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

NHTSA
www.nhtsa.gov

DOT HS 811 318

May 2010

Closing the Data Quality Assurance Gap in the Fatality Analysis Reporting System (FARS)

DISCLAIMER

This publication is distributed by the U.S. Department of Transportation, National Highway Traffic Safety Administration, in the interest of information exchange. The opinions, findings, and conclusions expressed in this publication are those of the authors and not necessarily those of the Department of Transportation or the National Highway Traffic Safety Administration. The United States Government assumes no liability for its contents or use thereof. If trade names, manufacturers' names, or specific products are mentioned, it is because they are considered essential to the object of the publication and should not be construed as an endorsement. The United States Government does not endorse products or manufacturers.

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT HS 811 318		

4. Title and Subtitle	5. Report Date
Closing the Data Quality Assurance Gap in the Fatality Analysis Reporting System	May 2010
	6. Performing Organization
7. Author(s) John McDonough and Christopher Smith	8. Performing Organization
9. Performing Organization Name and Address National Institute of Safety Research, Inc. 326 Hill Top Road Oakland, MD 21550	10. Work Unit No. (TRAIS)
	11. Contract or Grant No. DTNH22-07-D
12. Sponsoring Agency Name and Address State Data Reporting Systems Division, Office of Data Acquisitions National Center for Statistics and Analysis National Highway Traffic Safety Administration U.S. Department of Transportation 1200 New Jersey Avenue SE. Washington, DC 20590	13. Type of Report and Period Covered NHTSA Technical Report
	14. Sponsoring Agency Code

15. Supplementary Notes

Dennis Flemons is the Contracting Officer's Technical Representative of the NISR Contract.

16. Abstract

The Fatality Analysis Reporting System (FARS) is the National Highway Traffic Safety Administration's census of motor vehicle crash fatalities in the United States and its territories. This paper seeks to explain the implementation of the re-coding process, in which previously entered cases were sampled and re-coded by select, experienced FARS analysts to check the accuracy and performance of the original analysts. The goal of the re-coding process is to enhance the quality control process to maintain accurate, complete, and timely FARS data.

To test the feasibility of the re-coding process, 80 cases were selected by a random case generator tool for a pilot study. The pilot study results indicated that when all the re-coders coded the same value, there was an 85.4-percent agreement with the original coded value. The pilot study proved projected timelines, sample sizes, and the re-coding processes to be practical and the first systemwide re-coding process was conducted. Overall, the first systemwide re-code was successful and results indicated that there was 83-percent agreement between the re-coder's values and the original coded values for the Accident, Driver, and Person level variables. At the vehicle level, the re-coder's matched the original coding 98 percent of the time. Both the pilot study and the first systemwide sample had over 80 percent of cross-coder agreement, an acceptable level of uniformity. The re-coding process also identified complex data elements that require analytical skill to discern and code correctly.

17. Key Words	18. Distribution Statement
FARS, Fatality Analysis Reporting System, case re-coding, quality control, quality assurance, data accuracy, police reports, training, analysts	Document is available to the public from the National Technical Information Service www.ntis.gov

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	28	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

Table of Contents

Executive Summary

Recent appropriations have provided NHTSA with funding to improve Fatality Analysis Reporting System (FARS) data quality by enhancing its quality control processes. Changes in data collection at the State level prompted by the Safe, Accountable, Flexible, Efficient Transportation Equity Act: A Legacy for Users (SAFETEA-LU) legislation and corresponding changes in FARS have made improving data quality more challenging. Ensuring that FARS data is of high quality requires robust quality control processes. According to Redman (1996), a sound quality assurance process encompasses error detection and correction, maintaining most relevant data, clear methods of evaluation, and an operational plan for improvement.[1] As part of the larger quality control efforts that include training, documentation of system standards, and data monitoring, NHTSA has developed the FARS case re-coding process.

The FARS case re-coding process enables NHTSA to conduct ongoing sampling of FARS cases to check the accuracy of the original coding, assess analyst performance, and conduct State-specific training to address problems identified in the sampling. When fully implemented, the re-coding process will work with current file year data so that problems can be identified early and corrective actions can be performed quickly in conjunction with a targeted training response to improve data quality. This paper seeks to review the re-coding process during the pilot study and first systemwide sample.

Using a random case generator tool, cases were selected for the re-coding project. Case request lists were distributed to the FARS sites for compilation of case materials and then sorted and checked for completeness; if pertinent information was missing to thoroughly code the case supplemental documentation was sought. Cases were then shipped to the re-coding team to code. After the coding was completed data analysis began.

The pilot re-coding study sampled 10 crash cases from 8 States. Although it was a nonrepresentative sample, the pilot study sought to identify data elements that coders were having difficulty analyzing, examine cross-coder agreement for the systemwide sample, and to test the feasibility of the re-coding process before full implementation. The pilot study identified the more complex data elements that require advanced training, attention to detail, and experience to complete, and require further examination to discern the appropriate code. The pilot study results indicated that when all the re-coders coded the same value, there was 85.4-percent agreement with the original coded value. Coding discrepancies and errors found during the re-coding process were addressed at the annual FARS systemwide training. Within two weeks of requesting cases for the pilot study, all requested cases were received and forwarded to the re-coders.

The first systemwide sample consisted of 475 cases. The systemwide sample yielded some cases where critical information was missing from the case file. To better respond to cases with

[1] Redman, T. (1996). Data Quality for the Information Age. Boston: Artech House.

critical missing information, a library of source documents from all FARS sites that provides samples of missing documents and assists in clarifying data from poor or faded copies of source documents. Overall, the first systemwide was successful and results indicated that there was 83-percent agreement between the re-coders' values and the original coded values on the Accident and Driver level variables.

Possible implications resulting from the implementation of the re-coding process are: the need for additional coding training, categorizing data elements based on complexity, and scheduling the re-coding process to coincide with FARS data releases to facilitate rapid corrections to the current year file release. Both the pilot study and systemwide sample produced benefits for the re-coding process and confirmed its value in the FARS quality control process.

Background

To develop programs and policies that will be effective in reducing the number of motor vehicle fatalities, NHTSA relies on data. As NHTSA's principal source of data on motor vehicle fatal crashes, FARS must provide the highest quality data possible. It is critical that FARS provides NHTSA with high quality data. Data of high quality is integrated, complete, consistent, timely, accurate, and accessible, creating a foundation for decision-making that best uses and directs resources in a manner that saves the most lives.

The proliferation of new technology combined with crash data collection standardization in recent years has significantly influenced change in traffic safety data collection by the States. Evolving technology, such as the laptop or hand-held electronic crash reports and electronic data transfer standards are being adopted across the country in various stages. While States are in transition, data acquisition has proved to be a challenge. The Model Minimum Uniform Crash Criteria (MMUCC) was developed in an attempt to compile the necessary crash data to improve highway safety at both the State and national level. However, MMUCC has undergone two revisions since being adopted and States are at various degrees of implementation of MMUCC in their crash reports and collection procedures. Consequently, crash data report content varies from State to State.

In recent years, Congress has emphasized the importance of data driven performance management and strategies with outcome-oriented goals and measures, such as Section 408 of the SAFETEA-LU legislation. This has guided corresponding changes in the NHTSA data collection programs that rely on the data collected by the States.

Seeking greater operational efficiency and to enhance the usefulness of the data it provides, NHTSA has planned a consolidation of definitions and some operational activities of FARS and the National Automotive Sampling System (NASS) General Estimates System (GES) by 2010. Further synchronization of the data systems is planned for future years. While NASS GES differs from FARS in that it provides a nationally representative sample of all motor vehicle traffic crashes (not just fatal crashes; FARS does), both systems rely on the police crash report, thus making it advantageous to harmonize these two systems where applicable and appropriate. Additionally, NHTSA data systems are in the process of aligning with the changes in MMUCC elements to be able to match State data sources. This effort requires change in the systems that will be critical to monitor in order to maintain quality data.

The principal data source for a FARS case is the police crash report. In addition, analysts collect supplemental documentation such as vehicle registration, driver history, and vital statistics data on each FARS case from a variety of State agencies. In aligning FARS, NASS GES, and MMUCC, the MMUCC convention was adopted for FARS and NASS GES starting in 2009. Each data system has "data elements" under which are "attributes" that can be selected. Each attribute has a numeric "code" associated with it in the entry system or database. The attributes and codes combined are known as "element values." The element values are the selections available under each data element. To complete each FARS case, the analyst applies specific definitions and guidelines and inputs the appropriate element values for each data element into the data entry system. In this way all data contained in the FARS system are uniform,

eliminating State differences in collecting and maintaining relevant crash records. For this effort, all elements on the crash (Accident), Vehicle, Driver, and Person levels were re-coded and analyzed however, for reporting purposes the simpler data elements will be exemplified in this paper.

For FARS data to be most useful, it is critical that the quality of data is maintained. The size of the collection effort and the geographical disbursement of the FARS analysts make this task difficult. To help ensure the quality of the data, multiple activities are performed throughout the year. These activities generally include analyst training, documentation of system standards, and data monitoring.

Training for the field personnel includes a new analyst training program that provides a self-directed preparatory training followed by a five-day classroom session. Training issues identified throughout the year and changes to the system are addressed at FARS-NASS GES systemwide training. Ongoing coding assistance and guidance to FARS analysts is available through a FARS hotline. The data itself is controlled upon entry with the FARS data entry system edit checks. These edit checks are updated annually along with a FARS-NASS GES Coding and Validation Manual that provides definitions, rules, and guidance for each data element. The quality of a FARS case is also monitored for completeness, unknown values, and violations of edit check rules. Once in the database, the FARS data is also monitored through statistical control charts which identify deviations from expected trends in the data and indicates when an inconsistency in the data occurs.

While these activities help to ensure consistency in data acquisition, additional factors such as changes in the collection of the data in the States and corresponding changes in FARS make monitoring data quality more complex. When these changes occur it can limit the effectiveness of monitoring data using trend analysis to identify potential problems. To help address these issues, steps have been taken to develop additional means to support data quality that involves manual reviews of the case work coded by the FARS analysts – the FARS case re-coding process.

The FARS case re-coding process was developed to conduct annual case sampling and re-coding for data quality monitoring, analyst performance assessment, and training. The design combines the concepts of selected case re-coding with State-specific training. This new quality assurance process, when fully active, will sample from the current file year so that corrective actions to improve the quality of the data can be performed throughout the file year when inconsistencies are identified. The aim is to provide more immediate benefit from a case re-coding effort in the form of analyst training and have more tangible effects on data quality.

An ongoing effort of this scope had never been performed for FARS, consequently, many dynamics of this effort needed to be planned, tested, and refined. Therefore, a pilot study was conducted to measure feasibility and identifies logistical challenges. One of the larger issues that had to be resolved was the establishment of a case sample size that provided enough data to find problems, yet was still manageable to re-code in a timely fashion. The gathering of the case materials for the selected cases had to occur in each of the FARS sites and the size of the sample and timing of the request must not be too much of a burden.

Because of the analytical skill needed to properly collect some of the data in a FARS case, comparison of re-coded data to original data can in most cases only identify differences. This necessitates that a level of confidence be established in the accuracy of the re-coded cases among the people selected to perform the re-coding of the sample. A high confidence level is necessary to ascertain a benchmark for comparison that can streamline the analysis of the re-coded case data. The pilot study was necessary to establish this benchmark and resolve other issues before moving forward with the larger scale re-coding process.

Methodology

For the FARS Case re-coding efforts, a statistically representative sample from all the FARS sites was not feasible given time and cost constraints. Additionally, the scope of this effort was to supplement other quality control efforts and focus on identifying problem areas and respective resolutions rather than specific error detection and correction in the database. The primary goal of the pilot was to finalize the re-coding methodology before attempting the first nationwide sample. In terms of establishing the quality of the case work of the re-coding team, the cross-coder agreement sub-sample was not established with a specific error detection rate in mind. Instead, it was devised to identify coding inconsistencies between the team members with attention to the larger concern of limiting the number of total cases to be collected, coded, and analyzed.

Prior to initiating the pilot, input was sought from the FARS coding committee, a group of veteran FARS personnel. The coding committee's experience in the field provided valuable feedback in developing the pilot study re-coding process. It was determined that 10 cases could be requested from each site, and that it should take no longer than one week to gather the requisite material. The committee members believed that a request of 10 cases would only be a minimal disturbance to the day-to-day operations of a FARS site. A 10-case random sample done twice annually (20 cases for each FARS site) also met the initial proposed sample size based on project cost and workload constraints. Additionally, the small number of cases and use of the committee members provided for a steady line of communication between the program facilitator and the FARS sites that submitted case material.

To establish and measure the quality of the case coding for the re-coding team a "cross-coder agreement process" was developed. The pilot sample of 80 cases was distributed equally among four of the re-coding personnel (20 cases each) with the fifth person to serve as a second level of coding quality assurance. A randomly selected subset of 16 cases (20%) was chosen from the pilot sample of 80 and distributed among the re-coding personnel to be coded a second time. Then, that subset as a whole was coded a third time by the fifth re-coder. This established a process where three separate sets of re-coding could be compared among the re-coding personnel. This small sub-sample cross-coding process was devised to identify inconsistencies between the re-coders that needed to be addressed while limiting the number of total cases to be collected, coded, and analyzed. Additionally, the sub-sample of cases could be compared on all four levels of coding (the original data file plus three levels of re-coding) where the ideal result would be agreement across the four levels.

Using the random case generator tool, 80 cases were selected for the pilot study and instructions and case requests were sent to eight pilot States. The entry system that would be used for the compilation of re-coded data was developed and tested. All 80 cases were sorted and checked for completeness and prepared for the re-coders. The FARS data element report was also devised to help re-coders identify problem areas while re-coding cases.

Before re-coders were given their cases they received training and were briefed on their responsibilities. Coded cases were to be completed and returned for review within two months. A sub-sample of cases was selected for cross coding then that sub-sample was re-coded again. The cross checked cases were returned and entered into the electronic system for analysis.

Re-coding Process

The pilot study was launched in November 2007. An 80-case sample composed of 10 cases from each member of the FARS Coding Committee was selected from the 2006 data file. When comparing re-coded case results to the original field coding in the FARS data file it was imperative to establish a high level of confidence in the coding accuracy of the people selected to perform the re-coding. The re-coding personnel for the pilot consisted of a team of five former FARS analysts, each with more than 10 years of experience coding FARS cases. These people were still active in the coding of FARS cases because they provided various States with coding assistance. Their experience level minimized training needed to build a re-coding team.

Once the coded data enters the FARS database, it no longer contains any information that identifies a specific person. However, some of the source documentation used to obtain the data for the case, such as the crash report, does contain some personal information. Consequently, a second significant concern was developing procedures for securing the shipment and handling of case material. The procedures had to meet all published guidelines and ensure secure handling of any confidential information throughout the process.

To ensure that all the data contained in the case material was properly secured, comprehensive shipping instructions were developed for the FARS sites and re-coding personnel to follow. The pilot allowed a test of these procedures before attempting the distribution of cases.

FARS sites were given their individual case lists along with a copy of the shipping instructions. All requested cases had been received with no reported problems. The cases were then sorted, checked for completeness, and prepared for coding. A FARS Data Element Report for use by the re-coders was developed to identify problem areas when re-coding the pilot cases. This document was used to identify missing information in the case file as well as pinpoint information that was difficult to translate without input from the State. This “labeling” of missing or untranslatable information is critical to differentiate missing and unknown information in the original case from that in the re-coded version. Finally, the members of the pilot re-coding team were given their case material (20 cases) and briefed on their duties.

The cross-coder agreement sub-sample of 16 cases was identified and duplicated for redistribution among the re-coders. Copies of all 16 cases of the sub-sample were also prepared to be sent to the individual responsible for the third level of cross-checking. The first level of the

re-code process was scheduled to take no longer than a month and all cases of the first level were returned accordingly. Along with the re-coded cases, each re-coder submitted the data element report and documented any observations or problems that occurred during the process.

While the cross-coding cases were completed, a data entry system and means of comparative analysis were developed. The case entry system mirrors the existing FARS forms and outputs to a Microsoft Access database that matches the FARS Microcomputer Data Entry System (MDE) database fields. In the pilot, all cases were coded on FARS paper forms and forwarded for data entry. In the future, when it is available on machines for the re-coding personnel, the entry system will allow cases to be coded and sent electronically. This will reduce shipment of material and be more timely and cost effective when compared to the pilot process.

All cross-checked cases were returned to input and test the data entry system so that the analysis of the data could begin. Analysis of the pilot data measured the feasibility of the re-coding process, internal quality control between the re-coders, and the identification of missing data within case materials. Upon completion of the pilot, all case material was destroyed in accordance with plans for keeping all data secure.

As stated earlier, an electronic system was needed for data entry and storage for comparative analysis. A Microsoft Access-based entry system was built modeling the current FARS paper coding forms. Matching the electronic data entry forms with the paper forms was done to facilitate the entry process as the re-coding personnel were already familiar with the paper forms.

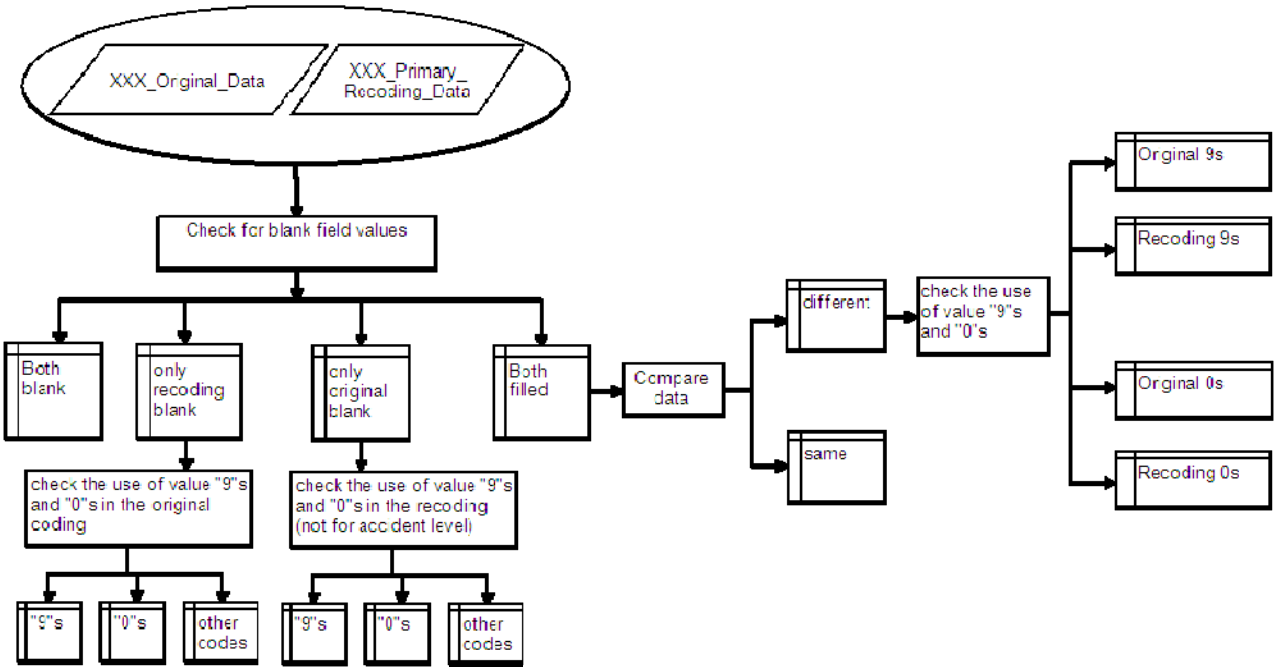
Re-coding Pilot Process – Analysis Results

The pilot analysis compared the values of each data element found in the original case file against the re-coded case values. This included identification of data elements left blank by the re-coding personnel. The comparison was made in absolute values, which meant that the data must match exactly to avoid being identified as an incongruity in our analysis. As a result of the structure of some FARS data elements, discrepancies that are not true differences in coding known as false positives were discovered during analysis. For example, the Related Factors data elements are multi-value elements where there is no set guideline for the order in which to fill the relevant factors. The analyst may choose to enter them sequentially, by relevance to the case, or the order in which they appear in the case material. Therefore, the analysis had to be refined so that the element could be viewed as a whole with no specific order tied to the sequence with which they appear. The pilot study helped to identify these analysis issues related to specific data elements, and provided the insight to develop resolutions before the full re-coding process was initiated.

Overall, the electronic entry and analysis system proved to be effective in identifying differences. The analysis revealed that many of the differences between the original coded case and that of the re-coder stem from the coding of “unknowns” and blanks within the case. As a result, it was determined that special codes would have to be developed for the systemwide re-coding to differentiate between a valid unknown coded value for the original analyst or re-coder, and an unknown where the re-coder did not have access to a corresponding piece of information. This distinction will become even more significant when the re-coding of partial file year cases

begins. To rectify this issue, following the pilot analysis a series of specialized codes were developed to identify various reasons for "unknown" data. The "X" code was created to represent unknown information because the related source documents were not included in the case file. A second code, "T," was selected for identifying instances where the related document was in the case file, but it could not be properly translated by the re-coder. Finally, the "R" code was for instances when the relevant piece of information was included in the cases but had been redacted by the FARS site prior to shipping. This necessitated a number of comparisons of the data be done in the process flow to enable actual comparison of coded values from the re-code and the original. The process flow tested in the pilot is illustrated in Figure 1.

Figure 1



The analysis of the pilot re-coding data was done on each of the FARS coding levels (Accident, Vehicle, Driver, and Person). The data was also examined for completeness (blank rate), differences between the re-coded results and the original file, as well as internal differences among the sub-sampled cases. In short, the computer analysis of the re-coded data only identifies differences. To provide resolution to the differences, one has to filter the data at a more detailed level. Some differences will be valid while others will be errors in the coding of the case. In some instances, only by going back to a source document is the correct coding of the case identifiable.

A portion of the cross-coder agreement is illustrated below in Table 1 as an example. This is a section from a spreadsheet that compares the re-code personnel (identified as A, B, C) across the cross-coder agreement cases that were each coded 3 times. An example of the complexity of analysis required is the FARS data element Trafficway Identifier. It appears that this element is causing difficulty for the re-code staff when you look at the raw numbers. However, this field is

an open field where an electronic comparison of the data would identify differences that are not necessarily coding issues. For example, State Street would be flagged as different from State St.

Table 1

Data Comparison								
	all same	A is different	B is different	C is different	all different	not all filled	all same/compared	all same/16
CITY	1	0	0	0	0	15	100%	7%
COUNTY	0	0	0	1	0	15	0%	0%
ACCDATE	14	0	0	0	0	2	100%	100%
ACCTIME	12	0	0	2	0	2	86%	86%
VFORMS	14	0	0	0	0	2	100%	100%
PFORMS	13	0	0	1	0	2	93%	93%
NMFORMS	14	0	0	0	0	2	100%	100%
NHS	5	0	0	0	0	11	100%	36%
RFUN	3	0	0	1	0	12	75%	21%
RTESIGN	7	0	0	1	0	8	88%	50%
TRAFID	3	1	0	2	7	3	23%	21%
TRAFID2	0	0	0	1	1	14	0%	0%
MILEPNT	0	0	0	0	0	16	N/A	N/A
SPEJUR	12	0	0	0	0	4	100%	86%
FHEVENT	13	0	0	1	0	2	93%	93%
MANNCOL	10	1	2	1	0	2	71%	71%
RELJUNC	13	0	1	0	0	2	93%	93%
RELRDWY	9	2	2	1	0	2	64%	64%
TRAFFLOW	10	2	0	0	0	4	83%	71%
TRAFLANE	11	1	0	1	0	3	85%	79%

As identified earlier, a high confidence level in the re-coded case data is necessary to streamline the analysis of the data. Testing the cross-coder agreement was a key component of the pilot. The four levels of coding in the sub-sampled cases were identified in the analysis as shown below.

- "O": The case coded by the Original FARS Analyst
- "A": The case coded by the Primary Re-coder
- "B": The case coded by the Secondary Re-coder
- "C": The case coded by the Independent Re-coder

In looking at the pilot data the most common difference between the original and re-coded cases was the frequency of blanks. This was mostly due to the lack of certain documents needed to complete the coding. In terms of the cross-coder agreement, the obvious goal is that all three re-coders code the same value. In the pilot data, when all re-coders are in agreement:

- 85.4 percent matched the coding of "O"- the original FARS Analyst;
- 11.8 percent of the values were blanks; and
- 2.8 percent of the values are filled, but not the same as "O."

Analysis of the pilot re-coding data was made more difficult by the lack of the predetermined codes for the re-coding personnel to use to identify that the source documents for a specific data element were missing, or untranslatable in the actual case material versus a valid blank or unknown coded value. This is especially the case when the re-coded information is compared to the original FARS data. The pilot used the 2006 final file with no blanks, but the ultimate goal will be to check a current file year where blank information is expected. Coding differences and

rates were also examined for each re-coder. This was done to identify areas for review and potential correction in the coding practices of the re-coding personnel. In reviewing this information, specific cases where significant discrepancies existed were reviewed and then used in the update training of the re-coding personnel. An example that identifies the value of the cross coding of a sub-set of cases is shown in Table 2 below using the FARS data elements Most Harmful Event (MHE) and Sequence of Events (SOE). The data element MHE identifies the event that produced the most severe injury or property damage pertaining to each vehicle. The SOE records the events that pertain to each vehicle time-wise (up to 6 events) as they occurred in the crash. Table 2 shows the comparison of the values recorded for Vehicle #1 in a two-vehicle crash in the original data and then across the three levels of the cross-coder agreement as re-coded in the pilot re-coding of this case.

Table 1

Recode ID	Veh #	Veh Man	MHE	SEQ Event 1	SEQ Event 2	SEQ Event 3	SEQ Event 4	SEQ Event 5	SEQ Event 6
0072	01	01	12	65	12	64	24	00	00
0072A	01	01	13	65	13	64	24	00	00
0072B	01	01	13	64	65	13	64	24	00
0072C	01	01	13	64	65	13	64	24	00

Table 2 identifies several differences among the original FARS analyst and the re-code team with regards to the coding of the MHE and SOE for Vehicle #1 in this case. In the original coding for this case the value "12 – Motor Vehicle In-transport on Same Roadway" was chosen as the MHE; however, all three members of the re-code team independently chose the value "13 – Motor Vehicle In-transport on Different Roadway" as the MHE. Examination of the crash report narrative clearly states "V1 left the roadway, crossed the unpaved median and entered the westbound traffic lanes where V1 struck V2 head-on." Almost the same example is provided in the FARS manual for use of the value "13 – Motor Vehicle In-transport on Different Roadway." The coding manual identifies this value should be used when "one motor vehicle in-transport travels across the median of a divided highway, entering oncoming traffic and is struck." The correct value for MHE for this case was "13 – Motor Vehicle In-transport on Different Roadway."

Table 2

Recode ID	Veh #	Veh Man	MHE	SEQ Event 1	SEQ Event 2	SEQ Event 3	SEQ Event 4	SEQ Event 5	SEQ Event 6
0072	01	01	12	64→	65	12	64	24	00
0072A	01	01	13	64→	65	13	64	24	00
0072B	01	01	13	64	65	13	64	24	00
0072C	01	01	13	64	65	13	64	24	00

This was not the only discrepancy in this case. The SOE also were not all coded the same across all the levels of coding (O, A, B, C). Table 3 above identifies the corrected coding of the SOE for this case.

In the example above, the coding of this element has one major inconsistency between the original analyst and the re-coding team. Both the "B" and "C" level re-coder selected "64 – Ran Off Road – Left" as the first event in the Sequence of Events. The Original and "A" re-coder selected "65 – Cross/Median Centerline" as the first event. Special guidance in the coding manual addresses this conflict and states:

> "On a divided highway; a vehicle can run off the roadway by leaving the roadway and entering the median. When this occurs, the proper "Run Off Road" code is always code "64 – Ran Off Road – Left." Code "64" will also apply in situations where the vehicle traverses the median and continues across the opposing roadway." (2006 FARS Coding Manual pg. 373)

Table 3 shows that the insertion of the value "64 – Ran Off Road – Left" would correct the SOE coding with the exception of the carry-over MHE error in the original data.

Since the pilot study used such a small sample, it is not possible to determine if the coding of these elements is an issue that would need to be addressed through additional training for all FARS analysts. However, the pilot did provide sufficient information to identify areas for review and clarification for the re-coding personnel. This example also illustrates that examination of the re-coded data in the systemwide sample can yield a measure of performance for specific data elements across the system where corrective action can be addressed in training, or as the re-coding process grows, the granularity can be achieved to address problems by State, or even by field analyst. Only by this level of detailed examination of the data can errors that fall outside the capabilities of the system edit checking be identified and corrected in a timely fashion.

Overall, the pilot study was successful and proved case selection plans to be feasible. It also identified some limitations that would be necessary to resolve before moving forward such as, identification of missing source materials; inconsistent interpretation of source documents, and a need for more detailed analysis of specific elements to clearly identify differences in the original and re-coded case data.

To resolve the limitations identified in the pilot re-coding process, several measures were implemented. To address the identification of missing source materials, special codes were developed to differentiate between a valid unknown and an unknown where the re-coder did not have access to a corresponding piece of information. These specialized codes could identify the reasons for "unknown" data in the analysis. For example, an "X" code was created to represent information that was unknown because the related source documents were not included in the case file. A training session was held to ensure consistent interpretation of source documents and a common understanding of coding guidelines among the re-coding staff. The training brought together the personnel involved in re-coding the cases to address items identified in the cross-coded cases from pilot. Lastly, analysis efforts were refined with initiatives such as the identification of elements that required more analytical skill to complete and deeper examination of problem data elements to identify the specific element values that caused the most concern.

Following the re-coding pilot study, the first nationwide case selection and re-coding was performed. This first nationwide sample was initiated and used cases from the 2007 FARS file. This first sample case selection planned for inclusion of all 50 States, the District of Columbia, and Puerto Rico, making the total cases selected for the sample 520. There were 4 States that did not participate. Also, due to pending litigation, 5 cases could not be included resulting in 475 cases to be re-coded. This is about the same workload as a medium-size State's annual FARS case load. Although the pilot was able to test many of the logistical aspects of the plan, it did not provide a measure of the actual workload of a nationwide sample in terms of materials handling, case distribution, re-coding time, and data analysis. The estimated times for collecting source data, re-coding the cases, and producing the analysis limited the total sample size to approximately 500 cases. For this reason, the same case count from the sample was maintained with the understanding that the sample size from each FARS site could be modified in the future.

After the re-coding was completed, analysis began. Part of the analysis focused on finding potential problem areas in the FARS analyst's case coding by looking at difference rates across the various data elements. The data elements with the most significant differences between the original case file and the re-coded case data were targeted. Additional analyses addressed the specific element values that were the likely problem areas. To address the coding discrepancies, revisions were made to the manual guidance for the 2009 Coding and Validation Manual and training material were developed and presented at the systemwide training for FARS analysts. In the analysis of the data, comparisons were also broken down by difference rate for each collection level (Accident, Vehicle, etc.) for each State. This analysis can identify a particular FARS site that may need individual training assistance or correction.

To enhance the FARS Case Re-coding process, several sample modifications were made to better detect errors in larger States and new training initiatives were added to circulate issues and resolutions to the field more rapidly. Future analysis will include categorizing data elements based upon the method or skill needed to collect that data. Because certain pieces of data are more complex to collect they require more skill on the part of the FARS analyst. As a result, collection of data can benefit greatly from ongoing monitoring and training. Sites that experience significant staff turnover may need additional training support until the novice analyst or analysts increase in skill.

First Systemwide Sample

In preparation for the request for the first set of cases from the 2007 FARS file, an announcement was sent to each FARS site preparing them for their roles in the re-coding process and alerting them that further instruction would be provided. The case requests were sent out with detailed instructions to the field on shipping and case preparation procedures. Five hundred and twenty cases were randomly selected from each site's pool of Level 1 cases. Cases with a Level 1 designation are nearly complete with few blanks or errors.

Overall, the response to the initiation of the systemwide re-coding process was positive and most sites were eager to participate. A few FARS sites did not participate in the first full sample due to various reasons. However, most sites sent their cases within two weeks of the request, and all participating sites had their cases submitted by the end of March when permitted by the State.

The first full nationwide sample of cases provided some new insights. The pilot re-coding process used the coding committee member States. Those involved were briefed in person on the project, and were involved in the discussion of source documents. Consequently, all the items needed to re-code the pilot cases were included in the case material submitted. When expanded to the nationwide sample, a substantial number of cases lacked critical information that had been requested in the detailed instructions sent to each FARS site. This required searching for the missing items. Future re-coding efforts will provide clearer description of the necessary cases material to the FARS units so as to improve the initial submission of material. Also, as a remedy, a library of source documents was created from all of the FARS sites. These documents include items such as blank crash reports, crash report code keys, and police instruction manuals. This will not only provide samples of missing documents but will also help to clarify poor or faded copies of source documents.

For this first full sample, an additional re-coding person was added for a total of six people. This was done to allow a more reasonable case load for each person given the time constraints. All cases were collected, sorted, and distributed to the re-code team. The total case count for participating sites was 475 cases. The total case load handled by these 6 people is equivalent to a medium size FARS site's yearly case load. For example, Minnesota had 463 cases in 2007. The primary re-coding of the 475 cases was completed by September. The target date to produce reportable results was the annual FARS-NASS GES systemwide training. Initially, the re-coding personnel were to enter their cases into the system remotely; however, it was identified that the laptops for the data entry portion where not going to be available for use or as planned. An alternative plan was formulated where the cases were redistributed to be entered by three of the re-coding personnel so that initial analysis of the data could be performed prior to the systemwide training.

After completing case entry, a summary analysis was conducted for all four coding levels (Accident, Vehicle, Driver, and Person) of the 475 cases. The analysis included comparisons, blank rates, and identification of areas with the highest impact or potential problem, such as a review of the data elements for coding alcohol-related data. Additionally, a review of missing information was prepared using the specialized codes established as a result of the pilot to more specifically identify missing data. An evaluation of the data available from all of the FARS sites

will highlight certain FARS data elements where the requisite source data is unavailable. These data elements could be omitted from future re-coding samples. Analysis was completed and the findings from the pilot study and the systemwide re-coding process along with future goals and projections were presented at the annual FARS-NASS GES systemwide training. In addition to presentation of the results at the national training, specific items were identified by secondary analysis for inclusion in follow-up training.

The data analysis results were also used for update training for the re-coding personnel. The update training incorporated the 2009 FARS changes and the cross-coder agreement items from the pilot into a 3-day session. At this meeting the re-coding process was discussed in-depth with members of the re-coding team with specific focus on future implications of the re-coding project. A new timeline for 2009 re-coding was developed with plans for two samples from the 2008 data file. The first will focus on January-through-June cases, and the second will focus on July-through-December cases.

First Systemwide Sample Analysis Results

The comparative analysis of the 475 re-coded 2007 cases was prepared using the electronic comparison tool developed in the pilot with improved processes. The full case sample made additional detailed analysis possible such as the observance of data trends across elements and States. One of the larger improvements was the incorporation of the special re-coding values to identify missing information. This allows differentiation of a valid coded "Unknown" from an unknown as a result of missing data which were not implemented in the pilot study.

Included below is a table that shows a selection of Accident level data elements to illustrate the effective resolution of blank information in the re-coded files that results from the use of the special X, T, and R code values. Note that the column labeled B* represents unresolved blanks at case entry. It is anticipated these blanks will diminish when all re-coders enter their own cases, identifying the uncoded elements during entry.

Table 4 shows the counts for each element in the re-coding file in comparison to the original FARS data. This is followed by a calculation of the rates for each element in the re-coded data. The breakdown section shows the specific special code that comprised the re-coding blanks. This enables identification of what data is missing, why it is missing, and facilitates comparison of valid unknown values to the original data.

Table 4

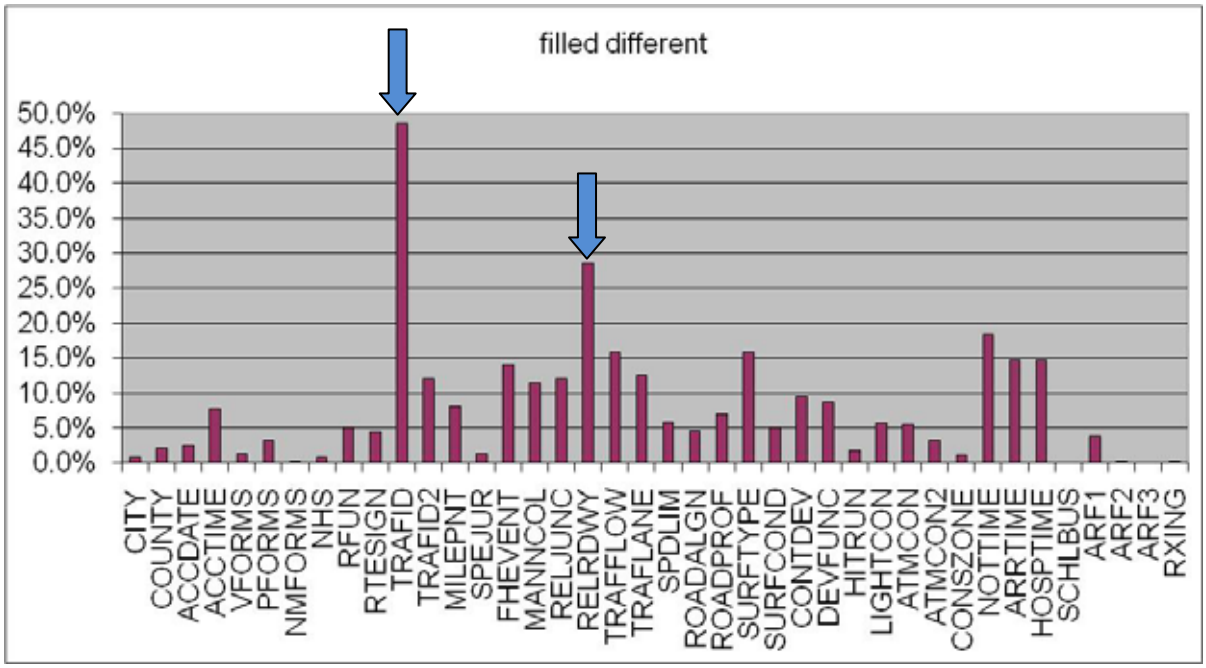
	Counts			Rates			Breakdown the "re-coding blank"			
	filled same	filled different	re-coding blank	filled same	filled different	re-coding blank	X*	T*	R*	B*
CITY	165	4	306	34.7%	0.8%	64.4%	83	223	0	0
COUNTY	44	10	421	9.3%	2.1%	88.6%	83	338	0	0
ACCDATE	464	11	0	97.7%	2.3%	0.0%	0	0	0	0
ACCTIME	439	36	0	92.4%	7.6%	0.0%	0	0	0	0
VFORMS	469	6	0	98.7%	1.3%	0.0%	0	0	0	0
PFORMS	460	15	0	96.8%	3.2%	0.0%	0	0	0	0
NMFORMS	474	1	0	99.8%	0.2%	0.0%	0	0	0	0
NHS	220	3	252	46.3%	0.6%	53.1%	222	29	0	0
RFUN	131	23	321	27.6%	4.8%	67.6%	280	39	0	2
RTESIGN	309	20	146	65.1%	4.2%	30.7%	127	18	0	1
TRAFID	273	194	8	57.5%	40.8%	1.7%	2	5	0	1
MILEPNT	93	38	344	19.6%	8.0%	72.4%	288	41	0	15
SPEJUR	368	6	101	77.5%	1.3%	21.3%	58	0	0	43
FHEVENT	404	67	4	85.1%	14.1%	0.8%	2	1	0	1

- X: Represents unknown information due to the related source documentation missing from the case file
- T: Related documents could not be properly translated by the re-coder
- R: Relevant information in the source documents had been redacted by the FARS site prior to shipping
- B: Represents unresolved blanks at case entry

A second area examined in the full sample analysis was the calculation of difference rates across each data element on all four levels of data collection in FARS (Accident, Vehicle, Driver, and Person). Figure 2 below is a bar chart of the Accident level data elements difference rates for the 475 cases in the re-coding sample. This was calculated as a percentage of filled values between the re-coding data and original data and included the special codes. This analysis gives insight into potential coding issues for each data element across the system as a whole. Once a potential problem element is identified further analysis can be performed to isolate the problem identification. This more specific analysis may identify issues such as; missing data for a particular data element, a high difference rate at a particular FARS site for an element, or problems across multiple FARS sites related to the same specific data element value(s). For example, note the spike for the data element Trafficway Identifier (marked with an arrow in Figure 2). The reason for this difference was identified in the pilot study and is the result of the data element being an open text field. The computer comparison examines each character, thus,

for this element it flags differences that may not be coding errors. For example, "State St." would be identified as different from "State Street" although likely the same roadway. This requires additional analysis of the data in the specific cases where these differences occur. Trafficway Identifier is an important data element to check because an error in identification of the appropriate trafficway can result in a cascading set of errors in other location information elements on the Accident level. Fortunately, this element has a number of edit checks built into the MDE system that help narrow the actual cases that need to be reviewed.

Figure 2



As discussed above, the difference rates identify areas that require additional analysis but not necessarily the specific problem. For example, the FARS data element Relation to Trafficway (RELRDWY) has a significant difference between the re-coded and original coding (28.9%) and is marked in Figure 2 above with an arrow. The data element Relation to Trafficway has 12 separate attribute selections available including the attribute "99 – Unknown." Further review of the data by comparative examination of the values entered revealed that the differences most frequently occurred in the use of values; "02 – Shoulder," "04 – Roadside," "05 – Outside Trafficway/Outside Right-of-Way," and "06 – Off Roadway - Location Unknown." In terms of the components of the trafficway, these attributes represent areas that are adjacent to one another with the exception of "06 – Off Roadway - Location Unknown." This attribute identifies that it is known that the First Harmful Event of the crash did not occur on the roadway (travel lanes), but it cannot be determined which of the other attributes would apply. Shown below in Table 5 is the breakdown comparison of the re-coded data values as compared to the original file values for the data element Relation to Trafficway that highlights these differences.

Table 5

		Original Values							
	Count	01	02	03	04	05	06	07	08
Re-coding Values	01	219	6	1	5	1	1	1	0
	02	3	6	0	4	1	1	0	0
	03	1	3	8	2	0	2	0	0
	04	4	22	0	88	13	20	0	1
	05	1	1	0	4	4	0	0	0
	06	2	2	1	17	10	10	0	0
	07	1	0	0	0	0	0	0	0
	08	0	1	0	2	0	0	0	0
	99	1	1	0	0	0	0	0	0
	TT	0	0	1	0	0	0	0	0
	XX	1	0	0	1	0	0	0	1

Relation to Trafficway Element Values

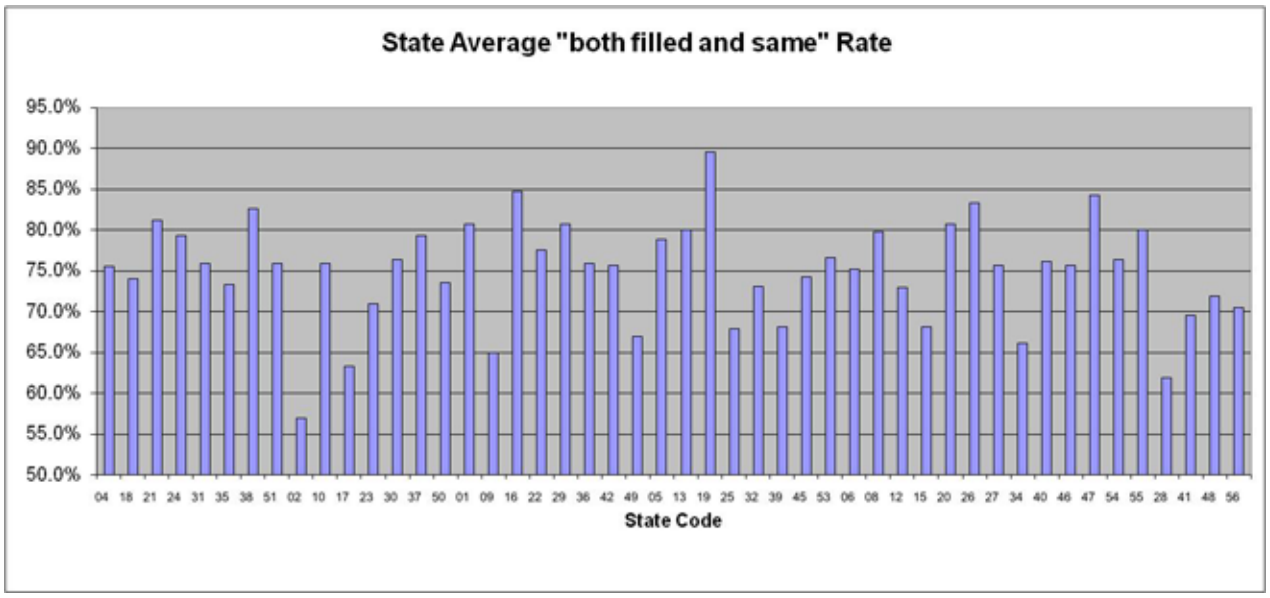
01 On Roadway
02 Shoulder
03 Median
04 Roadside
05 Outside Trafficway/Outside Right-of-Way
06 Off Roadway – Location Unknown
07 In Parking Lane/Zone
08 Gore
10 Separator
11 Two-Way Continuous Left-Turn Lane
99 Unknown

In review of the actual cases and materials, it appeared that both the original analyst and the re-code personnel were having difficulty with these attribute selections. For example, a case where a vehicle entered the shoulder, and struck an object on the roadside while still partially on the shoulder was not coded consistently. To resolve the issue, the FARS coding manual was reviewed to ensure that coding directives are clear for this element. Also, additional guidance and clarification was drafted in this area for inclusion in a follow-up training in early 2009. The additional training guidance drafted for this element was developed specifically for the attributes the drill-down analysis identified as problems, and will use the exact scenarios from the field cases that produced the discrepancies. New guidance on Relation to Trafficway was disseminated in early 2009 to address the identified issues as well as forestall potential future problems with a new 2009 data element, Location of Rollover, which shares the same attributes as Relation to Trafficway.

Another useful way to examine the re-coded data is to look at the agreement or "filled the same" rates by FARS site (State). Figure 3 below shows the difference rates for the Accident level by State. It provides an overall percentage of data that was coded with the same value in the FARS file and the re-coded data. The comparison can also be broken down by the four levels of data collection in FARS or detailed by data element for each State. This analysis can identify a need for targeted training updates for specific sites and can be useful to monitor a site that has experienced significant turnover. Sites that experience significant staff turnover may need

additional training support until the novice analyst or analysts increase in skill. This type of analysis of the 2007 re-coded data was used to help support corrections to the 2007 data file for the State of Alaska prior to the close of the 2007 final file. As identified below in the Accident level, Alaska (02) had a low agreement rate across all coding levels.

Figure 3



State Codes

01 Alabama	15 Hawaii	26 Michigan	37 North Carolina	48 Texas
02 Alaska	16 Idaho	27 Minnesota	38 North Dakota	49 Utah
04 Arizona	17 Illinois	28 Mississippi	39 Ohio	50 Vermont
05 Arkansas	18 Indiana	29 Missouri	40 Oklahoma	51 Virginia
06 California	19 Iowa	30 Montana	41 Oregon	52 Virgin Islands
08 Colorado	20 Kansas	31 Nebraska	42 Pennsylvania	53 Washington
09 Connecticut	21 Kentucky	32 Nevada	43 Puerto Rico	54 West Virginia
10 Delaware	22 Louisiana	33 New Hampshire	44 Rhode Island	55 Wisconsin
11 District of Columbia	23 Maine	34 New Jersey	45 South Carolina	56 Wyoming
12 Florida	24 Maryland	35 New Mexico	46 South Dakota	
13 Georgia	25 Massachusetts	36 New York	47 Tennessee	

Conclusion

The success of the pilot study demonstrated the viability and value of systemwide FARS Case Re-coding and provided a firm foundation for implementation. All participants involved responded to the data requests in a timely manner and follow-up contacts identified that gathering case material did not significantly disrupt their work. For the first systemwide sample, the FARS sites sampled all 50 States, the District of Columbia, and Puerto Rico.

The pilot was also a beneficial tool for assessing the validity of projected timelines, as most target dates were feasible. This information enabled better projections to be made regarding the systemwide re-coding process and confirmed that the re-code project work would not overtax the re-coding staff or prevent them from providing coding assistance to States. A relatively small number of cases (80) generated over a thousand pages of case material. The volume of case material grows rapidly when a case involves multiple vehicles or multiple fatalities. That amount will increase by roughly six times as cases are requested from all 52 FARS sites. This put greater emphasis on organization and record keeping so that the larger case volume of the full sample could be processed.

In the pilot study, the re-coding personnel did not have machines to enter the cases. In the field the actual FARS data is entered into the FARS MDE system, the paper forms are principally for organization and record keeping, not final data. Consequently, the experienced analysts used as re-coding personnel in the pilot all had individual ways of recording information on the paper forms as the entry into MDE is where the information is standardized. For example, one person may use a “9" followed by a slash for an “Unknown” field or another person may use just a slash through the field to represent information that was unknown. For another analyst this may have meant that the information was unavailable or blank on the crash report. In future samples, entry of the data will be done by the person coding the case and standard codes for missing information in the case file will be provided. These two initiatives will help resolve the problems that variation in reporting of missing and unknown data pose to the re-coding analysis.

The FARS Case Re-coding process could also be easily updated as the forms change from year to year. The current system is not without its limitations. In its current form, the re-coding entry system does not have the extensive edit/consistency check engine or case structuring routines that are part of the data entry system used by the FARS field for case entry. Without the automated edit checking the re-coding personnel are coding at a disadvantage in comparison to the FARS field personnel. This complicates the analysis of the data and requires extra work in sanitizing the data in that phase of the project. One possible solution is to enable the re-coding personnel to use a customized FARS MDE system.

To develop an effective re-coding process which identifies data inaccuracies and improves the quality of the data using an on-going re-coding of randomly selected FARS cases requires a staged approach. After pilot testing and executing the first year of development and refinement, the plan is to increase the frequency of the sample and include targeted over-sampling to address specific situations. These over-sample situations could include concerns such as novice analysts or “hot-button” issues such as alcohol data collection. A fully operational case re-coding process will strengthen FARS quality assurance by more rapidly identifying and correcting data errors and be more proactive in training FARS analysts, while the case is still in the field rather than

waiting for the final file. The ability to proactively address issues regarding coding conventions is critical due to the scope of the changes being produced by the FARS-NASS GES systems integration effort and other changes occurring in the States.

In terms of information security, the procedures developed to ensure the safety of the data during transport were very effective. At no point during the entire re-coding process was a case in jeopardy and all cases could be located at any time. Additionally, the pilot better identified what information was readily available and what information was unobtainable.

The re-coding process as discussed in this paper satisfied its goals of assessing analyst performance, conducting specific training to address problems identified in the samples, and data quality and monitoring. The re-coders in both the pilot study and systemwide sample achieved over 80 percent of agreement with the original coders. Typically, 80 percent agreement in re-coding is an acceptable performance.[2] The results from the re-coding were discussed at the 2008 FARS-NASS GES systemwide training and update training for the analysts was given to include the re-coding discrepancies. The analysis of the 2007 re-coded data was used to help support corrections to the 2007 data file for Alaska prior to the close of the 2007 final file. Consequently, future samples will be taken earlier in the collection year with the goal of having actionable analysis of the data prepared prior to annual training in the fall and then again prior to the close of the file year (that occurs in the spring of the following calendar year). Also, more detailed analysis of the data will be performed for each State with the capability to investigate how the coding differs between States and how to minimize differences. This analysis can identify a need for targeted training updates for specific sites, which would be particularly useful for sites that have experienced significant turnover.

Future analysis will include a categorization of data elements based upon the method or skill needed to collect that particular piece of data. Each FARS data element will be identified as a "transcribed/translational" element or an "analytical" element – the former generally being easier to code than the latter. Because certain pieces of data are more complex to collect they require more skill on the part of the FARS analyst. For example, identification of the Most Harmful Event is an element that can require some analytical skill. While relatively simple to identify in many cases, identification of the MHE can become complex when multiple vehicles or multiple fatalities are involved. As a result, collection of this type of data can benefit greatly from ongoing monitoring and training. The element categorization will be used in the next re-coding sample analysis in 2009 from the 2008 FARS data file. Also, more detailed examinations of State-specific issues are planned. This will require both identification of the potential problem and thorough review of the State's crash report and source documents.

Ensuring that the FARS data are of high quality requires a sound quality assurance process. Successful quality assurance will provide confidence in the data's suitability for its intended purpose. Change in the collection of the data in the States and corresponding changes in the FARS make controlling data quality more complex. When significant changes occur, it can limit the effectiveness of monitoring certain data using trend analysis to identify potential problems. As part of the larger quality control efforts that include training, documentation of system

[2] Lombard, M. (2008, October 3). *Practical Resources for Assessing and Reporting Intercoder Reliability in Content Analysis Research Projects*. Retrieved September 24, 2009, from http://astro.temple.edu/~lombard/reliability/

standards, and data monitoring, NHTSA has used some of the funding made available to it to develop an additional means to support data quality for the FARS.

Steps have been taken to develop a process to conduct on-going case sampling and re-coding of FARS cases. In response to the impacts of change on data quality, this added quality control initiative will seek to support data quality monitoring, assess analyst performance, and enhance training. This new quality assurance process, when fully active, plans to work with current file year data so that corrective actions to improve the quality of the end data can be performed throughout the file year. This will provide more immediate benefits through timelier problem identification and analyst training, thus more tangible effects on data quality.

Re-Coding Process Summary Timeline

August 2007: The initial plan in the task response was proposed in which a smaller pilot study was followed by a broader systemwide process. General guidelines and percentages were drawn up as well as target dates for deliverables.

September 2007: Specifics of the pilot study were planned including sample size, sites to be used, and personnel to be involved.

October 22, 2007: The pilot study was presented to the coding committee assembled before the systemwide training and its input was included in the final version of the pilot.

Late October 2007: Documents were prepared to provide step-by-step procedures to properly select and secure shipped cases in accordance with all known requirements for data security and privacy protection. Developed and began testing entry system and comparison database.

November 1, 2007: Using the random case generator tool, 80 cases were selected for the pilot study and instructions and case requests were sent to the pilot States.

November 16, 2007: All 80 cases were received, sorted, checked for completeness, and prepared for coding. A FARS data element report for use by the re-coders was developed to identify problem areas when re-coding cases.

November 19-20, 2007: Update training for coding and re-coding personnel was held in Reno, Nevada. The members of the re-coding team were given their cases and briefed on their duties.

December 21, 2007: All 80 cases of the pilot were re-coded.

January 8, 2008: A sub-sample of 16 cases (20%) were removed from the original 80, sorted and duplicated for use in a cross-check to establish inter-coder reliability and distributed to the re-coders. The sub-sample was also sent to the fifth recoder to serves as a blind re-code, providing the third round of cross-checking,

January 14, 2008: Successes of the pilot study were discussed and revisions for the full re-coding process were planned. Dates of the first sample of the 2007 files were moved back two weeks so as not to conflict with the early assessment of the 2007 file.

February 5, 2008: Initial pilot study cases were received for data entry.

February 8, 2008: All cross-checked cases were returned.

February 15, 2008: Received go-ahead to begin systemwide collection of 2007 cases. Data analysis began comparing all four levels of pilot re-code cases (original, primary re-code, cross-check, and blind re-code).

March 3, 2008: NHTSA announced the systemwide process to the States and 2007 case list requests were generated for all FARS sites.

March 7, 2008: Case request lists were distributed to the FARS sites. Six States were delayed until after the annual report file preparations so those States could concentrate on processing cases needed to meet the file preparation deadlines.

April 4, 2008: All requested cases were delivered with the exception of New Hampshire (privacy laws), the District of Columbia (unable to get permission), Rhode Island (staff turnover), and Puerto Rico (language barrier).

April - May 2008: Cases were sorted and reviewed for completeness. Additional case material was sought out where applicable. Cases were prepped for distribution

June 9, 2008: Cases were sent out via FedEx to the re-code team and re-coding began.

June 2008: Results of the analysis run on the pilot study were studied and improvements were made based on the findings.

July–August 2008: 475 cases (equivalent to a medium-size FARS site's yearly case load, e.g., Minnesota had 463 cases in 2007) were re-coded by the re-code team.

September 2008: Primary re-code was completed and data entry began.

October 10, 2008: Data entry completed and analysis began.

October 17, 2008: Initial data analysis completed.

October 27-31, 2008: Initial findings of both the pilot and systemwide re-coding process along with future goals and projections were presented at the annual FARS-NASS GES systemwide training. Items identified for inclusion in update training for the field.

December 8-12, 2008: Update training for the coding and re-coding personnel was held in Anaheim, California. The re-coding process was discussed in-depth with members of the re-coding team with specific focus on how to improve it in the next editions.

- State time constraints were considered to mitigate any impact on the States in development of a new timeline for 2009 with the intention of taking two samples from the 2008 data file. The first (a partial) will focus on January through June; and the second (complete) will focus on July through December.
- Laptop computers furnished by NHTSA were distributed for use in re-coding.
- Cross-coder agreement items were reviewed with the re-coding team and concerns were expressed and addressed with the intention of strengthening the re-coding process.

December 22, 2008: A list of 2008 Level 1 cases was run through the random case generator in preparation for next sample to begin.

January 5, 2009: The next edition of the re-coding process began as 10 cases were requested from each FARS site. The FARS Regional Office Managers (ROMs) for each of the 4 States where cases were not obtained in 2008 were contacted and work began to resolve that issue.

DOT HS 811 318
May 2010

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