

A horizontal decorative bar consisting of a blue gradient on the left and an orange segment on the right.

Wake County

Reporting Design Document

Version 1.0 07 July 2017

Contents

1	Executive Summary	1
2	Background and Challenges	5
3	Data Sources and Limitations	6
3.1	Wake County Jail Data	6
3.2	Homelessness Data.....	7
3.3	Emergency Medical System Data.....	8
3.4	Analytic Window	8
4	Entity Resolution	10
4.1	Individuating Process	10
4.2	Identifying Data Elements that Might Be Used for Matching.....	10
4.3	Standardization of Data across Sources	11
4.3.1	Gender/Sex	11
4.3.2	Race and Ethnicity.....	12
4.3.3	Social Security Number	14
4.3.4	Telephone Number	16
4.3.5	Date of Birth.....	22
4.3.6	Dummy Names and Name Standardization.....	22
4.3.7	Name Standardization and Name Match Codes.....	27
4.3.8	Address Data Quality and Standardization	28
4.3.9	Matching Rules.....	36
4.3.10	Rules for Matching within Jail Data	37
4.3.11	Rules for Matching within EMS Data	38
4.3.12	Rules for Matching within HMIS Data.....	39
4.3.13	Rules for Matching across All Three Sources.....	40
4.3.14	Individuating Surrogate Keys	41
5	Wake County: A Demographic Profile	42
6	Agency Data	2
6.1	Wake County Jail.....	2
6.2	Homeless Management Information System (HMIS)	5
6.3	EMS Interaction Data	11
7	Intersection of Jail, EMS, and Homelessness	16
8	Timeline of Events: A Case Study	21
9	Potential Interventions	26
10	Next Steps	30
Appendix A:	Terms, Definitions, and Acronyms	32
Appendix B:	Data Dictionary	33
Appendix C:	Wake Jail Medical Screening Questions	35

Information about This Document

Document Control.....	38
-----------------------	----

Contacts..... 38
 Revision History..... 38

Figures

Figure 1: Iterative, Data-Supported Decisions..... 2
 Figure 2: Wake County Jail Booking Process..... 7
 Figure 3: Analytic Study Period across the Three Data Sources 9
 Figure 4: Person Count of Homeless Program Participation by Gender and Age..... 8
 Figure 5: Service Count of Homeless Program Participation by Gender and Age 8
 Figure 6: Individual Counts of EMS Interaction by Age and Gender..... 12
 Figure 7: Definitions of EMS Utilization 13
 Figure 8: High EMS Utilizers EMS Interaction by Age and Gender 15
 Figure 9: Interaction of Wake County Jail System, EMS, and HMIS Homeless data..... 16
 Figure 10: Distribution by Age and Sex of Intersecting Population 17
 Figure 11: Demographic Distribution of Interacting Population by Age, Sex, and Race 17
 Figure 12: UCR Categories as a Percent of Bookings for the Intersecting Population, Jail Familiar Faces,
 and Jail Population..... 19
 Figure 13: Iterative, Data-Supported Decisions..... 26

Tables

Table 1: Data Elements for Entity Matching by Data Source..... 10
 Table 2: Source ID by Data Source 11
 Table 3: Sex Codes in the Jail File..... 12
 Table 4: Gender Codes in the EMS File 12
 Table 5: Gender Codes in the HMIS File 12
 Table 6: Race Codes in the Jail File..... 12
 Table 7: Race Codes in the EMS File 13
 Table 8: Race Codes in the HMIS File..... 13
 Table 9: Ethnicity Codes in the Jail File 14
 Table 10: Ethnicity Codes in the EMS File 14
 Table 11: Ethnicity Codes in the HMIS File 14
 Table 12: SSN Configurations and Standardization in the Jail File..... 14
 Table 13: SSN Configurations and Standardization in the EMS File..... 15
 Table 14: SSN Configurations and Standardization in the HMIS File 16
 Table 15: Telephone Number Standardization in the Jail File 16
 Table 16: Telephone Number Standardization in the EMS File 18
 Table 17: Telephone Number Standardization in the HMIS File 19
 Table 18: Dummy or Test Names removed from the EMS File..... 23
 Table 19: Name Standardization Examples 27
 Table 20: Last Name Match Code Handling Spelling and Data Entry Differences 27
 Table 21: First Name Match Codes for Typos, Spelling Variations, and Nicknames..... 27
 Table 22: Jail Addresses with No Street Number..... 29
 Table 23: Dummy and High Frequency Addresses from the Jail File..... 31
 Table 24: EMS Addresses with No Street Numbers 32
 Table 25: EMS Dummy and High Frequency Addresses 34
 Table 26: Address Standardization and Match Code Examples..... 36
 Table 27: Examples of Name and SSN Match Criteria and Chaining 37

Table 28: U.S. Census Bureau Statistics for Wake County, NC ^a	42
Table 29: Living Wage Calculations for Wage County, North Carolina ^a	44
Table 30: Annual Expenses Used to Calculate the Living Wage Estimate for Wake County	1
Table 31: Agency Definitions of High System Utilizers	2
Table 32: Distribution of Booking Counts	3
Table 33: Demographic Profile of Wake County Jail Population and Familiar Faces.....	3
Table 34: Count, Percent, and Relative Rank of UCR Categories by Wake County Jail Population and Familiar Faces	5
Table 35: Homeless Program Participation by Service Count and Number of Days.....	6
Table 36: Percentile Distribution of HMIS Homeless Program Service Counts and Cumulative Days of Participation	9
Table 37: Demographic Profile of HMIS Homeless Population and High Utilizers	10
Table 38: Interactions with Homeless Programs among High Utilization Groups.....	11
Table 39: Distribution of EMS Utilization Counts.....	13
Table 40: Demographic Profile of EMS Population and High Utilizers	14
Table 41: Booking Length of Stay in Wake County Jail (in Days)	20
Table 42: Intersecting Populations' Homeless Program Participation by Service Count and Number of Days	20
Table 43: Person Count by Jail and EMS Incident Combinations.....	21
Table 44: Timeline of Jail, EMS, and HMIS Events for Familiar Face Male between Ages of 56-65	24
Table 45: Mental Illness Screening Questions (1).....	35
Table 46: Mental Illness Screening Questions (2).....	37

1 Executive Summary

1.1 Background

Like many counties and municipalities across the country, Wake County faces the challenge of managing a population of individuals with disproportionately high utilization of county emergency medical, homeless, and jail services. The recurring interactions with various county systems are costly and, perhaps more importantly, overlook key needs of the individual and may not result in long-term, sustainable, and positive outcomes for the individuals involved. Community stakeholders have tried to identify those most vulnerable in the community and meet their needs, but currently use disparate and unconnected information systems.

Wake County seeks to leverage its various data systems to understand the characteristics and utilization of its most frequent users, or “familiar faces” of these systems in an effort to break the cycle of recidivism and provide more cost-effective services and interventions. Wake County partnered with SAS to develop insights into the data to better understand the “familiar faces” population and best deliver services to them.

SAS has extensive experience in developing enterprise level analysis to integrate data from disparate source systems and build insights from the cross-functional perspective of the data. This report will bring together three data systems – **jail, emergency medical services, and homeless services** – to provide a more complete picture of the “familiar faces” and understanding about the patterns of unproductive and unhealthy behaviors.

Leveraging the data and insights provided in the report, Wake County intends to ensure that the right services are available for the right individuals, proactively targeting programs such as long-term subsidized permanent housing, coordinated services and support, intervention, and diversion services for at-risk individuals in an effort to break the cycle and improve outcomes. Through proactive intervention and supportive services, the County can save taxpayer dollars by reducing jail incarceration and frequent, possibly avoidable, visits to local emergency departments. More importantly, perhaps, coordination of care and services may improve stability and self-sufficiency, providing a greater quality of life for these individuals.

1.2 Summary Findings and Interventions

Wake County is a rapidly growing geographic area with significant migration of people moving to the area both from within and outside of the State of North Carolina. Wake County is home for a highly educated and increasingly youthful population, an economy driven by technical, healthcare, education and financial organizations, and a relatively strong housing market with increasing average housing prices. As Wake County grows and disparity gaps in employment, income, and housing widen, the population of those needing supportive government services puts a greater strain on public systems.

SAS integrated and analyzed key data sources from jail incarceration records, Emergency Medical System (EMS) transport records, and Homeless Management Information System (HMIS) records for the county, to provide data driven insights to identify the population that frequently interacts with the police, jail, and health and social service systems. The [Findings](#) section of this report provides understanding of the high utilizer or “familiar faces” population for each of the individual data source systems.

SAS also evaluated the intersecting population between the HMIS, Jail, and EMS systems. **807** individuals were identified with at least one incident in all three systems. The intersecting population are 26 - 55 years old (70%), predominately male (75%), and disproportionately Black or African American men (46%).

Other insights provide better understanding of why certain individuals are disproportionately utilizing services. More than 70% of jail bookings for this intersecting group are misdemeanor level charges, often with charges likely related to homelessness, mental health or substance abuse (*trespassing, city ordinance violations, disorderly conduct*), or that appear to be technical violations or issues with probation or court requirements from prior criminal activity (*contempt of court, perjury, or court violation*). The resulting jail stays appear to be longer – 18 days on average – for the intersecting population compared to the jail’s population of familiar faces (11 days).

The intersecting population demonstrates unsettled living conditions as the majority of that group (more than 85%) had some interaction with the emergency shelter program, which tends to offer short-term assistance. This population showed far less participation in other housing programs that offer longer-term support which may reflect individuals with a criminal past who are prevented from or who choose to not access programs that offer longer-term support.

While the intersecting population data did not find key factors that differed from the high utilizers of EMS services, discussion of the findings with key stakeholders highlighted several insights. One can infer that individuals who have interaction with HMIS/Jail but not EMS may not have an acute or chronic health issue and conversely those who do have EMS interactions likely do have underlying health problems that may need to be considered in a coordinated case management plan.

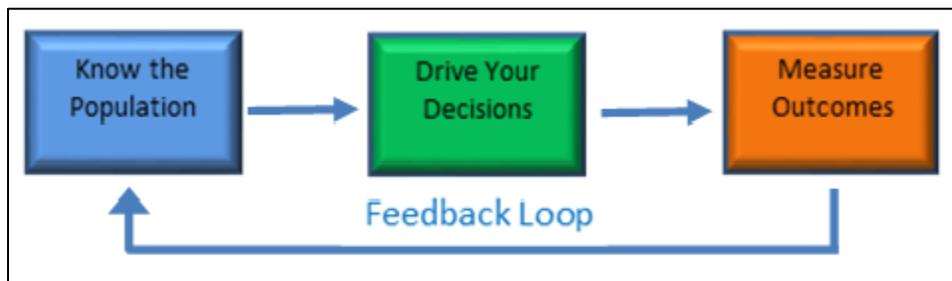
To identify an initial population of high utilizers across all three services, the study evaluated high incidents of Jail and EMS interactions (the 95th percentile) with associated high HMIS utilization. **Twenty-six** of the 807 have 5 or more Jail and EMS incidents. A case study was performed for a male from this population, with a total of 47 total interactions with the three agencies, to understand the recurring interaction with the jail system, chronic use of EMS services, and interaction with emergency shelters. For more detail on the case study, see section [8: Timeline of Events: A Case Study](#).

1.3 Recommendations

Wake County wants to bring together service organizations and systems that are currently challenged in sharing information and build a collaborative and coordinate approach to provide needed and necessary services to reduce costs and improve opportunities for stability and sustainability for Wake County’s most at-risk population.

To build this collaborative approach, Wake County needs to follow an iterative, data-supported decisions outlined in Figure 1 geared toward improving service outcomes:

Figure 1: Iterative, Data-Supported Decisions



This study laid the foundation for Knowing the Population so that Wake County can take proactive steps to target coordinated services to the individuals who are challenged with a variety of issues that impact their self-sufficiency and quality of life. Recommended interventions focus on ways to Drive Decisions and can be found in section [9: Potential Interventions](#).

Recommended interventions include:

- Who is at most risk for being or becoming a high-risk utilizer of costly county services?
 - Investigate the high utilizer population identified by jail id for potential coordinated services.
 - Develop case analysis of the **26** highest utilizers to understand their needs and the strategies required to reach and impact this population.
 - Pursue additional data to be collected and analyzed to enhance the understanding of the at-risk population.
 - Expand the scope of analysis for high utilizers. The County may want to consider expanding the analysis to consider family and intergenerational relationships and how these factor impact the reasons and patterns for a person's interactions with county systems.
- When can intervention result in better outcomes? (The sooner the better!)
 - Pursue coordinated support services and collaborative efforts with the court system to address interactions with jail that often begin with low level misdemeanor charges but overtime result in increasing occurrences of failure to appear, probation and parole violations, longer jail stays, and higher costs.
 - Pursue additional analysis into key population segments, such as young adult men between the ages of 19 and 25, whose use of emergency shelter far exceeds that of other programs, to understand potential intervention points that can reduce the frequency and cost of future interactions.
 - Encourage additional sharing of health information, including mental health information, in order to better target wraparound and case-management services and reduce costly ED and jail utilization.
 - Expand analytic data sources to gain further insight into key events that start cyclical high utilization.
- Where are the County's needs and resources?
 - Increase data collection requirements and incorporate additional data sources to enable reliable and up-to-date analysis and mapping of incidents, population needs, and service availability.
- How does the County use these insights to reduce recidivism, reduce costs associated with jail and EMS interactions, increase housing stability, and monitor and measure improvements in long-term outcomes?
 - Analyze dollar costs associated with the various services included in the current data sources – cost for a stay in jail, cost for an EMS interaction, cost for a stay in an emergency shelter. By approximating these costs, analytics can apply them to the high utilizer population, as well as individual subsets populations (EMS-HMIS, HMIS-Jail, Jail-EMS), providing Wake County with insights in the costs and savings of proactive supportive services versus reactive, cyclical utilization.

Next Steps:

To ensure that Wake County can meet its goals of 1) reducing recidivism and improving outcomes for the high utilizer population, and 2) being able to monitor and measure outcomes, the following steps are recommended:

1. Expand analytics to enhance insights by acquiring additional data sources as well as longer historical information. Key data sources would expand the accuracy of the high utilizer definitions, enhance understanding and management of the high utilizer population, and ensure the ability to assess the impact and outcomes of new programs such as supportive housing and wrap-around services. Key additional data sources are needed to confirm anecdotal evidence related to mental health, substance abuse, and other health related conditions.
2. Develop an expanded cross-sector data system that provides comprehensive, entity resolved, person-centric data for individuals who interact with one or more of the Wake County stakeholder systems of service to serve the purpose of coordinated case management, program analysis, and population research and understanding.
3. Convene a stakeholder community summit to review the findings associated with this study and to determine next steps for cross-system of service collaboration to meet the needs of the high utilizer population.

2 Background and Challenges

Wake County and its key stakeholders have long been at work to identify the most at-risk population in the community who interact with the jail, emergency medical system, and homeless services on a regular, recurring basis. These interactions with various county systems are costly and, perhaps more importantly, overlook key needs of the individual and may not result in long-term, sustainable, and positive outcomes for the individuals involved. The County's objectives in this effort are to:

- Develop a framework for a multi-sector data exchange for systems of service.
- Understand the characteristics and utilization patterns for the most frequent users of costly county systems of service.
- Inform the respective systems of services of “familiar faces.”
- Initiate interventions to break the cycle of recidivism.

To establish a data-driven approach to understanding the “familiar faces” population and enable proactive management of coordinated and targeted interventions, the analytic exercise will:

- Assess and report on data quality, content, and standardization for each of the Phase 1 data systems.
- Demonstrate the capability to match data across systems of service.
- Create a baseline profile of “familiar faces” for each system of service and across the combined population in comparison with a demographic profile of Wake County.
- Identify patterns of behavior among the “familiar faces.”

3 Data Sources and Limitations

Data points from each of the agency data sources are outlined in [Appendix B: Data Dictionary](#). In cases where the data elements were not used for this report, a reason for the omission is noted.

3.1 Wake County Jail Data

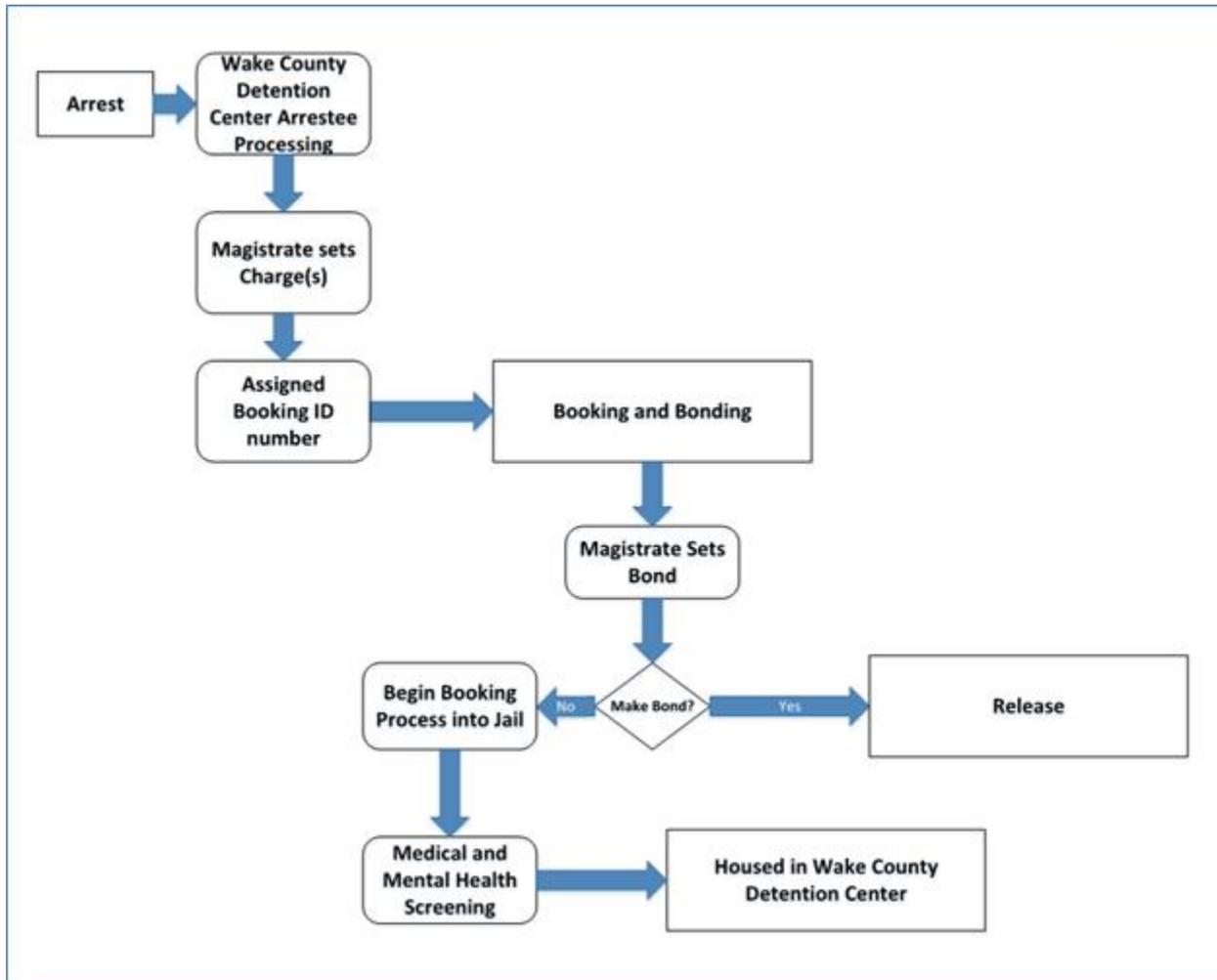
Jail data comes from the Wake County Sheriff's Office. The dataset covers jail bookings occurring between July 2013 and December 2016 and includes 61,365 distinct individuals as per the local name ID and 272,592 total bookings. In addition to name, address information, phone numbers, birthdate, and social security number (used for the purposes of resolving identities both within and across datasets), the data include basic demographics of age, race, and sex as well as booking details such as length of stay, arresting charge(s), bond status, and general release information. It is important to note that the term "booking" in this report includes all individuals who enter the facility through arrest processing outlined in Figure 2. Each "booking" number indicates an arrestee's entrance into jail. Once an arrestee is processed through the Raleigh/Wake City-County Bureau of Identification (CCBI) and seen by the magistrate to determine bond, custody is now with the Wake County Sheriff's Office. After the exchange of custody, the inmate has up to four-six hours to bond out before they are prepared to be housed.

The Sheriff's Office also collects medical and mental health screening information on individual detainees, which occurs once individuals are ready to be housed in jail as shown in Figure 2. The data from the assessments includes the question asked (see [Table 45](#) and [Table 46](#) in [Appendix C](#): for examples of the screening questions), Yes/No response, follow-up comments to questions, and the administration date and time.

While the screening data is particularly rich, there are several reasons why this data was not leveraged for this report. First, screening responses are self-reported by detainees, and cannot be fully substantiated. For example, an inmate's response to the question, *Are you homeless?*, is not verifiable in most cases and, in particular, for responses like "kinda-couch hopping around," which would go undetected by any third-party data source tracking formal homeless supportive services (i.e., emergency shelters or long-term supportive housing). Second, it's unclear whether assessment administrators take any liberties while recording responses. For instance, it's not uncommon to see conflicting information between a detainee's response (Yes or No) and the screener's comments. In a review of the *Are you homeless* question, there are several occurrences where the response is "N" (No) and the corresponding comment states otherwise; comments include: *Yes; Sleeps in car; Sleeps outside down town; Homeless*; and so forth. The screening data does not indicate who administered the assessment or whether the screener has some prior knowledge with which to override a detainee's response. Such conflicting information, found in both screening questionnaires, makes discerning response accuracy difficult.

In addition to issues of response validity, a final reason for leaving out the screening data is due to the proportion of the inmate population issued the assessment. Matching up the jail dataset with the screening assessment file produced only a 60% match. That means that nearly 40% of the jail dataset was not accompanied by a medical and/or mental illness screening. While it is understood that the gap in available data is primarily a result of posting bond before the screening could be conducted (see Figure 2), the fact that a sizable proportion of the jail population lacks screening data hinders any generalizations that might be made from the assessment data. Consequently, a decision was made to exclude the both medical and mental health screening data from the current report.

Figure 2: Wake County Jail Booking Process



3.2 Homelessness Data

Homelessness is defined by the Housing and Urban Development (HUD) as 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. Homeless Management Information System (HMIS) is a local information technology system used to collect client-level and data on the provision of housing and services to homeless individuals and families and persons at risk of homelessness.

The HMIS data received for Wake County covers a 20-month period, May 2015 through December 2016 based on exit dates, and includes 10,141 individuals as per the HMIS client ID and 80,125 instances of homelessness program services. The following data elements were exported from HMIS and included: housing program provider, type of housing, program entry and exit(s), past history of homelessness that include prior living situation, length of stay, and the number of times homeless over last 3 years. As with the Jail and EMS datasets, recipient information includes first and last name, date of birth, social security number, age, sex, and race and ethnicity. While the HMIS data captures address and telephone information, the data is not as well populated, which is logical given the housing fluidity of this population. Additional recipient characteristics include veteran status, family status, healthcare insurance status and source, employment status, and any Federal assistance programs such as the

Supplemental Nutrition Assistance Program (Food Stamps) or other Temporary Assistance for Needy Families (TANF) program.

Specifically, HMIS is used to produce an unduplicated count of homeless persons and understand patterns of service use. We would be remiss, however, if we failed to acknowledge that this data only tracks those who formally received housing provisions. Often times demand exceeds the supply, leaving many who are homeless unable to secure temporary and/or permanent housing assistance. Moreover, those at risk of homelessness who rely on informal supports, such as sleeping on a friend's couch, or those who take refuge in their car are likely to be missed in this data. It is important to note that homelessness runs along a continuum and that the HMIS data only captures a portion of the homeless population.

For the purposes of this report, HMIS data only covers a 20-month period, it is possible that some recipients of HUD services, while interacting with the Jail and EMS systems, were not captured in the HMIS data during the study period. For the purposes of identifying individuals at-risk or in a state of homelessness, future work may consider additional data sources for identifying homeless status as well as expanding the period of the HMIS data.¹

3.3 Emergency Medical System Data

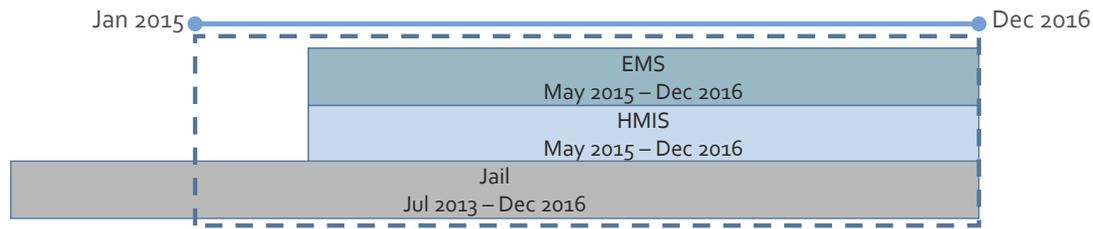
Usage records of emergency services come from the Wake County Emergency Medical System (EMS). The EMS data covers the same span of time as the HMIS data, May 2015 through December 2016, and covers 170,634 transports during the 20-month period. Data elements include date of the incident, disposition (e.g., Assist; No treatment, No transport; Transported No lights/Siren; etc.), GPS location of the scene, and when pertinent, the transported to destination. Patient information includes first and last name, date of birth, social security number, home address and phone, age, sex, and race and ethnicity.

It is important to note that the data received from Wake County EMS only tracks the immediate details of a patient at the scene (e.g., name, date of birth, address, etc.) and records a discrete outcome and, if transported, a destination. The dataset does not specify the reason for the EMS response, state of patient during the event, treatment provided, or any additional details following the patient's transfer to the transport destination. Undoubtedly, future analyses into the well-being of the high-utilizer population would be interested in the acute reasons for EMS services, and in particular, treatment following patients' transport to the hospital.

3.4 Analytic Window

The study period of this report is dependent on the alignment of dates in the three data sources. [Figure 3](#) shows the analytic window selected covers a 24-month period, with an additional 4 months of data being drawn on from the Jail data compared to the 20-months of data in the EMS and HMIS data respectively. For the purpose of account for data ranges pulled for each dataset, the Jail data was pulled based on the booking date, EMS on the incident date, and HMIS data on clients' exit date from the housing program.

¹ Due to reasons of privacy and client consent, the HMIS data cannot be obtain further back than May 1, 2015.

Figure 3: Analytic Study Period across the Three Data Sources

Given the modest period of time for following individuals' interactions with Jail, EMS, and homeless support systems, there are likely to be incidents within each system that we are unable to detect at this time. In fact, the cross-sectional nature of each data source offers a snapshot of each population rather than a carefully drawn representative sample of the population that might be tracked over time. As such, there will be 'event gaps' both within the scope of the data being analyzed and for events that follow these populations as they are dispatched to hospitals following an EMS transport or transfer out of the Wake County Jail system and into the Federal system. The goal of identifying these data shortfalls is to assist Wake County and future phases of this work to better understand its vulnerable populations.

4 Entity Resolution

A prerequisite for evaluating how segments of the population cycle through local jails, emergency medical services, housing and services to the homeless. Bringing these disparate datasets together requires finding records that refer to the same entity both within and across each data source, otherwise known as *entity resolution*. Entity resolution is necessary due to differences in how content is recorded as well as data completeness. For instance, while the Jail, EMS, and homeless data include a common shared identifier of social security number, given the circumstances surrounding this at-risk population (such as being homeless) or the conditions under which data is collected (medical emergency), social security numbers and other identifying information like names and addresses are subject to being recorded incorrectly or not at all. Potential inconsistencies for how data is collected not only across each system but within the same agency, requires entity resolution both within each data source as well as between. The entity resolution technique used to link records together in the Jail, EMS, and homeless data is called *individuating*.

Note: Entity resolution was not limited to the analytic window. The identification of common entities, both within and between data sources, was pursued irrespective of dates.

4.1 Individuating Process

Individuating is the process of logically grouping records from different sources, or even within the same source if there are multiple records and IDs for an individual, to build a more complete picture of an individual.

- The underlying records within a conglomeration are not physically merged, but they are only logically combined by assigning them the same individuating surrogate key value.

The goal is to identify element-based links between data sources. An entire conglomeration need not have all of the same elements in common. Rather, seek to identify a link between two sources, which might then link to another record in that source, or from another source.

4.2 Identifying Data Elements that Might Be Used for Matching

First one must identify the data elements that could be used to distinctly identify a person in each data source (See [Table 1](#)). This will allow us to define individuating rules using data elements found in all of the data sources. For example, because Middle Name is found only in the Jail data it is not a candidate for a matching rule between the data sources. Additionally, we see in the Jail data the street number is split from the street where the street and number and together in the other data sources so we would need to concatenate those in Jail. In the HMIS data we see we have address1 and address2 lines that would need to be standardized as well as a second set of address and phone information for some client ids.

Table 1: Data Elements for Entity Matching by Data Source

Data Source	Potential Identifying Data Elements
Jail	name_id, lastname, firstname, middlename, dob, race, sex, ethnic, ssn, streetnbr, street, city, state, hphone, mphone, wphone
HMIS	HMIS_client_ID, HMIS_fname, HMIS_lname, HMIS_DOB, HMIS_SSN, HMIS_race,

	HMIS_ethnicity, HMIS_gender, address1, address2, city, state, home_phone_contact_number, client_street_address, client_city, client_state, home_phone
EMS	PatientFirstName, PatientLastName, PatientDOB, SocialSecurityNumber, HomeAddress, HomeCity, HomeState, HomeZipCode, PhoneHome

The entire process of individuating hinges on the correct definition of the *source ID*, which is the level at which records from each source are identified as being the *same person*. The source ID can be defined using one or more fields in the source. It may or may not be the same as the primary key. If the primary key is defined at the level of the individual, then the *source ID* probably is the primary key. However, if the primary key is defined at an event level, then the *source ID* is probably a subset of the fields being used to define the primary key (leaving out the date or sequencing information used to define each event). Even when there is a primary key at the individual level checks should be done to ensure that this key does indeed identify only a single unique individual and that an individual does not have more than one key. For example, are parent and child given the same primary key or when an individual's name or address changes are they given a new primary key? [Table 2](#) shows the initial source ID variable(s) identified in each data sources.

Table 2: Source ID by Data Source

Data Source	Initial Source ID
Jail	Name ID
HMIS	Client ID
EMS	Data is event level – need to use Name, DOB, Gender and other attributes to identify an individual

4.3 Standardization of Data across Sources

As noted above, some alignment of data fields was required to match individuals across the data sources. These include gender/sex, race and ethnicity, social security number, telephone number, date of birth, names, and address. Additionally, we need to look at values within columns to make sure they align.

4.3.1 Gender/Sex

In the Jail data, for example, gender was coded as 'M' and 'F' but in the other data sources it was 'Male' and 'Female'. [Table 3](#) highlight the differences and standardization of gender codes across the Jail, EMS, and HMIS datasets. We also standardized names and addresses and created match codes so that the matching is less sensitive to small data entry differences.

Table 3: Sex Codes in the Jail File

Sex	Name ID Count	Booking Count	Standard Gender
F	16,637	26,105	Male
M	44,516	80,044	Female
U	1	2	U

Table 4: Gender Codes in the EMS File

Gender	Incident Count
Female	87,607
Male	71,961
Not Reported	6,244

Table 5: Gender Codes in the HMIS File

Gender	Client ID Count	Service (EE ID) Count
Male	5,750	70,611
Female	4,331	9,277
Transgender male to female	44	50
	10	161
Data not collected	3	10
Transgender female to male	2	9
Doesn't identify as male, female, or transgender	2	8

4.3.2 Race and Ethnicity

The jail data categorized Asian and Pacific Islander Race 'A', so in EMS and HMIS these two races were combined into 'Asian or Pacific Islander' in the standard race column to be consistent. [Table 6](#) shows the differences and standardization of race and ethnicity codes across the Jail, EMS, and HMIS datasets.

Table 6: Race Codes in the Jail File

Race	Name ID Count	Booking Count	Standard Race
A	269	326	Asian or Pacific Islander
B	31,759	61,334	Black or African American
H	162	232	H
I	45	50	American Indian or Alaska Native
O	65	109	O

Race	Name ID Count	Booking Count	Standard Race
U	10	13	Unknown
W	28,844	44,087	White

Table 7: Race Codes in the EMS File

Race	Incident Count	Standard Race
White	91,111	White
Black or African American	56,231	Black or African American
	8,347	
Other Race	8,084	Other Race
Asian	1,512	Asian or Pacific Islander
Unknown	385	Unknown
American Indian or Alaska Native	164	American Indian or Alaska Native
Native Hawaiian or Other Pacific Islander	161	Asian or Pacific Islander

Table 8: Race Codes in the HMIS File

Race	Client ID Count	Service (EE ID) Count	Standard Race
Black or African American (HUD)	7,199	60,803	Black or African American
White (HUD)	2,628	17,464	White
American Indian or Alaska Native (HUD)	87	844	American Indian or Alaska Native
Native Hawaiian or Other Pacific Islander (HUD)	35	293	Asian or Pacific Islander
Client doesn't know (HUD)	44	266	Client doesn't know
Asian (HUD)	45	203	Asian or Pacific Islander
Client refused (HUD)	33	71	Client refused
Other Multi-Racial	2	66	Other Multi-Racial
Data not collected (HUD)	22	64	Data not collected
	47	52	

Table 9: Ethnicity Codes in the Jail File

Ethnic	Name ID Count	Booking Count	Standard Ethnicity
	22,015	28,348	
H	4,010	5,860	Hispanic/Latino
N	35,110	71,912	Non-Hispanic/Non-Latino
U	19	31	U

Table 10: Ethnicity Codes in the EMS File

Ethnicity	Incident Count	Standard Ethnicity
Not Hispanic or Latino	150,702	Non-Hispanic/Non-Latino
Hispanic or Latino	7,686	Hispanic/Latino
	7,559	

Table 11: Ethnicity Codes in the HMIS File

Ethnicity	Client ID Count	Service (EE ID) Count	Standard Ethnicity
Non-Hispanic/Non-Latino (HUD)	9,547	75,750	Non-Hispanic/Non-Latino
Hispanic/Latino (HUD)	447	4,168	Hispanic/Latino
Client refused (HUD)	31	61	Client refused
	51	58	
Data not collected (HUD)	38	52	Data not collected
Client doesn't know (HUD)	28	37	Client doesn't know

4.3.3 Social Security Number

In the HMIS data, the social security numbers (SSN) had hyphens, but the Jail and EMS data did not. Additionally sometimes the last 4 digits of the SSN was provided on its own, but other times it was preceded by 5 zeros or *****. To standardize SSN we removed the hyphens and spaces and when the only digits were the last 4 digits of the SSN we proceeded those digits with xxxxx. Any SSN that were all repeating digits or not enough digits were left blank in the standardized SSN column and not used for matching. Examples of the SSN configurations and standardization in the 3 source files are shown in [Table 12](#).

Table 12: SSN Configurations and Standardization in the Jail File

SSN Pattern	Name ID Count	Booking Count	Standard SSN
#####	51,977	94,554	#####
	8,827	10,987	

SSN Pattern	Name ID Count	Booking Count	Standard SSN
####	117	167	xxxxx####
#####	95	118	#####
00000####	69	95	xxxxx####
000000000	47	61	
#####	18	26	#####
*****####	10	14	xxxxx####
#####	2	6	
###	2	5	
## #####	1	1	#####
0000	1	1	
00000	1	1	
00000000	1	1	
#	2	2	

Table 13: SSN Configurations and Standardization in the EMS File

SSN Pattern	Row Count	Standard SSN
#####	85,510	#####
	84,847	
999999999	147	
00000####	83	xxxxx####
000000000	28	
123456789	7	
#####	4	#####
111111111	2	
####	1	xxxxx####
#####	1	
#####	1	#####
555555555	1	
777777777	1	
987654321	1	

Table 14: SSN Configurations and Standardization in the HMIS File

SSN Pattern	Client ID Count	Service (EE ID) Count	Standard SSN
###-##-####	9,025	76,120	#####
	864	3,128	
--####	217	766	xxxxx####
000-00-####	18	93	xxxxx####
000-00-0000	8	8	
-##-####	3	3	xxxxx####
###-##-	1	2	
###--	2	2	
###--####	2	2	xxxxx####
-##-	1	1	
111-11-1111	1	1	

4.3.4 Telephone Number

For telephone number standardization (see [Table 15](#)), we strip out the characters, spaces and punctuation and then only populate the standardized phone number field where the remaining number is 7 or 10 digits. When the remaining number is only 7 digits the area code 'xxx' is prepended.

Table 15: Telephone Number Standardization in the Jail File

Phone Pattern	Name ID Count	Booking Count	Standard Phone
#####	52,566	94,338	#####
	45,064	65,984	
###	23,927	54,276	
### #####	322	483	
9190000000	44	125	
#####	52	93	
#####	31	79	xxx#####
##	17	39	
### #####	18	30	
### #####	15	26	#####
9999999999	13	26	
0000000000	6	22	
9	15	22	

Phone Pattern	Name ID Count	Booking Count	Standard Phone
#####	15	21	
9199999999	4	17	
### ####	9	13	xxx#####
####	4	12	
#####	5	11	
#####	7	10	
0	2	8	
#####	1	6	
### ###	1	4	
### Ne	1	3	
### #	2	3	
#####	2	3	
### #####	1	2	
#### #	1	2	
#####	2	2	
#####?	1	2	#####
1	1	2	
3	1	2	
### #	1	1	
### ##	1	1	
### ###	1	1	
### #####	1	1	#####
### #####	1	1	#####
### # #####	1	1	
### ###	1	1	
#####	1	1	
00000000	1	1	
111111111	1	1	
555888888	1	1	
919111111	1	1	
919333333	1	1	

Phone Pattern	Name ID Count	Booking Count	Standard Phone
9195555555	1	1	

Table 16: Telephone Number Standardization in the EMS File

Phone Pattern	Row Count	Standard Phone
(###)###-####	99,640	#####
	54,260	
#####	14,575	#####
() -	1,581	
(919)999-9999	266	
(999)999-9999	126	
(000)000-0000	46	
9199999999	23	
	20	
#####	19	#####
9999999999	15	
(919)000-0000	9	
(###)###-###_	8	
(###) -	8	
0000000000	5	
#####	4	
#####	4	
#####	2	
(919)999-999_	2	
5555555555	2	
9190000000	2	
#####	1	
#####	1	
#####	1	
#####	1	
(###)###-_____	1	
(###)888-8888	1	
(555)555-5555	1	

Phone Pattern	Row Count	Standard Phone
(919)111-1111	1	
0005555555	1	
1111111111	1	
1234567890	1	
2222222222	1	
5554443333	1	
9	1	
9191111111	1	
9195555555	1	
9999999999_____	1	

Table 17: Telephone Number Standardization in the HMIS File

Phone Pattern	Client ID Count	Service (EE ID) Count	Standard Phone
	10,121	79,925	
###-###-####	2,281	28,100	#####
none	638	10,853	
#####	291	1,957	#####
### ###-####	224	4,007	#####
none reported	182	4,562	
###-####	79	231	xxx#####
None Reported	70	1,639	
(###) ###-####	64	3,217	#####
n/a	61	876	
None	53	723	
N/A	52	1,232	
###.###.####	43	147	#####
None reported	34	789	
###-#####	21	174	#####
### ### ####	18	159	#####
(###)###-####	16	274	#####
#-###-###-####	15	45	
NONE	14	291	

Phone Pattern	Client ID Count	Service (EE ID) Count	Standard Phone
#####-####	12	19	
###/###/####	6	7	#####
###-###-####-cell	5	156	#####
none to report	5	74	
na	4	129	
### #####	4	39	#####
#####	4	10	xxx#####
Not reported	3	16	
###-###-####	3	15	#####
NONE REPORTED	2	41	
###-###-####*	2	10	#####
###-###-####, ###-###-####	2	4	
none given	2	4	
###/###-####	2	3	#####
###-###-	2	2	
no	2	2	
NA/	1	270	
#####-#### or (###)###-####	1	143	
None to report	1	99	
#####	1	80	
none report	1	69	
VM ###-###-####	1	55	#####
None to Report	1	49	
none reported	1	48	
none provided	1	46	
###-###-####	1	38	
###-###-####	1	31	#####
###	1	24	
###-###-####	1	23	#####
(###-###-####	1	22	#####
None Provided	1	18	

Phone Pattern	Client ID Count	Service (EE ID) Count	Standard Phone
###-###-##-####	1	16	
### ###-####/guardian #	1	11	#####
Nono	1	8	
###.###.#### x-###	1	7	
###-###-#### (work)	1	6	#####
#####	1	5	#####
(###) ###-###-####	1	5	
### ###-####-case manager	1	4	#####
#####	1	4	
###-	1	4	
###-###-####	1	4	#####
###-###-####/<Some Name>	1	4	#####
No	1	4	
###-####-####	1	3	
###-###-#### (H); ###-###-#### (C)	1	3	
###-###-#### text preferred	1	3	#####
###-###-####-	1	3	#####
###-###-####/###-###-####	1	3	
(###) #####	1	3	#####
None	1	3	
#####	1	2	
###-#####	1	2	
###-###-#### or	1	2	#####
###-###-#### or ###-###-####	1	2	
###-###-####(<Some Name>) mom	1	2	#####
###-###-####(<Some Name>)	1	2	#####
###.###.####- cell	1	2	#####
###/	1	2	
NA	1	2	
no phone as of #/##/##	1	2	
None	1	2	

Phone Pattern	Client ID Count	Service (EE ID) Count	Standard Phone
### ###-####-sister	1	1	#####
### ###--####	1	1	#####
#####/#####	1	1	
#####-####	1	1	
###-#####	1	1	
###-###-####-##/##/##	1	1	
###-###-####-lost phone	1	1	#####
###-###-####/###-###-####	1	1	
###-###-####/sister <Some Name>	1	1	#####
###-###-####;###-###-####	1	1	
###-###- #####	1	1	#####
###-##-####	1	1	
###/###/###	1	1	
###=#####	1	1	
(###-###-####)	1	1	#####
`###-###-####	1	1	#####
no phone number given	1	1	

4.3.5 Date of Birth

There appear to be no consistent dummy DOB in any of the three sources; that is we did not find a large number of entries in any of the three data sources that all had the same DOB. Instead it appears that DOB is left blank or null when not known. EMS has 4,019 rows with missing DOB and HMIS has 55 client_id with missing DOB. Jail had no missing DOB values.

4.3.6 Dummy Names and Name Standardization

The data may contain test rows, rows where the individual’s name is not known and other rows that should not be used for matching. To identify these, we used a combination of examining high frequency names found in the data as well as some identifiers that we have seen from past data exploration.

In the Jail data we did not find any of the obvious “dummy” names. The only potentially dummy name we found was “John Doe” but two of the rows had date of birth and SSN values, so we did not delete “John Doe” from the Jail data.

In the EMS data, we found several rows that appear to be associated with TEST records, cancelled calls, unknown patient name, or where there was no patient found at the scene. We removed rows with the names in [Table 18](#) from the EMS data. An additional 74 rows were removed that had blank last name because those would not be matched during individuating.

Table 18: Dummy or Test Names removed from the EMS File

First Name	Last Name	Row Count
JOHN	DOE	119
<Name>	UNKNOWN	102
NO	PT	83
DELETE	DELETE	68
NO PATIENT		61
NO	PATIENT	61
PATIENT	NO	55
JANE	DOE	46
NONE	NONE	23
CANCELLED	PTA	18
CANCELLED	PTA	18
PT	NO	14
CALL	CANCELLED	13
FOUND	NO PT.	11
FOUND	NO PT	10
CANCELLED	CANCELLED	8
DOE	JOHN	8
MALE	UNKNOWN	8
NO PATIENT	NO PATIENT	8
CANCELLED	CALL	7
CANCELED	CALL	6
ENROUTE	CANCELLED	6
CANCELLED ENROUTE	NO PATIENT CONTACT	5
NO PT	NO PT	5
ACADEMY	DELETE	4
CALL	CANCELED	4
CANCELLED	ENROUTE	4

First Name	Last Name	Row Count
DELETE	ACADEMY	4
EN ROUTE	CANCELED	4
<Name>	CANCEL	4
NO PATIENT CONTACT	CANCELLED ENROUTE	4
NO	PT.	4
CANCELLED	NO PATIENT CONTACT	3
DOE	JANE	3
ENROUTE	CANCELED	3
FEMALE	UNKNOWN	3
NO	FOUND	3
NO PATIENT CONTACT	NO PATIENT CONTACT	3
NO PATIENT FOUND	NO PATIENT FOUND	3
PTA	CANCELLED	3
ACADEMY DELETE	ACADEMY DELETE	2
CALL	CANCELLED PTA	2
CANCELED	CANCELED	2
CANCELLED ENROUTE	CANCELLED ENROUTE	2
DELTE	DELETE	2
NO PATIENT CONTACT	CANCELLED	2
NO PATIENT CONTACT	FIRE STANDBY	2
NO PATIENTS FOUND	NO PATIENTS FOUND	2
NO PT	FOUND	2
NO	PT FOUND	2
PT. FOUND	NO	2
PTA	CANCELED	2
UNKNOWN	FEMALE	2
UNKNOWN	PATIENT	2
UNKNOWN	UKNOWN	2
-	NO PT-	1
<Name>	(UNKNOWN)	1
<Name>	UNKNOWN LAST NAME	1

First Name	Last Name	Row Count
<Name>	[UNKNOWN]	1
ADMIN ONLY	NO PATIENT	1
ASSIST CODE BLUE	NO PATIENT CONTACT	1
ASSIST	PATIENT	1
BABY	DOE	1
BEFORE ARRIVING	CANCELED	1
CALL	CANCEL	1
CALL CANCELLED	NO PATIENT	1
CALL CANCELLED	NO PATIENT CONTACT	1
CANCEL	CANCEL	1
CANCELED	DISPATCH	1
CANCELED	EN ROUTE	1
CANCELED ENROUTE	FIRE STANDBY	1
CANCELED	FIRE	1
CANCELLED	NO PT	1
CANCELLED PTA	CANCELLED PTA	1
CANCELLED PTA	CARDIAC ARREST	1
CANCELLED PTA	HIGH LIFE HAZARD	1
CANCELLED PTA	MEDICAL ARRIVED	1
CANCELLED PTA	REPORTED FIRE	1
CANCELLED PTA	STOVE FIRE	1
CANCELLED PTA	WORKING FIRE	1
CANCELLED	UPON ARRIVAL	1
CODE 7	NO PATIENT	1
DELETE	DUPLICATE	1
DELETE	DUPLICATE CALL	1
DELETED	DELETE	1
EN ROUTE	CANCELLED	1
ENROUTE	CANCELLED WHILE	1
FALSE ALARM	CALL CANCELLED	1
FALSE ALARM	NO PATIENT CONTACT	1

First Name	Last Name	Row Count
FEMALE	UNKNOWN WHITE	1
<Name>	PT REFUSED	1
NO MPATIENT CONTACT	CANCELLED ENROUTE	1
NO	NAME	1
NO PATIENT	CANCELLED ENROUTE	1
NO PATIENT CONTACT	CALL CANCELLED	1
NO PATIENT CONTACT	CANCELLED BY CALLER	1
NO PATIENT CONTACT	DISREGARDED ENROUTE	1
NO PATIENT CONTACT	FIRE STANDY	1
NO PATIENT CONTACTS	STANDBY CHATHAM CO.	1
NO PATIENT	FALSE CALL	1
NO PATIENT	NO PATIENT FOUND	1
NO PATIENTS FOUND	FIRE STANDBY	1
NO PT. CONTACT	DUPLICATE CALL	1
NO	PT. INFO	1
NOT FOUND	PT	1
ON SCENE	CANCELED	1
PAGER	TEST	1
PRIOR TO ARRIVAL	CANCELLED	1
PT	ASSIST	1
PTS	NO	1
<Name>	CANCEL	1
STANDBY ONLY	NO PATIENT	1
STRUCTURE FIRE	NO PT	1
<Name>	NO	1
TO ARRIVAL	CANCELLED PRIOR	1
TRAUMA FEMALE	UNKNOWN	1
UNKNOWN	DEAD ON SCENE	1
UNKNOWN	MALE	1
WHILE EN ROUTE	CANCELED BY FIRE	1
`JANE	DOE	1

In the HMIS data, there was a single row with both first and last name equal to “TEST” that was removed. Because it also had dummy SSN and missing gender and date of birth it would not have been matched anyway. There were no other obvious dummy names found in that data.

4.3.7 Name Standardization and Name Match Codes

We do not replace names in the data with standardized names, but instead add columns with standardized names and name match codes to be used for individuating (entity resolution) in combination with other fields identified in the matching rules (See [Table 19](#) and [Table 20](#)). When name information is being recorded by someone other than the person to whom the name belongs, there is a greater chance of variation in spelling. We need standardization and match codes to overcome potential matching issues associated with data intake like spelling errors, typos and differences due to legal name vs nicknames like “Robert” and “Rob”. While Match Codes do not handle every first name typo, spelling variation, or nickname, the algorithmic match codes handle many of them (see [Table 21](#) for examples).

Table 19: Name Standardization Examples

First Name as Found in Data	Last Name as Found in Data	Standardized Name
John	Smith	John Smith
Smith	John	John Smith

Table 20: Last Name Match Code Handling Spelling and Data Entry Differences

Last Name as Found in Data	Last Name Match Code
ONEIL	#B7W\$\$
O`NEIL	#B7W\$\$
ONEILL	#B7W\$\$
O`NEIL	#B7W\$\$
ONEAL	#B7W\$\$

Table 21: First Name Match Codes for Typos, Spelling Variations, and Nicknames

First Name as Found in Data	First Name Match Code
MICHAEL	B73_
MICHEAL	B73_
MICHAL	B73_
JOHN	C@P\$

First Name as Found in Data	First Name Match Code
JOHNNY	C@P\$
JONATHAN	C@P\$
MARGARET	BYF7
PEGGY	BYF7
CASSIE	J&44
CASSONDRA	J&44
SANDRA	J&44
SONDRA	J&44
BILLIE	M7WW
BILLY	M7WW
WILLIAM	M7WW
JAMES	C&B_
JAMESON	C&B_
JAMEY	C&B_
JIM	C&B_
JIMMIE	C&B_
JIMMY	C&B_
JAMIE	C&B7
JAMIEE	C&B7

When creating match rules, one might consider creating a rule with an “or” condition; allowing either *first part of first name matches* or *first name match code equivalents* to handle cases like JAMIE and JAMES having two different match codes and CASSONDRA and SONDRRA, which have different first part of name, even though SONDRRA is a nickname of CASSONDRA (and same match code as seen above). Another alternative is to use last of the match code when matching – perhaps just the first three characters. Again, because this is only a part of the match rule – other data elements would also need to match – *entities would not be brought together on these first name matches alone*.

4.3.8 Address Data Quality and Standardization

For addresses standardization for matching, we need to have properly formed address information. In the Jail data (see [Table 22](#)), most of the dummy addresses had blank street number field. For all of those addresses, we set the standard address to blank and created no address match code. As a result, the address for these rows is not used in matching – instead other data elements must match. [Table 23](#) shows other dummy and frequently used addresses found in the Jail data. [Table 24](#) and [Table 25](#) demonstrate the blank street numbers and dummy and frequently used addresses found in the EMS data.

Table 22: Jail Addresses with No Street Number

Address	Name ID Count	Booking Count
	1,294	1,425
ANYWHERE	1,188	2,662
HOMELESS	127	371
UNKNOWN	74	88
ANY WHERE	11	12
NONE	5	5
NCDOC	5	16
SOMEWHERE	5	5
ANYWAY	3	4
HOMLESS	3	12
UNK	3	3
HEALING PLACE	3	3
EVERYWHERE	3	7
WCJ	3	9
ANYWHERE USA	3	9
SHELTER	2	4
ANYWHERE RALEIGH	2	2
ANYWHER	2	4
NOWHERE	2	8
MAIN ST	2	4
FBOP	2	3
ANYWHERE NC	2	2
NO WHERE	2	10
ANYWERE	2	2
HOMELESS SHELTER	2	5
ANTWHERE	2	2
WAKE COUNTY COURTHOUSE	1	1
UNEMPLOYED	1	4
SMITHFIELD RESCURE MISSION	1	1
SALVATION ARMY SHELTER	1	5

Address	Name ID Count	Booking Count
TENT BEHIND WALMART	1	2
URBAN MINISTRIES	1	1
ANYWHERE USA	1	1
FAIRYLAND	1	1
ANYWHERE	1	1
STAY IN HOTELS	1	3
STREETS OF RALEIGH	1	1
UNKNOWN	1	1
UNKNOWN ADDRESS	1	1
ANBYWHERE	1	1
ANYWHEREEE	1	2
ANYWHERE FUQUAY	1	1
ANYWHERE (REFUSED TO ANSWER)	1	1
ANYWHERE GARNER	1	2
NONE LISTED	1	1
NO RESIDENCE	1	1
ANYWHERE- HOMELESS	1	2
ANYWHERE	1	1
NO ADDRESS PROVIDED BY INMATE	1	1
ANYWHERE	1	1
S WILMINGTON (SHELTER) ST	1	8
NC DOC	1	4
SNOW AVE	1	7
HOME,LESS	1	1
ANYWHERE	1	2
THE HEALING PLACE	1	1
UN KNOWN	1	1
N/A	1	1
AHYWHERE	1	2
415 Other Street where missing street number	475	656

Table 23: Dummy and High Frequency Addresses from the Jail File

Address	Name ID Count	Booking Count	Notes
1420 S WILMINGTON ST	274	882	South Wilmington Outreach Center (Homeless) – 9 variations on the street name (S, S., South, Street, ST, STR)
220 SNOW AVE	204	779	Cornerstone Center (Homeless)
1491 US 70 HWY W	35	80	InTown Suites
118 E SOUTH ST	34	58	Shaw University
1251 GOODE ST	29	72	Healing Transitions (Homeless and Substance Abuse)
3520 MAITLAND DR	26	69	Raleigh Inn
2020 YONKERS RD	24	78	Department of Corrections
2800 BRENTWOOD RD	23	47	Hospitality Studios
3301 HAMMOND RD	22	45	Wake County Detention Center
3215 CAPITAL BLVD	22	40	InTown Suites
832 WAKE FOREST RD	20	67	Haven House (multi services – at risk youth and families)
330 S SALISBURY ST	18	33	Wake Public Safety Center
1401 BUCK JONES RD	16	35	Motel 6
312 TRYON RD	15	25	Cavalcorp Limited – Federal Halfway House
314 E HARGETT ST	14	18	Raleigh Rescue Mission
1315 OAKWOOD AVE	14	20	Saint Augustine's University
2641 APPLIANCE CT	13	30	Motel 6
3304 GLEN ROYAL RD	12	21	Healing Transitions
410 LIBERTY ST	12	20	Urban Ministries of Durham
3804 NEW BERN AVE	11	28	Knights Inn
501 NEW BERN AVE	11	27	New Bern Transition Housing
112 COX AVE	10	34	Women's Center of Wake County
1234 ANYWHERE	6	9	
1234 HOMELESS	3	3	
7610 SHELTER CV	2	2	
80 UNKNOWN	1	2	

Address	Name ID Count	Booking Count	Notes
123 SHELTER HAVEN PL	1	1	
2530 MILLS RD HOMELESS SHELTER	1	1	
1503 ANYWHERE	1	1	
0000 ANYWHERE USA	1	2	
314 ANYWHERE	1	2	
123 HOMELESS	1	3	
1234 UNKNOWN	1	1	
NONE ANYWHERE	1	3	
N/A UNKNOWN	1	1	
000 ANYWHERE	1	2	
000 UNKNOWN	1	1	
123 ANYWHERE	1	1	
1234 ANYWHERE ST	1	2	
8630 ANYWHERE	1	1	

Table 24: EMS Addresses with No Street Numbers

Address	Last Name Count	Row Count
	5,859	14,653
SOR	504	1,022
Streets of Raleigh	284	465
unknown	147	179
Unknown	111	145
homeless	75	88
Homeless	64	74
streets of raleigh	62	69
UNKNOWN	49	56
S.O.R.	44	54
S.O.R	40	43
STREETS OF RALEIGH	30	32
streets of Raleigh	29	34
sor	24	24

Address	Last Name Count	Row Count
Undomiciled	19	20
HOMELESS	13	15
Streets Of Raleigh	12	13
Indomecile	12	12
streets	9	10
none	9	10
Streets of raleigh	9	10
Streets of Cary	8	8
Streets of Zebulon	7	9
UTO	7	7
UNK	7	7
S. O. R.	7	7
Streets	6	6
Unkown	6	7
no address	6	6
unkown	6	6
unk	6	6
SOG	5	5
Streets of Garner	5	6
None	5	5
Streets of Morrisville	4	4
No Address	4	4
Anywhere	4	4
General Delivery	3	3
Unknown Address	3	3
UTA	3	3
Streets of Wake Forest	3	3
street of Raleigh	3	3
Unable to obtain	3	3
SoR	3	3
streets of cary	3	3

Address	Last Name Count	Row Count
Street of Raleigh	3	3
Healing Place	3	3
unable to obtain	3	3
unavailable	3	3
507 Additional distinct addresses without numbers. Generally street name only, intersections and variations on unknown, not provided or "Streets of"	528	541

Table 25: EMS Dummy and High Frequency Addresses

Address	Last Name Count	Row Count	Notes
220SNOWAVE	301	811	Cornerstone Center (Homeless)
1420SWILMINGTONST	296	705	South Wilmington Outreach Center (Homeless)
1300WESTERNBLVD	213	386	Central Prison
1034BRAGGST	156	235	N.C. Correctional Institution for Women
3304GLENROYALRD	54	84	Healing Transitions
3019FALSTAFFRD	54	59	Holly Hill Behavioral Health
1420SWILMINGTONSTREET	53	68	South Wilmington Outreach Center (Homeless)
220SNOWAVENUE	53	74	Cornerstone Center (Homeless)
1251GOODEST	53	67	Healing Transitions
1863CAPITALBLVD	51	82	Salvation Army Shelter
314EHARGETTST	48	80	Raleigh Rescue Mission
3301HAMMONDRD	34	34	Wake County Detention Center
401WCABARRUSST	34	50	Helen Wright Shelter for Women
112COXAVE	34	77	Women's Center of Wake County Transitional Housing
601 <Street Deleted>	31	45	Single Family Home
1012OBERLINRD	29	50	Interact
1000ROCKQUARRYRD	29	33	Wake Correctional Center
400FAYETTEVILLEST	22	27	Apartments

Address	Last Name Count	Row Count	Notes
1420SOUTHWILMINGTONST	21	26	South Wilmington Outreach Center (Homeless)
2910CAPITALBLVD	20	26	Hotel / Hostel
501NEWBERNAVE	19	31	New Bern House - Transitional Housing for Seniors
3520MAITLANDDR	17	20	Raleigh Inn
750BRIGHTCREEKWAY	17	45	Apartments
9177THAVE	17	23	Apartments
832WAKEFORESTRD	16	24	Haven House Outreach
3215CAPITALBLVD	15	18	Extended Stay Hotel (InTown Suites)
1420SWILLMINGTONST	15	21	South Wilmington Outreach Center (Homeless)
917SEVENTHAVE	14	15	Apartments
1315OAKWOODAVE	14	15	University
1420SWILMINGTON	14	14	South Wilmington Outreach Center (Homeless)
3939GLENWOODAVE	13	17	Apartments
1212 <Street Deleted>	13	46	Single Family Home
1209 <Street Deleted>	13	45	Single Family Home
116STMARYSST	13	15	Public Housing Apartments
3939WAKEFORESTRD	13	20	Apartments
2800BRENTWOODRD	13	22	Hospitality Studios
7700 <Street Deleted>	12	14	Single Family Home
510 <Street Deleted>	12	24	Single Family Home
1420SOUTHWILMINGTONSTREET	12	13	South Wilmington Outreach Center (Homeless)
220SNOW	8	8	Cornerstone Center (Homeless)

Note: Addresses have been capitalized and spaces and punctuation removed. Assisted living addresses have been dropped. Address Standardization and Address Match Codes

As with normalization process for names, addresses in the data are not replaced with standardized names, but instead columns are added with standardized addresses and address match codes to be used for individuating (entity resolution) in combination with other fields identified in the matching rules. We need standardization and match codes to overcome potential matching issues associated with data

intake like spelling errors, typos and differences with abbreviation like “S” and “South” or “St” and “Street” or leaving the street type off all together (for examples, see [Table 26](#)).

Table 26: Address Standardization and Match Code Examples

Address as Found in Data	Standardized Address	Address Match Code
1420 S WILMINGTON ST	1420 S WILMINGTON ST	ZSH \$\$ LWBPF 4 \$\$ \$\$\$\$
1420 S. WILMINGTON ST	1420 S WILMINGTON ST	ZSH \$\$ LWBPF 4 \$\$ \$\$\$\$
1420 SOUTH WILMINGTON STREET	1420 S WILMINGTON ST	ZSH \$\$ LWBPF 4 \$\$ \$\$\$\$
1420 SOUTH WILMINGTON ST	1420 S WILMINGTON ST	ZSH \$\$ LWBPF 4 \$\$ \$\$\$\$
1420 S WILMINGTON STR	1420 S WILMINGTON ST	ZSH \$\$ LWBPF 4 \$\$ \$\$\$\$
1420 S willmington St	1420 S WILLMINGTON ST	ZSH \$\$ LWBPF 4 \$\$ \$\$\$\$
1420 WILMINGTON ST	1420 WILMINGTON ST	ZSH \$\$ LWBPF 4 \$\$ \$\$\$\$
1427 S WILMINGTON ST	1427 S WILMINGTON ST	ZSH \$\$ LWBPF 4 \$\$ \$\$\$\$
1420 S Wolmington St	1420 S WOLMINGTON ST	ZSH \$\$ LWBPF 4 \$\$ \$\$\$\$
1420 S. Wilmingto Street.	1420 S WILMINGTO ST	ZSH \$\$ LWBPF 4 \$\$ \$\$\$\$
1420 S Wilmington St- Streets of Raleigh	1420 S WILMINGTON ST	ZSH \$\$ LWBPF 4 \$\$ \$\$\$\$
220 SNOW AVENUE	220 SNOW AVE	HHO \$\$ 4PL\$\$ & \$\$ \$\$\$\$
220 SNOW AVE	220 SNOW AVE	HHO \$\$ 4PL\$\$ & \$\$ \$\$\$\$
220 SNOW AVE APT 7	220 SNOW AVE APT 7	HHO \$\$ 4PL\$\$ & \$\$ \$\$\$\$
220 Snow Avn	220 SNOW AVE	HHO \$\$ 4PL\$\$ & \$\$ \$\$\$\$

4.3.9 Matching Rules

Once the *Source ID* was identified for each data source and the identifying data elements were determined, we then moved on to data quality. We found that we needed to remove dummy names, SSN, and addresses.

Additionally, we confirmed whether the *Source ID* uniquely identified a person in that source. To do that for each source, we first looked for *Source ID* with multiple values for name, SSN, date of birth or gender. Once we confirmed that Name Id was associated with just a single *name-SSN-date of birth-gender* combination, two important observations emerged:

1. It was determined that the individual data sources appeared to have multiple Source ID assigned to the same distinct person due to small differences in spelling, DOB or Gender. For example in the Jail data, Joe Smith with SSN 123-45-6789, date of birth 03-12-1974 at 123 Meadow Lane would have a different name_id than Joe Smith with SSN 123-45-6789, DOB 03-21-1974 at 123 Meadow Lane. Note, the difference between the two name_id records is the transposed day in the date of birth column. Thus, Name id was identified as representing a distinct combination of identifying information and *not* necessarily an individual person.

2. In the HMIS data, we found that often the parent SSN was used for the children and thus we needed stricter matching rules within HMIS than for the Jail data.

To overcome these matching challenges, matching was first performed *within* each data source with rules specific to that source, and then stronger rules for additional passes matching entities across the data sources.

For our initial matching to have the same individual cluster id, individuals must have either the same *Source ID* provided in the data, full SSN or same full name match code plus additional criteria. If none of those conditions are met, then an individual will not be added to the cluster. That is, for the scope of this project, even if the social security number is off by only one digit and the individuals have:

- Same first name, but not same last name and
- Same date of birth match, but the *Source ID* do not match

The individuals will not belong to the same individual cluster id. Only if these records can be ‘chained together’ through other records that have the same full SSN match as one of these and same name match as the other.

[Table 27](#) illustrates the minimum match criteria and chaining required for bring disparate records together within the same data source (for actual matching, dummy SSNs would not be used).

Table 27: Examples of Name and SSN Match Criteria and Chaining

Name	SSN	
Susy Baker	123-45-6789	First two rows are not candidates to be brought together because neither name nor SSN match
Susy Miller	122-45-6789	
Susy Smith	67-89-0123	Here the middle two rows are candidates to match because they share same SSN. The first row is a candidate to add to the cluster because it shares the same name as row 3. The last row is a candidate to add to the cluster because it shares same SSN as first row and same name as second row. In this way the 4 rows might chain together.
Susy Jones	66-89-0123	
Susy Smith	66-89-0123	
Susy Jones	67-89-0123	

4.3.10 Rules for Matching within Jail Data

For our initial matching within the jail data to have the same individual cluster id, individuals must have either same full SSN or same full name match code plus additional criteria. At the end, clusters would be forced together if they share the same *name_id* provided in the Jail data, but in this cut of data all individual rows with the same *name_id* have the identical full name, SSN and date of birth so no forcing together actually occurs. The matching is done in two passes; the first on SSN and then on name. In a

final step, the name match clusters are forced together where SSN matching put individuals into the same cluster.

Individuals with the same SSN will cluster together if date of birth is within one digit and one of the following conditions is also met:

- Full name match code are identical.
- Gender matches and full name match code or standardized full name have Levenshtein edit distance less than or equal to 2.

Note: The Levenshtein edit distance is a metric for measuring the difference between two string sequences. It can be understood as the minimum number of operations (insertions, deletions, or substitutions of a single character, or transposition of two adjacent characters) need to change one word into the other.

- First name exactly matches last name and last name exactly matches first name (first and last names flipped For example “Susy Smith” and “Smith Susy”)
- First name matches and last part of last name matches last name in other row (handles one row with hyphenated or compound name and other row not. For example one row last name is “Smith-Jones” and other row has last name “Jones”.
- First name matches last part of last name matches last name in other row and vice versa (handles one row with hyphenated or compound name and other row not, and first and last names flipped. For example one row has “Susy Smith-Jones” and other row has name “Jones Susy”.
- Female and first name matches and either phone number Levenshtein edit distance less than or equal to 2 or address match code matches.

Individuals with the same full name match code will cluster together if gender also matches, date of birth have Levenshtein edit distance less than or equal to 1 or a single pair of flipped digits and one of the following conditions is also met:

- Full SSN have Levenshtein edit distance less than or equal to 2
- One is a partial SSN and last 4 of SSN match
- Phone number Levenshtein edit distance less than or equal to 2
- Address match codes match

4.3.11 Rules for Matching within EMS Data

For our initial matching within the EMS data to have the same individual cluster id. Individuals must have either same full SSN or same full name match code plus additional criteria. The matching is done in two passes, the first on SSN and then on name. In a final step, the name match clusters are forced together where SSN matching put individuals into the same cluster.

Individuals with the same SSN will cluster together if date of birth is within one digit and one of the following conditions is also met:

- Full name match code are identical.
- Gender matches or is “Not Reported” and full name match code or standardized full name have Levenshtein edit distance less than or equal to 2.

- First name exactly matches last name and last name exactly matches first name (first and last names flipped, for example “Susy Smith” and “Smith Susy”).
- First name matches and last part of last name matches last name in other row (handles one row with hyphenated or compound name and other row not. For example one row last name is “Smith-Jones” and other row has last name “Jones”).
- First name matches last part of last name matches last name in other row and vice versa (handles one row with hyphenated or compound name and other row not, and first and last names flipped. For example one row has “Susy Smith-Jones” and other row has name “Jones Susy”).
- Female and first name matches and either phone number Levenshtein edit distance less than or equal to 2 or address match code matches.

Individuals with the same full name match code will cluster together if gender also matches or is “Not Reported”, date of birth have Levenshtein edit distance less than or equal to 1 or a single pair of flipped digits and one of the following conditions is also met:

- Full SSN have Levenshtein edit distance less than or equal to 2
- One is a partial SSN and last 4 of SSN match
- Full 10 digit phone number Levenshtein edit distance less than or equal to 2
- One or both of the phone numbers do not have area code and the last 7 digits match exactly
- Address match codes match

4.3.12 Rules for Matching within HMIS Data

For our initial matching within the HMIS data to have the same individual cluster id, individuals must have either same full SSN or same full name match code plus additional criteria. The matching is done in two passes, the first on SSN and then on name. In a final step, the name match clusters are forced together where SSN matching put individuals into the same cluster. If there had been rows with the same HMIS Client ID that were not placed into the same cluster by the matching they would have been forced into the same cluster.

Individuals with the same SSN will cluster together if date of birth is within one digit and one of the following conditions is also met:

- Full name match code are identical.
- Gender matches or is “Not Reported” and full name match code or standardized full name have Levenshtein edit distance less than or equal to 2.
- First name exactly matches last name and last name exactly matches first name (first and last names flipped, for example “Susy Smith” and “Smith Susy”).
- First name matches and last part of last name matches last name in other row (handles one row with hyphenated or compound name and other row not. For example one row last name is “Smith-Jones” and other row has last name “Jones”).
- First name matches last part of last name matches last name in other row and vice versa (handles one row with hyphenated or compound name and other row not, and first and last names flipped. For example one row has “Susy Smith-Jones” and other row has name “Jones Susy”).
- Female and first name matches and address match code matches.

Individuals with the same full name match code will cluster together if gender also matches or is “Not Reported”, date of birth have Levenshtein edit distance less than or equal to 1 or a single pair of flipped digits and one of the following conditions is also met:

- Full SSN have Levenshtein edit distance less than or equal to 2
- One is a partial SSN and last 4 of SSN match
- Full 10 digit phone number Levenshtein edit distance less than or equal to 2
- Address match codes match

4.3.13 Rules for Matching across All Three Sources

After the individualizing has completed within the individual data sources we stack together the results and standardize the values for gender and race as indicated above. As was done for the matching within separate data sources, matching of individuals *across* data sources is done in two passes; the first on SSN and then on name. In a final step, the name match clusters are forced together where SSN matching put individuals into the same cluster. Next, if there were individuals that were clustered together in one of the separate source matching passes, these too are forced together.

Individuals with the same SSN will cluster together if one of the following conditions on name and address are met:

- Full name match code or up cased, all spaces and punctuation removed full names are identical.
- Gender matches or is “Not Reported” and last name match code has Levenshtein edit distance less than or equal to 2 and first name match code has Levenshtein edit distance less than or equal to 1.
- Gender matches or is “Not Reported” and standardized full name has Levenshtein edit distance less than or equal to 2.
- Date of birth and first name match code match exactly and last name match code has Levenshtein edit distance less than or equal to 1 (this handles cases where we see nearly slight typo in name and matching DOB but different gender. For example Susy Jones, Female, 11-23-1979, 111-22-3333 and Susy Kones, Male, 11-23-1979, 111-22-3333).
- First name exactly matches last name and last name exactly matches first name (first and last names flipped, for example “Susy Smith” and “Smith Susy”).
- First name matches and last part of last name matches last name in other row (handles one row with hyphenated or compound name and other row not. For example one row last name is “Smith-Jones” and other row has last name “Jones”).
- First name matches last part of last name matches last name in other row and vice versa (handles one row with hyphenated or compound name and other row not, and first and last names flipped. For example one row has “Susy Smith-Jones” and other row has name “Jones Susy”).
- Female and first name matches and either phone number Levenshtein edit distance less than or equal to 2 or address match code matches.

Also, at least one of the following conditions on date of birth are met:

- Date of birth is a single digit off (for example 11-23-1974 and 11-23-1984)
- Day of birth is one day off and year and month match (for example 11-19-1974 and 11-20-1974)

- Month of birth is a single month off and year and day match (for example 09-23-1974 and 10-23-1974)
- Year of birth is a single year off and month and day match (for example 11-23-1969 and 11-23-1970)
- Only year and month match, but full name match code and gender also match (Susy Smith, Female, 11-23-1989 and Suzy Smith, Female, 11-01-1989)

Individuals with the same full name match code will cluster together if gender also matches or is “Not Reported”, date of birth have Levenshtein edit distance less than or equal to 1 or a single pair of flipped digits and one of the following conditions is also met

- Full SSN have Levenshtein edit distance less than or equal to 2
- One is a partial SSN and last 4 of SSN match
- Full 10 digit phone number Levenshtein edit distance less than or equal to 2
- One or both of the phone numbers do not have area code and the last 7 digits match exactly
- Address match codes match

4.3.14 Individuating Surrogate Keys

Once matching is complete, an individuating surrogate key is created by numbering the resulting clusters of individuals from 1 to the total number of individual clusters.

The individuating surrogate is used to bring together information from many different sources to get the most complete picture of an individual. That “complete view” is then used compute summary statistics, like the number of jail bookings or the total number of EMS calls over each individuating surrogate key as well as summarizations by attributes such as gender, race, and age.

We create master demography information by individually selecting the name, DOB, race, gender, and ethnicity most frequently associated with the surrogate key. It is this master information that is used in this report to show aggregate demographic information for our populations of individuals.

5 Wake County: A Demographic Profile

To provide decision makers with insight into vulnerable populations at risk for repeatedly cycling through local jails, emergency medical services, housing and services to the homeless, it is necessary to first understand the broader context at both the state and county level. Men, women, families, and youth fall victim to homelessness for a variety of reasons, such as a shortage of affordable housing, low-paying jobs, substance and alcohol abuse, mental illness, and family conflict. Criminal records, poor credit, inconsistent employment histories, and deficient independent living skills are additional causes.

Based on statistics gathered by the U.S. Census Bureau (see [Table 28](#)), the demographic profile of Wake County shows a rapidly growing, youthful, and diverse population compared to national statistics. For instance, population estimates show that Wake County has grown by 13.7% between 2010 and 2015 while the U.S. as a whole grew by only 4.1% over the same five-year period. Wake County is also more youthful compared the U.S. with nearly a quarter of the County's population under the age of 18 and 10.3% at 65 years of age and older versus national percentages of 22.9% and 14.9% respectively. The County displays greater racial diversity as well with higher percentages of both black/African Americans (21.3%) and Asian/Pacific Islanders (6.8%) compare to national figures.

Table 28: U.S. Census Bureau Statistics for Wake County, NC ^a

	Wake County	United States
Population		
Estimate as of April 2010	901,021	308,758,105
Estimate as of July 2015	1,024,198	321,418,820
Percent Change	13.7%	4.1%
Age (July 2015)		
Persons under 18 years	24.7%	22.9%
Persons 65 years and over	10.3%	14.9%
Sex		
Female	51.3%	51.3%
Race		
White	68.7%	77.1%
Black or African American	21.3%	13.3%
Asian/Other Pacific Islander	6.8%	5.8%
Other Race or Multi-Racial	3.2%	3.8%
Ethnicity		
Hispanic or Latino	10.1%	17.6%
Housing and Families		
Median value of owner-occupied housing units	\$234,000	\$178,600
Median Gross Rent	\$948	\$928

	Wake County	United States
Persons per Household	2.61	2.64
Education (persons age 25 or higher)		
High school graduate or higher	91.9%	86.7%
Bachelor's degree or higher	49.0%	29.8%
Income and Poverty		
Median household income	\$67,309	\$53,889
Per Capita income (past 12 months)	\$34,202	\$28,930
Persons in Poverty	11.1%	13.5%

^a <https://www.census.gov/quickfacts/table/PST045216/37183,00>

Despite the County's extraordinary growth, youthfulness, and diversity, the high cost of living and well-above-average educational levels of the populace, where nearly 50% of persons age 25 or older hold a bachelor's degree or higher compared to 30% nationally, makes a homeless person's transition to full economic self-sufficiency particularly difficult. We can get a sense of the minimum income an individual would require to support oneself and/or their family by looking at the *Living Wage Calculator* developed by [Amy K. Glasmeier](#) and the [Massachusetts Institute of Technology](#). [Table 29](#) shows the Living Wage Calculation for Wake County, which breaks down hourly pay by wage types and family size and composition. The living wage estimate is calculated from geographically specific expenditure data related to a family's likely minimum food, childcare, health insurance, housing, transportation, and other basic necessities (e.g. clothing, personal care items, etc.) costs. The total *gross* "living" income, detailed in [Table 30](#), is estimated for each family type and an hourly wage derived by dividing the total living income by 2,080 hours (hourly work-year; 40 hours per week for 52 weeks). According to the developers of the living wage model, it is:

a 'step up' from poverty as measured by the poverty thresholds but it is a small 'step up', one that accounts for only the basic needs of a family. The living wage model does not allow for what many consider the basic necessities enjoyed by many Americans. It does not budget funds for pre-prepared meals or those eaten in restaurants. It does not include money for entertainment nor does it does not allocate leisure time for unpaid vacations or holidays. Lastly, it does not provide a financial means for planning for the future through savings and investment or for the purchase of capital assets (e.g. provisions for retirement or home purchases). The living wage is the minimum income standard that, if met, draws a very fine line between the financial independence of the working poor and the need to seek out public assistance or suffer consistent and severe housing and food insecurity. In light of this fact, the living wage is perhaps better defined as a minimum subsistence wage for persons living in the United States.²

² Living Wage Calculator User's Guide / Technical Notes, Page 2. Available at <http://livingwage.mit.edu/resources/Living-Wage-User-Guide-and-Technical-Notes-2016.pdf>.

The poverty wage rate is calculated by converting the 2016 poverty thresholds by household size, defined by the Department of Health and Human Services, into an hourly wage.³ The North Carolina minimum wage is \$7.25 and thus the same for all individuals. The state legislature, however, is currently considering a bill that would increase the state's minimum wage to \$12 an hour by 2020; and \$15 an hour by 2022.

Table 29: Living Wage Calculations for Wake County, North Carolina^a

Family Composition	Living Wage	Poverty Wage^b	Minimum Wage
1 Adult	\$11.30	\$5.71	\$7.25
1 Adult 1 Child	\$23.25	\$7.70	\$7.25
1 Adult 2 Children	\$28.46	\$9.69	\$7.25
1 Adult 3 Children	\$35.15	\$11.68	\$7.25
2 Adults (1 working)	\$18.66	\$7.70	\$7.25
2 Adults (1 working) 1 Child	\$22.94	\$9.69	\$7.25
2 Adults (1 working) 2 Children	\$25.41	\$11.68	\$7.25
2 Adults (1 working) 3 Children	\$28.08	\$13.67	\$7.25
2 Adults	\$9.33	\$3.85	\$7.25
2 Adults 1 Child	\$13.20	\$4.85	\$7.25
2 Adults 2 Children	\$15.52	\$5.84	\$7.25
2 Adults 3 Children	\$17.94	\$6.84	\$7.25

^a <http://livingwage.mit.edu/counties/37183>

^b Poverty Thresholds are from the 2016 guidelines for the 48 contiguous states and the District of Columbia. To convert values from annual to hours, a work-year of 2,080 (40 hours per week for 52 weeks), annual income poverty threshold for each household size is divided by 2,080 hours.

³ Poverty data is available at <https://aspe.hhs.gov/poverty-guidelines>. The 2016 values are published in the Federal Register <https://www.federalregister.gov/documents/2016/01/25/2016-01450/annual-update-of-the-hhs-poverty-guidelines>

Table 30: Annual Expenses Used to Calculate the Living Wage Estimate for Wake County

	Food	Child Care	Medical	Housing	Transportation	Other	Gross Income	Net Income
1 Adult	\$2,983	\$0	\$2,128	\$7,728	\$4,401	\$2,458	\$19,698	\$23,501
1 Adult 1 Child	\$4,516	\$6,035	\$6,283	\$11,364	\$8,358	\$4,008	\$40,564	\$48,363
1 Adult 2 Children	\$6,722	\$9,808	\$5,996	\$11,364	\$10,918	\$4,826	\$49,634	\$59,201
1 Adult 3 Children	\$8,882	\$13,580	\$6,157	\$14,736	\$11,911	\$6,010	\$61,276	\$73,103
2 Adults (1 working)	\$5,469	\$0	\$4,928	\$9,816	\$8,358	\$4,008	\$32,579	\$38,817
2 Adults (1 working) 1 Child	\$6,921	\$0	\$5,996	\$11,364	\$10,918	\$4,826	\$40,025	\$47,719
2 Adults (1 working) 2 Children	\$8,888	\$0	\$6,157	\$11,364	\$11,911	\$6,010	\$44,330	\$52,862
2 Adults (1 working) 3 Children	\$10,799	\$0	\$6,015	\$14,736	\$11,951	\$5,474	\$48,975	\$58,405
2 Adults	\$5,469	\$0	\$4,928	\$9,816	\$8,358	\$4,008	\$32,579	\$38,817
2 Adults 1 Child	\$6,921	\$6,035	\$5,996	\$11,364	\$10,918	\$4,826	\$46,060	\$54,930
2 Adults 2 Children	\$8,888	\$9,808	\$6,157	\$11,364	\$11,911	\$6,010	\$54,138	\$64,582
2 Adults 3 Children	\$10,799	\$13,580	\$6,015	\$14,736	\$11,951	\$5,474	\$62,555	\$74,632

Note: Details on the estimates for the individual expense type are available in the Living Wage Calculator User's Guide / Technical Notes. Available at <http://livingwage.mit.edu/resources/Living-Wage-User-Guide-and-Technical-Notes-2016.pdf>.

The Living Wage calculations show that average household size and, in particular, family composition dramatically changes the hourly wage required to support a family. For instance, where the average household size is roughly 3 persons (2.61 according to U.S. Census estimates in [Table 28](#)), a living wage rate for 2 working adults with 1 child is estimated at \$13.20 or a before tax income of \$54,930 compared to a family comprised of 1 adult with 2 children where the hourly rate is set at \$28.46 or at net income of \$59,201. For a single working adult with children, a before tax income of \$50,000 or more is roughly the median annual income of a person with Associate's degree or higher according to the Bureau of Labor Statistics. For a single adult with no children, a work-year at minimum wage yields an annual income of \$15,080, which is above the poverty threshold of \$11,880 but well below the \$23,501 estimate of a gross income living wage gross income.

These incomes estimates provide a basis for understanding the significant hurdles persons in a state of homelessness face when attempting to achieve economic self-sufficiency. Housing represent one of the largest expenses American household face. A recent report produced by Wake County estimates that nearly 50% of renters spend more than 30% of household income on housing compared to only 29% of those who own their own home (Wake County by the Numbers: Wake Growth Highlights).⁴ In Wake County, where the median owner-occupied home value is \$55,000 higher than the national figures and rent is slightly higher at \$948 (see [Table 28](#)), housing represents a significant cost-burden to families, especially those struggling with homelessness.

Although Wake County offers a number of supportive services and programs for people experiencing homelessness, demand has outstripped the funding of these services. Recent figures estimate that over the course of year more than 4,500 people experience homeless in Wake County. With only 952 shelter beds available each night across almost a dozen different locations, many are left without the safety and protection from the elements that temporary shelters provide (Wake County By the Numbers: Wake Growth Highlights). Those who are homeless represent some of society's most vulnerable sub-populations, including children, single mothers, veterans, and those suffering from mental illness and/or substance use disorder.

To combat the occurrence of homelessness in Wake County, this report aims to more fully identify at-risk populations who are homeless and frequently interact with law enforcement and emergency services with the expressed purpose of ending the repetitive cycling between these costly systems. By understanding the characteristics, utilization, pattern of engagement with these systems among the most frequent users, County officials can more effectively identify intervention strategies.

⁴ More Than One Million: An Overview of Growth, Demographics, and What We need to Keep Our Eye On.

<http://www.wakegov.com/data/bythenumbers/SiteAssets/Pages/Forms/EditForm/1.%20Wake%20Growth%20Highlights%20Final.pdf>

6 Agency Data

This section explores the unique composition of each agency's data following the entity resolution process discussed in Section 3 and, where possible, how the populations within each dataset compare to Wake County. We also investigate the concept of "high utilization" or "familiar faces" within each system. That is, what criteria is used to define someone who frequently interacts with either homelessness services, the jail system, or EMS in Wake County and how does this high utilization group compare the general population. Table 31 outlines the agency specific definitions of frequent use; each includes a minimum number of incidents, 4 or more times, within a specific period of time. Because this report leverages cross-sectional data rather than longitudinal, the numbers of "familiar faces" observed is dependent (to a certain extent) on the span of data received for each system. That is, individuals may have contact before or after the analytic window that would meet the threshold for high utilization, yet our current view of system interactions falls short of such as designation. Consequently, frequent use is also considered within the data coverage, notably by the percentile distribution of the data population.

Table 31: Agency Definitions of High System Utilizers

Agency	High Utilization Concept	Definition
Wake County Jail	Familiar Face	An individual interacting with jail more than 4 times in a 24 month span.
Homelessness (HUD)	Chronic Homelessness	Head of household has a disability AND has been homeless for at least 12 consecutive months or has had 4 or more episodes of homelessness in the last in 3 years totaling 12 months or more.
Wake County EMS	High Utilizer	An individual that has utilized EMS services 4 or more times during a rolling 30 day period.

6.1 Wake County Jail

After applying entity resolution techniques discussed in Section 3, a total 36,665 distinct individuals were identified in Wake County Jail data with a booking on or after January 1, 2015. Over the 24-month analytic period, these individuals account for a total of 57,735 bookings and 151,368 criminal charges. The Wake County Jail system processes between 2,000 and 2,500 arrestees on a monthly basis, with nearly half of bookings resulting in no jail stay and 72% involve only misdemeanor charges.

We also see that most individuals in the jails data are not repeat offenders, where more than 70% of persons have one booking record. To further explore the number of times that individuals enter the jail system, Table 32 shows the percentile breakdown of booking counts in the jail data. Percentiles identify the value (in this case booking count) at which a certain percentage of population is at or below. For instance, based on the 24-months of jail data, 90 percent of arrestees were found to have 3 or fewer bookings. Considered differently, that means that at the 90th percentile, 10% of arrested individuals had entered the jail system on more than 3 occasions. At the 95th percentile, we see the booking count

increase to a count of 4 bookings, which is right at the threshold Wake County Jail defines as a “familiar face.” In this sense, approximately 5% of the jail population over this 2-year period meet the criteria as a high utilizer of the jail system. The maximum number of jail bookings observed in the data was 40.

Table 32: Distribution of Booking Counts

Percentiles	Number of Time Entering the Jail System
50th	1
75th	2
90th	3
95th	4
99th	7
Maximum Count	40

[Table 33](#) shows the demographic composition of the total Wake County jail population as well as for the sub-group of “familiar faces” – those who have entered the jail system more than 4 times. The peak age group of those who interaction with the jail system is between 26 and 35 years for both the total population and familiar faces, at 33% and 34% respectively. A notable difference among the high utilizers is the relative youth of this group with nearly 71% less than 36 years of age compared to just 63% for the overall jail population. The discrepancy by gender is also considerable relative to not only the County, which is 49 percent male, but also compared to the total jail population (73% male) and familiar faces (80% male). Consistent with National trends, black or African Americans are over-represented in the jail system at 53% compared to their 23% representation in Wake County (see [Table 28](#)). Among the group of familiar faces, the representation black or African American is even higher at 68% despite making up only 13% of the total U.S. population.

Table 33: Demographic Profile of Wake County Jail Population and Familiar Faces

	Jail Population	Familiar Faces
Distinct Individual Count	36,665	1,333
Age Categories		
18 years and under	3.89%	5.93%
19 to 25 years	25.57%	30.68%
26 to 35 years	33.32%	34.28%
36 to 45 years	19.32%	15.23%
46 to 55 years	12.22%	8.63%
56 to 65 years	4.61%	4.88%
66 to 75 years	0.95%	0.30%
76 or more years	0.11%	0.08%
Unknown	0.01%	N/A
Sex		

	Jail Population	Familiar Faces
Male	73.22%	80.12%
Female	26.76%	19.88%
Other or Unknown	0.02%	N/A
Race		
White	45.84%	31.51%
Black or African American	53.22%	68.19%
Asian/Other Pacific Islander	0.38%	0.08%
Other Race or Multi-Racial	0.56%	0.23%
Ethnicity		
Hispanic or Latino	6.54%	3.45%
Non-Hispanic/Non-Latino	60.99%	89.72%
Missing, Unknown, or Refused	32.46%	6.83%

[Table 34](#) examines the types of crimes committed, defined using North Carolina’s Uniform Crime Report (UCR) categories, by the general jail population and familiar faces. The UCR category most often violation among both the total population and “familiar faces” is for a generalized category of *contempt of court; perjury; court violation*, which for both group makes up about 20% of all crimes found on bookings. For the total jail population, the second most frequent UCR was for *Driving While Impaired* (DWI) at 9% of bookings, whereas among “familiar faces” DWI crimes was the 11th ranked UCR with less than 3% of bookings. Thus, DWI offenses does not appear to be endemic to “familiar faces.” Drug violations, however, are a frequent criminal offense found among “familiar faces” with nearly 10% of this group’s bookings involving drugs, and equally prevalent as the 3rd most frequent UCR among the total jail population.

While the remaining relative rankings of UCR categories are roughly similar between the Wake County jail population and Familiar Faces in [Table 34](#), *Parole & Probation Violations* stands out as the 4th ranked UCR for “familiar faces” (almost 9% of bookings), but only 8th for the total population (5.5% of bookings). Given the repeated interactions of “familiar faces” with the jail system, it makes sense that successive engagements with the law lead to compounding violations as new crimes break the terms of release of prior offenses. Yet, noted above, the vast majority of individuals in the 24-months of jail had only one booking (more than 70%) and thus are unlikely to have a parole violation.

As we continue to detail the HMIS and EMS transport data, and eventually the intersection of all three systems, we can begin to better understand the familiar faces and the set of events that underlie their interaction with each system.

Table 34: Count, Percent, and Relative Rank of UCR Categories by Wake County Jail Population and Familiar Faces

UCR Category	Total Jail Population			Familiar Faces		
	Count	Percent	Relative Rank	Count	Percent	Relative Rank
Contempt of Court; Perjury; Court Violations	14,202	18.66%	1	1,865	17.65%	1
Driving While Impaired	7,037	9.25%	2	265	2.51%	11
Drug Violations	7,034	9.24%	3	1,094	10.35%	2
Simple Assault	6,695	8.80%	4	720	6.81%	6
Larceny	5,284	6.94%	5	743	7.03%	5
Non-UCR Offenses	5,151	6.77%	6	385	3.64%	9
All Other (Includes drug arrests)	4,719	6.20%	7	927	8.77%	4
Parole & Probation Violations	4,294	5.64%	8	1,033	9.78%	3
Escape from Custody or Resist Arrest	2,908	3.82%	9	506	4.79%	8
Trespassing	2,300	3.02%	10	705	6.67%	7

6.2 Homeless Management Information System (HMIS)

Entity resolution of the HMIS data outlined in Section 3 identified 8,834 distinct individuals, out of the original 10,141 HMIS client IDs, who had received some level of housing assistance on or after January 1, 2015. There are 78,740 housing support services recorded for this group during the 20-month period, which include program types such as permanent supportive housing, homelessness prevention services, transitional housing, street outreach, emergency shelter stays, and other forms of assistance. Looking at participation across all programs, the average number of program interactions, or discrete service events, is 9 with a cumulative average of 122 days of involvement. Because some program interactions are more frequent than others, for instance emergency shelter stays which tend to be daily versus longer-term engagements like permanent supportive housing, it can be informative to view the median values. At the median values, the number of service events lowers to 2 and the cumulative numbers days is 42.

To further understand the participation in homelessness programs in terms of both number of instances and duration, Table 35 shows the average and median number of service events and cumulative participation days by program type. Emergency shelters is the most frequented program with over 6,000 distinct individuals making use of the program at least once during the 20-month period. The average number of services for emergency shelter stays among this group is 12, but the median is 2 indicating that a small proportion of individuals heavily rely on this form of homeless support. The cumulative length of stay at emergency shelters echoes this disproportionate use with the average stay being 43 days, but median at 14 days. For other programs, the number of individuals participating is far less

(between 100 and 1500) as well as the number of times individuals frequent these programs. For example, across the board the average and median service count for non-emergency shelter programs is 1 occurrence; meaning that individuals generally have only one recorded encounter with that homeless support program. The number of days participating in these programs, however, is much longer than for emergency shelters where the cumulative average/median number days range between 50 and more than 600 depending on the program.

Table 35: Homeless Program Participation by Service Count and Number of Days

Homeless Program	Distinct Individuals	Average Service Count	Median Service Count	Average Number of Days	Median Number of Days
Emergency Shelter ^a	6,132	12	2	43	14
Day Shelter ^b	510	1	1	338	327
PH – Permanent Supportive Housing ^c	360	1	1	672	432
PH – Rapid Re-housing ^d	1571	1	1	200	182
PH – Housing with Services ^e	102	1	1	349	365
Transitional Housing ^f	648	1	1	138	103
Homelessness Prevention ^g	602	1	1	148	135
Street Outreach ^h	243	1	1	277	257
Services Only ⁱ	425	1	1	213	138
Other Services	974	1	1	106	51

^a Emergency shelter: Any facility, the primary purpose of which is to provide temporary or transitional shelter for the homeless in general or for specific populations of the homeless

^b Day shelter: Offer daytime facilities and services (no lodging) for persons who are homeless.

^c PH – Permanent Supportive Housing: Supportive housing is an evidence-based housing intervention that combines non-time-limited affordable housing assistance with wrap-around supportive services for people experiencing homelessness, as well as other people with disabilities.

^d PH – Rapid Re-Housing: Rapid re-housing is an intervention designed to help individuals and families quickly exit homelessness and return to permanent housing. Rapid re-housing assistance is offered without preconditions — like employment, income, absence of criminal record, or sobriety — and the resources and services provided are tailored to the unique needs of the household.

^e PH – Housing with Services: Rapid Re-Housing with support services.

^f Transitional housing: Designed to facilitate the movement of homeless individuals and families to permanent housing within a reasonable amount of time (usually 24 months). Transitional housing includes housing primarily designed to serve deinstitutionalized homeless individuals and other homeless individuals with mental or physical disabilities and homeless families with children.

^g Homelessness Prevention: Activities or programs designed to prevent the incidence of homelessness, including, but not limited to: (1) short-term subsidies to defray rent and utility arrearages for families that have received eviction or utility termination notices; (2) security deposits or first month's rent to permit a homeless family to move into its own apartment; (3) mediation programs for landlord-tenant disputes; (4) legal services programs that enable representation of indigent tenants in eviction proceedings; (5) payments to prevent foreclosure on a home; and (6) other innovative programs and activities designed to prevent the incidence of homelessness.

^h Street Outreach: Essential Services related to reaching out to unsheltered homeless individuals and families, connecting them with emergency shelter, housing, or critical services, and providing them with urgent, non-facility-based care. Eligible costs include engagement, case management, emergency health and mental health services, transportation, and services for special populations. See 24 CFR 576.101.

ⁱ Services Only: When only support services are provided and not housing

We can further unwrap program participation, as shown in [Figure 4](#) and [Figure 5](#), by examining the differences by gender and age. [Figure 4](#) shows the individual count of program participation; counting each person-program interaction only once. Overall, we see that more men than women are participating in homeless programs and, consistent with the aggregate view, emergency shelter is the most frequented program for both men and women. Among men, emergency shelter use is greater for those between the ages of 26 and 65, with the greatest presence across all programs for men between the ages of 46 and 55. While young adult men between the ages of 19 and 25 are fewer in number, their use of emergency shelter far exceeds that of other programs. It is only for boys 18 years old and younger do we see greater interaction with permanent housing support in the form of rapid re-housing, presumably as part of a family unit versus the independence tends to come after the age of 18. Among women, interaction with homeless support programs is more varied across all age groups. While emergency shelter is the most often engaged, there is an uptick in participation for day shelters and rapid re-housing and across age groups relative to men.

Whereas [Figure 4](#) details the raw person counts, [Figure 5](#) displays the cumulative times individuals interacted with each homeless program. [Figure 4](#) shows how many times homeless persons, particularly men, frequently interact with homeless programs. An overwhelming majority of service counts in the HMIS data are for short-term emergency shelter visits. What is not captured, however, is the duration of time spent in these other programs. As evident in [Table 35](#), the span of days is much shorter for the emergency shelter program than other, longer term assistance programs, and thus more numerous in the data.

It is important to note that participation in these programs is not mutually exclusive (i.e., a person could be in more than one program at the same time) and, as shown in [Table 35](#), the duration in programs also varies considerably. As such, identifying someone as a high utilizer of homeless support services is less clear than for someone's frequent interaction with the jail system. The U.S. Department of Housing and Urban Development (HUD) defines chronic homelessness in cases where the Head of Household has a disability AND has been homeless for at least 12 consecutive months or has had 4 or more episodes of homelessness in the last 3 years totaling 12 months or more. The criteria HUD uses to identify and verify chronic homelessness, however, is precise and not all indicators can be pulled from the HMIS data, thus limiting our ability to exclusively determine whether someone is chronically homeless strictly from the current set of data. In many cases, the documentation of chronic homelessness may come from street outreach workers or other homeless service workers, or from non-HMIS participating agencies.

Figure 4: Person Count of Homeless Program Participation by Gender and Age

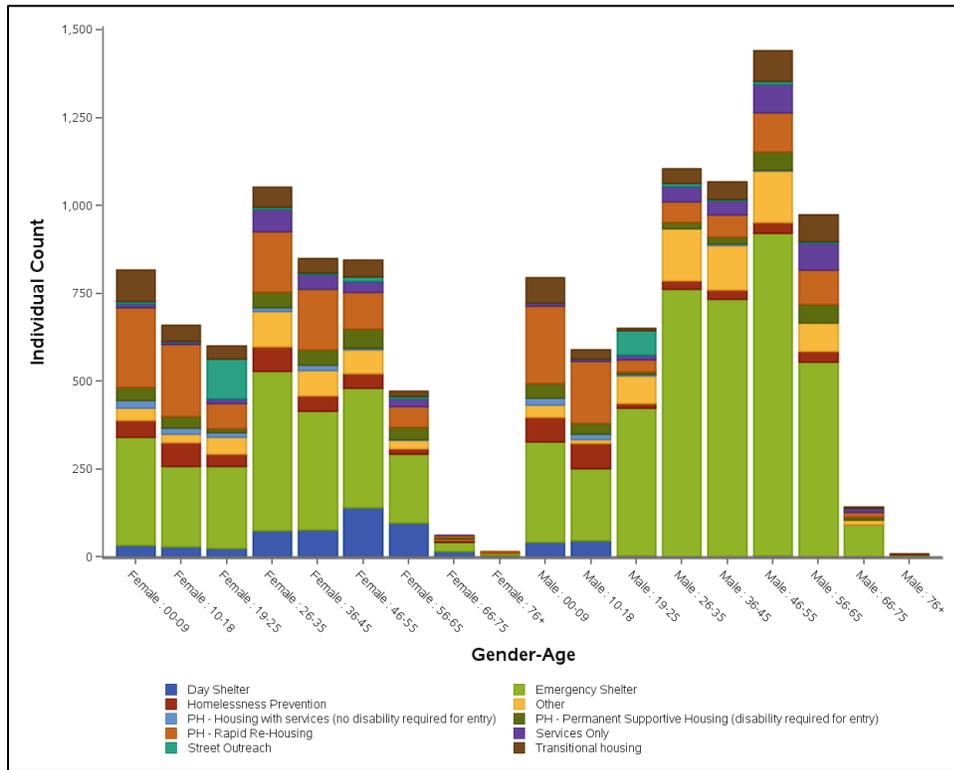
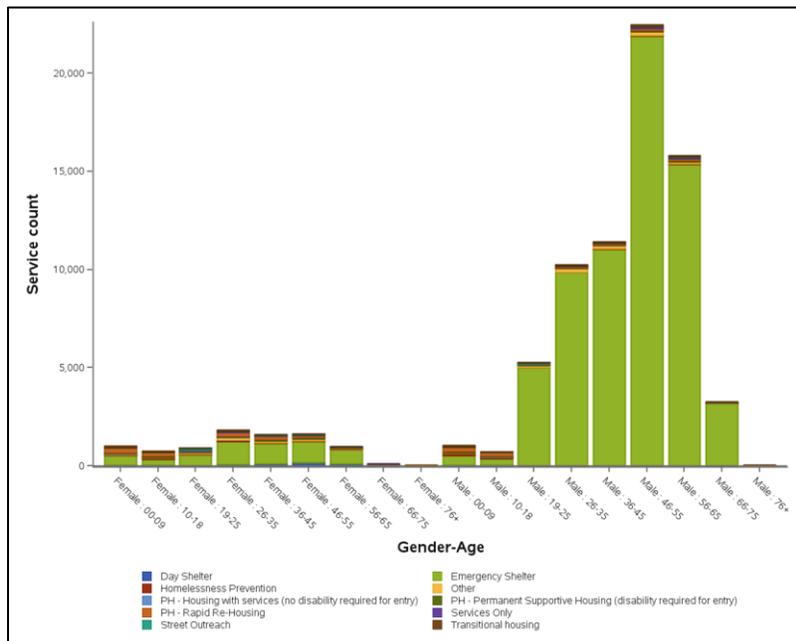


Figure 5: Service Count of Homeless Program Participation by Gender and Age



While it is beyond the scope of this report to evaluate chronic homeless using the HUD definition, we can achieve some understanding of high utilization of homeless services by leveraging the 95th percentile cut-offs for the count of services as well as the cumulative days across all programs. As [Table 36](#) shows, the 95th percentile values for services is a count of 36 and the cumulative days is 472. Using these cut-offs, a total of 800 individuals were identified as being above the 95th percentile for services

count (401 persons), cumulative days (359 persons), or on both dimensions of utilization (40 persons). While this report does not consider high utilization separately for men and women, percentile distribution of service counts and cumulative days by gender in [Table 36](#) further illustrate the disparate usage of homeless programs by men (more service counts and fewer cumulative days) and women (fewer service counts and more cumulative days).

Table 36: Percentile Distribution of HMIS Homeless Program Service Counts and Cumulative Days of Participation

	Service Count	Cumulative Number of Days	Female		Male	
			Service Count	Cumulative Days	Service Count	Cumulative Days
Percentiles						
50th	2	42	1	54	2	34
75th	4	153	2	194	8	127
90th	16	366	5	412	30	314
95th	36	472	7	539	62	426
99th	140	865	17	1102	200	694
Maximum	609	7,407	64	7,407	609	2,628

[Table 37](#) details the demographic differences between the high utilization groupings and the total identified HMIS population. For the High Services Count group, individuals tend to be of older age (75% are between the ages of 36 and 65) and almost exclusively male (98%). The High Cumulative Days group, on the other hand, is slightly more female (58%) and much younger (62% are 55 years and under with 25% under the age of 18). Both high utilization groups are comprised of roughly 10% veterans and 30% in the High Service Count group and 40% in the High Cumulative Days group have a self-reported diagnosis of some type of mental illness. When the two utilization groups (service counts and cumulative days) are combined into a single High Utilizers group, many of the defining characteristics of each group are averaged out.

[Table 38](#) further highlights the distinctness of each high utilization group by detailing the program interaction counts. For the High Services Count group, all individuals had interactions with the emergency shelter program, which is consistent with the significant presence of older males seen in [Figure 4](#). Interactions among the High Service Count group with other programs tended to be modest and dispersed. For the High Cumulative Days group, while more than half of the individuals had at least one interaction with the emergency shelter program, participation is far more varied across all programs and at much higher levels than compared to the High Services Count group.

These findings illustrate the important differences among homeless programs and the populations they support. Emergency shelter programs, for example, offer short-term support and thus frequented more often, especially by older-aged men. Other longer-term programs, such as transitional housing and permanent support housing programs, assist individuals for more sustain periods of time (in some cases up to 2 years) and mostly assist younger groups and women; and presumably families. Future analyses may consider high utilization within the various homeless programs as well as the HUD definition of chronic homelessness.

Table 37: Demographic Profile of HMIS Homeless Population and High Utilizers

	HMIS Population	High Services Count	High Cumulative Days	High Utilizers
<i>Distinct Individual Count</i>	8,834	441	399	800
Age Categories				
18 years and under	22.44%	0.00%	24.56%	12.25%
19 to 25 years	10.77%	7.26%	13.03%	9.63%
26 to 35 years	18.45%	14.06%	11.28%	12.75%
36 to 45 years	16.26%	17.23%	13.03%	15.38%
46 to 55 years	18.52%	34.01%	21.05%	27.75%
56 to 65 years	11.29%	23.13%	14.79%	18.88%
66 to 75 years	1.69%	4.31%	2.01%	3.25%
76 or more years	0.19%	0.00%	0.25%	0.13%
Unknown	0.41%	0.00%	0.00%	0.00%
Sex				
Male	58.38%	98.41%	41.85%	70.63%
Female	41.11%	1.36%	57.89%	29.13%
Other or Unknown	0.51%	0.23%	0.25%	0.25%
Race				
White	27.42%	23.81%	19.30%	21.38%
Black or African American	69.76%	73.92%	79.20%	76.88%
Asian/Other Pacific Islander	0.76%	0.45%	0.25%	0.38%
Other Race or Multi-Racial	2.06%	2.27%	1.25%	1.36%
Ethnicity				
Hispanic or Latino	4.56%	5.44%	3.26%	4.25%
Non-Hispanic/Non-Latino	94.09%	94.56%	95.99%	95.38%
Missing, Unknown, or Refused	1.35%	0.00%	0.75%	0.38%
Veteran Status	9.25%	10.20%	9.77%	10.13%
Mental Health Diagnosis	20.00%	29.93%	39.85%	34.50%

Table 38: Interactions with Homeless Programs among High Utilization Groups

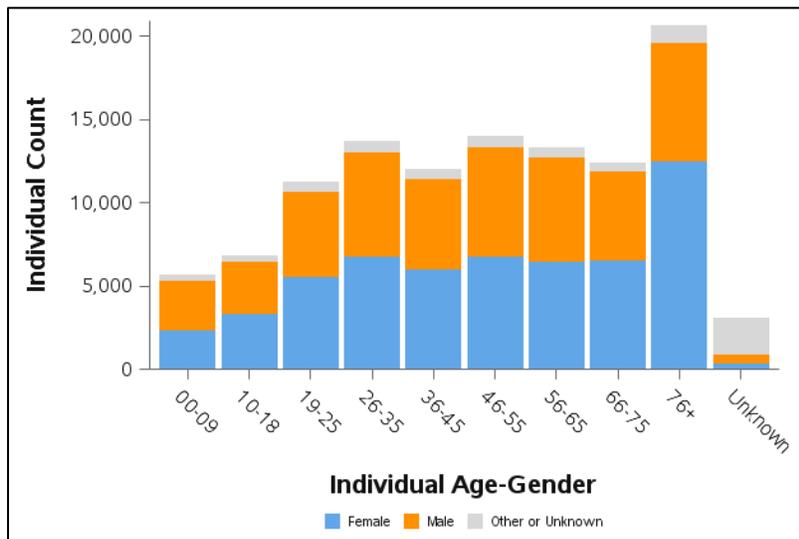
Homeless Program	High Services Count	High Cumulative Days	High Utilizers
Emergency Shelter	441	230	631
Day Shelter	6	105	107
PH – Permanent Supportive Housing	10	119	127
PH – Rapid Re-housing	28	159	180
PH – Housing with Services	0	15	15
Transitional Housing	32	64	91
Homelessness Prevention	2	26	28
Street Outreach	18	64	74
Services Only	43	104	132
Other Services	41	31	66

6.3 EMS Interaction Data

The EMS data, unlike the Jail and HMIS data, reaches a much broader segment of the population as interactions with EMS are more likely to occur by random events (i.e., an accident or injury). The breadth of the EMS data is evident by the 112,148 unique individuals identified who had at least one EMS interaction during the 20-month period of data.⁵ In total, the data covers 165,785 distinct incidents. And while random events and injuries increase the possibility of interacting with EMS services, [Figure 6](#) shows that older individuals, who are at increasing risk for medical emergencies, steadily interact with EMS more frequently until sharply peaking after 75 years of age. In terms of gender differences by age, with the exception of the 9 and under age group where boys are more likely to have an EMS incident than girls (54% versus 44%), females overall are slightly more likely to have an EMS incident with the greatest difference at among the over 75 age group (61% for females versus 35% for males).

⁵ Due to conservative logic requiring more than just name and date of birth to link individuals in the EMS data together (see Section 3 for details), including additional matching on sparsely populated measure for social security number, phone, or address, the same individual may be counted more than once. Because key data fields used for entity resolution are often missing in the EMS data or incomplete compared to the Jail or HMIS data, record linking was less complete.

Figure 6: Individual Counts of EMS Interaction by Age and Gender



Beyond examining who interacts with EMS, a slightly more difficult question is assessing how frequently persons rely on EMS services and, in particular, distinguishing between high episodic utilization and chronic use over time. For the purposes of this report, each distinct date/time interaction a person has with EMS is considered an incident. Applied to Wake County EMS's definition of high utilization, an individual who has 4 or more incidents during a rolling 30 day period would be considered a "high utilizer." However, this definition does not exactly capture chronic use, but instead what we might call high episodic utilization. In other words, several EMS incidents in a relative short period of time, but not necessarily frequent use over a sustained period of time.

The concept of episodic utilization is illustrated by *Person A* in [Figure 7](#) with 5 incidents in the first 30-day segment and then only 1 incident in two later 30-day periods. Thus, *Person A* meets the EMS definition of a high utilizer with 4 or more EMS services during a 30-day period and, as such, we can categorize it as 'high' episodic utilization. *Person B* in [Figure 7](#), however, does not meet the EMS criteria as there is no single 4-plus incident episode in a 30-day span but does have a total of 9 EMS incident (2 more than *Person A*) over the same length of time. We can classify this type of consistent interaction with EMS as *chronic* utilization. *Person C* offers an example of how both episodic and chronic use could present itself; one 5-incident episode and 7 incidents spread out over the next 150 days.

In this report, we are interested in the chronic use of EMS services rather than episodic interaction, thus utilization is calculated by the total number of incidents outside of high utilization episodes. A high utilization episode is defined using the EMS chronic utilizer criteria of 4 or more incidents in a 30-day rolling window. Our calculation down weights episodic occurrences and the clustering of incidents within by counting an episode as 1 incident.

The utilization count formula is the following:

$$\text{Utilization Count} = [\text{Total incidents}] - [\text{Number of Incidents within Episodes}] + [\text{Number of Episodes}]$$

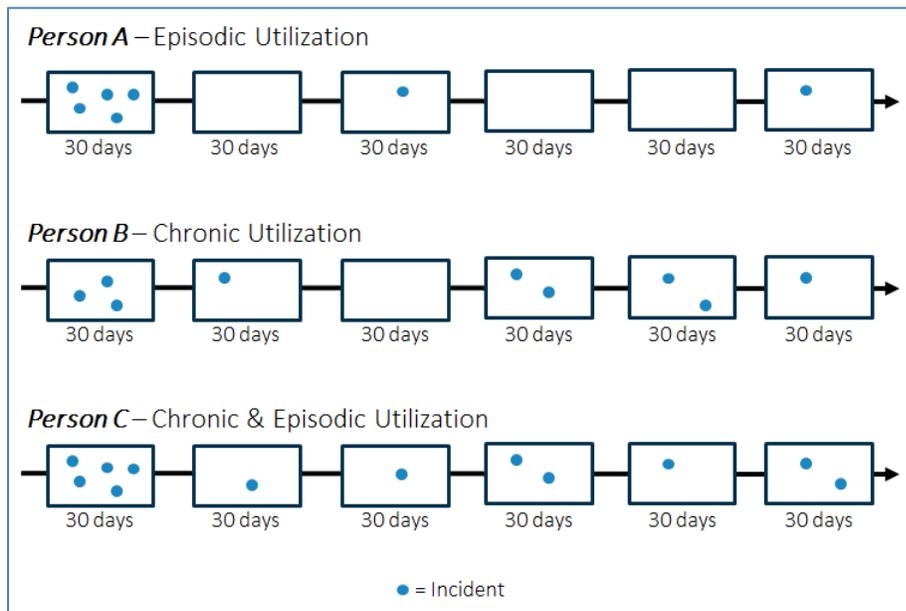
Where:

- *Total incidents* is the total number of distinct interactions with EMS during the 20-month period
- *Number of incidents within episodes* is the cumulative total incidents that fall within all identified episodes

- *Number of episodes* is the total number of groupings of 4 or more incidents in a 30-day rolling window

Applied to *Person C* in [Figure 7](#), for example, there are 12 total incidents, 5 incidents within episodes, and 1 episode for a utilization count of 8 ($12 - [5] + [1]$). For *Person B*, who has 9 total incidents and zero episodes, the utilization count is 9 ($9 - [0] + [0]$). *Person A* has a utilization count of 3 with 7 total incidents, 5 incident within episodes, and 1 episode ($7 - [5] + [1]$).

Figure 7: Definitions of EMS Utilization



[Table 39](#) shows the percentile breakdown of the utilization count for the EMS population. While the utilization count measure takes into account the EMS high utilizer definition (termed in this report as episodic utilization), it is important to note the EMS definition identified just over 1,000 individuals or 1% of the EMS patients with 80% of the group having only 1 episode. For the full EMS utilization count calculation, 90% of the individuals in the EMS data had 2 or fewer interactions with EMS transports. At the 95th percentile, the number of EMS interactions increases to 4. As with the Jail and HMIS data, we used the 95th percentile value as the cut-off point for identifying high utilization where any individual with 5 or more EMS interactions was classified as a “familiar face.”

Table 39: Distribution of EMS Utilization Counts

Percentiles	EMS Utilization Count
50th	1
75th	1
90th	2
95th	4
99th	7
Maximum Count	26

[Table 40](#) shows a demographic comparison between the high utilization group with 5 or more EMS interactions and Wake County EMS transport population. The high EMS utilization group consists of 3,532 out of the total 112,148 persons identified in the EMS data. Despite the group's more frequent use of EMS services and smaller size, most of the demographic characteristics of sex, race, and ethnicity remain proportionally consistent with the EMS population overall. For the high utilization group, as with the EMS population, an individual's age is a key factor associated with EMS services as older individuals are more likely to interact with EMS. The trend toward older individuals' usage for the high utilization group is more prominent as evident in [Figure 8](#), particular among the 76 years or more age category, who are 36% of the high utilization group compared to 18% for the general population. EMS usage of men and women within age categories consistent with the pattern observed for the general population; greater usage by women than men except for under 10 years of age group.

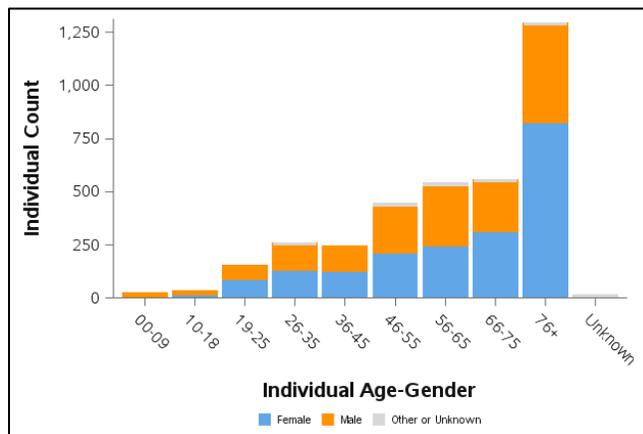
While beyond the scope of this report, future exploration of the EMS data may consider the geo-spatial distribution of the EMS calls and whether the high utilization group is clustered within certain areas of Wake County. Additional consideration should also be given to criteria for resolving person identities within the EMS data, where key measures used in the match process, such as social security number, address, and phone number, or often missing or incomplete and thus limit potential identity matches.

Table 40: Demographic Profile of EMS Population and High Utilizers

	EMS Population	High EMS Utilization
<i>Distinct Individual Count</i>	112,148	3,532
Age Categories		
18 years and under	11.01%	1.50%
19 to 25 years	9.97%	4.19%
26 to 35 years	12.16%	7.28%
36 to 45 years	10.64%	6.85%
46 to 55 years	12.41%	12.49%
56 to 65 years	11.80%	15.20%
66 to 75 years	11.00%	15.69%
76 or more years	18.32%	36.49%
Unknown	2.69%	0.31%
Sex		
Male	43.55%	42.98%
Female	51.57%	56.48%
Other or Unknown	4.88%	0.54%
Race		
White	54.27%	58.21%
Black or African American	31.97%	39.67%

	EMS Population	High EMS Utilization
Asian/Other Pacific Islander	1.23%	0.54%
Other Race or Multi-Racial	12.53%	1.59%
Ethnicity		
Hispanic or Latino	5.92%	1.22%
Non-Hispanic/Non-Latino	88.27%	98.19%
Missing, Unknown, or Refused	5.81%	0.59%

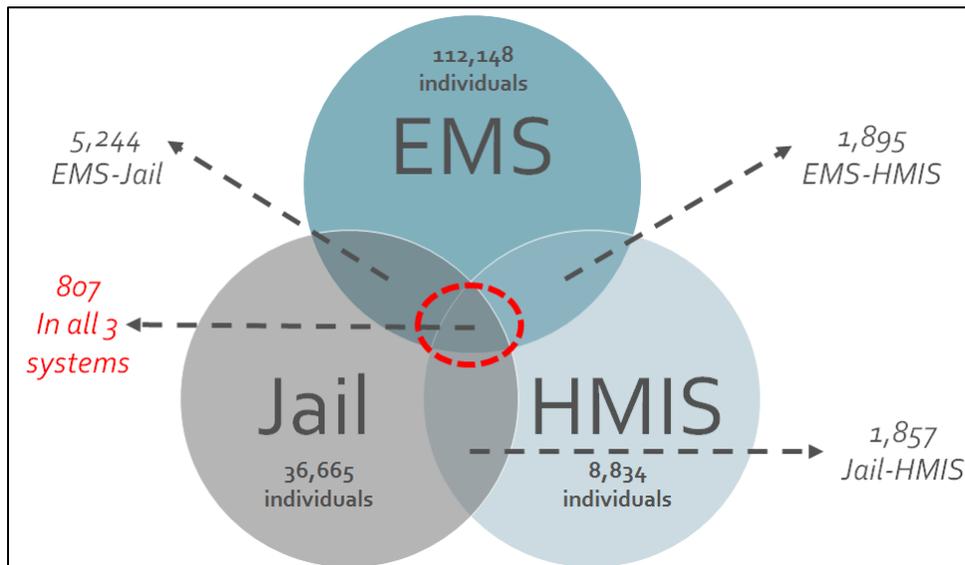
Figure 8: High EMS Utilizers EMS Interaction by Age and Gender



7 Intersection of Jail, EMS, and Homelessness

Having explored in Section 5 the high utilization of services within Wake County’s Jail system, EMS, and homeless assistance programs respectively, this section details the intersection of all three data systems, which includes a total of 149,458 distinct individuals. [Figure 9](#) summarizes the count of individuals within each system as well as the number of distinct individuals identified in two or all three data systems. The resolution of entities across the agency datasets identified a decent amount of overlap considering the disproportionate number of individuals in each system. The greatest amount of overlap was found between the EMS and Jail data with 5,244 individuals, followed by over 1,850 persons identified in the EMS-HMIS and Jail-HMIS data systems respectively. There were 807 distinct individuals identified with at least one incident in each the Jail, EMS, and HMIS data.

Figure 9: Interaction of Wake County Jail System, EMS, and HMIS Homeless data



While future analyses may consider an in-depth exploration of these two-system interactions, this section focuses on the 807 individuals found in all three data systems. [Figure 10](#) shows the breakdown of the intersecting population by age and sex. Exploring the distribution based on age categories we see that more than 95% of the intersecting population are between the ages of 19 and 65 and roughly 70% between the ages of 26 and 55. Also apparent are two peaks occurring for the 26-35 and 46-55 age groups, which together account for more than half of the intersecting population. In the first peak of 25-35 year olds, women appear to be driving much of the elevation, where they make up 35% of the age category – their greatest representation across all age groups. The presence of women declines thereafter going from 30% to 5% between the 36-45 and 56-65 age categories. The second peak, occurring for those ages 46-55, is driven by the large number of men who make up 85% of the age category and 29% of males overall. Men, who comprise 75% of the intersecting population overall, outpace their female counterparts across each age category (except for the nominally smaller 10-18 age group) by 60% or more. Thus, the intersecting population is made up predominately of men between the ages of 26 and 55.

Figure 10: Distribution by Age and Sex of Intersecting Population

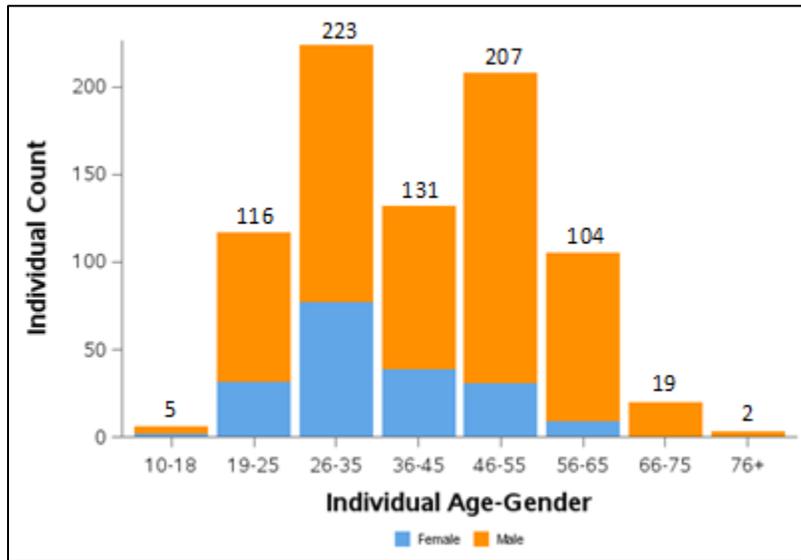


Figure 11 shows the same age category distribution for the intersecting population, but broken out by sex and race. Consistent with population dynamics observed in the jail and HMIS data, individuals found in all three data systems tend to be black or African American and these differences tend to intensify among older age groups, particularly for black or African American men who 46% of the overall intersecting population and close to or above 50% in each age category. The conditions and challenges that plague the black or African Americans community are well documented and beyond the scope of this report to explain. The findings in this report do, however, further illustrates the vulnerability of this population, particularly black or African American men.

Figure 11: Demographic Distribution of Interacting Population by Age, Sex, and Race

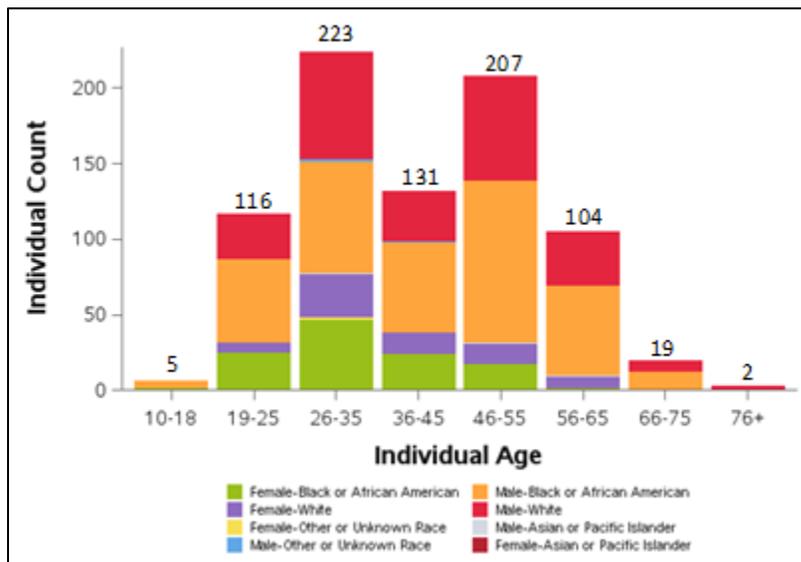


Figure 12 compares the types of criminal charges found among the intersecting population relative to the offenses discussed in Section 6.1 for those identified as “familiar faces” to the jail system as well as the total jail population. The charts show the UCR categories as a percent of bookings. Across all groups, bookings are generally for low level offenses with more than 70% bookings involving only misdemeanor

level charges. Comparing across the charts, there are several common charges across all three groups, such as a *contempt of court*, *perjury*, or *court violation* charges; *simple assaults*, and *larceny* which occur in similar percentages within each group at roughly 18%, 8%, and 7% respectively. In the case of *driving while impaired*, the intersecting population is similar to familiar faces with far fewer charges for DWI compared to the total populations (less than 2.5% versus 9% for the total population).

The intersecting population, however, is itself distinctive compared to the other groups where the percent of *drug violations* is less (6% versus 10% and 9%), but higher for charges of *trespassing* (11% versus 7% and 3%), *city ordinance violations* (6% versus 2% and 1%), and *disorderly conduct* (6% versus 3% and 2%) relative to familiar faces and total jail population. These charges are mostly likely the associated with being homeless or, at minimum, a consequence of an unstable housing situation that is likely to force individuals to spend more time on the streets and in public places. It is likely this increased public exposure is at least partially driving the higher rates of trespassing and ordinance violation charges.

Anecdotal evidence suggests that the intersecting population is more likely to have longer lengths of stays in jail as a form of temporary shelter and reprieve from the outside elements. As [Table 41](#) shows, this narrative is supported by the data where the average length of stay among the intersecting population's jail bookings is 18 days compared 11 days for both "familiar faces" and overall jail population. Even at the median values, which is less susceptible to data fluctuations, the intersecting population continues to show a lengthier jail stay of 8 days. In fact, moving across the distribution of jail stays, the bookings for the intersecting population involve longer stays in jail.

The unsettled living conditions that underlies the intersecting population is further highlighted by examining the types of homeless programs this group frequents. Consistent with the earlier findings in Section 6.2, [Table 42](#) shows that the majority of intersecting population (more than 85%) had some interaction with the emergency shelter program, which tends to offer short-term assistance based on the average cumulative days of participation, compared to other, longer-term programs which show far less participation. The data shows that not only was contact with emergency shelters the greatest among homeless programs, but that the frequency with which individuals utilized emergency shelters was much greater compared to other programs. For instance, the average service count for emergency shelters among the intersecting population was 24 versus other programs, which average a one-time count. Thus, from this summarized perspective, the intersecting population is most likely to have contact with emergency shelters and do so frequently.

What is not clear, however, is the chain of events that underlie our summarized understanding of the intersecting population. It is possible, for instance, that the reason for greater emergency shelter participation is because individuals with a criminal past are prevented from accessing programs that offer longer-term support. And, while we might correctly assume that persons who experience episodes of homelessness many seeks refuge in the jail system, especially during the heat of the summer or cold of the winter, cannot decipher these patterns from summarized analyses alone. An understanding of individuals at this level requires us to explore the timeline of events as part of a case study.

Figure 12: UCR Categories as a Percent of Bookings for the Intersecting Population, Jail Familiar Faces, and Jail Population

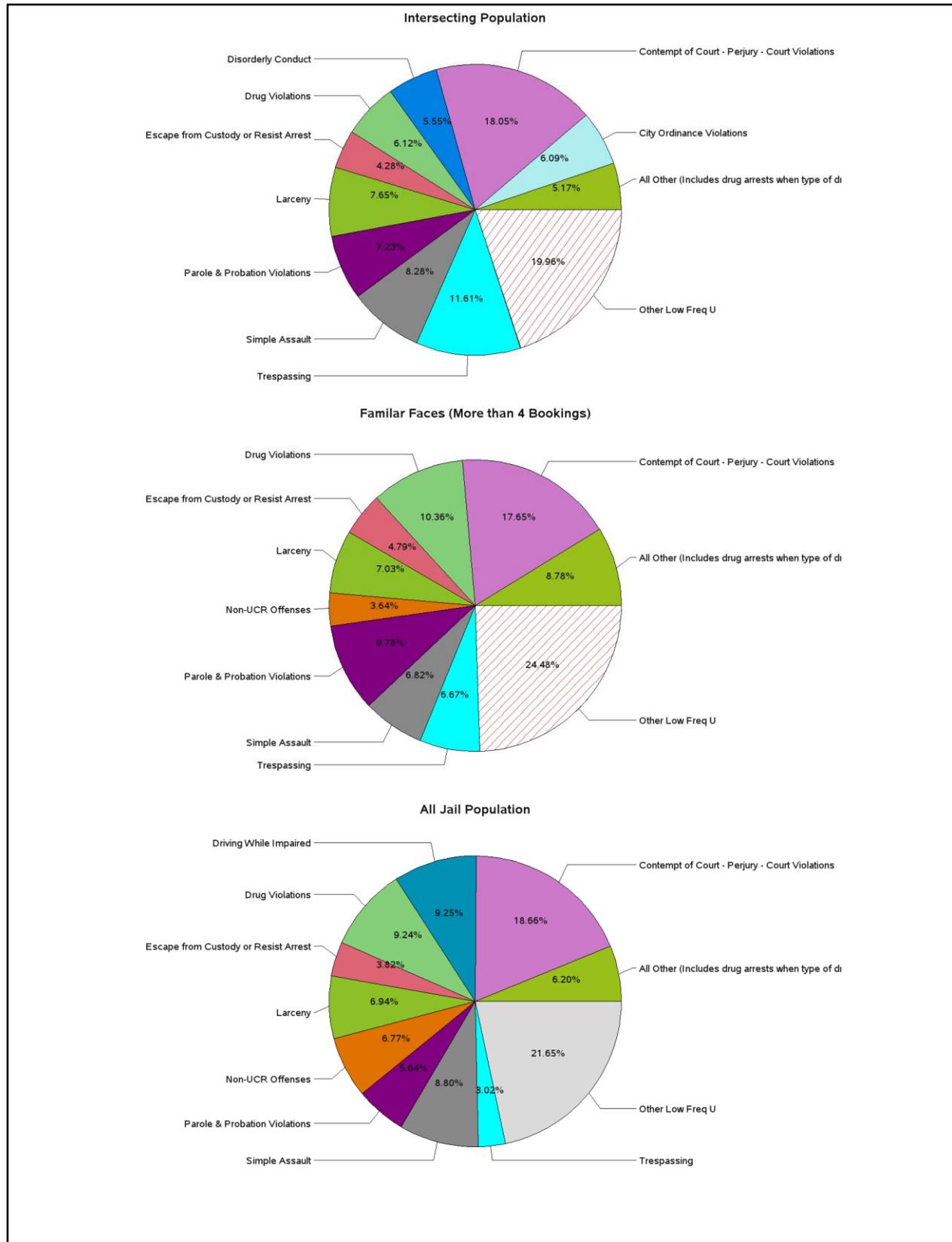


Table 41: Booking Length of Stay in Wake County Jail (in Days)

	Intersecting Population	Familiar Faces	Total Jail Population
Number of Bookings	2,214	9,114	57,735
Average Length of Stay (Days)	18	11	11
Percentiles			
50th (median)	8	2	1
75th	21	10	6
90th	44	29	28
95th	74	50	68
99th	150	120	163

Table 42: Intersecting Populations' Homeless Program Participation by Service Count and Number of Days

Program	Distinct Individuals	Average Service Count	Median Service Count	Average Days	Median Days
Emergency Shelter	707	24	6	50	18
Day Shelter	36	1	1	232	161
PH – Permanent Supportive Housing	20	1	1	369	186
PH – Rapid Re-housing	60	1	1	224	203
PH – Housing with Services	1	1	1	<missing>	<missing>
Transitional Housing	42	1	1	94	76
Homelessness Prevention	21	1	1	138	140
Street Outreach	32	1	1	333	324
Services Only	41	1	1	208	146
Other Services	116	1	1	82	20

8 Timeline of Events: A Case Study

Developing a detailed case study for each of the 807 individuals found in all three data systems is beyond the scope of this report. Instead, we aim to highlight the type of events and patterns we might see by examining the sequence of interactions with the Jail system, EMS, and homeless programs in the HMIS data. Future analysis will be able to produce richer timelines with expanded history from the existing data sources as well as the incorporation of data from other sources such as hospitals.

A primary objective of this report, however is to understand who is repeatedly cycling through local jails, emergency services, and homeless assistance programs. We can begin to narrow the intersecting group of individuals by leaning on the Jail and EMS definitions of high utilization, also referred to as “familiar faces.” A familiar face in the Jail data was anyone who had entered the Jail system 5 or more times, and for EMS a high utilizer was anyone with 5 or more incidents outside of episodic clusters (see Section [6.1](#) and [6.2](#) for further explanation). Utilization of homeless services was not leveraged because of the challenges of identifying high utilization given the dynamic interplay between program services counts and cumulative days spent in programs, which vary greatly by program and for men and women.

[Table 43](#) shows different Jail and EMS incident combinations and the number of individuals who meet those incident thresholds. In the first row, for instance, we see that there are 170 individuals out of the 807 found in all three data systems who have only 1 jail incident and 1 EMS incident. The incident combinations apply to both the Jail and EMS count should be read as *2 or more Jail incidents and 2 or more EMS incidents*. Also included are the average and median values for HMIS program service counts and cumulative days for each Jail-EMS incident combination. From a case management standpoint, one could examine the event histories starting at any of the **X or more** combinations depending on resources and time. The **5 or more Jail and EMS incidents** category is noteworthy, however, as 5 or more interactions was the high utilization threshold (i.e., above the 95th percentile) identified in Jail and EMS data, respectively. Thus, from the group of 800 in all three systems, there are 26 individuals who are high utilizers of both the Jail system and EMS services.

Table 43: Person Count by Jail and EMS Incident Combinations

Jail and EMS Incidents	Person Count	HMIS Program Services		HMIS Cumulative Days	
		Average Count	Median Count	Average Days	Median Days
1 Incident	170	18	3	107	32
2 or more Incidents	271	21	7	86	27
3 or more Incidents	127	22	9	69	21
4 or more Incidents	62	21	7	53	14
5 or more Incidents	26	18	8	44	11
6 or more Incidents	14	22	12	50	14

In narrowing the intersecting population down to 26 individuals, we must be mindful to protect the identities of this small group, which includes traits that may be combined with other personal information to identify an individual. Consequently, we do not provide a demographic breakdown of this group in order to ensure identities are not compromised. Moreover, discussion of timeline events with the different agencies are stripped of certain details. The type of information withheld, for example, from EMS events include the service location, time of incident, and transport destination. For the

homeless program events, the specific names of programs and shelter are suppressed. Jail incidents are limited to arrest date and charges as well as release date.

This report limits the exploration of timeline events to one case study as many of them share similar patterns between and within agencies (e.g., many with charges for trespassing, begging, and failure to appear on past charges), with some having longer timelines than others. By focusing one case study with a relative robust history of events, we are better able to highlight trends within and between agencies as well as note gaps in time between events.

Our case study example is for a *male between the ages of 55 and 65* and from the group of 26 identified as being high utilizers of both the jail system and EMS services. [Table 44](#) provides the complete history of interactions with the jail, EMS, and HMIS program services for our 24-month study period. As noted in [Section 3.4](#), the analytic data window includes 4 additional lead-in months of Jail data starting in January 2015, while the incidents from the EMS and HMIS data start in May 2015. Our male case study has a total of 47 total interactions with the three agencies, which includes 9 EMS incidents and 19 events apiece with the jail system and homeless services tracked by HMIS. Among the 19 jail events, there are 38 misdemeanor charges and no felonies with an average length of stay of 9 days with a maximum stay of 25 days. Of the 9 EMS incidents, 7 incidents resulted in transport with no lights or siren, 2 for assistance, and one incident with no treatment or transport. While our male case study is a high EMS utilizer with 5 or more services, there were no high episodic periods events (i.e., 4 or more incidents during a rolling 30 day period) when looking across the history of EMS event. Finally, among the 19 HMIS service events, all except one was for emergency shelter stays where the average length of stay was 2 days and maximum stay of 6 days.

A deeper examination into the events over time reveal a steady pattern of misdemeanor arrests for begging, intoxication and disruption, and trespassing; a total of a 7 jail bookings between mid-January and late July of 2015 with jail stays ranging from 0 to 12 days. In half cases, jail events are separated by 8 days or less while roughly 40 days separate the other booking events. The first EMS event occurs on July 30 with a transport but no lights/siren, which was 7 days after being released from jail following an arrest for trespassing. This first EMS event starts a five-event alternating pattern of EMS transports (no lights/siren) and bookings into jail on charges of being intoxicated and disruptive, ending with an EMS transport on September 22. During that nearly 2-month stretch, events were separated by as few as 4 days and as many as 18 days.

Our male case study does not appear in either of the three systems until 62 days after the September 22 EMS transport when booked into jail for 2 days on November 23 for a misdemeanor charge of *failing to appear in court for a prior charge*. This nearly 2-month gap is unaccounted for in our current view of the resolved data. There are a number of potential reasons for the absence of activity. On the one hand, it is reasonable that activity is simply going undetected because of interactions elsewhere. It is possible, for example, that our case study example was admitted to the hospital for all or some period of time following the last EMS transport. Another possibility is that identity of our case study is not being picked up in the data, perhaps because of the use of an alias or the strict criteria of the entity resolution techniques. However, we see another lengthy gap of 48 days following an EMS transport on February 14, 2016 and then a Jail booking on April 2, so perhaps some type of medical admittance is occurring that renders our case study invisible.

It is after this first 62-day gap after the EMS transport on September 22 that we see two separate bookings for charges on *failing to appear in court for a prior charge* and third booking for trespassing in the span of 22 days (between November 22 and December 15). The charges for failing to appear in court seems to indicate that our case study's criminal past is catching up with him, and that arresting charges are more the result of technical court violations than any immediate criminal activity. It is after this

booking on charges of trespassing and subsequent jail 7-day stay that we see the first interaction with homeless services for a recovery program on December 30. While the recovery program participation was a one-day event, we might suspect that it represents some attempt at intervention – perhaps a condition of release from jail given our case study’s repeated history intoxication.

Bolstering the notion that the first HMIS services program interaction was indeed an attempt to break a cycle of recidivism is a sequence of four consecutive emergency shelter events from January 9 to January 26, 2016. While none of the shelter stays were long (1 to 6 days), the days between each event (all at the same shelter) were within a few days suggesting some sort of reliable assistance was available. If in fact there was some type of housing support available, our case study’s remaining history shows participation was regularly interrupted with long bouts of jail stays and EMS events (late January through early May 2016), followed by a cluster of emergency shelter events (early to mid-June 2016), more jail stays and EMS events (late June to mid-October 2016), and finally an uninterrupted sequence of shelter stays (10 service events in all) from November 7 to the end of December 2016.

The jail events in between the groups of shelter stays show the same collection of charges that include intoxication and disruption and trespassing. However, the last grouping of jail events (late June to mid-October 2016), which include four separate bookings with charges of *failure to appear on prior charges*, resulted particularly lengthy jail stays that ranged from 7 to 25 days and totaled 72 days over 5 booking events. It’s not clear whether these longer jail stays were the result of repeated offenses and growing criminal record and/or an attempt to hold the individual in jail until the scheduled court appearance – avoiding additional ‘failure to appear’ charges.

Much can be understood from our case study. We were presented with a person who, based on the collection of arresting charges of begging, trespassing, and intoxication, is homeless and suffers from some form of substance abuse. A deeper examination of the HMIS data also reveals that our case study example has been identified as having some type of mental illness (the specifics of which are unknown). Trends of criminal offense over time do not suggest an escalating pattern offense, either of non-violent to violent crimes or misdemeanor to felony offense. We do observe, however, that unresolved criminal offenses, resulting in charges of failing to appear in court, has an escalating impact on being arrested in the future and longer jail stays. While there is some evidence of seasonality of jail and HMIS services events with clustering of activities during the cold and warmer months, additional data (more than our identified group of 26) is necessary to confirm such trends. We also see what looks to be an attempt to break a cycle of recidivism through participation in homeless support programs (i.e., recovery program and emergency shelter services), but the end result was oscillating periods of activity between the Jail and EMS systems and homeless supports. Section 9 discussed the potential for coordinated activities among the three agencies to better serve this vulnerable population.

In sum, by bringing the Jail, EMS, and HMIS homeless data together, we were able identify high utilizers within each agency, identify almost 150,000 unique individuals across those agencies, and then reduce that number down to roughly 800 who interacted at least once with all three systems. From the group of roughly 800 cross-agency individuals, we were able to reapply the high utilization thresholds found for each agency to identify a manageable set of individuals whose event history timelines offer insights on when and where to intervene

Table 44: Timeline of Jail, EMS, and HMIS Events for Familiar Face Male between Ages of 56-65

Agency	Event	Event Start	Event End	Event Length	Days Between Events
Jail	Arrest Misdemeanor: Aggressively Begging Begging w/o a Permit Intoxicated and Disruptive	14Jan2015	22Jan2015	8	--
Jail	Arrest Misdemeanor: Aggressively Begging Begging w/o a Permit Intoxicated and Disruptive Second Degree Trespassing	04Mar2015	12Mar2015	8	41
Jail	Arrest Misdemeanor: Intoxicated and Disruptive Second Degree Trespassing	16Apr2015	17Apr2015	1	35
Jail	Arrest Misdemeanor: Disorderly Conduct at Terminal Second Degree Trespassing	21Apr2015	07May2015	16	4
Jail	Arrest Misdemeanor: Intoxicated and Disruptive	15May2015	27May2015	12	8
Jail	Arrest Misdemeanor: Intoxicated and Disruptive	08Jul2015	16Jul2015	8	42
Jail	Arrest Misdemeanor: Second Degree Trespassing	23Jul2015	23Jul2015	0	7
EMS	EMS Transported: No Lights/Siren	30Jul2015	30Jul2015	0	7
Jail	Arrest Misdemeanor: Begging-Solicitation-Vehicle Interference Intoxicated and Disruptive	03Aug2015	06Aug2015	3	4
EMS	EMS Transported: No Lights/Siren	24Aug2015	24Aug2015	0	18
Jail	Arrest Misdemeanor: Intoxicated and Disruptive	03Sep2015	17Sep2015	14	10
EMS	EMS Transported: No Lights/Siren	22Sep2015	22Sep2015	0	5
Jail	Arrest Misdemeanor: Failure to Appear on Misdemeanor	23Nov2015	25Nov2015	2	62
Jail	Arrest Misdemeanor: Failure to Appear on Misdemeanor	07Dec2015	10Dec2015	3	12
Jail	Arrest Misdemeanor: Second Degree Trespassing	15Dec2015	22Dec2015	7	5
HMIS	HMIS Other: Men's Recovery	30Dec2015	31Dec2015	1	8
HMIS	HMIS Emergency Shelter	09Jan2016	10Jan2016	1	9
HMIS	HMIS Emergency Shelter	12Jan2016	13Jan2016	1	2
HMIS	HMIS Emergency Shelter	18Jan2016	24Jan2016	6	5
HMIS	HMIS Emergency Shelter	25Jan2016	26Jan2016	1	1
Jail	Arrest Misdemeanor: Second Degree Trespassing	26Jan2016	04Feb2016	9	0
EMS	EMS Transported: No Lights/Siren	05Feb2016	05Feb2016	0	1
EMS	EMS Transported: No Lights/Siren	14Feb2016	14Feb2016	0	9
Jail	Arrest Misdemeanor: Intoxicated and Disruptive Second Degree Trespassing	02Apr2016	14Apr2016	12	48
EMS	EMS Transported: No Lights/Siren	26Apr2016	26Apr2016	0	12

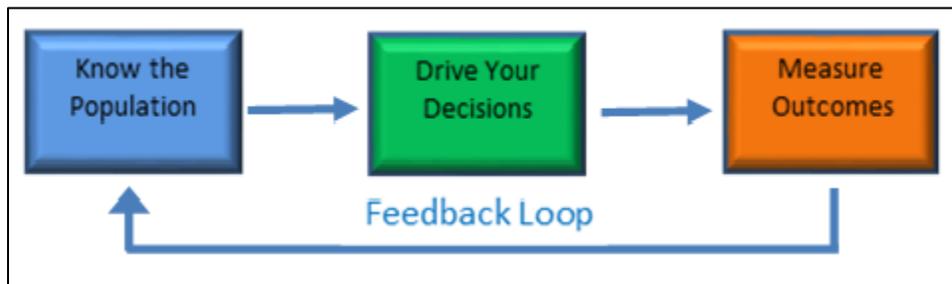
Agency	Event	Event Start	Event End	Event Length	Days Between Events
EMS	EMS Assist	09May2016	09May2016	0	13
HMIS	HMIS Emergency Shelter	05Jun2016	06Jun2016	1	27
HMIS	HMIS Emergency Shelter	07Jun2016	09Jun2016	2	1
HMIS	HMIS Emergency Shelter	10Jun2016	11Jun2016	1	1
HMIS	HMIS Emergency Shelter	13Jun2016	14Jun2016	1	2
Jail	Arrest Misdemeanor: Failure to Appear on Misdemeanor Intoxicated and Disruptive Second Degree Trespassing	20Jun2016	07Jul2016	17	6
Jail	Arrest Misdemeanor: Failure to Appear on Misdemeanor	21Jul2016	15Aug2016	25	14
Jail	Arrest Misdemeanor: Failure to Appear on Misdemeanor Intoxicated and Disruptive Second Degree Trespassing	15Aug2016	24Aug2016	9	0
EMS	EMS Assist	29Aug2016	29Aug2016	0	5
Jail	Arrest Misdemeanor: Second Degree Trespassing	01Sep2016	08Sep2016	7	3
Jail	Arrest Misdemeanor: Failure to Appear on Misdemeanor	29Sep2016	13Oct2016	14	21
EMS	EMS No treatment, No Transport	16Oct2016	16Oct2016	0	3
HMIS	HMIS Emergency Shelter	07Nov2016	08Nov2016	1	22
HMIS	HMIS Emergency Shelter	11Nov2016	13Nov2016	2	3
HMIS	HMIS Emergency Shelter	17Nov2016	18Nov2016	1	4
HMIS	HMIS Emergency Shelter	19Nov2016	22Nov2016	3	1
HMIS	HMIS Emergency Shelter	26Nov2016	27Nov2016	1	4
HMIS	HMIS Emergency Shelter	27Nov2016	01Dec2016	4	0
HMIS	HMIS Emergency Shelter	17Dec2016	18Dec2016	1	16
HMIS	HMIS Emergency Shelter	19Dec2016	20Dec2016	1	1
HMIS	HMIS Emergency Shelter	24Dec2016	26Dec2016	2	4
HMIS	HMIS Emergency Shelter	30Dec2016	31Dec2016	1	4

9 Potential Interventions

Wake County needs quality data to ensure a more complete understanding of the people who represent high utilizers of costly county services. With that understanding, the County wants to bring together service organizations and systems that are currently challenged in sharing information and build a collaborative and coordinate approach to providing the most appropriate services to the most at-risk individuals in order to reduce costs and improve opportunities for stability and sustainability for Wake County's most at risk population.

To build this collaborative approach, Wake County needs to follow an iterative, data-driven approach:

Figure 13: Iterative, Data-Supported Decisions



Who: Who is at most risk for being or becoming a high risk utilizer of costly county services?

The study found that the target population of high utilizers of jail, EMS, and housing programs are predominately men, disproportionately Black or African American and who struggle with homelessness and substance abuse based on the prevalence of charges associated with public nuisance, intoxication, begging and trespassing. While the data systems included in this study did provide some insights in mental health issues and veteran status, the limited information inhibited the ability to draw conclusions about those two characteristics.

Given the repetitive and chronic nature of incidents with this population, it would appear that these individuals may not be engaged in the workforce and lack the resources and stability to establish a reasonable quality of life. Individuals may have entered this cycle through circumstance outside their control such as chronic physical or mental health issues or as a result of their own actions and decision, such as criminal activity, or simple bad luck. From either entry point into the cycle, these individuals may find it difficult to establish eligibility and access supportive services.

Recommended Actions:

- Investigate the jail identification numbers, provided by SAS, of the high utilizers in the intersecting population. Wake County can convene a team of community stakeholders, including care managers from Alliance, to complete an in-depth review of the population for potential coordinated services. With a clear understanding of the at-risk population, targeted, proactive, coordinated case management actions can reduce costs, stop the cycle of highly reactive utilization of services, and improve quality and stability of life. Clear definition of the at-risk population will also enable trend analysis and comparison over time to monitor and measure the impact and outcomes of coordinated service management.
- Develop a case analysis of the relatively small subset of cross service high utilizers (26) to understand if this represents a reachable or unreachable population – individuals who may be

unwilling or unable to meet the requirements of the targeted interventions. Should the highest utilizer population be found to be unreachable, Wake County may want to use the analysis to:

- Target the next tier of high utilizers – additional jail identifiers can be provided.
- Target a population of utilizer that are trending toward the cycle but not yet clearly defined as high utilizers.
- Target different subsets by re-evaluating intersecting groups between EMS-HMIS, HMIS-Jail, and Jail-EMS.
- Pursue additional data to enhance the analysis of the at-risk population. Adding appropriate patient data from providers (e.g. hospitals) and payers (e.g. Alliance) would enable analysis that provides more insight into the correlation and impact of social and environmental factors on outcomes. With no systematic health screening or other health-related data, the current study was unable to provide health indicators related to the high utilizer population.
- Expand the scope of analysis for high utilizers. The County may want to consider expanding the analysis to consider the family relationship and its impact on utilization. As the County considers coordinated case management of services, understanding of family relationships, interactions across multiple service programs as the individual and as related to other family members, and intergenerational history can provide valuable insights into needs and services.

When: When can intervention result in better outcomes? (The sooner the better!)

As seen in the timeline case study, it can be inferred that as the frequency and recurrence of interactions with the County services increase, the opportunities for positive outcomes for the at-risk population may decrease without coordinated intervention of County services. Once a person gets into a cycle of homelessness, interactions with the criminal justice system, and associated EMS incidents, it may become difficult for a person to break the cycle without support to find employment and stable housing, and manage physical and behavioral health issues.

The data available for this study represent cross-sectional snapshots of each population rather than a representative samples of the population that might be tracked over time. Given the relatively short time period of data and lack of insight into events preceding and following the time period, there is the potential that a number of high utilizers were omitted from the study population. With each of these data sources, there is opportunity to improve the resulting analytic output by increasing the longitudinal time period for assessment and enhancing the consistency and completion of data collection. Incorporating additional data sources will enhance the ability to identify the “familiar faces” population and to provide broader insights into the entry points this cycle of high utilization. Longitudinal tracking will also improve the capability of testing the efficacy of new interventions.

The timeline analysis provides some key insights into the impact on both County services and the individual over time.

Recommended Actions:

- Pursue coordinated support services and collaborative efforts with the court system. Interactions with jail often begin with low level misdemeanor charges but overtime result in increasing occurrences of failure to appear, probation and parole violations, longer jail stays and higher costs. Coordinated services can help the individual meet the requirements of probation or parole and appear for required court dates, avoid further interactions with and cost to the jail system.
- Pursue additional analysis into key segments of the population to understand potential intervention points that can impact future interactions. One key area was the population young adult men

between the ages of 19 and 25 whose use of emergency shelter far exceeds that of other programs. With additional data sources such as juvenile justice and/or foster care data, the county could begin to understand the transition points at which this population may become the most at-risk and provide supportive services that can prevent a spiral into more serious health and social outcomes.

- Encourage additional sharing of health information, including mental health information, in order to better target wraparound and case-management services and reduce costly ED and jail utilization. Standardize and improve data collection of mental health screening data at key intervention points in the systems of service.
- Expand analytic data sources to gain further insight into key events that start cyclical high utilization. People may enter this cycle because they are incapable of handling a particular situation – unemployment, health crisis, homelessness. Health information, criminal justice data, foster care, juvenile justice, and even social service data may provide key insights into the events that cause a person to become caught in a cycle that grows more difficult to stop as time passes. These insights can help the County implement systems to detect this key event sequences and establish programs to intervene earlier in the process.

Expanded analysis will enable Wake County to understand the key indicators and behavioral patterns that indicate a person is in a cycle of high utilization or at risk of beginning a cycle of high utilization to enable proactive intervention before a person becomes chronically at-risk.

Where: Where are the County's needs and resources?

Geographic assessment of data was limited for this study. Better data collection and/or access to data about the location of arrests, combined with EMS call locations and location of homeless support programs may provide insight into geographic locations where additional housing and wrap-around support services may be needed.

Recommended Actions:

- Increase data collection requirements and incorporate additional data sources to enable reliable and up-to-date analysis and mapping of incidents, population needs, and service availability. This information will provide support in justifying funding needs, in resource allocation, and long-range planning.

This expanded analysis will enable Wake County to most efficiently leverage the County's limited resources to provide maximum services and interventions to the at-risk population.

How: How does the County leverage these insights to reduce recidivism, reduce costs associated with jail and EMS interactions, increase housing stability and monitor and measure improvements in long-term outcomes?

The timeline analysis provides an example of a typical high utilizer case. To better understand the resource impact on the County, there is an opportunity to assess the cost of every interaction that is reactive (repetitive jail interactions, emergency shelter, EMS incidents) against the cost of proactive and coordinated supportive services.

- Analyze dollar costs associated with the various services included in the current data sources – cost for a stay in jail, cost for and EMS interaction, cost for a stay in emergency shelter. By approximating these costs, analytics can apply them to the high utilizer population, as well as individual subsets populations (EMS-HMIS, HMIS-Jail, Jail-EMS), providing Wake County with insights in the costs and savings of proactive supportive services versus reactive, cyclical utilization.

This cost/savings analysis will help Wake County with funding justifications, long-term capacity and services planning and the ability to monitor and measure outcomes.

10 Next Steps

This study provides preliminary insights for Wake County's high utilizer population. To ensure that Wake County can meet its goals reducing recidivism and improving outcomes for the high utilizer population, and monitor and measure outcomes, the following steps are recommended.

1. Expand analytics to enhance insights from acquiring additional data sources as well as longer historical information. While some conditions could be identified or inferred through assessments from jail and HMIS data, as well as charge data, more comprehensive data is needed to understand mental health, substance abuse and other health issues. Key data sources would expand the accuracy of the high utilizer definitions, enhance understanding and management of the high utilizer population, and ensure the ability to assess the impact and outcomes of new programs such as supportive housing and wrap-around services:
 - a. Additional HMIS data based on coordinated intake system data with more consistent assessment data;
 - b. Behavioral health data (from Alliance?) to identify individuals challenged with behavioral health issues;
 - c. Hospital/ED data to understand conditions and disposition associated with EMS calls, as well as periods of hospitalization;
 - d. Substance abuse data to identify individuals whose addiction issues increase chances of becoming a high utilizer;
 - e. Court eviction data to provide insight into homeless status;
 - f. Foster care and/or juvenile justice data to indicate potential earlier intervention points (this data may need to be leveraged only in an anonymized manner);
 - g. Longitudinal data from these systems to begin to understand points of entry into the cycle, intergenerational patterns, and long-term outcomes;
 - h. And cost data to provide insights into the cost/benefit analysis for key housing and wrap-around service programs versus the cost of reactive jail, emergency housing, and EMS services.
2. Develop a cross-sector data system that provides comprehensive, entity resolved, person-centric data for individuals who interact with one or more of the Wake County stakeholder systems of service. This data system should provide:
 - a. Data Sharing Agreements and a governance model that ensure the privacy and security of all data sources, through access control for analytics, search and reporting;
 - b. Data collection standards that would improve existing data sources in terms of content, accuracy, and completeness;
 - c. The ability to search for an individual to determine appropriate course of action and treatment planning on an individual basis;
 - d. The ability to evaluate aggregate data to support coordinated case management and support team working at a programmatic level; and
 - e. The ability to analyze summary level data to support policy decisions, program funding justification, and outcomes analysis and reporting to key stakeholders and policy makers.
 - f. Example: Address information within Jail and HMIS data strongly suggests being homelessness.
3. Convene a stakeholder community summit to review the findings associated with this study and to determine next steps for cross-system of service collaboration to meet the needs of the high utilizer population. This effort may include:
 - a. Policy and procedure for coordinated services and case management;

- b. Development of cross-services team to initially address the target population of 26 high utilizers;
- c. Recommendations for a vision for cross-agency programming to ensure supportive transitions from jail, hospitalizations, and homeless programs.

Appendix A: Terms, Definitions, and Acronyms

Term	Definition
EMS	Emergency Medical System
HMIS	Homeless Management Information System
HUD	Housing and Urban Development
Levenshtein edit distance	Number of deletions, insertions, or replacements of single characters that are required to transform string-1 into string-2.
PH	Permanent Housing
SSN	Social Security Number
UCR	Uniform Crime Reporting

Appendix B: Data Dictionary

Data Point	Collecting Agency			Used in Analysis (Y/N)	Reason for Exclusion
	Jail	EMS	HMIS		
Name	X	X	X	Y	
DOB	X	X	X	Y	
SSN	X	X	X	Y	
Age	X	X	X	Y	
Race	X	X	X	Y	
Ethnicity	X	X	X	Y	
Gender	X	X	X	Y	
Family Status (Single, Married, Family)			X	N	Out of scope for report
Address	X	X	X	Y	
Phone	X	X	X	Y	
Charge	X			Y	
Misdemeanor/Felony	X			Y	
Pretrial/Sentenced	X			Y	
Veteran Status	X		X	N	Self-reported/Non-verifiable
Bond Status/Amount	X			N	Out of scope for report
Brief Jail MH Screen	X			N	Self-reported/Non-verifiable
Arrests	X			Y	
Mental Health Diagnosis			X	N	Self-reported/Non-verifiable
Service Provider			X	Y	
Homeless Status	X		X	Y	
Type of Homeless (Car, Shelter, Campsite etc...)			X	N	Not received
Type of Housing Need/Assistance			X	Y	
Housing Stability			X	N	Out of scope for report
Employment			X	N	Out of scope for report

Data Point	Collecting Agency			Used in Analysis (Y/N)	Reason for Exclusion
	Jail	EMS	HMIS		
Benefit Provider (Medicaid, SSI, etc...)			X	N	Out of scope for report
# of Days in Shelter			X	Y	
Waiting List			X	N	Not received
VI-SPDAT			X	N	Out of scope for report
Jurisdiction	X		X	N	Out of scope for report
Dates of Homelessness			X	N	Out of scope for report
# Times reported Homeless in previous 3 years			X	N	Out of scope for report
Prior Residence			X	N	Out of scope for report
Hospital Destination		X		N	Out of scope for report
EMS Disposition		X		N	Out of scope for report
EMS Call Location (Latitude)		X		N	Out of scope for report
EMS Call Location (Longitude)		X		N	Out of scope for report
Billing Data		X		N	Not received

Appendix C: Wake Jail Medical Screening Questions

Table 45: Mental Illness Screening Questions (1)

Question Number	Question
1	Temperature?
2	Weight?
3	Blood Pressure?
4	Pulse?
5	Respirations?
6	Blood Sugar?
7	Do you have any drug allergies? If yes, what are your symptoms?
8	Are you currently on any medications? If so, what medications are you currently taking?
9	Do you currently have Medical Insurance coverage? If so, who is the provider?
10	Does the inmate's behavior suggest the risk of suicide? If the answer is yes, then describe (Verbalizations, non-verbals)
11	Is the inmate conscious?
12	Does the inmate have obvious pain, bleeding, visible signs of trauma, or illness requiring immediate emergency or doctor's care?
13	Is the inmate handicapped in any way? Please specify.
14	Does the inmate have on any band-aids, dressings, a cast, ace wrap, or any artificial limbs?
15	Is there obvious jaundice (skin or eyes yellow), draining wounds, sores or evidence of infection which might spread through the jail?
16	Is the skin in good condition and free of vermin?
17	Does the inmate appear to be under the influence of alcohol, barbiturates, heroin or any other drug? Please specify.
18	Are there any visible signs of alcohol/drug withdrawal symptoms?
19	Does the inmate's behavior suggest the risk of suicide? If answer yes, then describe (verbalizations, non-verbals)
20	Does the inmate's behavior suggest the risk of assault to staff or other inmates?
21	Does the inmate appear to be psychotic? In what way? Delusions, hallucinations, incoherence. Please explain in your own words.
22	Is the inmate carrying medication or does the inmate report being on medication which should be continuously administered or available?

Question Number	Question
23	Are you presently taking medications for or had a history of diabetes, heart disease, seizures, arthritis, asthma, ulcers, high blood pressure, epilepsy, hepatitis, or history of tuberculosis?
24	Have you ever tested positive for tuberculosis? If yes, how, where and date?
25	Do you have night sweats?
26	Do you have a persistent cough?
27	Have you been coughing up blood?
28	Are you homeless?
29	Does the inmate report having a psychiatric disorder? If yes, specify?
30	Has the inmate ever attempted to kill him/herself?
31	Has the inmate had psychiatric treatment, history, diagnosis and/or medications?
32	Do you have a special diet prescribed by a physician? If yes, please explain
33	Do you have history of venereal disease or abnormal discharge?
34	Do you engage in unprotected sex?
35	Have you been tested or treated for HIV? If yes explain
36	Have you experienced recent weight loss?
37	Have you been hospitalized recently or been seen by a medical or psychiatric doctor for any reason?
38	Have you fainted recently or had a recent head injury?
39	Do you have a painful dental condition?
40	Do you have any other medical or psychiatric problems that we should know about?
41	Do you use alcohol?
47	Do you use drugs?
53	Is inmate female?
58	Chills?
59	Fever?
60	Sneezing?
61	Coughing?
62	Sore Throat?
63	Muscle Aches?
64	Who is your next of kin? List both name and telephone number.

Table 46: Mental Illness Screening Questions (2)

Question Number	Question
1	Do you currently believe that someone can control your mind by putting thoughts into your head or taking thoughts out of your head?
2	Do you currently feel that other people know your thoughts and can read your mind?
3	Have you currently lost or gained as much as two pounds a week for several weeks without even trying?
4	Have you or your family or friends noticed that you are currently much more active than you usually are?
5	Do you currently feel like you have to talk or move more slowly than you usually do?
6	Have there currently been a few weeks when you felt like you were useless or sinful?
7	Are you currently taking any medication prescribed for you by a physician for any emotional or mental health problems?
8	Have you ever been in a hospital for emotional or mental health issues?
9	Officer`s Comments/Impressions: Language barrier?
10	Officer`s Comments/Impressions: Under the influence of drugs/alcohol?
11	Officer`s Comments/Impressions: Non-cooperative?
12	Officer`s Comments/Impressions: Difficulty understanding questions?
13	Officer`s Comments/Impressions: Other comments?
14	Does this inmate need to be referred for further mental evaluation?

Document Information

Document Control

Title	Wake County Reporting Design Document
Version	1.0
Location	\\sso.vsp.sas.com\amd\Customers\United_States_RaaS_Business_Provider_RUSA\Solution_Development\System_Design\Reporting_Design\RUSA_Reporting_Design_v02.docx
Date	07-JUL-2017 1:52 PM

Contacts

If you have questions regarding this document, contact one of the following individuals:

Document Owner: Scott DeBurgomaster	Project Manager: Michael Weymer
Email: Scott.DeBurgomaster@sas.com	Email: Michael.Weymer@sas.com
Office Phone: 919 531 0194	Office Phone: 919 531 4564

Revision History

JIRA Ticket	Date	Version	Name	Description
RUSA-52	28-APR-2017	0.1	Scott DeBurgomaster	Initial draft
	01-MAY-2017	0.2	Will Butcher	Prepared draft for delivery
RUSA-72	28-MAY-2017	0.3	Tom Mock	Reviewed draft
	06-JULY-2017	1.0	Tom Mock	Finalized doc