

# Data Documentation for Linking Crash and Health Data in North Carolina

---

**Katherine Peticolas, MPS, PMP<sup>1</sup>, Katherine J. Harmon, PhD<sup>2</sup>, and  
Anna E. Waller, ScD<sup>1</sup>**

**August 2019**

<sup>1</sup>Carolina Center for Health Informatics, Department of Emergency Medicine, University of North Carolina School of Medicine

<sup>2</sup>University of North Carolina Highway Safety Research Center

## **Funding**

This project was funded by the North Carolina Governor's Highway Safety Program.

## **Acknowledgments**

The project was highly collaborative, with planning assistance from Carolina Center for Health Informatics (CCHI) staff, the Strategic Planning for MVC and Health Data Linkage team and stakeholders, and the TRCC. Because the compiled documentation required review, the data owners of each data source spent valuable time finalizing the documentation. We are very grateful to our many partners and data owners for their willingness to assist with this project.

In particular, we would like to thank Roger Akers (UNC Sheps Center), Matt Avery (NC SCHS), Clifton Barnett (CCHI), Eric Bellamy (NC DOT), Alan Dellapenna (NCDHHS IVP), Tony Fernandez (ESO Solutions), Eleanor Howell (NC SCHS), Amy Ising (CCHI), Robert Lee (NC SCHS), Dan Levitt (HSRC), Jane McNichols (UNC Trauma Registry), Alison Miller (NC OCME), Brian Murphy (NC DOT), Eric Rodgman (UNC HSRC), Sharon Schiro (NC Trauma Registry), Libby Thomas (UNC HSRC), and Samantha Wooten (IPRC).

# Contents

|  |    |
|--|----|
| Funding.....   | ii |
| Acknowledgments.....   | ii |
| Table of Tables.....   | iv |
| Table of Figures.....  | iv |
| List of Abbreviations .....  | v  |
| Background.....  | 1  |
| Purpose.....   | 1  |
| Aims.....  | 2  |
| Methods.....   | 2  |
| Aim 1: Describe the process used to document data sources that could be used in motor vehicle crash injury research .....  | 2  |
| Identify and prioritize the data sources.....  | 2  |
| Create a standardized documentation template.....  | 6  |
| Gather available documentation from each data source.....  | 9  |
| Apply the standardized format to each data source. ....  | 11 |
| Submit the final version to the data owner for approval.....   | 14 |
| Post the finalized documentation online, with data owner’s permission.....   | 15 |
| Summary of Aim 1.....  | 15 |
| Aim 2: Describe the results of the documentation and the challenges encountered, solutions found, and lessons learned..... | 15 |
| Results.....   | 15 |
| Challenges.....  | 17 |
| Variability in access to source documentation.....   | 17 |
| Variability in the quality and depth of source documentation .....   | 17 |
| Variability in the timeliness of the documentation .....   | 17 |
| Lack of responsiveness from the data owner .....   | 18 |
| Lessons Learned .....  | 18 |
| Summary of Aim 2.....  | 18 |
| Contact Information .....  | 19 |

## **Table of Tables**

|  |    |
|--|----|
| Table 1: Descriptions of data sources .....  | 3  |
| Table 2. Data elements for the overview template.....  | 7  |
| Table 3. Data elements for the data dictionary template.....                                 | 8  |
| Table 4: Source documentation.....   | 10 |
| Table 5. Field mapping from source data documentation.....                                   | 12 |
| Table 6: Additional data elements for Sheps Center Carolina Cost and Quality Initiative..... | 14 |
| Table 7. Summary of final documentation .....  | 16 |

## **Table of Figures**

|  |    |
|--|----|
| Figure 1. Table relationship diagram ..... | 11 |
|--|----|

## List of Abbreviations

|          |  |
|----------|--|
| BCBS     | Blue Cross Blue Shield of North Carolina               |
| CCHI     | Carolina Center for Health Informatics                 |
| CCQI     | Carolina Cost and Quality Initiative                   |
| DMV      | Division of Motor Vehicles                             |
| DOT      | Department of Transportation                           |
| ED       | Emergency Department                                   |
| EMS      | Emergency Medical Services                             |
| FARS     | Fatality Analysis Reporting System                     |
| FY       | Fiscal year (GHSP fiscal year is October – September)  |
| GHSP     | Governor’s Highway Safety Program                      |
| HSRC     | Highway Safety Research Center                         |
| IPRC     | Injury Prevention Research Center                      |
| MVC      | Motor vehicle crash                                    |
| NC       | North Carolina   |
| NCDHHS   | North Carolina Department of Health and Human Services |
| NCHA     | North Carolina Healthcare Association                  |
| NCTR     | North Carolina Trauma Registry                         |
| NHTSA    | National Highway Traffic Safety Administration         |
| OCME     | Office of the Chief Medical Examiner                   |
| PBCAT    | Pedestrian & Bicycle Information Center                |
| Ped/Bike | Pedestrian and bicyclist                               |
| SCHS     | State Center for Health Statistics                     |
| TRCC     | Traffic Records Coordinating Committee                 |
| UNC      | University of North Carolina                           |

## **Background**

The North Carolina (NC) Traffic Records Coordinating Committee (TRCC) included a goal in the 2013 Traffic Safety Information Systems Plan to evaluate the need for, and feasibility of, a statewide motor vehicle crash injury surveillance system. The ability to integrate information from a variety of sources has the potential to improve safety analysis, inform subsequent policy and program decisions, and support communication with the public and with transportation decision makers.

However, the 2017 NC Traffic Records Assessment by the National Highway Traffic Safety Administration (NHTSA) found the lack of data documentation for key data sources to be a major limiting factor. In addition, discussion with the NHTSA GO Team assigned to work with NC on developing an implementation plan for data linkage included a recognition of this lack of data documentation. The NHTSA GO Team endorsed our efforts to create the necessary data documentation to better inform data linkage efforts.

This project, initiated in 2018 with support and encouragement from the TRCC and the NHTSA GO Team, identified and documented key data sources in NC, providing a valuable resource for those who wish to use these data sources in research and intervention/evaluation projects. The project has worked in close collaboration with the Governor's Highway Safety Program (GHSP) funded Strategic Planning for Motor Vehicle Crash (MVC) and Health Data Linkage project and will be absorbed into that project for the 2019-2020 fiscal year.

## **Purpose**

In this report, we describe the data documentation project and its results. In Aim 1, we describe the process used to document data sources that could be used in motor vehicle crash injury research. In Aim 2, we describe the results of the documentation and the challenges encountered, solutions found, and lessons learned.

## **Aims**

Aim 1: Describe the process used to document data sources that could be used in motor vehicle crash injury research.

Aim 2: Describe the results of the documentation and the challenges encountered, solutions found, and lessons learned.

## **Methods**

The project was initiated in early 2018 with the majority of the work completed by the UNC Carolina Center for Health Informatics (CCHI) with support from Project Team members. When possible, funding to compensate for the time spent on the project was provided to the data owners.

### **Aim 1: Describe the process used to document data sources that could be used in motor vehicle crash injury research**

Documenting resources was a six-step process. Steps 3-6 were iterative for each data source.

1. Identify and prioritize the data sources.
2. Create a standardized documentation template.
3. Gather available documentation from each data source.
4. Apply the standardized format to each data source.
5. Submit the final version to the data owner for approval.
6. Post the finalized online, with data owner's permission.

Each step is described in detail below.

#### ***Identify and prioritize the data sources.***

Data sources were identified based on discussions with project team members and stakeholders. The list was prioritized according to both the perceived availability of the data sources and the anticipated usefulness of the data for research. Although the list was prioritized, the order in which the standardized documentation was finalized was primarily due to both the availability of existing

documentation and the responsiveness of the data owner. See Table 1 for the final list, in order of priority.

**Table 1: Descriptions of data sources**

| Priority | Data Source   | Description  |
|----------|---|--|
| 1        | Emergency Department (ED) visits in NC DETECT   | As mandated by North Carolina Statute § 130A-480 Emergency Department Data Reporting, NC DETECT receives daily extracts through the North Carolina Healthcare Association (NCHA), consisting of select data elements from all ED visits to 24/7 civilian acute care hospital affiliated EDs in NC.   |
| 2        | Crash Records from NC Division of Motor Vehicles (DMV)  | North Carolina crash records meeting at least one of the following criteria: the crash resulted in a fatality, the crash resulted in a non-fatal personal injury, the crash resulted in total property damage amounting to \$1,000.00 or more, the crash resulted in property damage of any amount to a vehicle seized, or the vehicle has been seized and is subject to forfeiture under G. S. 20-28.3.<br><br>In addition, a reportable motor vehicle traffic crash must occur on a trafficway (any land way open to the public as a matter of right or custom for moving persons or property from one place to another) or occur after the motor vehicle runs off the roadway but before events are stabilized. |
| 3        | Emergency Department and Discharge data from the North Carolina Healthcare Association (NCHA) | All ED visits and inpatient admissions/discharges to 24/7 civilian acute care hospitals in NC.   |



| Priority | Data Source   | Description   |
|----------|---|---|
| 4        | North Carolina Trauma Registry                                    | Records of all patients who:<br>1) Have an unplanned readmission related to initial injury within 30 days of last discharge date<br>2) Drowning or asphyxiation<br>3) Have a ICD-10-CM diagnosis code in the following range: S00-S99, T07, T14, T20-T28, T30-T32, T71, T79.A1-T79.A and sustained at least one injury outside of the range S00, S10, S20, S30, S40, S50, S60, S70, S80, S90, and one of the following:<br>a. Injury resulted in death<br>b. Patient was transferred to or from your hospital via another ED or hospital using EMS or air ambulance<br>c. Patient was taken to the OR from the ED<br>d. Patient was admitted as an inpatient or observation patient as a result of traumatic injuries |
| 5        | Death registration data   | Data from death certificates filed with the North Carolina Vital Records office   |
| 6        | EMS data in EMSPIC  | All Patient Care Reports (PCRs) submitted to the State EMS data system  |
| 7        | Hospital Discharge Data at the State Center for Health Statistics | Quarterly extract from the North Carolina Healthcare Association's dataset consisting of all Emergency Department visits and inpatient admissions/discharges to 24/7 civilian acute care hospitals in North Carolina  |
| 8        | Fatality Analysis Reporting System (FARS)                         | Fatal motor vehicle crashes involving a motor vehicle traveling on a trafficway customarily open to the public, and resulting in the death of a motorist or a non-motorist within 30 days of the crash.   |
| 9        | Office of the Chief Medical Examiner (OCME) data                  | Data from multiple documents filed with the Office of the Chief Medical Examiner: The Report of Investigation by the Medical Examiner (RIME), the death certificate, the autopsy report, and the toxicology report.   |

| Priority | Data Source   | Description   |
|----------|---|---|
| 10       | Sheps Center Hospital Discharge data                                    | The Hospital Discharge data include Inpatient, Ambulatory Surgery/Outpatient and Emergency Department discharge records from North Carolina hospitals. These data include patient characteristics, clinical information such as diagnosis and procedure codes and length of stay, payer information, charge information, admission source and patient status.   |
| 11       | Sheps Center Carolina Cost and Quality Initiative (CCQI)                | <p><b>Medicaid data:</b><br/>NC Medicaid data are from the North Carolina Department of Health and Human Services Division of Health Benefits (DHB).</p> <p><b>BCBSNC data:</b><br/>The BCBSNC database contains claims data on health care services reimbursed by BCBSNC. The database includes claim records from insured groups, administrative services only (ASO) groups, and individual market and Affordable Care Act Exchange Plans.</p> <p>Both datasets include claims information (including doctor visits, inpatient and outpatient care, prescription medications, dental treatments, charge amounts, dates of service, and diagnoses and procedures), provider information (including specialty and location) and member information (including date of birth, gender, county/ZIP code of residence).</p> |
| 12       | Highway Safety Research Center Ped/Bike (Pedestrian and Bicyclist) data | Two datasets based on North Carolina Department of Motor Vehicles (DMV) crash records representing all pedestrian-motor vehicle crashes or all bicycle-motor vehicle crashes reported to the NC Division of Motor Vehicles by investigating officers for the relevant crash years. The UNC Highway Safety Research Center (HSRC) staff reviewed each bicycle and pedestrian crash report form, including diagrams and narratives, and coded every crash verified as involving at least one bicyclist or pedestrian to a specific crash type. Other crash factors, such as age, sex, alcohol indicators of the bicyclist/pedestrian and driver, and roadway and environmental factors, were obtained from NCDOT's Crash Database, and added to the data.   |

| Priority | Data Source  | Description   |
|----------|--|---|
| 13       | Highway Safety Information System (HSIS): North Carolina | <p>The following types of files are available for North Carolina:</p> <p>Crash files contain basic crash, vehicle, and occupant information on a case-by-case basis. Typically, this information includes type of collision, types of vehicle, sex and age of occupants, fixed object struck, crash severity, and weather conditions.</p> <p>Roadway Inventory files contain information about roadway cross sections, types of roadway and other roadway characteristics. Data include number of lanes, lane and median width, shoulder width and type, rural or urban designation, and functional classifications.</p> <p>Traffic Volume files list Annual Average Daily Traffic (AADT) data. Additional information on hourly volumes and truck traffic percentages is also available in selected States and/or locations.</p> |

***Create a standardized documentation template.***

The goal of the standardized template was to have a user-friendly design that contained the data elements which would be the most important to researchers. With this in mind, we surveyed different methods of data documentation before designing our template. There was no standard template found for all data sources, so elements of different formats were chosen. Some of the formats reviewed include:

1. New Jersey Office of Information Technology Enterprise Data Services templates
2. California Fatal, Nonfatal Hospital, and Nonfatal ED data dictionary
3. Utah Motor Vehicle Crash Report Data Dictionary
4. Western Australia Emergency Department Data Collection Data Dictionary Version 1.0
5. Wisconsin Codes Custom Reporting System Data Dictionary

We based our documentation template primarily on the New Jersey template.

**Final Documentation Template**

Documentation of each data source required the creation of two types of documents: an overview document and a data dictionary.

The overview document provides general information about the data, including a description of the data, how it can be accessed, and any known data quality issues. See Table 2 for the complete list of data elements captured in the overview template.

**Table 2. Data elements for the overview template**

| <b>Data Element</b>   | <b>Description</b>   |
|---|--|
| <b>Overview</b>   | General description of data source   |
| <b>Data owner</b>   | Owner of the data  |
| <b>Data description and collection criteria</b>                           | Brief description of the data and the requirements for inclusion in the dataset                                      |
| <b>Type of data: source or compiled/abstracted</b>                        | Whether the data were entered by the original data entry source or were abstracted or compiled from multiple sources |
| <b>Are the data available to outside parties for analytical purposes?</b> | Indicates the availability of the data for research  |
| <b>Process to obtain the data for research</b>                            | The process for researchers to request access to the data  |
| <b>Website</b>  | Web address for the data source  |
| <b>Contacts</b>   | Key contact(s) in accessing the data   |
| <b>Data flowchart</b>   | A flowchart indicating the source and flow of the data   |
| <b>Who enters the original data</b>                                       | The profession(s) of the persons who enter the source data   |
| <b>Injury classification method</b>                                       | The injury classification method found in the data (e.g. ICD-10-CM, KABCO, etc.)                                     |
| <b>Collection timeframe</b>   | Timeframe for when the data were entered after event   |
| <b>Years available</b>  | Description of timespan for which data are available   |
| <b>Data history</b>   | Changes in data that might affect research   |
| <b>Is a data dictionary available?</b>                                    | Whether the data owner provided a publically available data dictionary   |
| <b>Dictionary</b>   | Name of the data dictionary used to populate the template  |
| <b>Source documentation field map</b>                                     | A mapping of how the labels found in the source documentation were translated into the template                      |
| <b>Additional fields available in source documentation</b>                | Any data elements in the source documentation which were not retained in the template                                |

|                                  |  |
|----------------------------------|--|
| <b>Known data quality issues</b> | Known issues with the data (e.g., race data may be incomplete or unreliable) |
|----------------------------------|--|

The data dictionary lists all the data elements *potentially* available; their availability for each data source was dependent on what was present in the source documentation. See Table 3 for the complete list of potential data elements captured in the data dictionary. The data dictionary can be exported from the Access database in different formats, including as a PDF or as an Excel document.

**Table 3. Data elements for the data dictionary template**

| <b>Data element</b>      | <b>Description</b>  | <b>Example value</b>  |
|--------------------------|---|---|
| <b>Field</b>             | Descriptive name of the field   | Non-Motorist Contributing Circumstances   |
| <b>Field - literal</b>   | The literal element name either in the source database or in SAS  | TYP_CD 8  |
| <b>Table or category</b> | The table or category of the field, depending on the source documentation   | Person: Non-Motorist  |
| <b>Description</b>       | A description of the field  | The actions of the non-motorist that may have contributed to the crash.   |
| <b>Format</b>            | The format of the field (alphanumeric, datetime, numeric, etc.)   | Numeric   |
| <b>Length</b>            | The allowable length of the field in digits   | 3   |
| <b>Required (Y/N)</b>    | Whether the field is required or not  | N   |
| <b>Unique key (Y/N)</b>  | Whether the field value is unique and can be used as a key field  | N   |
| <b>Source comments</b>   | Comments or background from the original source beyond the definition of the field. Source documentation which provided more extensive background for each field may have multiple data elements concatenated for this field. | Source: Refer to blocks 8 - 9 on DMV-349 Form.<br><br>Rationale: Important for evaluating the effect that dangerous risky non-motorist behavior has on motor vehicle crashes. |

| <b>Data element</b>            | <b>Description</b>   | <b>Example value</b> |
|--------------------------------|--|----------------------|
| <b>Retired (Y/N)</b>           | Whether the field is retired from use  | Y                    |
| <b>Retired Date</b>            | The date the field was retired   | January 2015         |
| <b>Option list: Code</b>       | Non-descriptive code for the option available for that field   | 008                  |
| <b>Option list: Code value</b> | Descriptive value of the option  | Wrong Side Of Road   |
| <b>Option list: Comments</b>   | Comments from source regarding that code/value pair  | Max of 2             |
| <b>Available for research</b>  | Whether the field can ever be returned to the researcher through the data request; used only for the North Carolina Trauma Registry; | Y                    |
| <b>Institutional</b>           | Presence in provider file; used only for Sheps Center CCQI data  | Y                    |
| <b>Professional</b>            | Presence in provider file; used only for Sheps Center CCQI data  | Y                    |
| <b>Pharmacy</b>                | Presence in provider file; used only for Sheps Center CCQI data  | Y                    |
| <b>Dental</b>                  | Presence in provider file; used only for Sheps Center CCQI data  | Y                    |
| <b>Capitation Management</b>   | Presence in provider file; used only for Sheps Center CCQI data  | Y                    |
| <b>Facility</b>                | Presence in provider file; used only for Sheps Center CCQI data  | Y                    |

***Gather available documentation from each data source.***

Source documentation was found via internet search and/or by contacting the data owner directly. Most data sources had some form of documentation that could be translated into the template. See Table 4 for a list of the source documentation.

**Table 4: Source documentation**

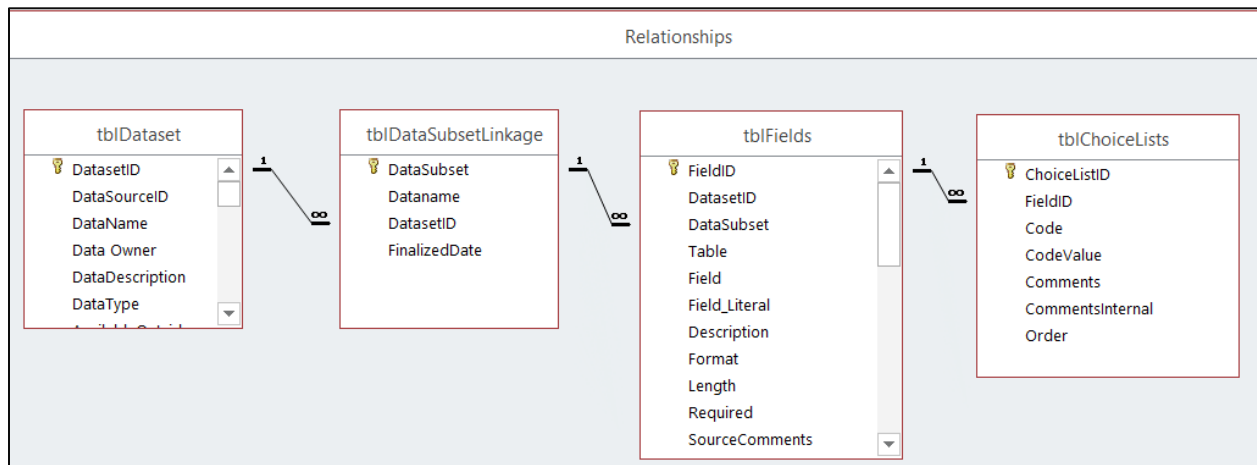
| Data source                          | Source documentation  |
|--------------------------------------|---|
| ED data from NC DETECT               | NCHES Cookbook,<br>CIHAtoNC_DETECT_fileformat_DRAFT_20170822 (Excel)  |
| Crash data from the NC DMV           | 2015 DMV 349 Crash Report Data Element Dictionary   |
| North Carolina Trauma Registry       | North Carolina Trauma Registry (NCTR) Data Dictionary, Effective 1 Jan 2018   |
| Death registration data              | 2014-present Mortality File Layouts_Detail (Excel)  |
| EMS data                             | EMSPIC NEMSIS v3.3.4 State Dataset-NC NEMSIS (Excel),<br>NEMSIS_FieldOptions (Excel)  |
| SCHS ED/hospital discharge data      | dised_layout.xls sent 5/9/2018  |
| FARS data                            | 1975-2017 Fatality Analysis Reporting System (FARS) Analytical User's Manual  |
| OCME data                            |   |
| Sheps Center Hospital Discharge data | 2016_NC_ED_DataDictionary_researcher<br>2016_NC_Inpatient_DataDictionary_researcher<br>2016_NC_Outpatient_DataDictionary_researcher   |
| Sheps CCQI                           | Medicaid:<br>Medicaid_codebook_Claims_201809.xls<br>Medicaid.Data_codebook.Member.9.18.2017.xls<br>Medicaid_codebook_Claims_201709_Summary-1.pdf<br><br>BCBSNC:<br>CLAIMS_USER_CB_V1_Dental.pdf<br>CLAIMS_USER_CB_V1_Facility.pdf<br>CLAIMS_USER_CB_V1_Pharm.pdf<br>CLAIMS_USER_CB_V1_Prof(1).pdf<br>MEMBER_USER_CB_V1.pdf<br>PROVIDER_USER_CB_V1.pdf |
| HSRC Ped/Bike data                   | Bike_Ped2012_2016_Dictionary_draft (Excel)  |
| HSIS data                            | Highway Safety Information System Guidebook for the North Carolina State Data Files, retrieved June 2019<br><a href="https://www.hsisinfo.org/guidebooks/north_carolina.cfm">https://www.hsisinfo.org/guidebooks/north_carolina.cfm</a>   |

When more than one potential document was available, the data owner was consulted as to which to use.

Some data sources had more than one dataset, so these sources were broken up into separate subsets. For example, the Highway Safety Research Center has separate data dictionaries for the pedestrian and bicyclist data, although they have very similar data elements.

An Access database was used to capture all data elements for each data source. Figure 1 illustrates the relational table structure used to capture the data elements.

**Figure 1. Table relationship diagram**



***Apply the standardized format to each data source.***

The elements of the source documentation were mapped to the template (see Table 5) and then imported or entered into the Access database. This mapping is captured in the overview document. The data can then be exported into the overview document (Word) and the data dictionary (PDF and/or Excel) as drafts.



**Table 5. Field mapping from source data documentation**

| Data Source                    | Table or category    | Field               | Field-Literal | Description       | Source comments                              | Format    | Length    | Required (Y/N)                   | Sensitive (Y/N)                       | Unique key (Y/N) | Retired Field (Y/N)    | Retired Date           |
|--------------------------------|----------------------|---------------------|---------------|-------------------|--|-----------|-----------|----------------------------------|---------------------------------------|------------------|------------------------|------------------------|
| ED data in NC DETECT           |                      | Field Name (edited) | Field Name    | Description       | Other Notes + Answer Options (if applicable) | Data Type | Length    | Required?                        |                                       |                  |                        |                        |
| Crash Records from NC DMV      | Number and heading   | Element             |               | Definition        | Source + Attributes (if needed) + Rationale  |           |           |                                  |                                       |                  |                        |                        |
| North Carolina Trauma Registry | Section + Subsection | Variable            |               | Definition        | Summary table, Additional information        |           |           | Summary table: Required by state | Summary table: Available for research |                  | History, Summary table | History, Summary table |
| Death registration data        |                      | Variable            | SAS           | Name              | Code structure description                   |           | Format    |                                  |                                       |                  |                        |                        |
| EMS data                       |                      | Element name        |               | Definition        |  |           | Maxlength | Usage                            |                                       |                  |                        |                        |
| SCHS Hospital Discharge        |                      | Field Name          | SAS Name      | Field Description |  |           |           |                                  |                                       |                  |                        |                        |

| Data Source                     | Table or category                        | Field          | Field-Literal     | Description                          | Source comments                              | Format        | Length          | Required (Y/N) | Sensitive (Y/N) | Unique key (Y/N) | Retired Field (Y/N)                   | Retired Date           |
|---------------------------------|--|----------------|-------------------|--------------------------------------|--|---------------|-----------------|----------------|-----------------|------------------|---------------------------------------|------------------------|
| FARS                            |  | Data element   | SAS               | Definition                           | Remarks                                      | Format        | Format          |                |                 |                  | Discontinued note in Definition field |                        |
| OCME                            |  |                |                   |                                      |  |               |                 |                |                 |                  |                                       |                        |
| Sheps Center Hospital Discharge |  | Label          | Variable          | Label                                |  | Type          | Len             |                |                 |                  |                                       |                        |
| Sheps Center CCQI               | Provider (BCBSNC only), Member or Claims | Variable Label | Variable Name     | Valid Values                         |  | Variable Type | Variable Length |                |                 |                  |                                       |                        |
| HSRC Ped/Bike                   |  | Variable       |                   | From DMV 349 Data Element Dictionary | Revised from DMV 349 Data Element Dictionary |               |                 |                |                 |                  |                                       |                        |
| HSIS                            | SAS Variable File                        | Description    | SAS Variable Name | Definition                           | Additional Information                       | Format Type   | Format Type     |                |                 |                  | Additional Information                | Additional Information |

Some source documentation included data elements outside the scope of our template, but the data elements did not seem important to include. For example, the EMS file had the following data elements that did not seem critical to include for researchers: minLength, pattern, fractionDigits, totalDigits, and minExclusive.

However, two data sources, the North Carolina Trauma Registry (NCTR) and the Sheps Center for Health Services Research had data elements that were not part of our template, yet appeared to be important for researchers. The NCTR had an “Available for Research” field which indicated whether the field could be included in an extract or only used as selection criteria.

The Sheps Center CCQI data had indicators when the data elements were present from different types of provider files. See Table 6 for a description of the data elements in the two subsets, Medicaid and BCBS. These indicators were added to the template for this dataset.

**Table 6: Additional data elements for Sheps Center Carolina Cost and Quality Initiative**

| Type of provider             | Medicaid | BCBS |
|------------------------------|----------|------|
| <b>Institutional</b>         | X        |      |
| <b>Professional</b>          | X        | X    |
| <b>Pharmacy</b>              | X        | X    |
| <b>Dental</b>                | X        | X    |
| <b>Capitation Management</b> | X        |      |
| <b>Facility</b>              |          | X    |

In these cases when the data elements seemed important to the final documentation, exceptions were made to the template to include the data elements.

***Submit the final version to the data owner for approval.***

The draft documentation was sent to a contact person at the data source by email for review.

***Post the finalized documentation online, with data owner's permission.***

Finalized documents were returned to the data owner. If the data owner permitted the documentation to be posted online, it was posted on the [CCHI Data Sources for Motor Vehicle Crash Injury Research in North Carolina webpage](#).

Two data sources, crash report data and FARS, were discussed with data owners, but were not formally reviewed and approved. There was verbal agreement that our approach with the source documentation was correct, but they indicated there were too many competing priorities for a review of the compiled documents. Given that these data are critical for motor vehicle crash injury research, we chose to create documentation using their source documents as per their instructions and post the results online with a note that they should be considered provisional.

## **Summary of Aim 1**

We found that the process to create standard documentation was effective. The creation of a database of all the documentation kept the information consolidated and easily retrievable. There was considerable variation in the source documentation, resulting in differences in the amount of descriptive data elements that were available for the data dictionaries. Given the variability of the source documentation, a flexible approach with a goal of usability for researchers was important.

## **Aim 2: Describe the results of the documentation and the challenges encountered, solutions found, and lessons learned.**

### **Results**

Twelve data sources were documented with the standardized template (see Table 7). One data source, the North Carolina Healthcare Association, declined to participate in the project. As of August 2019, documentation was posted online for seven of the data sources.

**Table 7. Summary of final documentation**

| Data source                          | Data subset count | Field count | Field choice list count | Date finalized | Compensatory funding given | Primary Contact(s)           | Documentation Status             |
|--------------------------------------|-------------------|-------------|-------------------------|----------------|----------------------------|------------------------------|----------------------------------|
| ED data from NC DETECT               | 1                 | 74          | 51                      | 6/18/18        | No                         | Clifton Barnett              | Available upon request           |
| Crash data from the NC DMV           | 1                 | 234         | 983                     | 1/10/2019*     | No                         | Eric Bellamy<br>Brian Murphy | <a href="#">Available online</a> |
| North Carolina Trauma Registry       | 1                 | 250         | 853                     | 2/9/2019       | Yes, FY 2018               | Sharon Schiro                | <a href="#">Available online</a> |
| Death registration data              | 1                 | 118         | 115                     | 8/21/18        | Yes, FY 2018               | Matt Avery                   | <a href="#">Available online</a> |
| EMS data                             | 1                 | 414         | 2,165                   | 9/20/18        | Yes, FY 2019               | Tony Fernandez               | Available upon request           |
| SCHS ED/hospital discharge data      | 1                 | 44          | 0                       | 4/15/2019      | Yes, FY 2018               | Robert Lee                   | Available upon request           |
| FARS data                            | 1                 | 415         | 3,842                   | 9/21/2018*     | No                         | Eric Bellamy                 | <a href="#">Available online</a> |
| OCME data**                          | 1                 |             |                         |                | Yes, FY 2019               | Alison Miller                |                                  |
| Sheps Center Hospital Discharge data | 3                 | 129         | 64                      | 10/16/18       | Yes, FY 2019               | Roger Akers                  | <a href="#">Available online</a> |
| Sheps CCQI                           | 2                 | 495         | 1,239                   | 6/4/2019       | Yes, FY 2019               | Roger Akers                  | <a href="#">Available online</a> |
| HSRC Ped/Bike data                   | 2                 | 147         | 698                     | 8/30/2019      | No                         | Libby Thomas                 | <a href="#">Available online</a> |
| HSIS data                            | 1                 | 324         | 1,968                   | 7/23/2019      | No                         | Ana Eigen                    | <a href="#">Available online</a> |

\*Date compiled; data owner indicated they were not able to review

\*\* Work is ongoing to finalize OCME data documentation.

# Challenges

## ***Variability in access to source documentation***

Some of the data sources had publically available documentation, others had ready documentation they were willing to share upon request, and others had neither. Some data source owners had reservations about sharing their documentation or did not consider their documentation to be shareable. One data owner chose to withdraw from the project after expressing concerns about sharing their documentation.

Our approach to these differences was to accept whatever documentation the data owner was willing to provide. When it was unclear whether any usable documentation existed, other resources, such as forms used by the organization to gather data, were used to create draft documentation for the data owner's review.

## ***Variability in the quality and depth of source documentation***

Source documentation ranged from a list of data elements with no descriptors to extensive reviews of each data element, including the history of the data element and every change which had been made to the data element since the data were gathered.

Our approach to working with different levels of documentation was to customize the data dictionary format to show whatever data elements were available that could be mapped to our template. When important data source-specific data elements were outside the template, the template was modified to include them.

## ***Variability in the timeliness of the documentation***

In addition to the variability of the source documentation, there were also differences in the frequency in which the documentation was updated. For the data sources with annual revisions to their documentation, this represented a challenge for maintaining an up-to-date version of the standardized documentation, given that revising the documentation on an annual basis was not

anticipated as part of this project. The following five data sources update their data documentation and post it online annually.

1. North Carolina Trauma Registry
2. FARS
3. Sheps Center Hospital Discharge data
4. Sheps CCQI

Our approach to this issue is to provide links to the data sources and to include the dates our documentation was last compiled on the [CCHI website](#). Potential users of the data will thus have access to the timelier source documentation, if needed.

### ***Lack of responsiveness from the data owner***

Given the need for assistance from the data owners, this was often a delaying factor in completing the documentation. However, we understood the project represented unanticipated work for data owners. When requested, funding was provided to reimburse for their time spent reviewing the compiled documents. However, in the setting of many competing higher priority projects, there was often an understandable delay in response time.

Our approach to other communication delays was to send reminders as needed. When a particular issue needed addressing, another successful approach was to discuss it in person at meetings.

## **Lessons Learned**

Although there was considerable variation in the ways data were documented across data sources, using a standardized template created a common documentation language across many data sources, which we hope will increase its usability. By having a centralized database of documentation, the standardized template should also be easily customizable in the future.

## **Summary of Aim 2**

Twelve data sources were documented using a standardized template for the project. Flexibility in working with different kinds of documentation and with the data owners themselves was key to the success of the project.

## Contact Information

If you have any questions about this report, please contact:

Anna Waller, ScD, Principle Investigator, [awaller@med.unc.edu](mailto:awaller@med.unc.edu)

Kathy Peticolas, Project Coordinator, [kathy\\_peticolas@med.unc.edu](mailto:kathy_peticolas@med.unc.edu)

### **Carolina Center for Health Informatics**

Department of Emergency Medicine, University of North Carolina at Chapel Hill

Floor 1, 100 Market Street

Chapel Hill, NC 27516