

Working with NCCEH Data as a Researcher



*Initial Draft by
Mike Dolan Fliss, PhD & Esther Chung, PhD
UNC Injury Prevention Research Center
7/25/2022*

Table of Contents

Overview of Federal HMIS	3
Key Resource Links	3
Overview of NCCEH HMIS	3
Statewide	3
HMIS in North Carolina (split (+timeline))	3
Continuum of Care (CoCs)	4
NCCEH Governance structure	4
NCCEH Datasets	5
Key NCCEH Partners & Related NC Organizations	5
NCCEH HMIS Data	6
Overview & key resources	6
Exporting Databases Tables	6
NCCEH Data Collection	6
Table Structure	8
Key Concept: Person-Enrollment	8
Key Concept: Person-Place-Time	9
Details on Cleaning Data	10
Table Fields - key considerations	10
Dates	11
Sub-assessments	13
Details on Flattening Data	14
Key Supplemental Tables	15
Linkage of NCCEH Data	15
The Importance of Statewide HMIS Self-Linkage	15
Past Linkage Projects	15
Linkage Types & Key Concepts	16
Key linkage fields	16
Special Linkage Considerations	17
Future Research Opportunities	17
Updates, Feedback, & How to Help	19

Overview of Federal HMIS

The Homelessness Management Information System (HMIS) is a client-level electronic data collection system for people experiencing homelessness. At the local level, homeless assistance providers use HMIS to coordinate care, manage and improve operations, and serve clients. At the federal level, the Department of Housing and Urban Development (HUD) uses HMIS for its annual report to Congress on homelessness in the US. Each HMIS system must collect information consistent with the HUD HMIS Data Standards. Continuum of Care (CoC) programs are responsible for designing and managing an HMIS system. A CoC is the group that coordinates services and programs for those experiencing homelessness in a specified geographic area. For an introductory background to the CoC program, see this [guide](#), and you can find more details on the HUD [website](#).

Key Resource Links

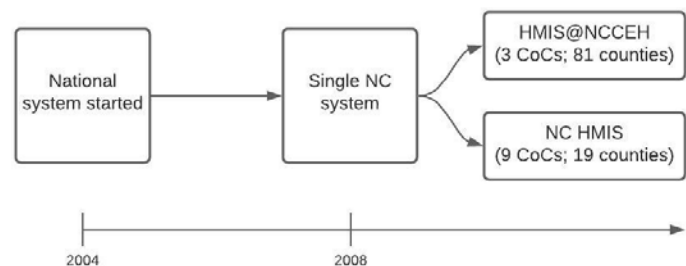
- [Brief history of HMIS](#)
- [HMIS Data Standards](#) (for current year)
- [HMIS Data Standards Logical Model](#) (2020 version)
- [HMIS Data Standards Logical Model](#) (2022 interactive)

Overview of NCCEH HMIS

Statewide

HMIS in North Carolina (split (+timeline))

Prior to 2008, there was a single statewide HMIS system. Currently, the North Carolina state-wide HMIS is managed by two distinct entities: HMIS@NCCEH (hereafter NCCEH) and NC HMIS. Both systems consulted with [Dennis Culhane](#) (Professor, University of Pennsylvania), a longtime expert in both HMIS instances and data governance, to assist with their instance relationships, data management, and governance decisions and the possibility of integrating data. However, at this point, researchers interested in using NC HMIS data will need to contact the appropriate governance boards at both NCCEH and NC HMIS. NCCEH has a formal, centralized advisory board and makes central decisions about research projects. Historically, NC HMIS also has a board of representatives, but makes decentralized, individual decisions; this, historically, has required county-by-county decision making for research projects, making it unworkable (as of yet) for state-wide research projects.

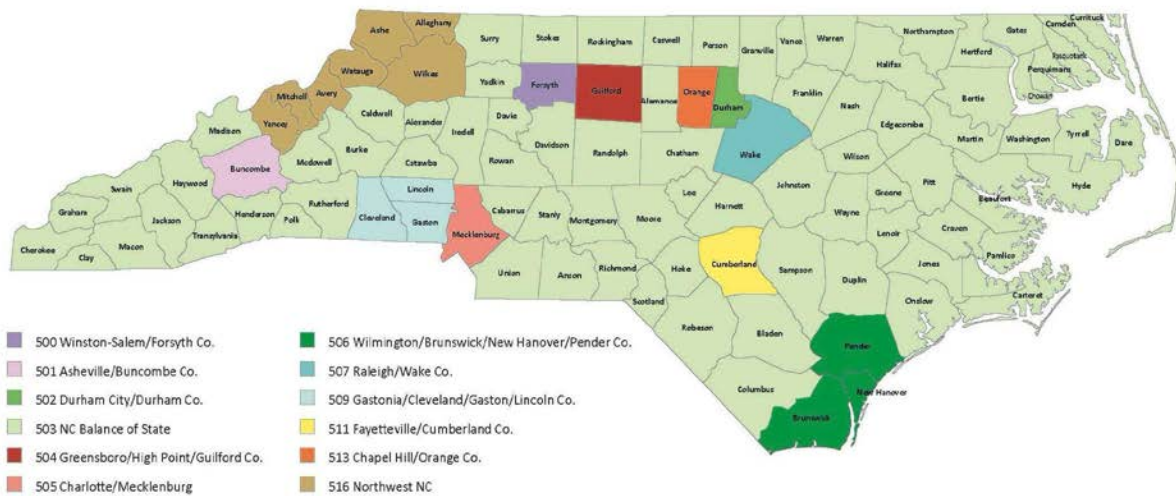


A combined system for statewide information on homelessness has not yet been created, and efforts to do so have previously failed. However, there have been discussions with the [CDC's Clinical and Community Data Initiative](#) (CODI) to link individual health-related information across healthcare and social systems. This may involve relinking the two systems to create a deduplicated statewide HMIS first, but discussions are ongoing.

Continuum of Care (CoCs)

As of 2021, North Carolina has 12 CoCs depicted below in the map. NCCEH manages three CoCs which covers 81 counties: Balance of State (503), Durham (502), and Orange (513). NC HMIS manages the remaining nine CoCs and 19 counties, including a number of large cities like Asheville in Buncombe, Charlotte in Mecklenberg, and Raleigh in Wake.

It is important to remember that this management and software implementation space is an ever-changing environment. Most recently, in 2022, there has been discussion of one instance leaving their NC HMIS instance to move to a different software platform (effectively creating a third instance of a single county). In 2022, NCCEH is also clarifying, with their board, the process, timelines, and considerations for welcoming other CoCs to join their implementation. This ever-changing environment can create challenges for standardized, statewide, longitudinal research.



NCCEH Governance structure

Data Governance includes the concepts including legal dataset and record protections, data security requirements, data suppression and release thresholds, consent to collect data practices, data culture practices and organizational goals, ethical research standards and review, and data destruction details. Researchers of NCCEH data should be prepared to respond to typical data governance details, including how they plan to safely receive, protect, store, and dispose of the data itself, as well as expectations for review of data products, like presentations and journal articles. Researchers may choose to share some more detailed results with NCCE directly, and reserve higher level observations for public release.

Regarding data use, NCCEH is the designated HMIS lead and meets regularly with the [HMIS Advisory Board](#), a group of representatives from their three CoCs (Balance of State, Durham, and Orange), NCCEH, and regional and state partners. Users who are interested in working with NCCEH HMIS data will need to complete the NCCEH Data Use Agreement (DUA) that describes the project intent, data protection controls, and public benefit. NCCEH will respond to appropriate data requests meeting their research objectives with the DUA and may provide a DUA online in the future. Note that data linkage projects may need to contend with multiple organizations, and therefore the different data governance rules, practices, and culture of those organizations.

NCCEH Datasets

NCCEH manages two major datasets on homelessness: Point-in-Time (PIT) Count and HMIS. This report primarily focuses on HMIS data, but we provide a brief summary of PIT. The PIT count is a count of sheltered and unsheltered individuals experiencing homelessness, using the HUD definition of “people who are living in a place not meant for human habitation, in emergency shelter, in transitional housing, or are exiting an institution where they temporarily resided.” In North Carolina, counts of unsheltered individuals occur every year on a single night. HMIS contributes counts of sheltered individuals to the PIT count. More information on how NCCEH collects PIT count data can be found [here](#).

Key NCCEH Partners & Related NC Organizations

- **NC HMIS** is the sister organization to NCCEH, managing the HMIS implementation for the 19 counties not in the 81 county NCCEH instance. Communication should be to their board chair or to their county-specific data leads.
- The [NC DHHS State ESG Office](#) is a key funder of homelessness shelter housing & services. State grants award annual Emergency Solutions Grant (ESG) Program funds from the US Department of Housing & Urban Development (US HUD) to local governments and non-profits, including NCCEH.
- The [Institute for Community Alliances](#) (ICA) manages the technical implementation of the NCCEH HMIS instance.
- [NC Housing Coalition](#) is a non-profit focused on homelessness and housing insecurity related policies. It does not manage shelter data (unlike NC HMIS instances).
- [NC Housing Finance Agency](#) provides financing for affordable housing opportunities for North Carolinians whose housing needs are not met by the market. It was created in 1973 by the NC General Assembly (NCGA) and has financed over 300,000 homes and over \$30 billion dollars. It reports its budget and activities through the NC Office of State Budget and Management (OSBM).
- **NC [Clinical and Community Data Initiative \(CODI\)](#)** aims to create a system to link an individual's data throughout the community in a single record using privacy preserving record linkage, a technique initially developed by the Center for Disease Control & Prevention (CDC). NCCEH is in ongoing discussions with [NC CODI](#) to provide shelter enrollment & service data for carefully chosen linkage projects.
- [NC DHHS Division of Public Health](#) (DPH) includes sections and branches who focus on key public health content areas, such as smoking and tobacco, childhood health, cancer, injuries, the environment, infectious diseases (like COVID-19 and STIs), and vital records (births, deaths, etc.). People experiencing homelessness often have unique and disparate experiences, exposures, and outcomes for these public health areas. NC DHHS DPH may be a partner for specific homelessness-related projects.

NCCEH HMIS Data

Overview & key resources

- [HMIS@NCCEH Data Tree Structure](#) (last page)
- [Required Data Elements](#)
- [HMIS Data Collection Forms](#)

Exporting Databases Tables

If making a request for aggregate data, NCCEH may provide tables directly to you. If making a request of record-level data, your data may need to be exported in bulk from the system by the data contractor. Data are available as CSV or XML files, and specifications for these files can be found [here](#). Data may be segmented by specific CoCs into multiple tables (e.g., Enrollment, Exit, Project). Data export issues have occurred in the past, but at current date are largely fixed, allowing export of entire tables. Exports typically require a group of projects (if not the full HMIS@NCCEH implementation and an enrollment start date and end date range. Exported tables may need to be carefully joined back together for certain questions. Joining individual tables together is described later under [Table Structure](#) and [Details on Flattening Data](#).

Even if you are not working with record-level data, it can be useful to understand some of the HMIS data structure to better communicate with NCCEH - and their vendor / support team at the Institute for Community Alliances (ICA) - about your data request.

NCCEH Data Collection

Data may be collected on clients at multiple points during an enrollment. An enrollment is any stay at a given HMIS project, which is any group, organization, or provider of homelessness services that collects HMIS data. Depending on the type of project, an enrollment may last for several years or still be ongoing. An explanation of the data collection stages is provided [here](#) and NCCEH has training videos related to data entry that may be useful to review [here](#). Briefly, projects may follow up with clients at multiple stages during an enrollment (see figure below). These data collection points include Entry Assessment (project entry), Interim Assessment (Interim update, annual assessments), and Exit Assessment (project exit).

Data Collection Stages

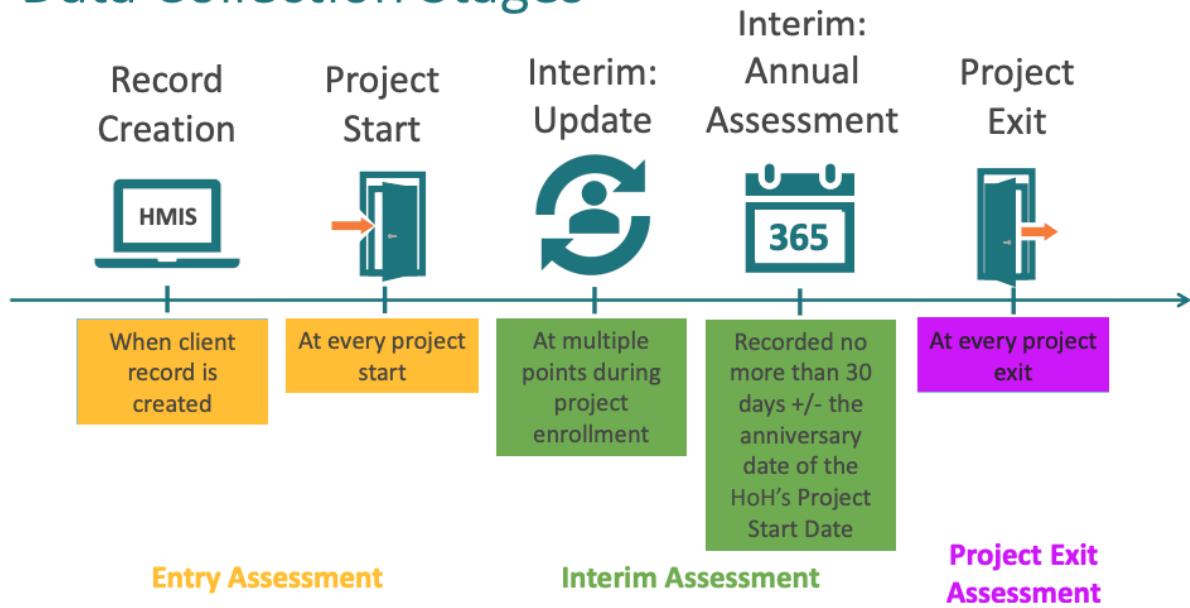
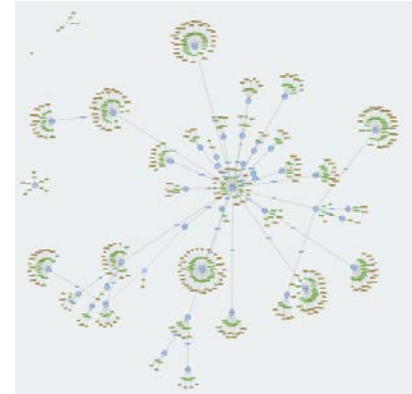
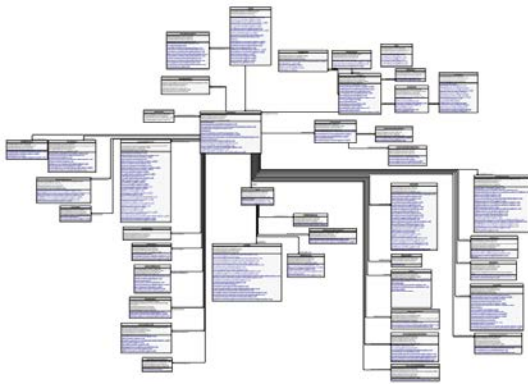


Figure. From [NCCEH Training](#)

Table Structure

Once you are ready to join tables, refer to the most up to date HMIS Logical Model [here](#), which will provide information on the relationships between tables. A simplified Entity Relationship Diagram (ERD) is shown in the next section ([Key Concept: Person-Enrollment](#)). The full ERD of HMIS data can be found [here](#) and an interactive version [here](#).



Figures: Full (and complex) ERD models of interrelated HMIS data, above.

Key Concept: Person-Enrollment

An enrollment is the central event in HMIS. A record is created when a client is enrolled at a project by that project's staff. All data are linked to a person-enrollment, and almost all tables in HMIS are directly connected to an enrollment. Important table joins to the Enrollment table include the following: Client, Project, and Exit.

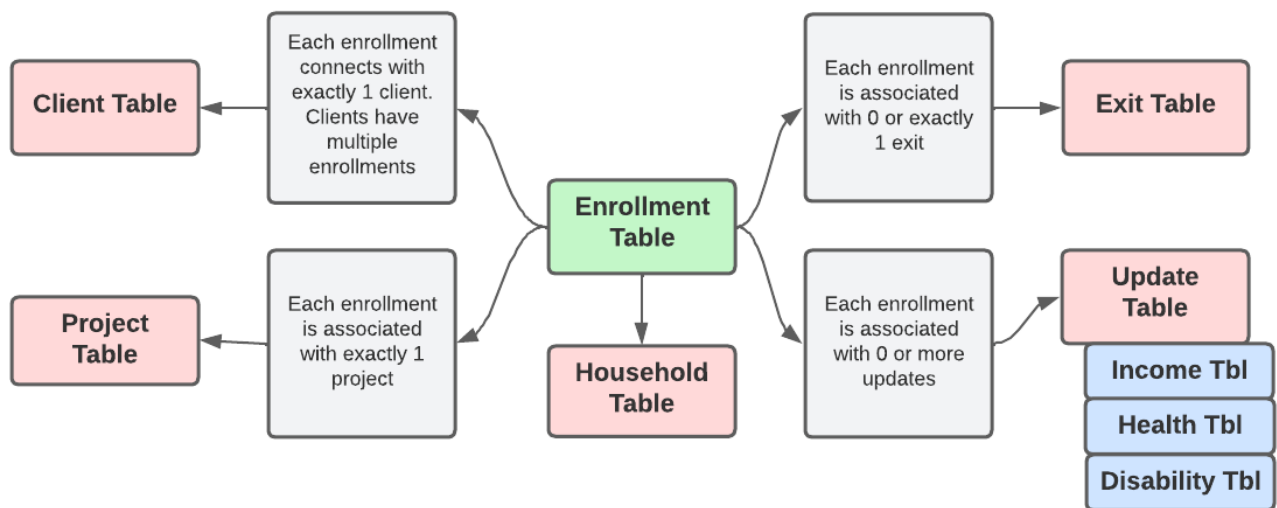


Figure. Simplified HMIS ERD for researchers

The **Enrollment table** contains information related to the period in which a person is considered a client of a project (e.g., entry date, prior living situation). Enrollments are the central event in HMIS. All enrollments describe project entries with an associated Entry Date and DataCollection Stage (1 = entry). To join to other tables, use EnrollmentID foreign keys in those tables. There can be multiple enrollments for a client. The unique ID for the Enrollment table is *EnrollmentID*.

The **Exit table** contains information related to a client leaving a project (e.g., exit date, where the client plans to stay). There are exactly zero or one exits associated with an enrollment. Zero exits can indicate open enrollments in which clients are still active at a given project. Any record in the exit table has a DataCollection Stage (3= exit). Depending on the project type, there may be missing exit dates. For example, there are enrollments at permanent supportive housing projects that are 10+ years. See below for more information. Joins to the Enrollment Table are made with the *EnrollmentID* and *PersonID*.

The **Client table** contains person-level information (e.g., name, date of birth, race, gender). Every enrollment must be associated with exactly one client, but a client can have multiple enrollments. There are some fields that are person-associated that do not appear to change over enrollments, such as gender. Joins to the Enrollment Table are made with the *PersonID*. Unlike from Income and Disabilities (see Enrollment-linked Update table), any changes made to demographic information are time-invariant (changes overwrite previous records).

The **Project table** contains information related to the specific project the client is enrolled in (e.g., project name, type of project [permanent supportive housing, street outreach]). NCCEH has 3 CoCs representing 81 NC counties, and ~250 projects (by unique ProjectID and name). Some of those projects may be closed. There are 8 project types (e.g., permanent supportive housing, emergency shelter). Emergency Shelter and Rapid Re-Housing represent nearly 4/5 enrollment events. Each project is associated with a project type. Some project types require more data entry / table completion than others. Every enrollment is associated with one project. Joins to the Enrollment Table are made with the *ProjectID*.

An **Update table** can provide updated information for enrollment events (see Figure above). There is not one single table that contains all information related to updates of an enrollment. Rather, this information is collected in separate, closely-related tables that are linked to an enrollment event. All updates associated with their own InformationDates, which necessarily have a DataCollection Stage (2=update, 5=annual assessment). These update dates can help replace missing *ExitDate* information on enrollments without *ExitDate*. "Updates" also include time-specific demographic information about services and identities. More information about some of these update tables can be found in [Sub-assessments](#).

Key Concept: Person-Place-Time

An important concept for data linkage is person-place-time. Correct research and linkage projects entail understanding who, where, and when a person and event occurred. Researchers using HMIS data should carefully consider each of these components.

Person: Given the HMIS system in NC is split between two entities (NCCEH and NC HMIS), it is important to consider that across an individual's history, they may have project enrollments in both systems, but these enrollments may not appear across both NCCEH and NC HMIS. This may lead to duplicate person-enrollments across both systems or more likely, an undercount of persons experiencing homelessness when working with just one instance of HMIS data. When working with NCCEH data, it may be important to examine this by looking at variables such as the enrollment date, project type, and CoC/county.

Place: In HMIS data, it is important to understand that place is a function of enrollment and **not** person. In other words, since enrollments are the central events in HMIS, place is tied to the enrollment and the project that the client is accessing. Depending on the research question, place may be defined as the client's current location, prior living situation, history of homelessness, or destination, etc. Given the HMIS data structure is

centered around an enrollment, it is important to consider whether your concept of place can be identified. For example, a client's current location defined as their "home address" is likely not possible to identify in NCCEH given the housing instability experienced by HMIS clients and the listed address information is typically related to the project and **not** the client.

Time: Like place, time is a function of an enrollment, not a person. There are multiple time constructs that may be of interest, such as length of stay for a given enrollment, time between enrollments, etc. Calculations of these time point and duration variables should be carefully considered given the data structure. We recommend consulting with the NCCEH team to confirm such date and duration calculations.

Details on Cleaning Data

To prepare data for analytical use, it will be important to reference the latest versions of the HMIS [Data Dictionary](#). The Data Dictionary will provide information on some, but not all, of the variables in the HMIS data, including variable type (e.g., numeric, factor, character, etc.) and response options. Further details on most of the variables can be found in the [Data Standards Manual](#). When cleaning and/or recoding variables, extra care should be taken to review both documents. The most up-to-date versions of these documents can be found [here](#).

Table Fields - key considerations

There are hundreds of fields in HMIS. Below we review a subset of those fields we believe will be of common interest to many researchers asking public health and housing related questions. Data field names **do not** necessarily correspond to the wording of a question; it is critical that users review the [HMIS Data Collection Forms](#). For example, *lengthofstay* in HMIS does not refer to the current length of stay for an enrollment, but rather the previous length of stay.

In NCCEH, there are a number of data elements that are required for an enrollment, which can be found [here](#). Some of these elements are project-specific such as "date of engagement", which is only used for Street Outreach projects. Data completeness may depend on the *Project Type*, *Head of Household status*, and funder.

Project Type

There are multiple types of projects in HMIS, such as Permanent Supportive Housing, Emergency Services, and Coordinated Entry. Understanding the flow of data collection by *Project Type* will help in understanding what data are available. Depending on the project type, different forms may be used for enrollments, updates, and exits. See [HMIS Data Collection Forms](#) for all forms organized by project type. As described below in [Dates](#), a missing *ExitDate* may be accurate for clients utilizing Permanent Supportive Housing who are in an "open" enrollment. In contrast, a missing *ExitDate* for Emergency Services that is beyond 30/90 days from the date of entry is likely an error in reporting.

Race and Ethnicity

In HMIS, race is measured using self-identification of one or more of the five racial categories established by the Office of Management and Budget (OMB): American Indian/Alaskan Native, Asian, Black/African American, Native Hawaiian/Pacific Islander, White. Ethnicity is captured by asking clients whether they identify as Hispanic or Latino (Yes/No). When cleaning these variables and creating a race and ethnicity variable (e.g., Non-Hispanic Black, Hispanic White), it is important to consider how to code bi- or multi-racial clients. For example, more people identify as American Indian and Black than American Indian alone. Using the mutually

exclusive OMB categories will be flawed for these individuals. Researchers should consider other options, such as a mutually exclusive category that combines racial groups or create a “Multi-racial” category. The problem with the Multi-racial group is that it encompasses many identities and it is impossible to examine individual multi-racial identities if they are all collapsed into one. However, the degree to which this is important depends on the focus. If the goal is linkage, then the focus on coding race and ethnicity should be on harmonization across data sources, in which case a Multi-racial category may be adequate. When smaller group totals may not be shareable . Researchers should remember that race is a socially constructed concept, with consequences including: people may (and do) identify differently over time and place, and self- and other-ascribed race and ethnicity may and do differ.

Dates

Date of birth (DOB) can be found in the Client table and should be used to calculate a client’s age on a given date.

As described [earlier](#), data may be collected on clients at multiple points during an enrollment. An explanation of the data collection stages is provided [here](#) and NCCEH has training videos related to data entry that may be useful [here](#). All enrollments are entries into a specific project and *EntryDate* reflects the date of enrollment in a given project. All exits are project exits and *ExitDate* describes the date of exit from a given project. An *ExitDate* may be missing for a given enrollment. Depending on the project type (e.g., Permanent Supportive Housing), this can indicate the enrollment is still open. For example, clients who are at Permanent Supportive Housing projects can have open enrollments since the 1990s. Depending on your analyses, it may be necessary to impute the *ExitDate* as the date when data were pulled in *ExportDate* in the Export Table. We recommend talking with NCCEH to determine whether missingness for *ExitDate* is reasonable or not given the project type.

The most common scenario is that a client will have an enrollment and exit. When projects follow up with clients during an enrollment with an interim update or annual assessment, this information will be captured in other HMIS tables, such as Disabilities, IncomeBenefits, HealthandDV, and EmploymentEducation (see [Sub-assessments](#) for more information). These tables will have *DataCollectionStage* and *InformationDate* variables, which correspond to the stages of data collection in the following table.

<i>DataCollectionStage</i>	<i>InformationDate</i>	Number of possible records
1 (entry)	Match <i>EntryDate</i> in Enrollment table	Only one
2 (update)	Date of the update	Multiple records are possible
5 (annual assessment)	Date of the annual assessment	Multiple records are possible
3 (exit)	Match <i>ExitDate</i> in Exit table	Only one

Table. Stages of data collection, information data, and record multiplicity relationships

Updates within Disabilities Table

HMIS Data from NCCEH. June 2022

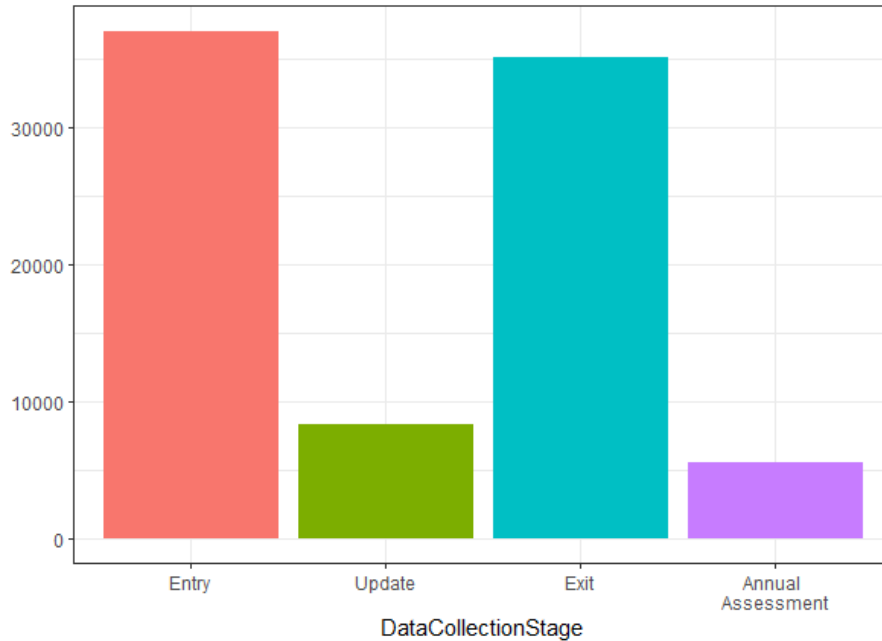


Figure. Updates within an enrollment within the Disabilities Table.

Below is an example of how these dates may appear in the Disabilities Table. Here, we have a client (*PersonalID*=17490) with two enrollments (*EnrollmentID* = 70651 and 1062692). For the first enrollment (70651), the client entered the project on 07/01/2006 (*DataCollectionStage* = 1), had two annual assessments on 07/06/2016 and 07/01/2017 (*DataCollectionStage* = 5), and then exited on 03/01/2018 (*DataCollectionStage* = 3). The *InformationDate* for entry and exit dates for these two enrollments also match those captured in the Enrollment and Exit Tables. The second enrollment (1062692) also included an entry, update, and an exit.

disabilities_id	enrollment_id	personal_id	information_date	data_collection_stage	disability_type
70651_1_M	70651	17490	2006-07-01	1	9
38250_5_M	70651	17490	2016-07-06	5	9
60084_5_M	70651	17490	2017-07-01	5	9
70651_3_M	70651	17490	2018-03-01	3	9
1062692_1_M	1062692	17490	2018-03-01	1	9
118858_2_M	1062692	17490	2018-03-02	2	9
1062692_3_M	1062692	17490	2018-06-29	3	9

(A) Disabilities Table

enrollment_id	personal_id	entry_date	project_id	exit_id	enrollment_id	personal_id	exit_date
70651	17490	2006-07-01	1698	70651	70651	17490	2018-03-01
1062692	17490	2018-03-01	7634	1062692	1062692	17490	2018-06-29

(B) Enrollment Table (C) Exit Table

Figures. (A) Disability, (B) Enrollment, and (C) Exit Table examples

Finally, in HMIS, there is a *lengthofstay* variable that reflects a client's prior length of stay at an HMIS project. To calculate the length of stay for the current project enrollment, users must generate this and calculate the time between *ExitDate* and *EntryDate*.

Sub-assessments

During an enrollment, there are certain “gateway” questions (i.e. conditional branching or skip logic) that when answered “Yes” lead to more detailed questions to be filled out. We provide details on the Disabilities and IncomeBenefits tables below. More information can be found in the [NCCEH training videos](#) and their [sub-assessment guide](#).

Disabilities Table (“Gateway” question for disability table)

A client’s disabling condition(s) can be collected at various *DataCollectionStages* (entry, update, annual assessment, exit), though it must be collected at enrollment. Details on the HUD definition of a disabling condition can be found in the [Data Standards Manual](#). In the **Enrollment table**, the gateway question to the Disabilities table is “Does the client have a disabling condition (No/Yes)?” According to the [2022 HMIS Data Standards Manual](#), only one value is allowed for the duration of a project enrollment. Changes should be made to the client’s record, which “should always reflect the known status of a client’s disabling condition.”

- If “Yes”, the sub-assessment for the Disabilities table will be filled out with information regarding the disabling condition (or disabling condition).
- If “No”, the sub-assessment is not filled out and the client and enrollment (*PersonIID*, *EnrollmentID*) will not appear in the Disabilities table.

When joining the Disabilities table to other tables (e.g., Enrollment), it is important to account for the following:

- 1) Only clients with disabilities will appear in the Disabilities table,
- 2) The information can be collected at multiple *DataCollectionStages* (entry, update, annual assessment, exit). See [Dates](#) for an example of a client with updates and annual assessments in the Disabilities Table.
- 3) Each row in the Disabilities Table is a person-enrollment-event-disabling condition, where an event can be an entry, update, annual assessment, or exit. In other words, for each enrollment-event when the gateway question is answered “Yes”, there is a row for every possible disabling condition in HMIS (e.g., physical or developmental disability, chronic health condition, mental health problem, substance abuse, etc.). There are potentially many empty rows of data. For example, a client may have a physical disability at project entry and will have empty data for the other disabling conditions. Then, at an update or annual assessment, they may report a new chronic health condition, for which the row for person-enrollment-update-chronic health condition would have new information. Depending on the research question, it may be possible to flatten these data before merging with enrollments (see [Details on Flattening Data](#)).

IncomeBenefits Table

A client’s health insurance, income, and non-cash benefits are recorded in the IncomeBenefits table. The gateway questions are as follows.

- “Is the client currently covered by health insurance?”
- “Does the client currently have any income from any source?”
- “Does the client have any non-cash benefits from any source?”

A “Yes” answer to any of these gateway questions will prompt the associated sub-assessment form. These sub-assessment forms are then combined into the IncomeBenefits table. Similar to the Disabilities table, the IncomeBenefits table can be collected at multiple *DataCollectionStages*. When merging the IncomeBenefits table to other tables, be sure to account accordingly.

Following our same client (*PersonID* = 17490), they also have updates and annual assessments in the *IncomeBenefits* table. The *InformationDates* are the same as the dates in the *Disabilities* Table. For the first enrollment (70651), their total monthly income increased from project entry (\$603 to \$733) at subsequent annual assessments and then decreased at project exit back to the project entry level.

enrollment_id	personal_id	information_date	data_collection_stage	total_monthly_income
<chr>	<chr>	<chr>	<chr>	<chr>
70651	17490	2006-07-01	1	603.00
70651	17490	2016-07-06	5	733.00
70651	17490	2017-07-01	5	735.00
70651	17490	2018-03-01	3	603.00
1062692	17490	2018-03-01	1	603.00
1062692	17490	2018-03-02	2	603.00
1062692	17490	2018-06-29	3	603.00

Figure. *IncomeBenefits* Table screenshot of example client

Other possible Sub-assessments include the *Health and Domestic Violence* Table and *Employment and Education* tables.

Details on Flattening Data

Given the complexities of the NCCEH tables, it may be common to “flatten” related tables to the concept of interest. We provide some examples below. In essence, there are time-varying elements to some of the tables that you may want to summarize at the person-level (or CoC level, etc.), such as disability patterns, average length of stay, minimum / maximum dates. Depending on the concept, flattening data may occur within each table before joining with other tables.

Examples

High-level linkage example: Typically, when linking one dataset to another (e.g., NCCEH and death records), we are linking people, not events. The data pull needs to support the linkage in question, which may require more or less specificity on person-place-time. For instance, death linkage requires mostly person-level data and little place and no time data. Linkage of an event (like an eviction) to an enrollment event requires aggregating HMIS data at a different level of detail (enrollments) and linking not just person, but also time and possibly place.

Person-level example #1: For person-level questions (e.g., the demographics of individuals with an active enrollment before a certain date), enrollment events will need to be joined to individual clients, then collapsed / combined back to person-level records. For example, if you want to know the gender breakdown of individuals for a certain CoC in 2020, that translates to: (1) select all enrollments from the Enrollment table that have any time (end date >= Jan 1 2020 or start date <= Dec 31, 2020); (2) merge the Client table to these enrollments; (3) collapse the enrollments to unique *ClientID*, thereby flattening the data to person-level records; (4) summarize the gender breakdown.

Person-level example #2: In the HMIS-death certificate linkage, we were interested in obtaining the dates of any possible interaction with clients after enrollment. In most cases, this was only the *ExitDate*, but for those actively enrolled in a project (e.g., clients in permanent supportive housing), they did not have an *ExitDate*. However, these clients may have had interim updates (*DataCollectionStage* = 2) or annual assessments (*DataCollectionStage* = 5) in which *InformationDate* would provide an indication of an interaction (and therefore suggest the person did not die before that date). Instead of creating a very long person-event table where an event could be any of the possible values of *DataCollectionStage* (entry, update, annual assessment, or exit),

we did the following: (1) obtain the earliest and latest *InformationDate* from each of the following tables: Disabilities, IncomeBenefits, HealthandDV, and EmploymentEducation; (2) join these date variables into the Enrollment table; (3) join the Exit table to the Enrollment table; (4) take the maximum date out of all the date variables.

Key Supplemental Tables

Another important table for consideration is the Export table, which contains information related to the actual export of data, including the date range of the exported data. For manual or periodic data exports, this will be important to review for overwriting data and confirming the prompts used to pull the export.

For other important tables, review the logical model. The HMIS CSV Format Specifications provides an overview of key tables, but it is not comprehensive. The [Data Dictionary and Data Standards Manuals](#) will provide more information.

Linkage of NCCEH Data

The Importance of Statewide HMIS Self-Linkage

The ideal linkage of HMIS data to other (e.g., public health) datasets would first include an HMIS self-linkage, that is, linkage between HMIS@NCCEH and NC HMIS instances to create one statewide HMIS system database across all instances that contains de-duplicated information on people experiencing homelessness across all 100 counties in NC. This is not yet possible, but should continue to be a priority. Without a single deduplicated state dataset, any current linkage with one instance will necessarily have important limitations. This is because people experiencing homelessness may receive service and shelter across multiple CoCs since, by definition, they do not have a single CoC “home.”

Past Linkage Projects

To our knowledge, no linkage projects have been completed with NC HMIS (19 county) shelter records. The following linkage projects involved collaborations with NCCEH (81 counties as of 2021):

- 1) **Duke Electronic Medical Records** - Debbie Biederman was the lead, making a linkage of NCCEH records to Duke Electronic Medical Records (EMR) data to understand and improve care.
- 2) **Death Certificates** - Mike Dolan Fliss and Esther Chung of UNC Injury Prevention Research Center (UNC IPRC) conducted a linkage of NCCEH records (across 81 counties) with NC death certificates and compared mortality patterns between people experiencing homelessness and the general NC population. Findings showed people experiencing homelessness had higher all-cause and cause-specific mortality rates compared to the general NC population. Importantly, people experiencing homelessness had higher injury-specific causes of death, particularly with med-drug overdoses, firearm assaults, and suicides. They hope to move the project to DHHS DPH to sustain it annually.
- 3) **COVID-19 Vaccination** - Erika Ferguson of the DHHS Data Office leads an ongoing linkage of the NCCEH shelter records across three CoCs (Orange, Durham, Balance of State) to NC Immunization Registry (NCIR) vaccination records. Results suggested that despite significant challenges, homelessness service organizations were able to provide access to COVID-19 vaccination for hundreds of individuals experiencing homelessness. Still, the vaccination rates for this NCCEH population was lower than the vaccination rate for the general population.

Linkage Types & Key Concepts

Deterministic linkages typically require a key set of fields to exactly match between two datasets. In this way they provide high quality “true positive” linkages, but by their strictness, miss pairwise linkages that are close but not exact. **Probabilistic linkages** establish variable frequency relationships mathematically and model the likelihood that two records are the same person. Hybrid techniques include mixing of deterministic patterns and maximum variable distance filters (for instance, exact match on all elements, but age must be within a few years in both records) or multi-stage, hierarchical linkage that links exact matches first, removes them from linkage consideration, then relaxes linkage requirements on subsequent steps.

“Fuzzy matching” enables imperfect matching of commonly string variables to overcome small typos. As example, instead of two records names matching or not, they may measurably closer (John Smith vs. Jon Smiths) or further (John Smith vs. Stacy Jones). Soundex methods convert strings in English or other languages to approximate and standardized sound syllables, which can aid in linkage of names and places.

Depending on the project, conducting a **hand-review of the linked records** can help to identify ways to improve / tune the linkage. A hand-review of the linked records will also allow researchers to develop a quality threshold so that a final linked dataset would only include high-quality links.

Key linkage fields

For most linkage projects, key fields related to person-place-time are required for accurate linkage. Depending on the research question, some fields are more important than others. It is also important to harmonize linkage fields between datasets before conducting the linkage. For example, when linking on First Name, there should be a character variable for First Name in each dataset. Similarly, when linking on Race and Ethnicity, there should be a factor variable for Race and Ethnicity in each dataset. We provide some details on the linkage fields we used in our NCCEH and death certificate linkage below.

- Person
 - First name
 - Last name
 - Age
 - Gender
 - Race and ethnicity
 - Veteran status
- Place
 - City, zip code, and state of last known residence
 - Homelessness-related place of death.
 - In death records, we searched for locations in place of death for indications of homelessness, such as HMIS project addresses, sidewalk, abandoned buildings.
- Time
 - Among NCCEH records, dates of entry and exit from projects
 - We defined person-time experiencing homelessness as the entire period from the date of first entry into a project until the date of death or end of the study period

Special Linkage Considerations

When conducting linkages with other datasets, it is important to consider how certain populations may or may not be accurately linked. As example, in a linkage between HMIS data and death certificates, transgender individuals were found to be likely to have gender (lived gender and sex assigned at birth) and name mismatches. If that linkage required gender-related variables to exactly match (blocking) to speed up the linkage, it would have systematically left transgendered individuals out of the linkage.

It is important to tune linkages to ensure such populations are not misrepresented, especially those that may be already marginalized. By thinking critically about groups that may be disproportionately under-linked, teams can either redesign linkages to better serve those populations or be explicit about known linkage limitations.

Inclusion / Exclusion Criteria

HMIS includes data on non-housing-related support (supportive services only) and many housing-related events, including stays in short term shelters, emergency supportive housing, long term stay housing, and rent support. Studies will need to consider both inclusion and exclusion criteria of HMIS program types and make those filtering decisions either before the data pull is made, after the data pull is made, or a combination of both. This information is best captured in the *ProjectType* field in the Project table.

Future Research Opportunities

Linkage of HMIS data and health-related datasets at the state level can help to inform programmatic and policy decision-making. Some individual state datasets do not have homelessness indicators (e.g., emergency department data) and linkage to these datasets can be informative. We provide ideas for future linkage opportunities below.

Sustaining & Improving Past Linkages

Death Certificate Linkage

To improve upon our current linkage with NC death certificate data, future opportunities include:

- (1) A deeper investigation between experiences of homelessness and causes of death beyond the primary cause of death (ACME)
- (2) Literal text examination for themes and key notes
- (3) Utilize Social Security Numbers from death certificate data for improved linkage
- (4) Include deaths happening outside NC to PEH considered to live in NC
- (5) Attempt again to achieve a statewide HMIS data by approaching the other instances.

To our knowledge, as of 2022, DHHS is still maintaining the COVID-19 vaccine-related linkage to understand vaccine uptake in this important population. Improvements may be possible there, as well.

New State Datasets

A complete survey of health-related datasets at the state - from the perspective of their ability to hold and be linked to homelessness - is beyond the scope of this report. However, we offer a collection of example datasets below from our past research experience. Though we task a future project with a more comprehensive state survey, we particularly highlight datasets related to the disproportionate causes of death of people experiencing homelessness, including violent and overdose injury outcomes, alcohol and tobacco use, and chronic diseases. Some of these datasets may be able to scaffold off the pairwise linkages already complete; for example, VDRS and SUDORS can be linked to death certificates using unique IDs.

- (1) NC DPH IVPB Violent Death Reporting System (VDRS)
- (2) NC DPH IVPB State Unintentional Drug Overdose Reporting System (SUDORS) SUDORS
- (3) Birth records
- (4) NC Electronic Disease Surveillance System (EDSS) & NC COVID-19
- (5) NC Disease Event Tracking and Epidemiologic Collection Tool (DETECT), the state's syndromic surveillance system, particularly looking at the emergency department data
- (6) Hospital discharge data
- (7) Veterans' affairs system
- (8) Medicaid and Medicare claims
- (9) Legal encounter / incarceration records from DPS or DOJ
- (10) NC Clinical and Community Data Initiative (CODI) aims to create a system to link an individual's data throughout the community in a single record using privacy preserving record linkage. NCCEH is in ongoing discussions with NC CODI to provide data for linkage. Though NC DPH is not explicitly (to our knowledge) participating in NC CODI at this stage, researchers may be able to bridge this gap and request state or hospital datasets for linkage support.

A future task can include a literature review of major public health themes and linkage efforts in other states.

Other datasets

HMIS is not a comprehensive method for capturing homelessness as it mainly captures individuals experiencing sheltered homelessness. Efforts to partner with local, regional, or state housing partners should be explored. This may help with identifying or verifying homelessness status or history with other sources. Efforts to track homeless populations require collaborations with housing partners who do not receive federal funds to share data and conduct ongoing linkages. In the Los Angeles County DPH in California, this was particularly helpful during the COVID-19 pandemic in which homelessness status was captured using multiple data sources including HMIS, COVID-19 outbreak testing sites, and housing partners. Researchers found that homelessness status was ascertained mostly through non-HMIS data sources (CSTE 2022 presentation, yet to be released).

Example Research Questions

Individuals providing direct support to PEH have provided a few immediate themes for timely data work.

- **Heat related illness (HRI) and exposure-related injuries using NC DETECT emergency department visit data.** This may also include heat-related injuries like sunburns, but also burns from unprotected sterno heating devices during winter months (e.g., lighting tents or sleeping bags on fire). ED visit narratives could be used to build a keyword based definition to (anonymously) flag ED visits among people experiencing homelessness.
- Both **VDRS and SUDORS** have homelessness related variables, but these variables may be incomplete. A datasets specific study of these variables followed by "chain-linkage" to the NCCEH-death certificate linkage may provide new information on these deaths.
- Excessive alcohol use can lead to many kinds of death – not just direct causes of death, like drunk driving and alcohol-associated liver diseases, but fractions of cancer and other causes. CDC and research partners have released Alcohol-Related Disease Impact (ARDI) tools to count this fractional effect across many causes of death. Linked death certificate data could be passed through an ARDI analysis to understand the burden of excessive alcohol use among on PEH.
- Research elsewhere suggests **Housing First** strategies – the idea that providing stable housing, even before "solving" personal health challenges like addiction and substance use, is a high impact strategy - have health protective benefits. NC researchers may design studies to demonstrate this effect in support of policies that most effectively promote the health of PEH.

Updates, Feedback, & How to Help

As NCCEH is perpetually aiming to support improved data quality and collection in its member CoCs, and also to support the impactful, ethical use of its data for research that improves the lived experiences of people experiencing homelessness, we expect this document on using NCCEH data for research will need regular updates. If you have improvements or lessons learned in working with NCCEH data, please share those ideas with the NCCEH team you are communicating with on your research project.