT F1700.7 (8-72) Reproduction of completed page authorized

LIST OF AUTHORS

Preparation of this document was a collaborative effort involving persons from the 16 reporting CODES states, NAGHSR and NHTSA as listed below:

DTNH22-96-H-39017	Karl Kim	University of Hawaii, Department of Urban & Regional Planning
DTNH22–98-H-27086	Dick Harmon	Iowa Department of Public Health Bureau of EMS
DTNH22–98-H-17086	Michael Singleton	KY Injury Prevention & Research Center
DTNH22-00-H-77012	Karl Finison	Maine Health Information Center
DTNH22-00-H-07012	Tim Kerns	University of Maryland National Study Center for Trauma/EMS
DTNH22-00-H-87012	Ming Qu	Nebraska Department of Health
DTNH22-97-H-57015	Gina Thaxton	University of Nevada School of Medicine Department of Surgery
DTNH22-00-H-37012	Tom Hettinger	New Hampshire Department of Safety, Div EMS
DTNH22-96-H-37266	Shereen Brynildsen	New York State Department of Health
DTNH22-97-H-17015	Ann Lunde	North Dakota Department of Transportation
DTNH22-00-H-57012	Ross Clarke	University of Oklahoma, Center for Health Policy
DTNH22-00-H-27012	Mike Allen	HealthcareData.com,LLC, Mechanicsburg, PA
DTNH22-00-H-17012	Mary Pease	South Carolina State Budget and Control Board
DTNH22-97-H-67015	Kareen Dougherty	University of South Dakota Business Research Bureau
DTNH22-00-H-67012	Mike Dean	Intermountain Injury control Research Center
DTNH22-00-H-47012	Wayne Bigelow	Univ. of Wisconsin Ctr Health Systems Research & Analysis
DTNH22-94-H-07364	Sandra Johnson, editor Consultant, National Assoc. of Governors' Highway Safety Representatives	
	Dennis Utter, CODES Contracting Officer's Technical Representative National Highway Traffic Safety Administration	

TABLE OF CONTENTS

THE IMPLEMENTATION OF CODES

Background

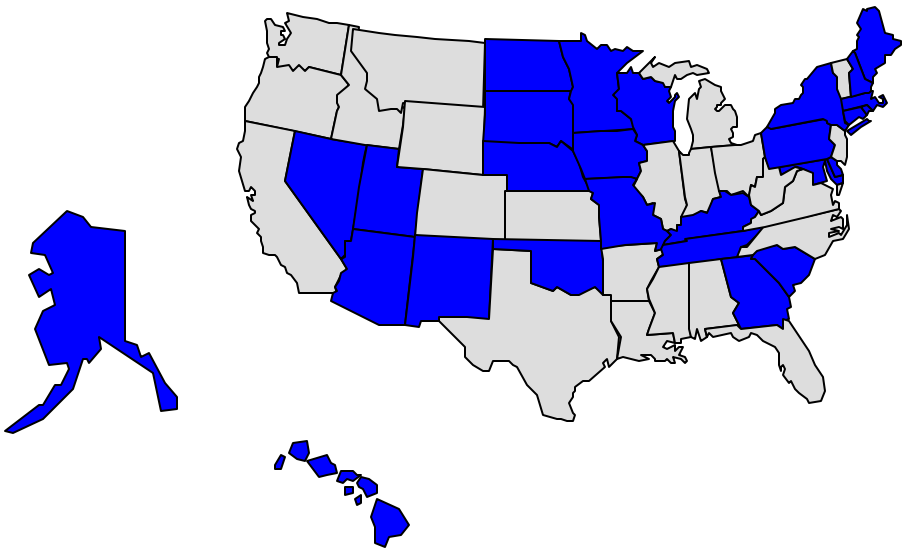
The National Highway Traffic Safety Administration (NHTSA) initiated development of Crash Data Outcome Evaluation Systems (CODES) because of the limitations of crash data alone to indicate the medical and financial outcome of motor vehicle crashes. In response to a Congressional mandate to evaluate the effectiveness of safety belts and motorcycle helmets on mortality, morbidity, injury severity and health care costs, NHTSA determined that outcome data could be obtained only at the state level. Thus, a group of states was funded to link crash and injury state data in a standardized format. NHTSA later merged the state-specific linked data to generate the safety belt and motorcycle helmet effectiveness information needed for the Report to Congress[1].

Since 1992, 27 states have been funded to develop CODES and/or develop applications for highway traffic safety using linked crash and injury outcome data. Only the first group of CODES states was required to generate the linked data in the standardized format for NHTSA. A second group of states, DEMO1, was funded to demonstrate state-specific applications, later published by NHTSA as examples for other CODES states. The applications include three studies related to highway safety[2,3,4], four studies related to traffic safety[5,6,7,8], two studies related to health care costs [9,10], two studies related to injury control[11,12] and one set of management reports[13]. Subsequent groups of states were funded by CODES to develop the data linkage capability and to focus on applications that would have an immediate impact on state-specific highway traffic safety decision making.

Each of the CODES groups is listed by funding year and group in Table 1.

Table 1: CODES States Presented by Group and Year of Funding by NHTSA for FY1992-FY2000		
FY92	CODES1	Hawaii, Maine, Missouri, New York, Pennsylvania, Utah, and Wisconsin
FY96	DEMO1	3 CODES States (New York, Pennsylvania, Wisconsin) and three new states (Alaska, Connecticut, New Mexico)
FY97	CODES2	Connecticut, Maryland, Nevada, New Hampshire, North Dakota, Oklahoma, South Dakota.
FY98	CODES3	Iowa, Kentucky, Massachusetts, Nebraska, and South Carolina
FY99	CODES4	Arizona, Delaware, Minnesota and Tennessee
FY00	CODES5	Georgia and Rhode Island
CODES - Crash Outcome Data Evaluation System NHTSA - National Highway Traffic Safety Administration		

Figure 1: NHTSA Funded CODES States - October 2000



As displayed in Figure 1, more than half of the states have been funded to generate and/or use linked data for highway traffic safety purposes.
The CODES and Demonstration (DEMO) states are distributed among all of the

NHTSA regions. Table 2 shows that in seven of the ten NHTSA regions, 50 percent or more of the states have implemented CODES.

Table 2: Status of CODES and DEMO Funding by NHTSA Region* as of October 2000

Reg 1	83%	Reg 6	40%
Reg 2	50%	Reg 7	75%
Reg 3	50%	Reg 8	50%
Reg 4	63%	Reg 9	75%
Reg 5	33%	Reg 10	25%

*Excludes Puerto Rico, Virgin Islands, Indian Nation, American Samoa, Guam, Mariana Islands from the denominators for Regions 2, 6, and 9 as appropriate. Includes the District of Columbia in the denominator for Region 3. Includes New Mexico and Alaska, both DEMO1 states, in the calculations for Regions 6 and 10 respectively.

Note: CODES - Crash Outcome Data Evaluation System
DEMO - CODES Demonstration funding
NHTSA - National Highway Traffic Safety Administration

By the end of 2001, the 27 CODES states will have generated about 91 years of linked data for the years 1990-1999. In order to take advantage of this unique source of routinely generated, population-based crash outcome data, NHTSA created the CODES Data Network. This Network will develop standardized processes, in compliance with state confidentiality and data release policies, to facilitate access by NHTSA analysts to the CODES linked data. At the same time, the additional funding will help states institutionalize CODES and continue the improvement in the quality and linkage of state crash and injury data.

The first group of CODES Data Network states, funded in FY00 include the following:

CODES1	CODES2	CODES3
Maine	Connecticut	Nebraska
Pennsylvania	Maryland	South Carolina
Utah	New Hampshire	
Wisconsin	Oklahoma	

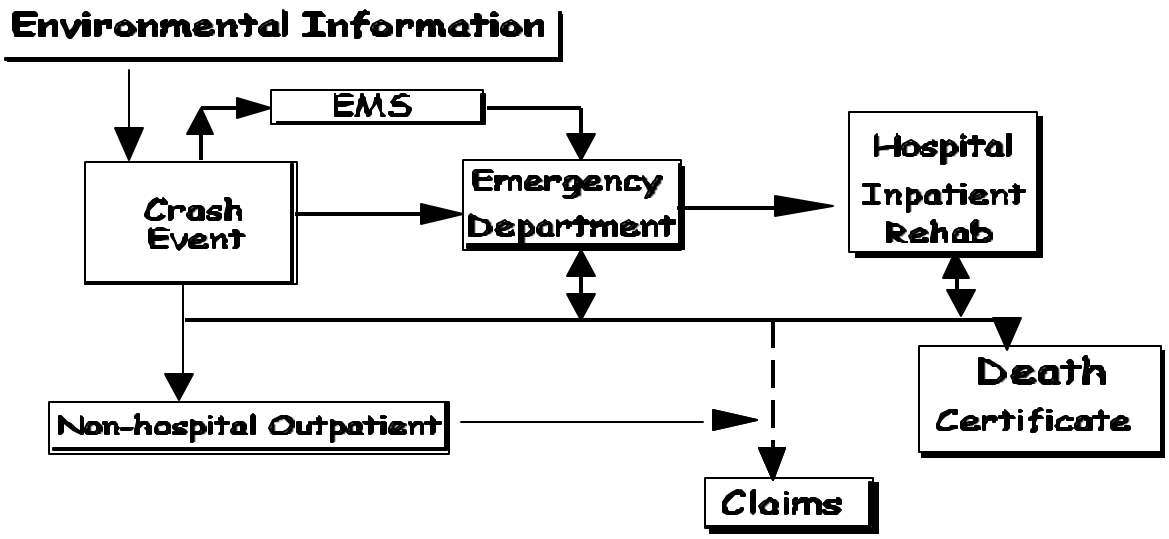
The Data Network states have generated 45 years of state-linked crash-hospital data for the period 1991-1999. Most of the 45 state-linked data years have been expanded to include some type of outpatient data as follows: 9 of the 45 state-linked data years include both EMS and ED data, 22 include EMS only, 3 include ED only, and 11 include insurance data instead of EMS or ED data. Most of the Network states also have expanded their state-linked data to include at least one of the other types of data, such as death certificate, trauma registry, driver licensing, vehicle registration, citation/conviction and/or roadway data. The Data Network states will perform future linkages using new software, CODES 2000, which was developed for NHTSA when the previous linkage software was removed from the marketplace. In addition to responding to the NHTSA data requests, the Data Network states will continue to develop CODES applications that have an impact on traffic safety decision making at the state level.

The CODES Model

CODES uses linked electronic data to track persons involved in motor vehicle crashes from the scene, and, if injured, through the health care system to a final destination. Figure 2 displays the types of data and linkages used to accomplish this task. When person-specific crash data are linked to injury data, characteristics of the event, person and vehicle involved in the crash are matched to their specific medical and financial outcomes. Use of probabilistic techniques makes it possible to work with large statewide data files which include all persons involved, injured and uninjured. Thus, sufficient records linked to outcome information are generated to determine statistically which highway safety counter measures are most effective for reducing injuries and deaths from motor vehicle crashes. With this information, NHTSA, the states, and other highway safety stakeholders can target resources where they will have the most impact on reducing mortality, morbidity, injury severity and health care costs.

Figure 2: The Data Sources and Linkages for CODES

CODES states must perform several functions. First, they must develop an



administrative structure to promote collaboration and share authority, because different entities are responsible for the crash and injury data displayed above. Second, the different data files must be converted to person-specific files where necessary and linked using probabilistic linkage techniques. For the linkages to be successful, the state data must include sufficient crash and person identifiers to discriminate between the crashes and the persons involved in each crash. Each state must link any two calendar years of statewide crash and injury data, and validate the results. Third, applications based on the linked crash and injury data must be designed to enhance the state's highway traffic safety decision making. Finally, CODES must be institutionalized so that the linked data are routinely available over time.

Although each state funded for CODES has encountered obstacles trying to implement these requirements, all of them, in spite of differing circumstances, have implemented CODES successfully.

Format for the Report

The format for this report describes the problems and solutions experienced when implementing CODES. The CODES activities are categorized into three types: administrative, linkage and application. For each of the three activities, the grant requirements are presented first, followed by the state-specific implementation problems and solutions. The final section of the report presents the recommendations which the states considered most important to successfully implement CODES. These recommendations also are categorized according to the administrative, linkage and application categories.

Sixteen of the 27 CODES states originally reported this information at the CODES Technical Assistance 2000 meeting held in Portland, Maine June 19-21, 2000. The sixteen include the following CODES states funded during the first three rounds of funding:

CODES1	CODES2	CODES3
Hawaii	Maryland	Iowa
Maine	North Dakota	Kentucky
New York	New Hampshire	Nebraska
Pennsylvania	Nevada	South Carolina
Utah	Oklahoma	
Wisconsin	South Dakota	

The state-specific problems, solutions and recommendations are presented "as reported" by the states with the elimination of all identifiers. In addition, some of the information was edited to eliminate grammatical errors and duplicate information.

Administrative Issues

List of Requirements for the Administration of a CODES:

- MAINTAIN AN ADMINISTRATIVE STRUCTURE TO PROMOTE COLLABORATION AND SHARE AUTHORITY

A. CODES Board of Directors
   1. Includes the owners of the state data
   2. Responsible for all decisions related to confidentiality, management and release of the linked data.

B. CODES Advisory Board
   1. Includes the data owners and major users of the linked data
   2. Reviews and advises on applications of the linked data.

C. State Agency with highway safety responsibilities as management entity.
   1. Oversee staff experienced in working with the crash and injury state data during the linkage process.
   2. Cross-train sufficient staff to ensure institutionalization of the data linkage capability.
   3. Obtain and maintain dedicated computer resources for linkage.
   4. Document the file preparation, linkage and validation processes to facilitate more efficient linkages in the future.
   5. Maintain a CODES Web site
   6. Facilitate teleconferencing and interdisciplinary meetings to ensure broad participation by all stakeholders.

D. Institutionalize CODES within the state to ensure routine linkage of the crash and injury state data and continued development of state-linked data applications that are useful for highway traffic safety decision making.

Problems and Solutions for Implementing the CODES Administrative Requirements:

Collaboration is the key to the successful administration of CODES. Communication must be maintained between the data owners. They share decision making authority for linking the state crash and injury data, for developing policies for confidentiality and release of the linked data and for institutionalizing CODES to support highway traffic safety and injury control. The data owners also must maintain communication with the major data users to ensure that the linked data are useful and available when needed. The organizational entity they designate to house CODES must have the capability to balance competing administrative priorities which inevitably exist in an environment where authority is shared.

The administrative implementation problems and solutions experienced by the 16 reporting CODES states are organized below under the themes of maintaining communication, supporting a collaborative source of authority, developing policies for confidentiality and release of CODES linked data, managing CODES (contracts, the organizational entity, personnel), and institutionalization.

MAINTAINING COMMUNICATION	
Problems	Solution
*Poor information flow to Board of Directors early in the project	Conducted more frequent board meetings, some during conference calls, to reinvigorate enthusiasm for continuing CODES.
*Maintaining effective communication among the various participating agencies	Allocated considerable staff time to support ongoing interagency communications as the project gained momentum and additional practical applications for the data were identified.

*Keeping all data contributors on board	Data contributors required a benefit for participating in CODES: a direct benefit is obtaining the linked data; an indirect benefit is support for a shared goal such as community health. It was important to demonstrate that CODES provides valuable information for the data contributor or its clients. For example, while motor vehicle crashes may represent 1% of hospital admissions, they are nearly always preventable and often affect people who are "in the prime of their life." Working to improve highway safety can generate community goodwill toward hospitals. The data contributor was kept informed about the use of their data and assigned credit whenever their data were used. All data contributors were required to participate in data release decisions. But they were not overly burdened with complicated data manipulation or complex tasks such as writing a data release policy. These tasks were handled by CODES staff (like you) and submitted for review by the data contributor when necessary. As the *sine qua non* of CODES, it was important to keep the data contributors HAPPY.

SUPPORTING A COLLABORATIVE SOURCE OF AUTHORITY	
Problems	Solution
*Goals of Advisory Committee may conflict with state agency/department	Assured routine and frequent communication with the members of the BOD/advisory committee.
*Problem with collaborative approach when department priorities conflict or resources are insufficient	Periodically renewed agency commitments through formal letters of agreement.
*Assuring full support and participation by all of the major data owners in the CODES	The focus was always on people rather than technology as the most important resource. Individuals, and their programs/agencies who would be most helpful in building a strong multi disciplinary interagency working group were identified. We sought representatives who had both an intimate knowledge of their data as well as the authority to make senior level policy decisions. This group ultimately formed as the foundation for our CODES Governing Board of Directors. Each potential member was asked to develop a prioritized wish list of ways that linked data could benefit their operation and to indicate available databases and resources which could benefit CODES. The focus was on practical applications.
*Development of interagency cooperative agreements to share data	Maintained close communication with agency board of directors to demonstrate the increased power of integrated data sets and how analyses improved with accurate and complete data.
*Problem developing a good working relationship with the Board of Directors and Advisory Council	Held meetings on a monthly basis using teleconferencing technology to facilitate greater participation. Progress was reported and questions and concerns were discussed. Board members were active decision makers.

*Authority for CODES Board of Directors and Advisory Committee functions split among existing councils and committees	The functions of the CODES Board of Directors were split between the existing Data Oversight Council and Injury Surveillance Advisory Council. The Advisory Committee functions were split between the existing Traffic Records Steering Committee and Injury Surveillance Advisory Council. Use of existing councils and committees that have data review responsibilities was expected to facilitate the institutionalization of CODES over time.
*Political – Turf and Scope of Authority. Who are you to do this project in our state?	Invited all stakeholders to participate and agreed to stay out of sensitive areas such as dangerous roads.
*"Little big man syndrome" - Hospital organization may want to assert its role in the data excess decisions.	Demonstrated the value of the linked data early in the project.
*Concern about who owns the CODES linked data	Established a policy that data bases, whether unlinked or linked, are owned by their original owners.

DEVELOPING POLICIES FOR CONFIDENTIALITY AND RELEASE OF CODES DATA	
Problems	Solution
*Developing data release policies	Used the Board of Directors to decide who could use the data, what data could be released, what mechanisms were necessary to protect the data owners, if there would be a fee to handle special data manipulation tasks and in which formats the data could be released. Questions were answered such as: Do you provide the data to anyone who requests it? What about lawyers on fishing expeditions? On the other hand, do you have criteria for determining who should get which data elements? All of these were particularly thorny issues but they were worked out by involving data contributors and users in developing data release policies and procedures. Deciding what data should be released was another issue. As researchers, we always want more data, but we also know that many data elements that are collected are not reliable or may be subject to gross mistakes in interpretation by other data users. A detailed data dictionary was developed. CODES staff were made available to discuss the use of linked data and appropriate research methodologies with CODES data users.
*Generating a set of guidelines for data usage and distribution that is not too limiting but maintains personal privacy	As a beginning, reviewed all existing data release policies maintained by the individual data contributors.

*Maintaining data confidentiality before and after linkage	Used the Board of Directors to determine policies that protected the patient and the self-interest of the data contributors. Breaches of confidentiality can destroy a data contributor's willingness to participate in CODES and could even have legal ramifications. CODES data are vulnerable to violations of the established confidentiality protocols of individual data contributors. For example, the hospital discharge data provider may have a standing policy to not release dates of hospital admission. However, by linking to the crash data and releasing the date of the crash, you have effectively provided the date of hospital admission. Special types of hospital admissions are often intentionally obscured in hospital data (psychiatric problems, alcohol/drug use, abortions, AIDS, etc.). If other data elements (the date and location of the crash) make it possible to identify an individual, you may be inadvertently releasing highly confidential data about individuals without their consent. Individual hospitals may not wish to have their own treatment performance scrutinized by researchers and results published without their consent. In a similar manner, the Department of Transportation may not be overjoyed that you have publicly identified particularly hazardous crash locations or even set the Department up for a lawsuit by individuals who have been injured at these locations. It was thus important to involve all data contributors in establishing protocols to protect data confidentiality and in turn their own interests.

MANAGING CODES: ORGANIZATIONAL AND FISCAL MANAGEMENT	
Problems	Solution
*Agency responsible for CODES changed three times since original grant was awarded in 1992.	Identified an agency that was committed to planning, implementing and maintaining CODES-related initiatives.
*Contracting mechanisms require considerable staff time and effort	Renewed agency commitments through formal letters of agreement.
*Bureaucratic red tape that impeded fiscal administration	At the end of the grant period, the location of CODES was changed to another agency, thus eliminating the fiscal management problems,

MANAGING CODES: PERSONNEL	
Problems	Solution
*Linkage process interrupted because of other responsibilities	Obtained administrative approval to schedule time away from the office to complete future linkage processes.
*Lack of dedicated personnel; staff faced with competing priorities and deadlines	Renewed agency commitments through formal letters of agreement.
*Staff turnover and use of "donated" staff present problems in maintaining a suitable knowledge base or schedule	Renewed agency commitments through formal letters of agreement.
*Back up for CODES staff not available	Renewed agency commitments through formal letters of agreement.
*Long learning curve for AutoMatch	Switched to CODES 2000 software when the new software became available.
*Personnel changes hampered progress	An acting administrator was assigned to the project until a permanent administrator was hired six months later.
* Implementing a large-scale project with minimal staff resources and ultimately having to cope with losing our primary staff person early in the project, with no possibility of replacing him	The commitment of existing staff who were willing to work overtime on the linkage plus efficient use of email, mail merge documents and telephone enabled the Project to stay on schedule and the Board members to receive frequent updates.

INSTITUTIONALIZATION	
Problems	Solution
*Lack of or insufficient dedicated and/or long term state funds for institutionalization	Attracted more customers by publicizing results to all users, in and out of government.
*Unable to create permanent position for the data analyst	Lobbied to create a permanent position within HHSS. In the meantime, the data analyst position was contracted out.

Linkage Issues

List of Requirements for the CODES Linkages:

- PROVIDE ACCESS TO DATA RESOURCES THAT ARE:
  - POPULATION-BASED
  - COMPUTERIZED STATEWIDE
  - INCLUDE EVERYONE INVOLVED

- DATA RESOURCES REQUIRED FOR CODES LINKAGE

A. Crash	Collected by police at the scene
B. EMS	Collected by EMTs who provide treatment at the scene and en route
-or-	
C. Emergency Department	Collected by physicians, nurses and others who provide treatment at the emergency department, in the hospital or outpatient setting.
D. Hospital	Collected by physicians, nurses and others who provide treatment after admission as an inpatient.

- OPTIONAL DATA RESOURCES FOR LINKAGE THAT ARE IMPORTANT FOR INSTITUTIONALIZING CODES

A. Death	Collected at the time the death is certified.
B. <u>Other Person Data</u>:	
Driver license	Collected when the driver is licensed
Citation/conviction	Collected when the driver is cited or convicted
Insurance Claims	Collected when the occupant files a health insurance claim
HMO/managed care	Collected when the occupant receives outpatient care
C. Roadway	Collected when inventory of roadway segments is created.
D. <u>Other Vehicle Data</u>:	
Vehicle registration	Collected when the vehicle is licensed.
Insurance Claims	Collected when the occupant files an automobile insurance claim

- **PROVIDE EVENT AND PERSON IDENTIFIERS TO DISCRIMINATE BETWEEN THE CRASHES AND THE PERSONS INVOLVED IN EACH CRASH.**

  A. Purpose
     1. Identify persons involved and injured in a motor vehicle crash from the scene.
     2. Trace the injured persons through the health care system to determine medical and final outcomes.

  B. Identifiers
     1. Use with probabilistic linkage techniques to identify valid pairs without the need for exact matches when it is uncertain which records should match.
     2. Include indirect (date, time, location of crash, birth date, gender) and sometimes direct (name, social security number, etc.) identifiers.

- **LINK CRASH AND INJURY DATA FOR ANY TWO CALENDAR YEARS AND VALIDATE THE RESULTS.**

  A. Software
     1. Use CODES 2000, a new probabilistic linkage software to determine the probability that a pair of records located in different data files represents the same person.
     2. All of the Data Network states will convert to CODES2000.

  B. Validation
     1. Document the significance of the false positives, false negatives and missing data
     2. Verify that the linked data are representative and generalizable for highway traffic safety purposes.

Problems and Solutions for Implementing the CODES Linkage Requirements:

In most states, the manual collection of medical and financial outcomes for specific environment/event, person and vehicle characteristics during a crash is not feasible. So linkage of these data electronically expands their usefulness without the expense and delay of additional data collection. The key to successful linkage is complete, accurate statewide data with sufficient identifiers to discriminate between both the crashes and person(s) involved in a specific crash.

The linkage process poses some problems. Records must be converted to person-specific to track those injured in the crash from the scene and through the health care system to final destinations. Data quality may be a problem when state data are not routinely edited, subjected to routine logic checks or tested for compliance with reporting thresholds. The linkage process itself highlights additional problems, unknown even to the data owner, with the quality of the data. Fortunately, probabilistic linkage techniques do not require exact matches to locate the valid pairs. However, not knowing which crash records should link to a medical record and vice versa complicates the validation process.

Different organizational entities are responsible for the different data files. Collaboration is necessary to get the job done given the available staff time. Multiple part-time staff may be needed since usually no one person has the necessary computer expertise and experience working with the state data. However, part-time CODES staff must deal with the frustrations caused by conflicting work priorities. Access to the data may be delayed when confidentiality policies vary among the organizational entities. Policies are usually more restrictive for accessing injury data, though some states also limit access to specific types of crash data, such as high frequency locations. The existing policies for each of the crash and injury data files participating in the linkage control, as a minimum, access to the linked data. All of these problems can be resolved: stakeholders can be educated; data sets can be improved and manipulated to facilitate the linkage; and all of the CODES states will convert to CODES 2000, a more user friendly version of the probabilistic linkage software.

The linkage implementation problems experienced by the 16 reporting CODES states are organized under the themes of data access, data quality/preparation, data linkage and validation.

DATA ACCESS: Some EMS Data Not Electronic	
Problems	Solutions
*Not all statewide EMS data are electronic	We obtained access to archived run sheets for 1996 and 1997 from the state EMS branch. A data entry operator examined all run sheets, and computerized those related to motor vehicle crashes. We entered 18,500 run sheets for 1996 and 14,000 for 1997.
*Lack of assurance of the availability of EMS data	CODES staff applied for other funds for the EMS agency to perform the EMS data entry.
*New legislation may eliminate case by case reporting	Decided to enter EMS records manually for years for which they exist (currently through June 2000) and to monitor the situation with the new state EMS board. Decided to explore the possibility of working with local Safe Community coalitions to obtain emergency department data for linkage.
*Two counties have separate EMS data collection systems	Used data resulting from the linkages for the rest of the state to show the value of CODES to potential data contributors who failed to participate because of infrequent communication with the project.

DATA ACCESS: Important data sets are missing:	
Problems	Solutions
*No access to Indian Hospital Services hospital data	We met frequently with IHS data personnel to discuss access to the IHS data.
*Outpatient non emergency records not provided to us initially	After negotiations with our hospital data provider, we were able to acquire the outpatient non emergency records for individuals with motor vehicle e-codes.
*Difficult to convince hospitals to provide electronic discharge data in the absence of a centralized state data file	Worked with state Assn. of Healthcare Organizations and Hospital Information Management Assn. (HIMA) to develop most effective way of getting data directly from each hospital.
*No hospital data for victims of motor vehicle crashes covered under the state's no fault insurance system	Because of the no-fault insurance system for motor vehicle crashes, the inpatient/outpatient data set was designed to include only patients who have a primary payer of Blue Cross/Blue Shield, Medicare, Medicaid, and several small private companies. Auto insurance is virtually always the primary payer for hospital inpatient/outpatient claims by victims of crashes. Neither the state insurance office nor the state hospital association could provide the missing hospital data for victims of crashes. We contacted other CODES states with a similar barrier and obtained copies of letters, data format specs, etc. which were useful for convincing hospitals to provide data for the linkage.
*Difficulty accessing hospital data because of statutes prohibiting the use of unique personal identifiers.	We relied on AutoMatch to perform our probabilistic linkages. We then invested significant time in examining random individual matches to satisfy us that these records, in fact, should have been matched.

DATA ACCESS: Delay to Acquire the Data:	
Problems	Solutions
*Lack of knowledge about the crash file content and database structure	Dedicated time and resources from DOT information services to educate the CODES linkage group about the crash data file.
* It took more than two months to fulfill a secondary request from the holder of the EMS data for a large county	Verified at the time of the request that all data elements required for linkage were included in the data request. When there was doubt whether a field was needed or not, the field was included. It was much easier to eliminate a field than to go back later when you are in the middle of the linkage.
*Gaining permission to access the data- even finding the databases	Access was obtained by having a clear purpose for the need for access – to study motor vehicle safety, etc. There were usually more data available than could be used; we discovered databases in virtually every public agency we approached.
*Maintaining the original ownership, understanding who owns the CODES database	Our policy established that databases are owned by their original owners, not by the CODES project.
*State law and regulation prohibits access to several identifying personal variables in health related data	Agreement was reached whereby state agency staff created the data extracts needed for linkage, including restricted data elements. The actual linkage occurred on site at the state agency. All current and future data linkages must be approved by the relevant state agencies. All persons utilizing linked data must sign confidentiality agreements.
*Confidentiality Issues: real or red herring?	It was not necessary to surrender before we started. Confidentiality and privacy were protected without giving up access to identifying information by controlling our own behavior.

DATA QUALITY/PREPARATION: CRASH DATA	
Problems	Solutions
*Separate crash data files (master, unit, operator, occupant, pedestrian) needed to be merged -the separate data sets contained identical fields with different variable labels	We cleaned the data. Also, we renamed non-uniform variables within each data set according to common labels. In the 'data merge' process, some newly-created linkage variables were implemented *before* the merge; some were added *after* the merge.
*Crash data do not include complete information for passengers reported as uninjured	Urged state officials who collect crash data to support complete reporting for all individuals in motor vehicle crashes.
*Crash file did not contain safety equipment use by non injured drivers	Included documentation of safety equipment used by drivers of all vehicles involved as part of the Crash Report Form revision for 2001.
*Crash file contained the names of only the injured drivers	Reporting of all injured names implemented in 1997 and included in crash file for 1997 linkage.
*Valuable data fields were not available. Very few personal identifiers included in the Department of Transportation data file	Convinced DOT to collect date of birth beginning with 1997 crash data.

*Multiple data coding schemes, for example for coding 2,500 minor civil divisions.	Entered each coding system into a spreadsheet and lined up the corresponding codes horizontally. (This may take a few hours but it is much easier than trying to write an individual line of code for each code you want to convert). Used this spreadsheet in a mail merge-like function in Word (or any other word processing program) to generate the source code. A useful tool in MS-Word is the catalogue function under Mail Merge to automatically write thousands of lines of error-free code at the push of a button. This tool can also be used to write code to make flat files and write data dictionaries and even match parameters for those using AutoMatch.
Location variables on crash data truncated	Location of crash was truncated due to the cost of data entry. GIS staff developed a method of cross-linking the location variable with geocoded addresses and state route numbers to improve the location data. GIS staff worked with Department of Public Safety through the CODES Board and Advisory Committee structure to revise the crash reporting form to improve quality of location variables.
*High occurrence of missing unit numbers in the 1995 crash data	The state DOT cleaned the data and resubmitted it to us.
*Finding the best way to utilize Driver History in the linkage while maintaining consistent information	Worked with Department of Driver Licensing to develop a database that would link consistently and represent the individuals involved in crashes.

DATA QUALITY/PREPARATION: EMS DATA	
Problems	Solutions
*EMS file did not contain service program times due to a problem with the software used by EMS providers to report data	Worked with EMS software vendor to fix problem.
*EMS data often incomplete and many records missing altogether	Encouraged a stricter policing of the data generated by a new EMS reporting form created for implementation in 2000.
*Incomplete/Inaccurate coding of EMS run data (unwritten coding changes)	Detailed edit reports were developed for data partners Department of Public Safety and DHEC-EMS. Edit reports contained two parts, a check for valid responses and a series of logic checks to ensure data consistency. Quarterly meetings with partners were instituted to review these data reports and identify solutions to problems. Additionally, ad hoc meetings served as a way to provide technical input into "quirky" coding and create a permanent solution.
*State EMS data was unusable. These data represented 15 of the 17 counties (16% of the state's population).	Lobbied the state agency responsible for rural EMS activities for changes to existing data collection forms to improve quality of the data collected at the scene.

*Multiple EMS responses for same patient	Aggregated EMS data into separate files for those patients transported directly to a hospital, those transferred to another EMS provider, and for those patients transported by air ambulance. These records were then linked to one another. The record for the EMS provider that transports the patient to the hospital was kept in the CODES database. In the final CODES data set, however, a relational database record ID number was established for multiple EMS records. This way, we had access to multiple EMS runs for the same patient if this information was needed but could still keep the number of EMS data fields to a reasonable number. This process was necessary because different skill levels respond at different points in time to a call for EMS producing multiple EMS records per patient. Often the multiple records are simply repeats of the same information and thus may make the data unnecessarily voluminous or confusing. On the other hand, accounting for multiple responses is an important question in making decisions about the allocation of EMS resources, in examining EMS triage patterns and even for record linkage itself (inconsistent recording of hospital destinations).

DATA QUALITY/PREPARATION: EMERGENCY DEPARTMENT DATA	
Problems	Solutions
*Large percentage of outpatient (emergency and non emergency) records not reported by hospitals	Efforts by the providers of the hospital data to improve reporting prior to CODES has produced more complete reporting in recent years.

DATA QUALITY/PREPARATION: HOSPITAL DATA

Problems	Solutions
*Change in data ownership and resultant differences in database maintenance	The most significant linkage challenge involved a change in ownership of the hospital data file at the state agency level. This resulted in changes in how some data fields were reported, thereby posing linkage problems. Resolution of this problem was expedited by working with the new owners of the data to help them meet their data reporting requirements to the new agency.
*Name fields contain multiple names: e.g., John/Jane Doe, John and Jane, etc.	The problem of multiple names in a name field was solved by the creation of a second record for the additional name. While this creates an additional record, it creates a secondary problem by increasing the amount of missing data. Data fields were manipulated and new data fields created to strengthen future linkages. This was a minor inconvenience for small data sets, but more cumbersome and time consuming for larger data sets.
*The utilization of different name formats by submitting entities in the hospital discharge data file: e.g., First, Middle, Last; Last, First Middle; etc.	The name format problem with the hospital discharge data file was resolved by identifying hospitals utilizing specific formats and building a standardized name field. Future problems may be alleviated by Department of Health personnel working with hospitals to standardize name formats.
*In the 1996 hospital data, some of the records had physician names in the 'hospital name' field	The Health Department assisted us in producing a list of physician names and 'probable hospital referral sites.'
*Low use of e-codes in the hospital data	To compensate for the lack of e-codes for linkage, we increased reliance on bodily location of injury (e.g., head, leg) and type of injury (e.g., fracture).

*Missing e-codes and discharge hours on hospital data set adversely affected the linkage of hospital cases.	We have reported our data findings to the hospital data advisory committee, which is pushing for more complete and accurate UB92 reporting. Also, we have attempted, through various channels, to educate healthcare providers about the importance of E-coding to injury research.
*Incorporating information from Indian Health Service when no patient charge information is available.	We worked on developing approximate charges for specific injuries, if appropriate.
*Inappropriate use of zero balances in hospital charge fields thereby implying that a given service has been provided	The problem of hospitals failing to report charge data resulted in the presence of zero balances for a significant number of individual records. This was particularly a problem for the 1996 and 1997 hospital discharge data. As a result, the Department of Health has informed providers that reporting of costs information is required with submissions.
*Standardizing unlike hospital records into usable format for linkage and analysis	With the Hospital Information Management Association's guidance, a standard template was used. Data received from individual hospitals were organized to conform to the developed standards as they arrived.

*Multiple hospital admissions for same patient	Used a unique patient identifier (e.g., SSN or a pseudo ID) across admissions to link these records together in SAS. This was accomplished by first sorting the records by the patient identifier and admission date (so the first admission date for each patient occurs first in the file). This file was then put into a flat (non-SAS file) and then read back in as an array. The array compared the patient identifier number on each line with the number on the subsequent line. If the numbers matched, the records were then concatenated and a new variable was incrementally established for each record that was concatenated together. If the patient ID numbers did not match, each record was read as a separate record. The reason you have to first put the file into a flat file is because SAS is not capable of "looking at" two records in a SAS data set at the same time. An estimated 10% of patients have more than one record. (One patient actually had 19 separate admissions for a crash). Multiple admissions for the same patient result from transfer to a higher level of care, readmission after discharge, or even admission before/after the crash for an entirely different reason. In some hospital record systems, patients admitted in one month may have a new record established if they are still in the hospital in a subsequent month. For example, a patient admitted on the 25th of May and discharged on the 6th of June, may have two hospital records, one for the period May 25-31 and one for the period June 1-6. It is important to link these records since these cases are often the ones in which greater than average injury severity and costs are incurred.

FILE PREPARATION	
Problems	Solutions
*Difficulties getting data organized and formatted in time for AutoMatch training	Relied on help from NHTSA experts during the first two days of training to standardize and format the data properly for use with AutoMatch.
*Our initial method of downloading data was automatically placing character variables into numeric format	Specified character format for the fields in question to get around the software default when re downloading the files.
*Multiple records in medical files	Linked to EMS files to get birth dates for occupants to identify duplicate records and assist in linking records correctly.
*Duplicate records	

DATA LINKAGE: MANAGEMENT	
Problems	Solutions
*As data needs change, the separate files change necessitating annual revision to the CODES linking and data analysis programs	Became as familiar with file, variable definitions and coding as possible for each year of data linked.
*Could not complete the linkage process from start to finish before being put on another project	Obtained administrative approval to schedule time away from the office to complete future linkages.
* Incompatibility of data software packages between us and our CODES peer state	We translated peer state's programming syntax into SPSS (our data package). Also, we received technical assistance from SPSS company support staff.

DATA LINKAGE: PROBABILISTIC TECHNIQUES	
Problems	Solutions
*Assuring consistent results	We hired an outside consultant to assist us in developing an Access97-based data processing engine which would accept raw data in a variety of formats and standardize them to comply with the database structure we had devised for our CODES. We felt that by automating the bulk of the standardization process, we would increase the probability of producing consistent results.
*The crash and hospital discharge files are huge making working with them time consuming and difficult	Kept the linkage as simple as possible.
*Matching Process Taking Too Much Time - Some of our databases were too large	By utilizing cause of injury (E-Codes) and diagnosis (ICD-9) codes that were crash related, we were able to cut down the number of hospital records from 353,000 to 101,000. By dividing our DOT database into three distinct data sets using "county" as criteria, linking to EMS data became more efficient.
*Finding a good point of linkage between EMS and Crash Records when records are incomplete and inconsistent	Requested and received additional information from the major ambulance services so we could use location of services to approximate county of crash.
*Understanding probabilistic linkage. Not a black box - does it work?	We learned about it by using fake databases, real databases, etc. Learning here can only occur by actual time at the computer.
*When we had exact match links to more than one record.	With limited personal and location identifiers, multiple links were expected. We chose to not use any records with multiple links (approximately 3.5% of the matches). Date of birth was added to the crash data file to minimize this problem in the future.

DATA LINKAGE: FAILURE TO LINK	
Problems	Solutions
*Duplicates-especially among vehicle occupants	By removing non injured bus occupants from the linkage process (~5,000 per year), we reduced the number of duplicate matches.
*Linkage success in large urban area	To improve the linkage success in a large urban city, a quadrant variable was added. Each crash and responding EMS agency was assigned to one of four city quadrants.
*Success of crash-to-EMS as first link	To improve the overall success rate, we first matched the crash file to the hospital discharge file. Since these two files had more discriminating variables in common, the crash-hospital-EMS linkage was improved by 23% over the crash-EMS-hospital route. We also used the state's trauma registry to add scene location and person type to hospital data.
*Transports in/out of state	
*Late arrivals at hospitals	
*Dates in crash, EMS and hospital data were not in the appropriate format	We computed new data fields with yyyymmdd format.
*Victims of evening crashes may have been admitted (hospital) the following day	We created a new date variable where evening (e.g., 8PM or later) crash victims were assigned a 'next day' value.
*Lack of strong patient identifiers (names, SSN) in all data sets	To compensate for the lack of strong identifiers, we created additional geographic indicators.
*Injury information in the EMS and hospital data was not uniform. -EMS ~ 90 dichotomous variables -Hospital ~ nine ICD-9 codes	We created ten new 3-digit injury variables: First two digits indicated body location (e.g., head, legs); third digit indicated injury type (e.g., fracture, blunt).

*Subset data into categories where there was an expectation for linking (can't link the crash to a hospital if the hospital is out of state)	Through the edit reports, the CODES staff developed expertise in the content of each variable in the files. Detailed 'brain storming' sessions were held to identify the records from each file that had a probability of being linked. For example, ambulance run reports that transported crash victims out of state only have a probability of being linked to crash data and not hospital data. Crash records that have no name, no birth date or other event identifier have a very low probability of being linked. Ambulance run reports that identify a specific receiving hospital have a high probability of linkage to that hospital's data.

VALIDATION	
Problems	Solutions
*False positives	A variety of studies were undertaken to evaluate the potential of false positives: -a random sample of hospital medical discharge records was cross-referenced and evaluated by comparison to the linked data. -A study of Medicaid eligibles was undertaken to determine whether appropriate linkages occurred. Additional personal identifier variables, not typically released to the public, were acquired to improve the linkage process.
*False negatives	Evaluation of missing links utilizing hospital e-code data was undertaken. Estimates were made of potential victims transported to hospitals for which no crash data were available. Despite the lack of identifiers, such as name and address, and the unavailability of data for most out of state discharges, missing cases did not lead to biased results with respect to outcomes evaluated utilizing linked data.
Initially kept record pairs with questionable weights as matches.	We redefined a “match” using revised cutoff weights based on new calculations.

Application Issues

List of Requirements for CODES Applications Using CODES Linked Data:

- DEVELOP AND INSTITUTIONALIZE THE DEVELOPMENT AND PRODUCTION OF APPLICATIONS FOR THE LINKED DATA THAT HAVE AN IMPACT ON TRAFFIC SAFETY DECISION MAKING

A. State-Specific Applications

1. Routine Standardized Reports displaying rows and columns of totals, percent and rates describing the outcome, medical and financial, for specific event, vehicle, or crash characteristics.

2. Fact Sheets to broadcast some of the results reported in the Routine Standardized Reports.

3. CODES Web site to increase public access to the information generated by CODES.

4. Research analysis to define priority highway safety issues.

5. Option to develop or incorporate CODES into a geographic information system

B. CODES Data Network

1. Funds provided for at least .5 FTE to CODES states with at least two years of linked data

2. Facilitate access by NHTSA analysts to CODES linked data to support NHTSA research priorities

Problems and Solutions for Implementing the Requirements for CODES

Applications:

Collaboration is the key to developing applications that will have an impact on highway safety decisions. Partnerships must be developed with the users of the linked data to facilitate compliance with existing confidentiality policies. Whatever review process is implemented should not prevent use of the data in a timely manner to target resources that reduce mortality, morbidity, injury severity and costs.

Timeliness also depends upon the availability of staff at the time the linked data are available, not always possible when staff share CODES with other competing priorities. Hiring contract staff to compensate may not be possible because of existing personnel policies.

Before the linked data can be used, their statistical implications must be understood. The limitations of routinely collected data may restrict case selection. It is crucial to know if bias exists and its source. And it is important to resist the temptation to attempt detailed and finely tuned research.

Presentation of the results must be at an appropriate level for the audience. Developing a CODES Web site that provides access to aggregated data reports or to a query system for customized reports increases public access to the linked data. It also saves personnel time.

The application implementation problems experienced by the 16 reporting CODES states are organized below under the themes of statistical issues, personnel, confidentiality issues, limitations for case selection, production issues, decision-making, and Web-site development.

STATISTICAL ISSUES	
Problems	Solutions
*Need for improved spatial/statistical tools for analyzing CODES data	Used ArcView to map spatial patterns Used SAS or CrimeStat to analyze spatially linked data Used AutoMatch or GIS software for geocoding
*Need to integrate different spatial databases (TIGER, DLG, PARCEL, CRASH, INJURY OUTCOMES)	Integrated land use data and aerial photographs in analyses.
*Need to devise appropriate statistical measures	Developed the statistics first, the mapping next.
*Accurate interpretation of data and results by outside agencies	Data results and statistics were interpreted in a number of ways. We followed the approach taken by other CODES states by having data requests come through the linking agency and then responding to them in accordance with guidelines developed by the Board of Directors. This provided users of the information with a single contact point for questions and clarifications.
*Missing data	Missing data caused under reporting of the impact of automobile crashes. In some instances, it was necessary to eliminate records with missing data and to perform the analysis using a reduced sample size for linked data sets. Sample size reductions not withstanding, sample sizes were usually adequate for extrapolation to the general population.

*Selecting samples for comparative analyses	Dependent variables in one set of records were commonly compared with the same dependent variables in another similar set of records for analysis. For example, to evaluate the effectiveness of seat belts, injury outcomes were compared for those using seat belts with those not using seat belts. Other independent variables (the type of crash, seating position, age, sex, vehicle speed, angle of impact, etc.) were taken into account using regression modeling such as logistic or linear regression or through an analysis of variance. But what happened when the selection criteria were themselves a source of bias? For example, a logistic regression accounting for the independent variables (age, sex, seating position, angle of impact, etc.) was performed for records comparing those for whom an air bag was deployed with records for those without deployment. When this approach was used, it first appeared that a crash in which an airbag deployed was *ipso facto* more serious than a crash in which an airbag was not deployed causing deployment to be associated with injury. This potential bias was resolved by restricting the case selection to those records for drivers of vehicles involved in high or moderate speed frontal impact crashes. By refining the selection criteria, it was possible to develop a more homologous group of records to analyze.

*Detecting subtle differences	Simply put, we identified both the power and the limitation of CODES data for analysis. CODES data, generated from the real world rather than from a carefully controlled clinical trial or scientific experiment, were not collected for the purpose of doing detailed and finely tuned research. Crash and EMS data were collected under difficult field conditions by hundreds (if not thousands) of observers all of whom may at times apply their own subjective interpretation to the data they collect. "Serious" and "moderate" injuries were defined differently by different police officers. Crash victims had an incentive to lie about safety belt use or vehicle speed, while other data elements were reconstructed from imperfect human memory. Moreover, "probabilistic" linkage meant that we know a percentage of linked records were in fact false positives. Because of these characteristics, the large volume of CODES data was most useful for detecting dramatic differences among crash victims (how effective are seat belts?) while at the same time they offset the impact of inaccurate data. Detecting subtle differences was more difficult with inaccurate data or small numbers of records. Sometimes, the problem of small numbers of records for analysis was overcome by using multiple years of data, but we had to account for differences in the data over time. For example, the availability of air bags changed dramatically over the last ten years, becoming mandatory only in 1994.

*Development of a "drill down" method for injury reporting	Community injury assessment required the inclusion of data for all injuries, not just motor vehicle crashes. The ability to place motor vehicle crash injuries in the context of total injuries provided a baseline to develop intervention strategies. CODES staff worked with local Safe Communities groups on developing and refining an injury profile. The "drill down" method of injury assessment allowed the users to identify statistically different injury rates, identify the specific sub populations at risk, and body parts that were injured. CODES staff relied on medical, public health and public safety expertise to facilitate the development of these reports. This effort required extensive staff time to coordinate activities.
*Common Definitions: What is totaled? Which speed number is important?	We developed common definitions at TA meetings.
*Common models: What covariates are important? - alcohol versus time of day.	We developed common models at TA meetings which were crucial for understanding the important covariates and interactions.
*Statistical Methodology. How simple is logistic regression to learn? Compared to log linear analysis? Are the observations independent? SAS?	The statistics used in the CODES applications were not trivial and most analysts did not have extensive experience with them. Logistic regression was often used for CODES data analysis. Yet log-linear analysis, which is more difficult, may be more appropriate. We addressed issues of nesting, eg, multiple occupants in a multiple vehicle crash are dependent observations! Finally, SAS was a really big playground (and sometimes available with an educational discount).

*Statistical methods to be used for different outcomes	Ordinary least squares methods were compared to logistic regression methods for results with censored variables and binary outcomes. Results led to the use of more appropriate multivariate techniques when such outcomes were studied.
*Lack of standardized element coding	A great deal of effort was put into ensuring that common codes resulted for all data sets included in the linked data.
*Over reporting of safety belt use resulted in overestimates of seat belt safety impacts	Seat belt estimates were adjusted utilizing new seat belt use variables developed using multivariate techniques.

PERSONNEL ISSUES	
Problems	Solutions
*Shortage of on-staff expertise in traffic safety research applications	We developed relationships with traffic safety experts in agencies across the state, including the Federal Highway Administration, State Police, and the state's Transportation Center. Their input was vital to our CODES project. Studies by CODES1 and CODES2 states have been invaluable sources of information for applications.
*Determining how our CODES could have the greatest positive impact on traffic safety with our limited project staff resources.	Thanks to the active participation of and guidance from our Board of Directors, this task was relatively easy. Our Board felt very strongly that we should concentrate on developing real world applications for our data analyses, and that we should design our reports to meet specific local community needs. As a result, every data extract, study, and report produced since the beginning of the project has been in direct response to a specific request from a local community, government agency, or individual citizen.

CONFIDENTIALITY ISSUES	
Problems	Solutions
*Confidentiality Policies	As the public release data set became finalized and more data analysis was released to the public, we attempted to resolve the problem of variables that may be considered confidential in one data set but not another (i.e., county). Our solution to this problem was to exclude identifiers and reduce the ability to make comparisons (i.e., response times) that are sensitive.
*Determining the most direct way of distributing information without compromising confidentiality	Worked with Board, the state's Association of Healthcare Organizations and the Hospital Information Management Association to determine guidelines acceptable to both data owners and users.
Providing too detailed tabular or other information from linked data may allow individuals to be identified	Tables and reports using linked data were developed utilizing the same constraints that govern use by the relevant state agency owning the health outcome data.

LIMITATIONS FOR CASE SELECTION	
Problems	Solutions
*Hospital discharge records in ICD-9 format and death records in ICD-10 for 1999	Maintained data bases from year to year with the same data items.
*Use of study results	Targeted studies toward topics which are of current concern.
*Existing definitions of rural and urban not well suited for our study of rural and urban crashes	We explored several existing definitions for rural and urban, but none seemed well suited for the state. With the assistance of our data management team, we came up with a definition that we felt suited us well.
*Data fields were not available in the DOT database for many of the applications requested:	
-injuries to passengers, especially children: No personal identifiers for any passengers	There was little we could do without passenger identifiers, including their location in the vehicle, gender, age, and either a birth date or social security number.
-injuries to passengers riding in the back of a pickup truck: No way to identify truck passengers or their location	
-injuries to bicyclists involved in crashes with a motor vehicle: Bicyclists were not identified	Characterizing the outcome of child occupants was critical to support new legislation to increase restraint use. The same can be said for legislation to increase bicycle safety on public roads.

*Reporting of hospital charges as opposed to hospital costs.	The financial information submitted by providers was charge data as opposed to cost data. Given that different providers have differing rationales for pricing decisions, analysis results tended to overstate the problem of motor vehicle crashes. Since length of stay was obtainable for the linked data set, it was utilized in analysis as a surrogate measure for hospital costs.
*Lack of clearly defined data keys/ data dictionaries.	While most data identifiers were obvious, the interpretation of certain keys (e.g., driver/ pedestrian position and driver/pedestrian keys can yield conflicting results) was unclear and made analysis difficult. Resolution of this problem was aided by working closely with state agency personnel to clearly interpret and communicate the meaning of data identifiers.
*The collapsing of field keys into usable categories (e.g., the vehicle type field has passenger cars listed in two categories)	In an attempt to capture more accurate data, state agencies tended to categorize data into highly specific categories. This often led to data that were not useful for the purpose of analysis. The problem was resolved by collapsing data into fewer specific categories.
*Incorrect entry of date fields	For the 1995 data set, date fields were entered as string fields that were not convertible to date fields in their present form. This required that individual day, month and year fields be created and then concatenated for purposes of creating a date field for reports.

PRODUCTION ISSUES	
Problems	Solutions
*Lack of planning caused many problems with the first application	Because we did not plan well for our first application, we ended up doing many rewrites that cost us valuable time. We are now planning all details of our applications prior to the start of any analysis.
*Process of determining components of our first application was too time consuming	We have revised our editing process in order to make it more efficient
*Process of reviewing and revising the report took much longer than expected	
*Disseminating a voluminous project report (1081 pages) in a user-friendly format which would facilitate browsing and quick searches for specific information	As the number of requests for CODES studies and reports continued to grow, it became evident that our first project report would be too large for conventional distribution as a printed document. Because we had developed in-house expertise in computer graphics and interactive CD production, we elected to produce the entire report as a totally self-sufficient interactive CD ROM. Therefore, in addition to text, we included numerous photographs, other graphics, and PowerPoint slide presentations, together with all of the software necessary to view any of the CD's contents. We also incorporated programming that would automatically run the project report menu and guide users through learning how to use the various resources on the CD. The response to the interactive CD was overwhelmingly positive, to the extent that we have distributed well more than 100 CDs since September of 1999.

*Keeping up with the increasing statewide demand for CODES reports and services	We still have only one staff person (Project Coordinator) and that person cannot devote his full-time efforts to CODES. Therefore, we continue to rely almost exclusively on email for receiving requests for reports as well as for distributing them, generally as attached Adobe Acrobat PDF files or PowerPoint slide shows. Not only have we been able to deliver all requested reports/data extracts on time, but we have also added NHTSA's Safe Communities program to the list of community-based initiatives which we actively support. Our active involvement in local community projects has gained statewide recognition and support for CODES. As we continue to establish our EMS Web page, we will expand the CODES section to include many of the reports already produced as well as a mechanism for requesting services from the CODES project.

DECISION-MAKING	
Problems	Solutions
*Developing partnership with organizations that can use linked data in an advocacy role	Worked with agencies providing information (i.e., KIDS Count, Safe Communities, the Maternal and Child Health Study). Continue to look for ways to provide data to safety advocates.

WEB-SITE DEVELOPMENT	
Problems	Solutions
*Underestimating the cost	Development of a Web site required extensive knowledge about our data users prior to the estimation of cost for the project. Data user needs were not fully evaluated prior to developing a cost estimate. The CODES staff envisioned a Web site that had limited database query capability and a static map. An evaluation of the data user needs revealed that they needed the ability to generate ad hoc reports and maps customized to their location. To incorporate the ability for dynamic mapping and on-line query capability into a Web site greatly increases not only the cost but also the staffing requirements for the project. Hardware, software (both database management and mapping) and staffing requirements to meet the needs of the data users far exceeded the funds allocated in the CODES grant. The original vision for the Web site will be completed and serve as the "reports" section for the enhanced Web site. Currently, the CODES staff is seeking funding for the enhanced Web site development.

RECOMMENDATIONS

The following three sections include the recommendations, described as the most important by the 16 reporting CODES states, for implementing the administrative, linkage, and application requirements for CODES.

Administrative Recommendations

The administrative recommendations focus on the CODES Board of Directors, collaboration, priorities, communication and project management.

CODES Board of Directors	*A successful CODES needs a Board of Directors that can be expanded as necessary.
	*Develop strong interagency trust for the rough times.
	*Decide who "owns" the CODES data
	*Establish written policies for release of the linked data
	* Inform data contributors of data releases
	*Develop the CODES Board of Directors and Advisory Structure using existing councils and committees, where possible.
	*Use the Board of Directors and Advisory Structures to address issues related to procurement of data, confidentiality, and data dissemination
	*Develop strong working Board and ask for input and assistance whenever possible. Most board members have contacts and resources that can be very helpful
	*Develop clear policies about distribution early to prevent concern and confusion about how linked data will be used.

Collaboration .	*Obtain firm commitments from participating agencies, such as memoranda of understanding, and include commitments to provide both human and financial resources.
	*Develop strong interagency agreements for the sharing of data sets with significant emphasis on confidentiality
	*Give data contributors credit
	*Show data contributors the value of CODES
	*Develop contacts with existing CODES states.
	*Partner with groups such as the state's Association of Healthcare Organizations and Hospital Information Management Association. While they may not be data owners, these groups can provide invaluable information and assistance.

Priorities	*Develop a strategy to disseminate results when in the planning stages of the study.
	*Select one or two topics to focus on; don't try to answer all the questions at once.
	*Have a small number of people involved who are also potential users of the data and study results.
	*Involve data analysts and statisticians as well as administrators from agencies that provide the data.
	*Take an active role in all activities related to highway traffic safety in order to expand the scope of CODES

Communication	*Keep upper management informed and involved in CODES
	*Keep data contributors informed
	*Work closely with agency stakeholders for the purpose of obtaining data sets.
	*Impress upon stakeholders the importance of accurate and complete data sets in order to increase the power of analysis.
	*Provide useful information to stakeholders for improvement of service delivery.
	*Invite all possible stakeholders to the table. If someone cares, invite them. Feed them, establish trust

Project Management	*From the start, hire a full time CODES administrator
	*Design your CODES so that the operation does not depend too much on any single staff member or position. Be sure to have a "Plan B"... and perhaps even a "Plan C," in case you lose critical project staff or other resources. Improvise, adapt, and overcome.
	*Be sure that the agency which will be your fiscal agent has a proven track record with similar projects. Don't be afraid to shop around for a good fiscal agent. Ask other CODES states or NHTSA for suggestions if you run into trouble.

Linkage Recommendations

The linkage recommendations focus on data access, data quality/preparation and probabilistic linkage.

Data Access	*Learn about the structure of your data sets as soon as possible. Contact other CODES states to find others that have similar data set structures.
	*Create a "dream list" of the ideal data sets and elements. Having this list ahead of time can be beneficial when you approach your data holders. Instead of waiting to see what they give you, show them the most important fields ahead of time. They may not have the particular field you want in their file, but may be able to tell you where to find it from another agency. When you actually find what data sets and fields are available in your state, it's more than likely your "dream list" will be heavily edited.
	*Create a game plan and a time-line for acquiring data sets. Whom are you going to get them from? What data elements are available? When can you get them? Are they going to be usable? If not, is there a way to overcome any limitations? Are there other data sets available to augment your basic data sets? For example, if your crash file only contains a driver's license number for a personal identifier and your EMS file only contains the person's social security number, is there an ancillary data set that may have both? If your state DMV data set has both driver's license number and social security number, you now have a method to augment both your crash file and your EMS file.
	*Negotiating the use and availability of data and data elements from owners is one of the most critical components of establishing a CODES system. This will become even more important with the passage and implementation of HIPPA. States need to invest a large amount of time establishing working relations and intergovernmental "trust" with relevant data owners.
	*To find data and obtain access, personally visit the owners and provide them with a single page executive summary of the project.

Data Access (Cont.)	*Accept any data format, but make a suggestion. Many providers of data could easily match your suggestion and this will save time later.
	*Seek all possible identifiers on all files. Ask for names – they might say "yes."
	*Finally, don't hesitate to ask a question here or contact anyone after you leave. We have all gone through quite a learning curve and (from personal experience) at times have felt we must be the only state experiencing problems with data and the linking process. What we have learned is encountering problems is more the norm rather than the exception. Ask a lot of questions and take a lot of notes. If you already have an idea about your data, start asking questions now.
	*Determine the quantity/quality of patient identifiers in your data. If ID information is scant, seek advice from other CODES states with a similar problem.
	*Assess your state's use of e-codes. If use is low, place greater emphasis on use of existing identifiers and the development of new ones.
	*Seek the advice and assistance of data stakeholders in overcoming barriers that arise.
	*Invest ample time in cultivating relationships with CODES data owners.
	*Maintain a good working relationship with the owners of the data sets you use.
	*Work closely with data owners to fully understand databases. A clear understanding of the data will prevent duplication and improve quality of linkage. Review reports and information from crash records thoroughly.
	*Ensure that CODES staff and data stakeholders are in agreement with the steps entailed in the data acquisition process. Each stakeholder should address their own legal requirements for releasing data as soon as possible to ensure proper lead time to file Data Use Agreements.

	*Obtain approval to access the data prior to initiating the CODES project. Ensure that all data stakeholders understand the role and importance of CODES.
	*Recommend additional variables to be collected in core data sets (i.e., date of birth, location).
	*Check each year for changes to the data sets, both new/ discarded data items, as well as changes in values of continuing data variables.
	*Understand who will use data and how it will be used before linking data. This will help to organize information more clearly and allow for more ready access.

Data Quality/ Preparation:	*If you decide to computerize significant amounts of data manually, don't underestimate the time and effort required to do the job well.
	*Work with data collectors and owners to improve data linkage by completing fields that have already been established (i.e., EMS run number on crash and hospital data).
	*Use a data transformation software package suitable for you. Contact CODES states using same software for advice. Also, seek out other avenues of assistance, such as software technical support. Learn how they managed their obstacles.
	*When receiving data sets, ensure they are accurate and complete. Check for unusual amounts of missing data in the crash 'unit number' field.
	*Ensure that your download method is not modifying the data sets.
	*Contact other CODES states for programming syntax that recodes dates into the required format. Be aware of this issue. Compute this new variable in preparation for its use in the data linkage process.
	*Assess the extent to which data sets have common fields. Some fields between data sets may only appear to be held in common. Assess the quantity of missing values in the data sets.
	*Assess the extent of compatibility of injury data between data sets. Consult other CODES states to develop strategies for increasing the compatibility
	*Conduct preliminary checks to ensure sufficient hospital data for MVC victims; the analysis of e-codes and/or payment source would assist this effort.
	*Identify the needs of data partners that may be helpful in understanding the data structure and provide a solution that will be mutually beneficial.
	*Identify experts in hospital billing, crash reporting, and EMS reporting to provide you with the little coding "nuances" that will save you hours of work.

	*Provide technical expertise to data partners relative to the revision of forms.
	*Develop an intimate knowledge of variables from each file to address the issues of reliability, validity and consistency of these data.

Probabilistic Linkage Issues:	*Educate data owners on the importance of complete and accurate data for the linkage process. In our experience, successful data linkage depends primarily on the quality of the data collection.
	*Understand limitation of data. This will minimize time wasted on matching fields that are incomplete or unsuitable.
	*Streamline your data sets before attempting any linkage. Why link 500,000 medical files if only 80,000 are motor vehicle related? If there is a way to extract only motor vehicle related incidents from a major file, do so as soon as possible. As part of your game-plan when you speak with data holders, ask them if they can give you only motor vehicle related records or if they can tell you what field(s) will help identify motor vehicle related records.
	*DO NOT BELIEVE probabilistic linkage until you completely understand what it is doing. It is exceedingly easy to make errors with linking software and have invalid links.
	*Consider an industrial size database at the outset of your project and use SAS views into that database. SAS is a terrible database engine.
	*Hire a computer nerd or maybe two. Read the previous suggestion again. Do it or your project will fail.
	*Use CODES-2000 to standardize your data processing and linkage processes. This will enable you to more easily manage dissimilar databases from a variety of sources. It will also make it easier for you to incorporate new databases into your CODES as the project grows. It will also be easier to collaborate with other CODES states if we are all using the same software.
	*Use the same strategies in processing and analyzing your data from year to year. Be sure that you get the same results every time from the same data no matter who performs the analysis.

	*Studies need to be made to evaluate the impact of "false negatives" and "false positives" on data available through a CODES linked data system. This is needed to ensure that reports and analyses developed using CODES data do not provide biased results.

Application Recommendations

The application recommendations focus on statistical issues, formats, decision-making, production and web-sites.

Statistical Issues:	*Use of appropriate statistical techniques is critical when performing analysis using CODES linked data. If a person with statistical training is not part of the CODES core staff, such a person needs to be added to the staff.
	*Given over reporting of seat belt use, and its consequent impact on measuring seat belt effectiveness, care needs to be taken in interpreting results when self reported seat belt use is included in analyses. While no simple solution is available, providing a range of estimates of seat belt effectiveness would provide a more appropriate picture for policy makers.
	*Keep in mind that proper data analysis is more dependent on the quality than the quantity of data (i.e., the quality of the linkage is more important than the number of links).
	*Beware of biases in your data that can influence your results.
	*Before releasing any data, verify numbers with relevant reports produced by the data owners.
	*Welcome external evaluations of your work.
	*Know the power and limitations of CODES.
	*Be aware of multiple records.
	*Find a statistician for your project, but try to read the statistics books yourself as well. The statistician cannot do all the analyses.
	*Learn to use SAS. Eventually you will have no choice.
	*Clearly define and explain all results, tables, graphs, etc.; never assume readers will “figure it out” on their own.

	*Don't try to reinvent the wheel. Take advantage of the knowledge acquired by the other CODES states. Different states have different methods for analyses of their data. These methods are a function of staffing and the availability of data sets, data elements and software. It is unlikely that one state will be able to answer all the problems you are going to have. A state that has strong personal identifier fields in its crash file may lack personal identifier fields in its medical files. Another may have the opposite.

Decision-making	*Get to know your state's key agencies and individuals involved in traffic safety research. Show them how CODES data can be of use to them.
	*Use analysis of the CODES data to help local agencies and public health groups improve prevention and intervention strategies.
	*Use CODES data as feedback to the police, EMS and hospital personnel that are doing the initial data collection.

Production:	*Work with your Board and data owners to develop a policy to provide requested reports and analysis.
	*Keep your data owners happy.
	*Keep in mind all the possible uses for your data and include those users in your planning whenever possible.
	*Make your studies as timely as possible.
	*In your planning process, clearly define the issues to be studied and stick to them.
	*Creating a computer program that produces standardized reports will save you tremendous amounts of time; these will serve as a starting point for just about any application as well as providing summarized information.

Production: (Cont.)	*Stay in close contact with NHTSA in order to stay informed of current goings on in motor vehicle safety and to remain on the same page with NHTSA.
	*Know your objectives before implementing any CODES activities.
	*Keep good relationships with all data users.
	*Be sure to give your data owners something useful in return for their contributions. Perform studies and analyses for them or design projects that will provide them with useful information, which they can use in their operations. Give credit where credit is due.
	*Community assessments of motor vehicle injuries must be completed in the context of all injuries.
	*Motor vehicle injury data provided to Safe Communities, Safe Kids and other local programs must be in a format that addresses local issues in a timely manner. Present the data in a format easy to understand. A picture is worth a thousand words.
	*For all applications, survey (update prior surveys) data users to identify current data needs.
	*Develop cost estimates that reflect a range of services that meet the identified data users needs.
	*Work with advocacy groups early on to determine what type of information they will want.
	*Keep in contact with data owners to make certain that usage is acceptable to them.
	*Participate in outreach activities to continue to make others aware of applications for linked data.
	*Absolutely stick to the original proposed question for at least the first year. Establish a track record for respecting data owners.

Web Site	* Invest whatever resources are necessary to develop interactive CD production capabilities in support of an active web page.
	*Use your Web site and email capabilities to manage the bulk of requests for service, relying on snail mail and personal presentations only when they are likely to further promote the CODES.

REFERENCES

1. National Highway Traffic Safety Administration. *Report to Congress on Benefits of Safety Belts and Motorcycle Helmets, Based on Data from the Crash Outcome Data Evaluation System (CODES)*. DOT HS 808 347. Washington, D.C.: Department of Transportation, February 1996.

2. Allen, M., Weiss, H. *Using Linked Data to Evaluate Collisions with Fixed Objects in Pennsylvania.* DOT HS 808 800. Washington, DC: Department of Transportation, National Highway Traffic Safety Administration, October 1998.

3. Finison, K., DuBrow, R. *Analysis of 1996 Maine Crashes Involving Vehicles that Ran Off the Road.* DOT HS 808 889. Washington, DC: Department of Transportation, National Highway Traffic Safety Administration, April 1999.

4. Zuckier, G., Jacobs, L., Thibeault, L. *Using Linked Data to Evaluate Severity and Outcome of Injury by Type of Object Struck (First Object Struck Only) for Motor Vehicle Crashes in Connecticut.* DOT HS 808 973. Washington, DC: Department of Transportation, National Highway Traffic Safety Administration, September 1999.

5. Finison, K., DuBrow, R. *Analysis of 1996 Maine Crashes Involving Young Adults.* DOT HS 808 887. Washington, DC: Department of Transportation, National Highway Traffic Safety Administration, April 1999.

6. Finison, K., DuBrow, R. *Analysis of Seat Belt Use and Outcomes in 1996 Maine Crashes.* DOT HS 808 888. Washington, DC: Department of Transportation, National Highway Traffic Safety Administration, April 1999.

7. Moore, M. *Comparison of Young and Adult Driver Crashes in Alaska Using Linked Traffic Crash and Hospital Data.* DOT HS 808. Washington, DC: Department of Transportation, National Highway Traffic Safety Administration, January 20, 1998.

8. Zuckier, G., Jacobs, L., Thibeault, L. *Using Linked Data to Evaluate Motor Vehicle Crashes Involving Elderly Drivers in Connecticut.* DOT HS 808 971. Washington, DC: Department of Transportation, National Highway Traffic Safety Administration, September 1999.

9. Allen, M., Weiss, H. *Using Linked Data to Evaluate Hospital charges for Motor Vehicle Crash Victims in Pennsylvania.* DOT HS 808 799. Washington, DC: Department of Transportation, National Highway Traffic Safety Administration, October 1998.

10. Zuckier, G., Jacobs, L., Thibeault, L. *Using Linked Data to Evaluate Medical and Financial Outcomes of Motor Vehicle Crashes in Connecticut.* DOT HS 808 972. Washington, DC: Department of Transportation, National Highway Traffic Safety Administration, September 1999.

11. Castle, S., Woods, B. *Using Linked Data to Evaluate Traumatic Brain Injuries in New Mexico.* DOT HS 808 798. Washington, DC: Department of Transportation, National Highway Traffic Safety Administration, October 1998.

12. Karlson, T., Bigelow, W. Beutel, P. *Serious Lower Extremity Injuries from Motor Vehicle Crashes, Wisconsin 199*1-1994. DOT HS 808 791. Washington, DC: Department of Transportation, National Highway Traffic Safety Administration, September 1998.

13. Finison, K. *Standardized Reporting Using CODES.* DOT HS 809 048. Washington, DC: Department of Transportation, National Highway Traffic Safety Administration, April 2000.

DOT HS 809 200
February 2001


NHTSA
People Saving People
http://www.nhtsa.dot.gov


NHTSA
People Saving People
http://www.nhtsa.dot.gov