Uniform Application of NASS CDS Child Safety Data Definitions and Data Extraction

Ana Maria Eigen
JoAnn L. Murianka
National Highway Traffic Safety Administration
United States of America

Paper 09-0550

ABSTRACT

The authors hope to provide an intermediate method of data extraction, taking advantage of the improvements in child passenger data collection and recording. The authors also wish to highlight the importance of appropriate usage of the data. As suggested in the Eigen 2007, the enhanced data set SAS files, also known as the 30-file data set, will be contrasted with the 11-file data set format, the traditionally available NASS CDS SAS files, and analysts will be referred to the NHTSA web site for supplementary information. Further, frequently asked questions will be addressed to provide uniform information dissemination to all users. The primary data source will be the National Automotive Sampling System (NASS) Crashworthiness Data System (CDS). As conclusion, the authors propose a three-step extraction methodology to be used until the enhanced data files can be released. This includes traditional data extraction to retain weighting factors, extraction of the enhanced variables, attributes, and associated graphics, and manually integrating the two data sources.

INTRODUCTION

Available Data Evolution

NASS started recording crash data in 1979. From that moment, paper forms recorded crash environment, vehicle damage, occupant demography, and injury. As technology advanced, NASS CDS kept pace with the electronic data reporting. Beginning with 1997 qualified crashes, the data forms were uploaded from pen computers used by the researchers and after careful review by Zone Center staff, the electronic data became part of the web site query data. With the enhancements offered in the enhanced 30-file SAS data set, the 11-file SAS format was not able to be adapted to the greater formatting needs and new variables and attributes. The two SAS data sets will be described in the Definitions section. For this reason, data analysts have been obligated to adapt existing data for suitable analysis. The delayed release of the data has been attributed to the stringent quality control practices that have characterized the data system.

Since 2000, with variable and attribute development for the 2002 child safety data enhancements, variables and attributes were designed taking into account state-of-the-art child restraint technologies and state-of-the-practice educational and installation guidelines. This expertise was inherent in the continual contact with child restraint manufacturers, child safety education networks, and hands-on-familiarity with technologies pursuant to a long trajectory of experience in the child occupant safety field. This high degree of scrutiny was present in the interactions, especially with the Zone Centers, owing to the complexity inherent in accurately coding a highly detailed data set. Discussion of the annotation fields and their use will further elucidate this point. It is also noted that this degree of oversight was present through crash year 2007 with data released in 2008.

It is envisioned that the enhancements and practices put into place will guide subsequent years of data collection and codification. These enhancements are contemplated in the enhanced 30-file data set. Stringent quality control practices have allowed for the release of two years of enhanced data. In the absence of the enhanced data, this work discusses strategies to bridge its absence by using the traditional 11-file SAS format supplemented by NHTSA case viewer query.

Research Questions

1. How can one use the enhanced 30-file NASS CDS data set for meaningful data extraction?
2. How to bridge the current absence of this enhanced data set?

DEFINITIONS

During the course of this paper, two data repositories for the NASS CDS data will be discussed and noted in the References. These are a NHTSA file transfer protocol (ftp) site and NHTSA web site case viewer. Interwoven in this discussion will be the 11-file data set available since crash year 1988 and the expanded, 30-file data set, available since 2002 but to date available only for data years 2002 and 2003. Both the 11-file and 30-file data sets are compiled in SAS format. The 11-file set is the original means of translating and condensing the vast amount of data stored in an Oracle system. With improvements in data collection and increased storage needs, the 30-file set is the Oracle-look-alike file compiled in SAS and allowing the end user the advantage of most

researcher-collected information. The 30-file data set will be referenced throughout the paper as the ***enhanced data set***. The data provided since 1988 in the format known by all CDS data users will be referenced as the ***11-file data set***.

The National Center for Statistics and Analysis (NCSA) has been the purveyor of crash data since its inception to support rulemaking and research within NHTSA and provide researchers with internationally recognized tow away crash data. NCSA continually seeks to improve their data collection practices by keeping pace with technology and addressing user data needs. Since 1988, NASS CDS has been made available to researchers. Originally the data was made available by compact disk disseminated by The Volpe Center to requestors. Most recently, NCSA has worked to make its data universally accessible by way of the World Wide Web ftp site and two state-of-the-art case viewers.

In 1995, the data files were supplemented by an accident summary, accident type, vehicle type, and person description text files, forming the current ***11-file data set***. For convenience, data compiled from 1988 through 1994 and available via ftp site also will be referenced as ***11-file***, yet, this was a 7-file format without the text files, later added in 1995. Although many modifications have occurred in terms of data collection, the ***11-file data set*** is constrained by an inflexible database structure, which was cutting edge at its moment of inception. Advancements in vehicle and child safety restraint technologies, and rollover data collection, have gone unnoticed owing to the inability to adapt the older data set structure to the exigencies of more robust attributes for existing variables and newer variables.

With this in mind, provision of the ***enhanced data set*** structure was initiated for crash year 2002. Although the data sets are still published in SAS format, this new structure allows for an Oracle look-a-like data set, which more closely resembles the complete data set collected by the crash investigators. To date, crash years 2002 and 2003 have been published. Delays in publication of the subsequent data years have been attributed to funding issues. It is important to note that NASS CDS data undergoes a thorough review process before publication. At the writing of this paper, crash data through calendar year 2007 has been available under the older, ***11-file data set*** structure. This paper is intended to aid researchers in working around the issues inherent to the delays, as no firm commitment exists for the completion of quality control and subsequent release of the ***enhanced data set*** for data year 2004 and beyond.

In 1997, NASS CDS transitioned from paper form data collection to electronic data collection. This information could be uploaded easily to a central quality control repository. The on-line case viewer became a reality which not only allowed data found in the SAS format to be viewed but also made available the photographs, scene diagrams, and mannequin injury sketches. This provided the final phase in complete reporting so necessary in clinical analysis. In 2004, an enhancement was made to the case viewer. This allowed for cases, with all associated graphics, to be downloaded in an Extended Markup Language (XML) file facilitating aggregation of data for analysis of a body of crashes. This format was also made available for crash year 2005. In tandem, though, the traditional case viewer was released for those years. In an effort to provide the greatest amount of data in the most accurate fashion, a number of 2008 cases have been reviewed and placed on a preliminary case viewer.

These database formats and additional data will be referenced throughout the paper. Relating these sources and data types is envisioned to allow researchers to undertake analysis using a variety of sources to reap the maximum benefit from the exceptional data collection and coding efforts of the NASS CDS field researchers.

BACKGROUND

Data Collection History

Before 2002, the codification efforts of NASS CDS found their strength in crashworthiness of the vehicle. The scene geometry, inclusive of road side furniture and forensic review of the roadway segment, were well-reported. Disaggregation of crash data to the vehicle level has always been of paramount importance to the accurate reporting. Equally important but lacking real-time information was the human element. Basic demography was generally well collected and recorded and when appropriate complete medical information was included. This data, however, was collected at the crash level, rather than attributed to a given event. Further NASS CDS, has kept pace with vehicle modification, with respect to model year changes and hardware. Most child safety seats are not factory-installed, and might be best described as retro-fit items. For this reason, their continued study and success must be accurately tracked and reported.

As understanding of the technologies increased and the number of child safety restraints increased, NASS CDS continued their commitment to data collection and recording excellence. During 2001, an in-depth

inventory of available technologies was undertaken. The NASS CDS field researchers were polled with respect to their understanding of these technologies. It was found that a vast gap in knowledge existed with the state-of-the-art practice in child safety restraint technologies.

Since then, an intensive modification of the reporting has been designed, supported by a complete overhaul of the data collection instruments. To support the enhanced data collection, intensive training has been provided to field researchers, resulting in one researcher being dedicated to coding child seat cases from each primary sampling unit.

DISCUSSION

Changes in Child Restraint Variables and Attributes

Prior to 2002, there were only a limited number of variables regarding child restraint type, use, and installation recorded in CDS, several of which were outdated and no longer reflective of current child restraint types and installation usage techniques.

The predominant child restraint data collection tool used by field researchers was a one page interview consisting of eight questions which was part of the main NASS interview. Normally, the interview was conducted over the telephone and in the majority of cases the child restraint was no longer available, which resulted in much of the child restraint data coded as "Unknown". Nonetheless, there were rare occurrences when the child restraint remained in the vehicle and/or with the crash occupants which allowed for identification and coding of the attributes available at that time. In those instances when the child restraint was available and permission was granted for doing so, photographs of the child restraint were taken and combined with the interview information.

From 1988 through 2001, the majority of child restraint types were coded as "Unknown/Other", due in part to the lack of information in the field, for example, the child restraint was no longer available coupled with the fact that the crash occupants were unfamiliar with the seat and were unable to recall identifying details.

Seat Belt Installation Considerations

The Proper Use/Misuse information regarding a child restraint was captured in the seat belt variables. There has never been a single variable or attribute which provides the overall proper/improper use of a child restraint, yet this attribute was at times relied upon to do just that.

The Proper Use/Misuse variables were defined as: Proper Use of Manual Belt (used properly with child safety seat – indicated when the manual belt was installed so as to comply with the manufacturer's directions); and Proper Use of Automatic Belt (used properly with child safety seats – indicated when the automatic belt was installed so as to comply with the manufacturer's directions). Proper/improper child seat installation is difficult to determine even when the child seat is available and remains installed in the vehicle as it was prior to the crash, because the pre-crash environment itself had changed. In addition proper/improper information was not coded regarding the child seat's type and use for the respective child occupant, which again lead to misinterpretation, so the attribute was removed beginning with the 2003 CDS file.

Rationale for Enhancement with Retrospective of Nine Years of Crash Data

Prior to 2002, the child restraint list/"pick list" used by field researchers to identify child restraints was populated with approximately 30 different makes of child restraints covering approximately 120 different child seat models belonging to five child seat types. The pick list was outdated, no longer reflective of current child restraint types and designs. The original attributes for Seat Type were Infant, Toddler, Convertible, Booster, Integral, Other and Unknown.

The original child restraint interview asked limited questions about the child restraint itself, one question about how the restraint was used and two questions pertaining to the child restraint's securement to the vehicle.

Some of the child restraint attributes coded harness design but not how the harness was used. Photographic guidelines did not focus on obtaining a picture of the child restraint's make, model and date of manufacture label, so often this information was not obtained during inspection of the child restraint.

Prior to 2002, the field researchers coding child restraint information received very basic child restraint and seat belt education during normal entry level training.

Steps Taken to Reach the Current Levels of Enhancement

In 2002, the child restraint attributes were changed to reflect current child restraint use/marketing terms:

The onus is upon each SAS analyst to determine a system to reap maximum benefit from the enhancements, while performing statistical analysis using the weighting factors. The subsequent section provides the location of currently available data, which in some cases supersedes the data in the ***11-file data set***. This is evident in Table 1, where the older formatting continues to report two booster seat categories available as attributes, one of which no longer exists, as discussed in Murianka 2005. Also, relevant restraint data developed for the 2002 data onward that could not be accommodated in the ***11-file data set*** is also highlighted in Table 1 and discussed in subsequent sections. Focus will be given to the Child Safety Seat, Manual Restraint Use, Component associated with Manual Restraint Use tabs found on the NHTSA web site NASS CDS case viewer.

Table 1.

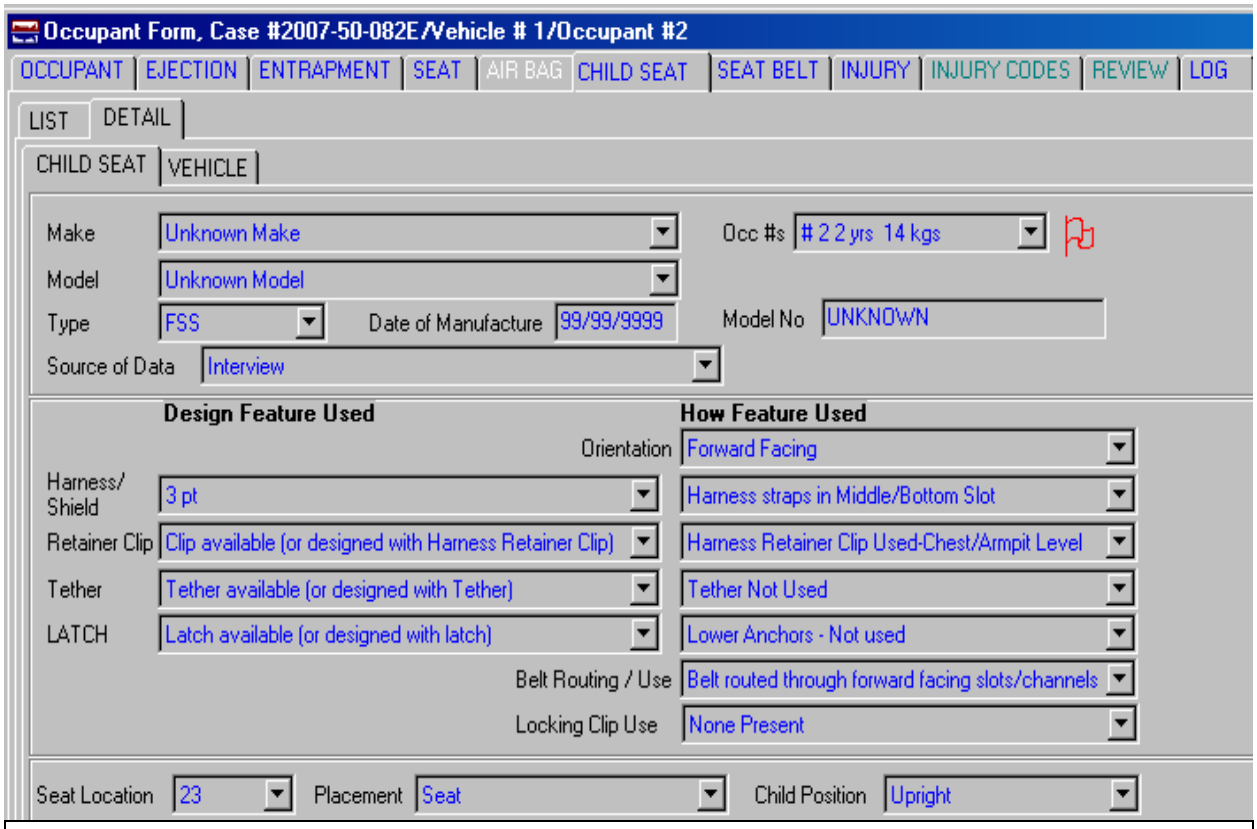
SAS child seat types, by data definition years

Pre-2002, ***11-file data set***		**2002 to Present,** ***Enhanced data set***	
SAS Format Code	**Type Definitions**	**SAS Format Code**	**Type Definitions**
1	Infant	1	Infant
2	Toddler	2	Convertible
3	Convertible	2	Convertible
4	Booster with Shield	3	Forward Facing Only
5	Booster w/o Shield	4	Booster Seat
5	Booster w/o Shield, 2 Toddler	5	Booster Seat/Forward Facing Only
5	Booster w/o Shield, 3 Convertible	6	Booster/ Convertible
7	Other Type	7	Integrated Seat
7	Other Type	8	Harness
7	Other Type	9	Vest
7	Other Type	10	Special Needs
7	Other Type	98	Other
8	Unknown Type	99	Unknown

For illustration, the internal data viewer, known as NASS MAIN, is used to show the extent of the ***enhanced data set***. The public case viewer, found on the NHTSA web site, is generally similar to the NASS Main but contains less information. Some of the additional information presented in the next section will not be made available until the ***enhanced data set*** is released for 2007.

Figure 1 highlights the location of the child restraint information within the context of the NASS CDS viewer, available on the NHTSA web site, and its correspondence with child seat information found in the ***11-file data set***. The Occupant Form, shown in Figure 1 and analogous to information found in the Occupant Assessment File of the ***11-file data set***, outlines demographic, restraint system, and crash outcome information, by occupied seating location. This tab places the occupant within the vehicle and describes the safety devices available to him. Further, crash outcomes such as vehicle entrapment or occupant ejection from the vehicle are described. From this information, it is also possible to ascertain whether the occupant departed the vehicle under his own power or was aided. In the case of occupant ejection, the degree of occupant expulsion is considered. Occupant injury outcome, analogous to information found in the Occupant Injury file of the ***11-file data set***, is discussed with respect to crash survival and the injuries are detailed with respect to their nature and severity. Figure 2 places the restraint technology in a seating position and identifies the factory-installed restraint usage and its interaction with the child safety technology. Figure 3 focuses on the components of the factory-installed restraint technology. Belt retractor and pretensioner technologies are among those reported for the safety belts and related back to the secured child safety seat.

Figures 1 through 3, with their insets, are meant to show the presence of the ***enhanced data set*** variables and their interrelationship with highlighted tabs and families of tabs. Further, this is also a means by which the counts provided in the ***11-file data set*** might be integrated with the complete case viewer data. The case viewer does not provide weighting factors, for this reason the ***11-file data set*** must still be queried to obtain estimates based upon the year, primary sampling unit, case identification number, and occupant number. The concatenation of these values represents the unique identifier subject to the crashes weighting factor. This can be merged with data extracted from the case viewer. For the data years 2004 and 2005, an XML file can be saved for easier transfer of the data elements. Finally, the correspondence of the ***11-file data set***, variables and attributes, and their placement within the case viewer, with the ***enhanced data set*** allows for the analyst to optimize NASS CDS information usage.



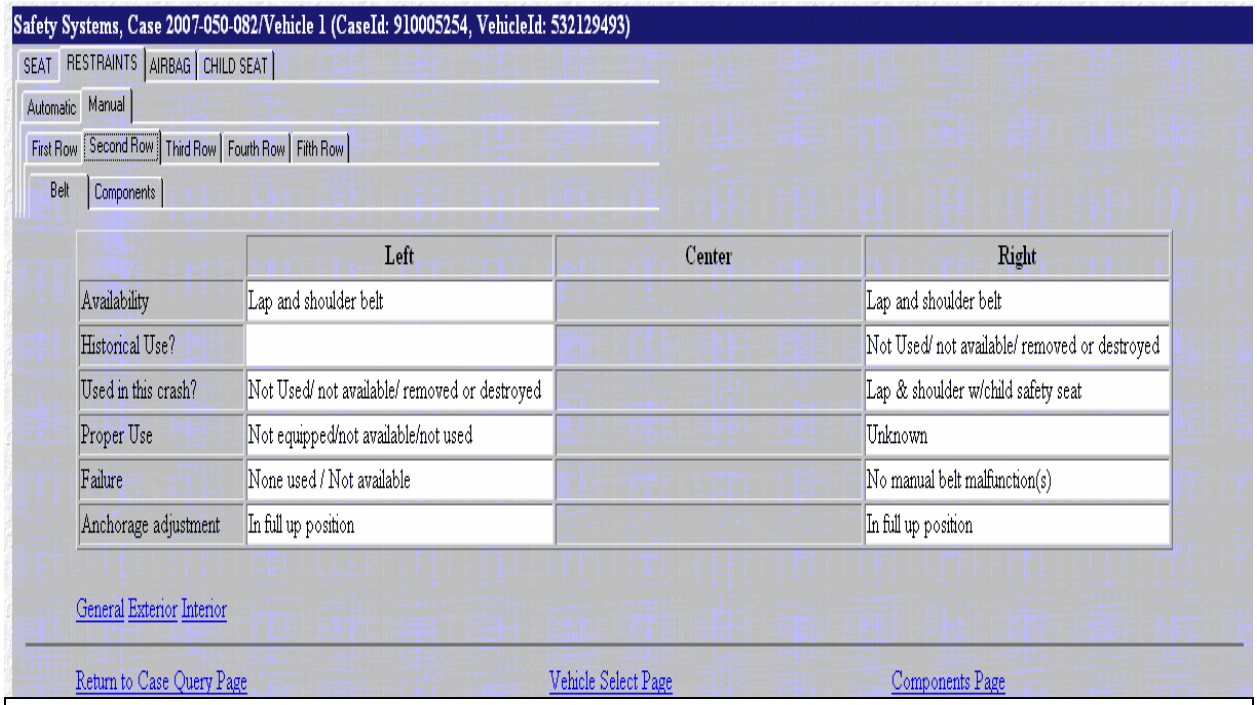
Inset 1: Child Seat Tab Contents

In contrast to the pre-2002 child safety seat types, the current child seat tab allows the data analyst to ascertain the precise type of seat, with current terminology, accurately identified by the NASS CDS field researcher. The child seat tab summarizes the enhanced child restraint variables collected currently.

In this case, a forward facing safety seat (FSS) was reported. Constrained by older formatting, the ***11-file data set*** identified a toddler child seat type but confirms the unknown make. A carefully researched listing of the newest child seat makes and models was compiled through 2007. On this occasion the technology might have reached market before it could be included in the listing or the seat was unavailable at the time of the vehicle inspection, as the source of data for this tab was the interview. Since the NASS Researcher is constrained to specifically identify known elements, the make and model were identified as "Unknown" in this case. When the seat is present, information on date of manufacture and model number might also be collected. The manufacture data is particularly important when attempting to identify a specific child restraint make, model and type in addition to its date of manufacture. With age, prolonged exposure to extreme temperature, plastic deformation or debilitation possible from a previous crash or forces imparted by the belt over the fastening surfaces might be identified as producing decrements in seat performance. An exhaustive inspection of the vehicle and child restraint coupled with the data gathered from each component specific interview will yield the design feature used and the manner in which it was used. This is another point where the expertise of those imparting quality control is crucial, as subtleties exist in the design of given types of child safety restraints.

As a reference back to the injury data and safety systems data, the occupant number, age, and weight are provided as a quick demographic summary when analyzing the child seat tab. Finally, many data analysts wish to establish the occupant kinematics throughout the crash. The child position provides a baseline for those recreating the crash.

Figure 1. Sample Occupant Form, Child Seat Tab



	Left	Center	Right
Availability	Lap and shoulder belt		Lap and shoulder belt
Historical Use?			Not Used/ not available/ removed or destroyed
Used in this crash?	Not Used/ not available/ removed or destroyed		Lap & shoulder w/child safety seat
Proper Use	Not equipped/not available/not used		Unknown
Failure	None used / Not available		No manual belt malfunction(s)
Anchorage adjustment	In full up position		In full up position

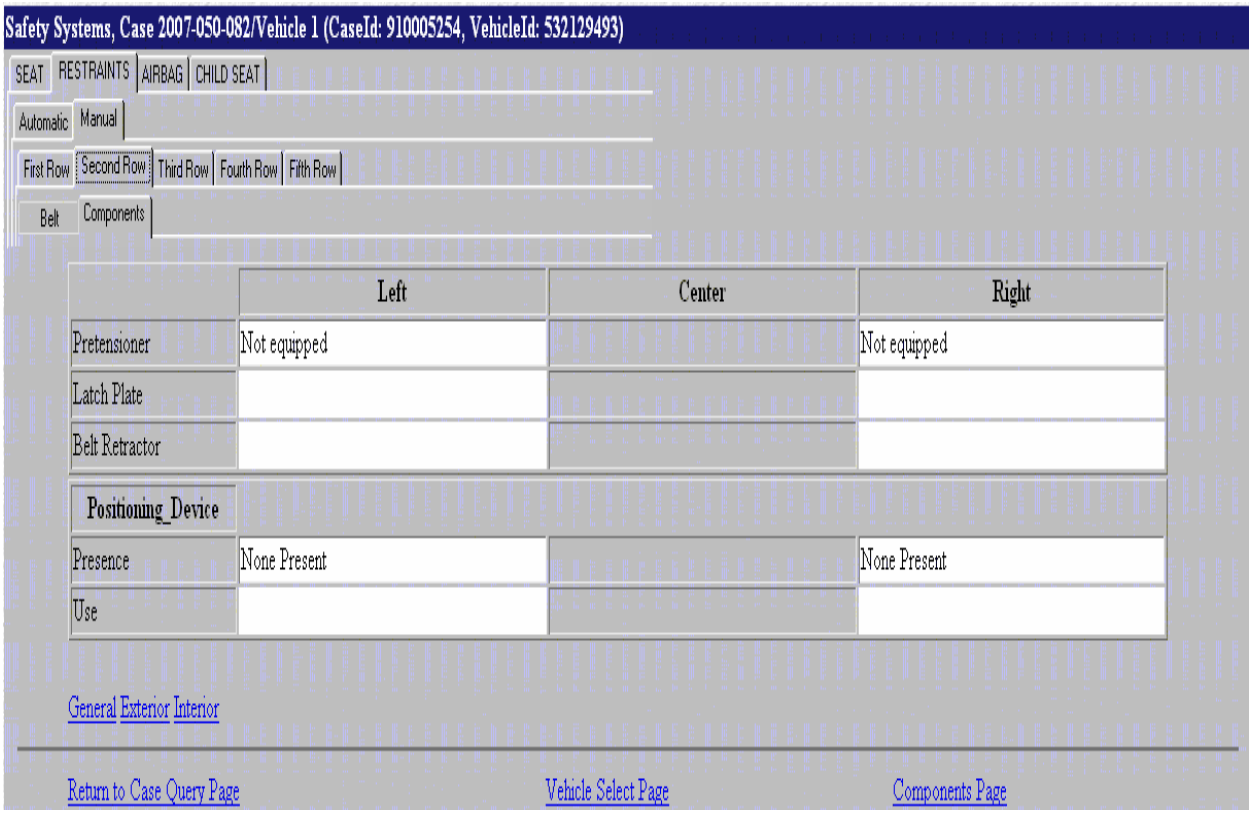
Inset 2: Manual Restraint Use Tab, Interaction with Child Safety Technology

The child restraint does not act independently within the vehicle. Instead, it is dependent upon factory-equipped, vehicle restraint technologies. These technologies are rarely designed in consideration of interaction with a retrofit device, including the child safety seat, and its integration into a safety technology suite. Correct installation is subject to the degree of knowledge that the installer has regarding the vehicle's restraint system and the respective installation techniques required for various types of child restraint systems.

The NASS CDS field researcher must be equally knowledgeable of various vehicle restraint systems so that the method used to install the child restraint is accurately captured and coded within the case. This is achieved by in-depth specialized training in the families of child restraint devices, their installation in a variety of vehicles, and their subsequent identification. With the training and the enhancement of the variables and attributes, the most complete, nationally representative child safety data set may be published.

For the right occupant, the type of manual restraint can be identified, as can its interaction with the child safety seat. This is particularly important for determining that in the right second seat position the vehicle installed technology was fastened, that the child restraint was secured with the lap and shoulder belt, and that the position of the anchorage adjustment was in the full up position. This level of detail is absent in the ***11-file data set*** but populates the ***enhanced data set***.

Figure 2. Manual Restraint Tab for Child Safety Seat Position, with Right Seat Position Example, Full Tab shown.



Inset 3: Components Tab, associated with Manual Restraint, Safety Belt

More new technologies associated with the factory-equipped vehicle safety belts are found in this tab. Presence of pretensioners, types of latch plate, and belt retractors can be noted, if present. Finally, any belt positioning technology, present in the vehicle and its status at the time of the crash might also be noted in this tab.

Please note that in the newest vehicles, especially those that have been inspected, the data provides the analyst with a clear picture of the child safety seat and factory-installed restraint equipment. Again, keeping pace with the newest technologies has given rise to these new variables and attributes but uniformity of database structure applied to the older cases might give the impression of incomplete information. This is not the case; instead, it is an indication of the completeness of the newest cases. This information is absent in the ***11-file data set*** but populates the ***enhanced data set***.

Figure 3. Manual Restraint Components Tab for Child Safety Seat Position

A search of the NASS CDS case viewer may be made selecting from the parameters set forth on the query page, as seen in Figure 4. The sample query might be framed seeking fatally injured children from birth to four years old restrained by a child safety seat. An excerpt of the sample query results are seen in Figure 5, in a format similar to that found in the ***11-file data set***. Upon saving the Hypertext Markup Language (HTML) information in an EXCEL spreadsheet or text document, or ACCESS database file, this information can be imported into SAS. This file will form the basis of the search of the ***11-file data set*** to obtain the requisite weighting factors for the analysis. The strength of this exercise is yielded upon manually reviewing elements such as those found in Figure 6, replicating the ***enhanced data set***, but absent in the ***11-file data set***.

As mentioned previously, Figure 6 is a screen shot of NASS Main, which, in some cases, has more detail than the public viewer, and forms part of the sample query framed above. It is the anticipation that this information will be contained in the ***enhanced data set*** to be released to the public. When consulting the "Source of Data" tab in Figure 6, the information regarding the child seat came from an interview with the occupant's Mother. The interviewee indicated that the Child Seat Type was a FF, forward facing only seat, and that it had a 3-point harness. If the seat was indeed a forward facing only seat then it likely had a 5-point harness or a shield. The case must be coded according to what was identified and/or what was actually seen/found by the field researcher, not what should be. In other words attributes are coded as described in the interview process or when the child seat itself is available, as found by the field researcher. When something unusual is coded, there is normally an annotation made to the case for explanation. The annotation field is available to the NASS researcher to provide comments outside the scope of the coded elements. An annotation was made in this case referencing the source of the child restraint information and the choice of Seat Type. For this reason, there may be instances when some of the attributes coded for design features or how feature used in a case do not make sense to an analyst.

Anomalies Explained

There are occasions when data do not seem to make sense. It should be noted that it is by design owing to the rigorous edit checks in place, through the 2007 crash year. The subsequent example provides some insight regarding one such example. The screen shots found in Figures 1, 2, and 3 were taken from the sample case, with Figure 6 containing excerpts of that case.

The annotation field is an esoteric, many times chaldaic device, whose understanding is confined to the Zone Center governing the usage of this field. It has been reported that the annotation has been used as an elucidating vehicle for unique elements defying the rigidity of the coding structures. Further, the annotation has also been used by Zone Centers to impart instructions or direct corrections to the researchers. Owing to the disparate motivations for including the annotations, the framers of NASS CDS have opted to exclude this information. The exclusion is based upon both consistency and cost-saving. Consistency is important to the system, as the two Zones Centers must work seamlessly to sample and report crashes. The cost-savings enters owing to the sanitization to which this field might be subjected. Unlike the crash summary, which is not subjected to edit check but has general guidelines governing the type of information to be included, this field has no proper rules to govern its preparation or the elements deemed appropriate for inclusion.

There have been instances where organizations have approached NCSA for clarification of seemingly incompatible entries as described above, these could form the basis for a published compilation of explained anomalies. Currently, the explanation is only beneficial for the individual or organization that has received the clarification, nevertheless, the case information remains confusing for others who might lose confidence in the data set owing to questionable query results but who have not known to seek help in interpreting the data. Data analysts have suggested that queries of this sort be published in a frequently asked questions section of the NCSA web site or in supplements akin to an Analytical Users Guide.

It should be noted that edit checks are in place for the wide variety of incompatibilities known to exist. This ongoing identification of new technologies and their interactions with the various restraint and vehicle technologies underscores the strength of NASS CDS data and its value to the child safety community. The preceding section, however, deals with ever-developing situations subject to identification by headquarters and Zone Center staff. The annotation field has served to bridge what might have been told to the researcher, which must be coded, and what might reasonably have been. To reinforce, the NASS CDS researcher may only code based upon their sources, ranging from a full-vehicle inspection with child safety seat resident in vehicle

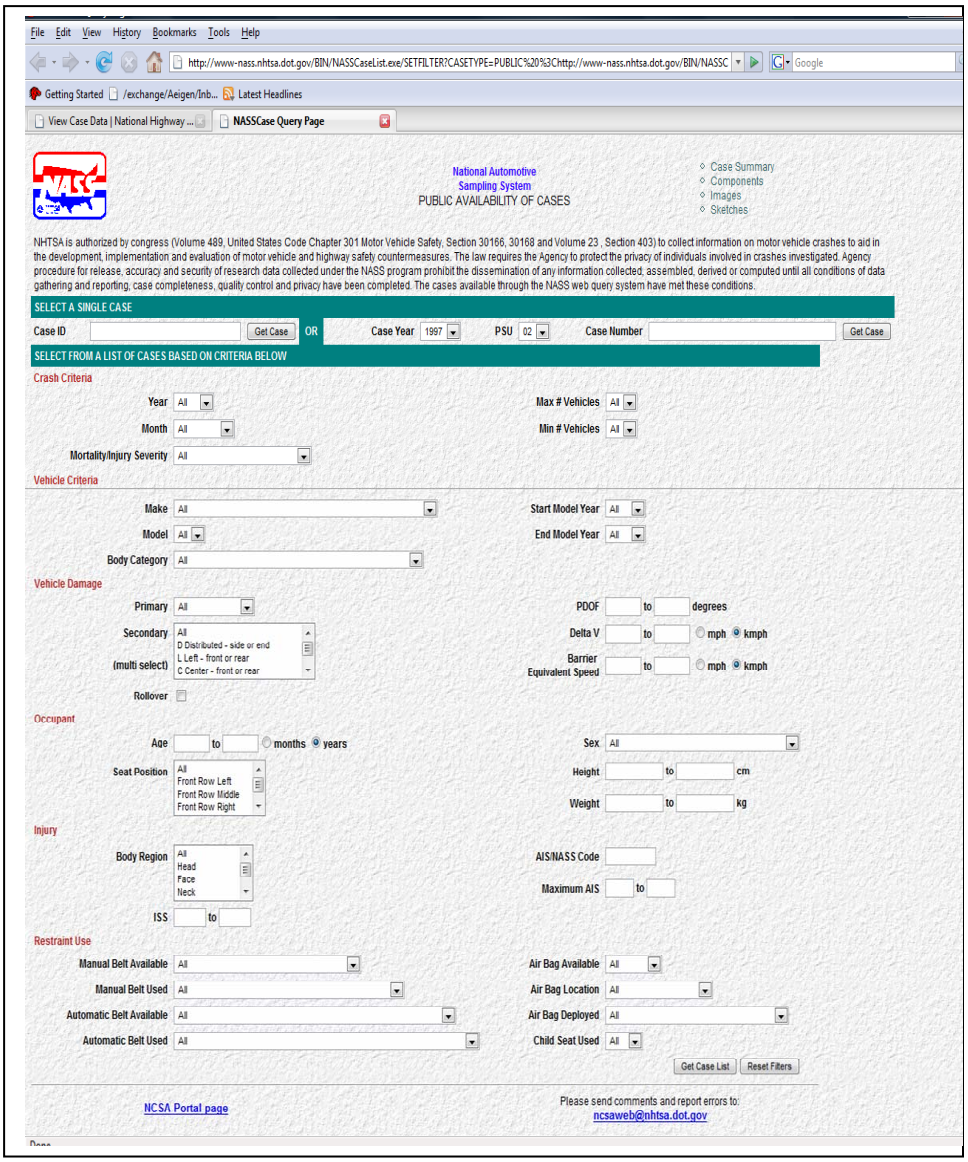
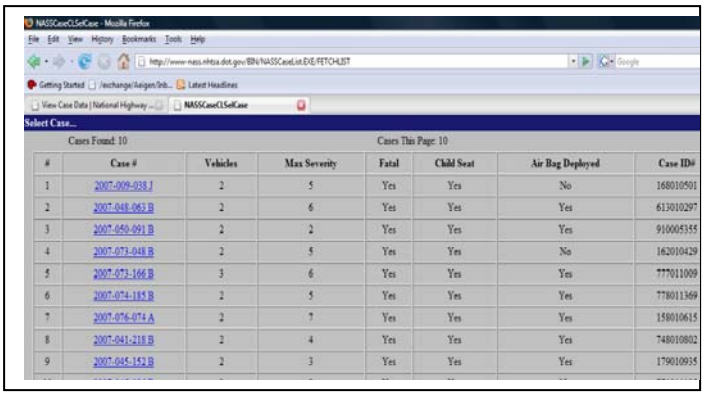


Figure 4: NASS CDS case viewer query screen



#	Case #	Vehicles	Max Severity	Fatal	Child Seat	Air Bag Deployed	Case ID#
1	2007-009-038 J	2	5	Yes	Yes	No	168010501
2	2007-048-063 B	2	6	Yes	Yes	Yes	613010297
3	2007-050-091 B	2	2	Yes	Yes	Yes	910005355
4	2007-073-048 B	2	5	Yes	Yes	No	162010429
5	2007-073-166 B	3	6	Yes	Yes	Yes	777011009
6	2007-074-185 B	2	5	Yes	Yes	Yes	778011369
7	2007-076-074 A	2	7	Yes	Yes	Yes	158010615
8	2007-041-218 B	2	4	Yes	Yes	Yes	748010802
9	2007-045-152 B	2	3	Yes	Yes	Yes	179010935

Figure 5: Partial Result of Query Page Search on fatally injured children 0 to 4 years old, restrained by a child safety seat.

with an interview as background information to merely a telephone interview with someone familiar with the crash, generally a related or unrelated caregiver.

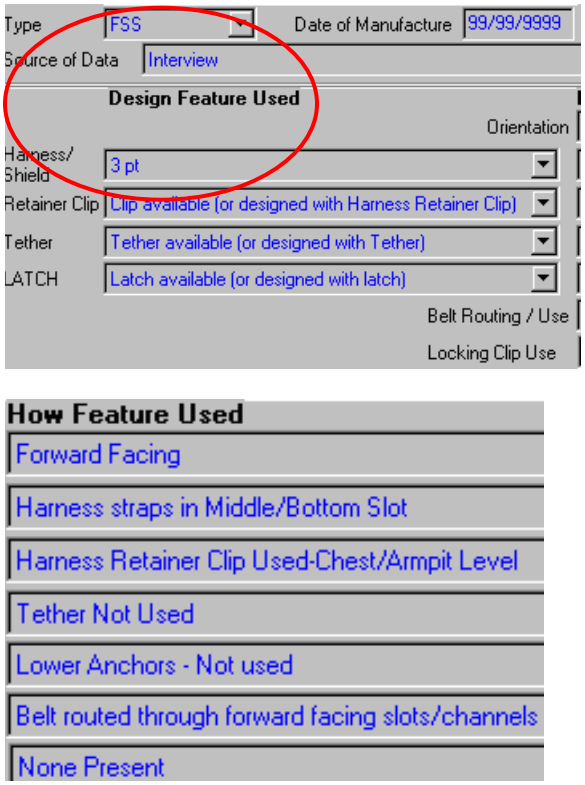


Figure 6: Excerpts from Figure 1.

Data Set Size versus Relevance

Suggestions for meaningful usage of pre-2002 and 2002 data onward

Although some problems exist with the pre-2002 data, this should not deter use on an aggregate level. The aggregation allows for an accurate, nationally representative estimate of child seat usage, in all forms. Finer disaggregation before 2002 might include some form of age or weight distinctions to clarify the limited attributes. Unfortunately, the advent of the enhanced seat data cast doubt on the validity of the previous years. The indication should be that greater detail, based upon informed child safety parameters, was included after 2002. The prior years might be better considered a counting of overall child safety incidence but not of less value and another example of perennial adjustments made to keep apace with technology and data recording advances.

Deciphering the Hidden Mysteries of the Coded Information

Child Seat Tab, Added Attribute - Source of Data

A uniform means of accessing and interpreting the data is provided. By omitting the annotations made by NASS researchers, possible inconsistencies are introduced. Coded elements are actually accurate but may only make sense after consulting published data elements, such as accident summaries, and unpublished data elements, such as annotation fields. A NASS-wide standardization of the annotation field usage must be instituted to guarantee a feasible means of complete information provision. This is the final installment in the series of the ***enhanced data set*** introductory papers prepared by the authors, as an interim approach is available to query the child safety data. This is, nonetheless, an inadequate means of query and only with the release of the ***enhanced data set***, at pace with the ***11-file data set***, can meaningful and efficient analysis be performed. As of the paper submission deadline for the 21st International Technical Conference on the Enhanced Safety of Vehicles Conference (ESV), only two years of data will have been made publicly available though the ***enhanced data set*** collection took effect for crash year 2002. To date, the enhanced variables must be sought out case-by-case through the NASS CDS case viewer.

Reprise of Archaic Formatting after 2002

It was evident that the formatting was an issue of archaic database structures rather than errors in data collection and codification, per Murianka 2005. NASS CDS continues to keep pace with evolving technologies and providing state-of-the art training from subject matter experts and industry representatives. The enhanced data collection is well documented in the enhanced data sets and on the NHTSA web site. The ***11-file data set***, however, is unable to support this new data and has suffered from lagging data provision not only in the area of child safety but also in the areas of rollover and vehicle-installed restraint systems.

To date, the general public has been granted access to the ***enhanced data set*** for crash years 2002 and 2003. The data were released after exhaustive quality control. Data for 2004 through 2007 are pending and the motivation for this document is to provide supplementary means for data extraction in the absence of the expanded files for these years.

CONCLUSION

The child safety data analyst will benefit from the newest data collection by performing an integrated data search. For years 2004 onward, the stringent quality control procedures are ongoing.

In the absence of the ***enhanced data set***, an interim methodology for data extraction was proposed. This included the following steps.

1. ***11-file data set*** SAS query – The traditional case identification ensures the retention of weighting factors necessary in any NASS CDS analysis.
2. NHTSA web site NASS CDS query – With the cases identified in Step 1, an analyst will be able to pull up a case file containing enhanced variables, attributes, and associated graphics. The case viewer is available from 1997 through 2007. An XML case viewer, with cases from 2004 and 2005, is also available facilitating data extraction.
3. Upon identifying relevant case information, this will either be manually entered from the case viewer or automated using XML output.

The three-step extraction process is cumbersome, although, it is a remedial activity until the ***enhanced data set*** can be released. Also, as shown in the sample query output for Figure 5, the cases might be identified via case viewer query, on the NHTSA web site and merged with the ***11-file data*** set to obtain available vehicle and occupant parameters, with case weighting factors.

It should be noted that sophisticated case extraction tools have been written but in their absence the three-step approach is the one used by the authors. It is also reinforced that without the publication of the annotation fields seemingly incompatible results might appear; nevertheless, these are the result of coding what is known by the researcher not what should be.

At the suggestion of the NHTSA Rulemaking, this document will be made available to the National Center for Statistics and Analysis for supplemental publication, at their option, as a transitional data usage manual until the ***enhanced data set*** is released in its entirety.

REFERENCES

Eigen, A., Murianka, J., Four Years of Enhanced Child Restraint Data Collection in NASS CDS, ESV 07-0235, Lyon, 2007.

Murianka, J., Eigen, A., Enhanced Child Restraint Data Collection in NASS CDS, ESV 05-0216, Washington, DC, 2005.

Association for Advancement of Automotive Medicine, Abbreviated Injury Scale (AIS) 1990 - Update 98, Barrington, 1998.

National Automotive Sampling System Crashworthiness Data System 1997 – 2007 case viewer, http://www-nass.nhtsa.dot.gov/BIN/NASSCaseList.exe/SETFILTER?CASETYPE=PUBLIC%20%3Chttp://www-nass.nhtsa.dot.gov/BIN/NASSCaseList.exe/SETFILTER?CASETYPE=PUBLIC%3E

National Automotive Sampling System Crashworthiness Data System 2004 – 2005 case XML viewer, http://www-nass.nhtsa.dot.gov/BIN/NASSCaseList.exe/SETXMLFILTER

National Automotive Sampling System Crashworthiness Data System 1988 – 2007 file transfer protocol web site, ftp://ftp.nhtsa.dot.gov/NASS/,

National Automotive Sampling System Crashworthiness Data System 2002 – 2003 file transfer protocol web site, ftp://ftp.nhtsa.dot.gov/NASS/,

United States Department of Transportation, National Highway Traffic Safety Administration, National Automotive Sampling System, Crashworthiness Data System, 2002-2005, NHTSA, Washington, D.C.

United States Department of Transportation, National Highway Traffic Safety Administration, "National Automotive Sampling System, Crashworthiness Data System, 2000 Coding and Editing Manual", NHTSA, Washington, D.C.