Leveraging Integrated Data to Examine Youth Homelessness

Prepared for

U.S. Department of Housing and Urban Development

Office of Policy Development and Research

Grant Number H-21693CA

Prepared by

The Center on Poverty and Community Development

Case Western Reserve University

August 16, 2024

Acknowledgements

The authors acknowledge all community partners for sharing insights about their agency data systems. They include:

- Cuyahoga County Office of Homeless Services
- Cuyahoga County Department of Job and Family Services
- Cleveland Metropolitan School District
- Cleveland Metropolitan Housing Authority
- Cuyahoga County Juvenile Court
- Cleveland Food Bank
- FrontLine Service
- Cleveland-Cuyahoga Continuum of Care providers

We also acknowledge the Ohio Department of Education and Workforce for providing us with McKinney Vento data and the Ohio Department of Health for allowing us to use lead testing data for this project. The Ohio Department of Health lead-testing data are part of our integrated data system and are accessed under an appropriate IRB-approved protocol. This should not be considered an endorsement of this study or its conclusions by the Ohio Department of Health.

Project Team

Rob Fischer, Ph.D, Professor, PI

Francisca García-Cobián Richter, Ph.D., Research Associate Professor, Co-PI

Dana Prince, Ph.D., Associate Professor

Meagan Ray-Novak, MSSA, Research Associate

Michael Henderson, Ph.D., Senior Research Associate

Stephen Steh, Ph.D., Senior Research Associate

Joseph Andre, MS, Research Associate

Sara Arrojo Montilla, Ph.D., Research Assistant

Jiayi Sun, MSW, Doctoral Student

Ashley Hajski, PhD., Researcher

Morgan Ashley, MSSA, Research Assistant

Table of Contents

Executive Summary

A more accurate assessment and understanding of the population of youth facing homelessness is crucial for effective federal and local policy-making that prevents or interrupts homelessness among young people. With a better understanding of the size and characteristics of this population, policymakers can make informed decisions regarding resource allocation and interventions that address root causes of homelessness.

Though much progress has been made to better estimate the number of youth experiencing homelessness via homeless services administrative data, surveys, and point-in-time counts, it is clear that many youth are missing from the data. Estimates based on these data alone may not only undercount youth, but they may be biased towards excluding marginalized youth that are less likely to connect with social services.

As data integration and technologies have become more accessible to local communities, this study sought to explore the value of integrated data systems (IDS) and institutional knowledge in producing more comprehensive counts of youth facing homelessness that may provide valuable planning information for a more effective Continuum of Care. For this purpose, we leveraged a county IDS, the Child Household Integrated Longitudinal Data (CHILD) system, spanning over 35 linked data systems from social service agencies in Cuyahoga County, Ohio.

After a systematic review of the literature, we assessed the administrative data held in CHILD, which includes the Homeless Management Information System (HMIS). We acquired additional data such as the McKinney-Vento school data and implemented semi-structured interviews with key staff from agency data providers. The information gathered was used to develop a baseline registry of youth ages 13 to 25 years old identified in HMIS or administrative school data as housing unstable for each of the years from 2017 to 2019.

This baseline registry was then extended using a novel Address List method that leveraged multiple other linked administrative data in CHILD along with insights from data managers regarding data entry practices. The Address List method flagged administrative records of social services agencies where the address field was indicative of homelessness or housing instability. Through this process, additional youth were identified leading to a final IDS registry of our study population. Finally, we proceeded to estimate the number of youth facing homelessness but unaccounted for in the registry via Multiple Systems Estimation (MSE) methods. These larger estimates are contextualized with benchmarks derived from survey and census data, as well as community knowledge.

Our findings show the incremental value of data integration for communities that use HMIS within the Continuum of Care. In 2018, HMIS data identified 2,820 youth. Integrated McKinney Vento records from schools increases our count estimates by approximately

73% (4,871 youth). Next, by exploiting address-level information across systems informed by community knowledge, we further increase estimates by 27% (6,198 youth). Lastly, using MSE methods applied to all lists in our registry, estimates dramatically increased by over 200% (21,360 youth). This suggests that for every young person identified in administrative data as facing homelessness, there are as many as three young people who faced housing instability but were not documented in administrative data. These annual prevalence estimates are plausible, given the census estimates of low-income youth in the county.

Administrative data held by organizational partners contain substantial value for understanding youth experiencing homelessness in a region. Communities beginning down the path of data integration are advised to prioritize school records as the first source to consider expanding on the value of HMIS. Address list techniques also provide a meaningful addition once data is available and integrated across multiple partners. In our case, SNAP and Food Bank records proved to be important contributors to the registry, due to robust address data in these two administrative systems. Importantly, drawing from past records for youth in the registry allows community partners to identify points of contact that can serve to support housing unstable youth before they experience homelessness. We found SNAP enrollment a year prior to homelessness to be prevalent among about 65% of youth in the registry. Finally, multiple systems estimation offers a novel approach to better assessing and validating the vulnerable population of youth likely to be facing issues of homelessness and housing instability.

Collaborating with data partners was essential to gain a better understanding of the data, whether we had ample experience working with it (such as HMIS) or were unfamiliar with it (such as McKinney-Vento school data). In both cases, data partners provided invaluable support. Insights from partners and youth with experiential knowledge suggested that marginalized youth facing homelessness may be less likely to seek formal services than non-marginalized youth, leading to patterns of missingness in the data that are not random. Missingness may reflect safety concerns or distrust in the system by some of the most vulnerable young people who experience homelessness. This had implications for the way in which we implemented our estimation models, choosing to estimate models by strata formed to reduce heterogeneity within groups.

This project underscores that the relationship between data and youth-tailored programs to address homelessness is bidirectional. High quality, integrated data can inform planning and policies for better services. At the same time, these improved services may be more suited to engage with and support youth facing homelessness, which would improve the quality and coverage of administrative data, thus completing a virtuous cycle.

The collaborative process to integrate community knowledge into data analysis is continuous. Sharing findings and seeking feedback from community data partners and people represented and not represented in the data is essential to promote a reinforcing cycle of better data and better programs to prevent and address homelessness, thus advancing equity in housing stability for young people.

1. Introduction

Accurately estimating the number of youth facing homelessness is crucial for effective federal and local policy-making as well as resource allocation. With a more accurate understanding of the population of youth experiencing homelessness, policymakers can make informed decisions that address root causes and provide appropriate support to those at risk of or in homelessness. However, multiple complex factors challenge the accurate enumeration of youth experiencing homelessness (figure 1).

First, the experience of homelessness is an inherently unpredictable and dynamic state impacted by changes to individual, social, community, regional, economic and political factors that independently and collectively interact with each other over time (O'Flaherty, 2012). Locating and identifying those experiencing homelessness can prove challenging. Difficulties collecting accurate data through interviewing, geographic differences, and level of oversight necessary to ensure reliable enumeration are known barriers (Cackley, 2020).

Second, the definition of homelessness varies across federal agencies, so administrative data associated with service provision will track youth experiencing varying levels of housing instability. HUD defines four categories of homelessness[1]: (1) literal homelessness (unsheltered, in emergency shelter or in transitional housing); lacking resources and support networks to obtain permanent housing while being either (2) at imminent risk of homelessness or (3) fleeing domestic violence. A fourth category includes youth not counted under any of the previous categories, but homeless under other statutes. On the other hand, the McKinney-Vento Homeless Children and Youth Act defines homelessness as *a state in which a person lacks a fixed, regular, and adequate nighttime residence* (National Center for Homeless Education). Based on this latter definition, the Education Department expands the HUD criteria for service provision to include youth doubling up, living in motels or cars, and other unconventional living arrangements as experiences of homelessness. Additionally, administrative data associated with non-housing services may also capture experiences of homelessness in the address field of their data systems. As the definition used directly impacts funding allocation and policy implementation, how homeless status is conceptualized matters (Sullivan, 2023).

[1] https://files.hudexchange.info/resources/documents/HUDs-Homeless-Definition-as-it-Relates-to-Children-and-Youth.pdf

Third, the developmentally appropriate tasks of individuation - personality development, establishing autonomy, and withdrawal from adult-directed interactions that youth experience- may reduce their interactions with formal housing support systems, further complicating efforts to accurately calculate the population. Youth have been found to utilize formal shelter services less frequently than their older adult counterparts (Samuels et al., 2019; Suchting et al., 2020). Youth-specific barriers to engagement include perception of safety, fit of older adult services with youth-specific needs, a desire to differentiate themselves from older adults, shelter policies that conflict with youth identity such as restrictions on cell phone use and curfews, and lack of shelter access for those under age 18 (Ha et al., 2015). Further, youth have been found to rely heavily on relationships with each other and trusted older adults in pathways out of homeless experiences (Rice et al., 2023).

Finally, the service systems may struggle to support youth facing homelessness in ways that meet the emotional and tangible needs of a socially diverse subpopulation. Youth with marginalized identities experience homelessness more frequently than their peers with dominant identities and may not seek services or may engage less frequently with them (Morton et al., 2018; Petry et al., 2022; Robinson, 2021; Wilson et al., 2020). When youth do not receive services, they do not appear in the administrative records of social system providers, increasing the challenge of identifying, characterizing, and counting the population.

Leveraging youth social capital and communication needs, connecting with youth in spaces they frequent, and providing services during times youth are available have been documented as facilitators to engaging with providers (Ha et al., 2015; Rice et al., 2023). However, these types of approaches are nuanced and may be challenging to staff by service providers.

Given the dynamic nature of homelessness, we refrain from using the term ‘homeless youth’ choosing to refer to *youth facing homelessness* during a given period, which is mostly set as a year in this study. Given the various ways in which agencies operationalize the definition of homelessness, when we refer to a person, “facing homelessness,” we include the high levels of housing instability captured by HUD or McKinney Vento, or any of those administrative data systems used in our analysis. A diagram depicting the linkage of administrative data systems used in this study is provided in appendix 6.3.

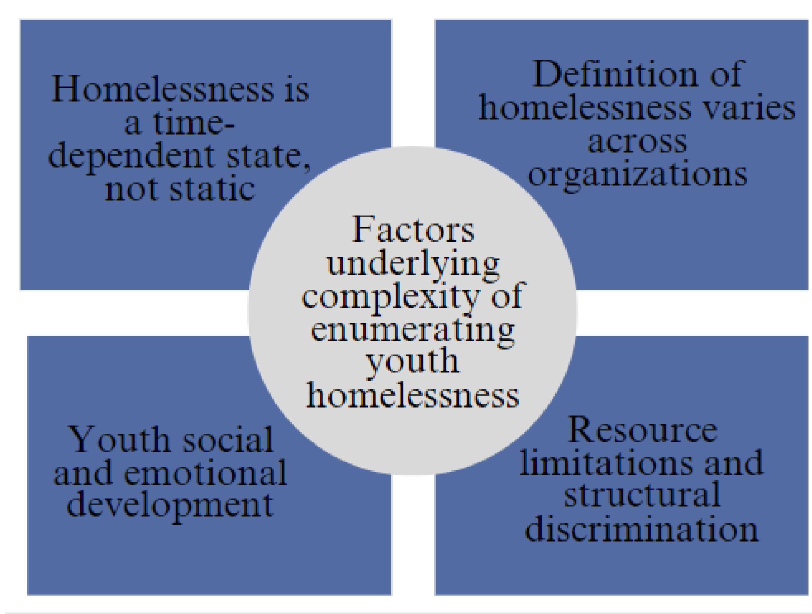


Figure 1. *Factors underlying the complexity of enumerating youth with homelessness experiences.*

Despite Continuum of Care (CoC) efforts to tailor programming to youth needs, enumerating the population of youth experiencing homelessness has remained challenging. Counts based on point-in-time estimates and survey-based methods have been found to underestimate the true number of homeless youth (Anthony, Fischer, & D'Orazio, 2016). Estimates of unaccompanied youth and those experiencing family homelessness have improved collective understanding of youth diversity of experience but have been difficult to replicate, relying on national cross-sectional data sets and statistical techniques which are unfamiliar to policymakers and service providers (see Cutuli et al., 2020; Morton et al., 2018).

Research has documented the interrelationship between homelessness and other acute social and health problems such as food insecurity, intimate partner violence, and untreated chronic health conditions (Barnes et al., 2021; Dzubur et al., 2022; Hargrave et al., 2024). This suggests that administrative data from related systems may provide additional information on the housing experiences of youth. As access to administrative data grows, there is a need to explore how integrated administrative data systems can enhance the accuracy of youth homelessness estimates. Leveraging the power of administrative data, especially regional data, allows communities to mine existing sources for planning, funding, and implementing programs tailored to this population.

With funding from the U.S. Department of Housing and Urban Development (HUD), The Center on Poverty and Community Development (the Poverty Center) at Case Western

Reserve University assessed the capacity of regional administrative data to improve the estimation of prevalence and incidence of youth experiencing homelessness. The project leveraged the value of an existing regional integrated data system holding individually-identifiable data from 35 public and private entities, the Child-Household Integrated Longitudinal Data (CHILD) system. CHILD is a proprietary integrated data system developed and maintained by the Poverty Center, used to conduct research to inform and support data-driven decision making on regional social policy (Fischer, Richter, Anthony, Lalich, & Coulton, 2019).

The rationale for a regional data integration strategy is threefold. First, homeless service responses are organized at the regional level through the Continuum of Care, making population estimates most relevant at this level. Second, regional-level data integration is feasible because organizational partners are familiar with each other and may be more willing to share data. Finally, building on existing data integration efforts maximizes previously shared data with new data that is of specific value to planning and regional policy.

The project informs regional responses to youth homelessness by illustrating how administrative data can best enhance counts and support strategies regarding practice and policy.

This study sought to 1) understand how youth homelessness is documented and defined across community service providers, 2) leverage the integration and analysis of data to produce more timely and comprehensive prevalence -based counts of youth facing homelessness, 3) identify administrative data systems that are most beneficial to count and describe youth experiencing homelessness, and 4) use study findings to provide recommendations of planning and programmatic changes.

The study's main objectives were:

1. To investigate the extent to which integrated administrative data systems can offer additional value to the task of estimating the number of youth facing homeless over a period of time, identifying the specific types of administrative data sources that provide the most benefit in counting and describing the population of youth experiencing homelessness
2. Develop a model for replication purposes in other communities, offering guidance to those just starting to coordinate data sharing agreements among partner agencies, as well as to those with already developed integrated data systems informing social and public policy in their communities

In order to accomplish these objectives, we explore and systematize the literature on integrated data systems and youth homelessness (Section 2), and we address the following questions in section 3.

1. How is youth homelessness currently captured and defined across agencies in our community?
2. How can we leverage the integration and analysis of data to produce more timely and comprehensive prevalence-based and incidence-based counts of youth with an experience of homelessness compared to using only Homeless Management Information System (HMIS) data?
3. What specific types of administrative data sources provide the most benefit in counting and describing the homeless youth population?

We conclude in section 4 with a discussion on the potential and limitations of integrated data models to inform policy and planning for agencies who offer services to individuals at risk of or experiencing homelessness. We transfer insights from this work into an accompanying community guide to support the work of community partners engaged in addressing and preventing homelessness among youth.

2. Review of the Literature

This project is informed by previous research and accounts of organizational efforts to better track the number of youth facing homelessness. We organize the review of the literature into two components: (1) approaches to enumerating youth facing homelessness via administrative data and surveys, and (2) a systematic review of the literature that concerns itself with estimating the population of youth facing homelessness via Integrated Data System approaches -the focus of this work. Within this section, we pay attention to the way in which studies contextualize their estimates and offer insights about who is seen and not seen in the integrated administrative data.

2.1. Enumerating youth facing homelessness via administrative data and surveys

The majority of studies dedicated to youth experiencing homelessness use surveys as a primary strategy for estimating the population size of homelessness. Yet, survey implementation is challenging due to factors like lack of permanent residence and reluctance to participate (Glasser et al., 2013; Meyer et al., 2021). A study conducted in the Netherlands highlighted that using standard sample survey methods to approach youth facing homelessness is likely to have low response rates leading to biased results (Coumans et al., 2017).

As an alternative to standard survey methods Anthony and Fischer (2016) designed a survey strategy in Cuyahoga County, Ohio in collaboration with multiple community agencies and following a youth-centered approach. They found that youth are more likely to participate if approached by someone they know well. However, there is some evidence that such surveys and counts do not fully capture the “hidden homeless” who do not interact with service providers (Metraux et al., 2016). The timing of surveys is also crucial; Anthony and Fischer (2016) noted that their study missed late-night or weekend data collection, which could lead to an undercounting of youth.

While surveys to count people facing homelessness have been conducted or commissioned by HUD since the 1980s, the first Point-in-Time (PIT) count appears in the 2007 Annual Homelessness Assessment Report to Congress (U.S. Department of Housing and Urban Development Office of Community Planning and Development, 2007). The PIT count provides a single-night count of sheltered and unsheltered people facing homelessness as reported by Continuum of Cares across the country. While PIT counts provide year-to-year snapshots of the population facing homelessness, they are not able to capture the full extent of homelessness (Schneider et al., 2016; Smith et al., 2019). The dynamic nature of homelessness makes it extremely challenging for some individuals to be accounted for, especially if they are living in cars, abandoned buildings, or other deserted places (U.S. Government Accountability Office, 2021). Moreover, common stereotypes influence our understanding of homelessness (Snow et al., 1986), and the PIT count often undercounts homeless students and employed individuals who are marginally housed and do not fit these stereotypes. Additionally, some homeless individuals report avoiding social services and refusing to participate in PIT surveys (Smith et al., 2019). Bird and her colleagues (2018) noted that censuses of homeless individuals within a specific district and timeframe, such as PIT counts, generally result in an undercount of the actual population. Given these issues, simply aggregating PIT results from local Continuums of Care (CoCs) to determine the number of homeless individuals in a specific area is likely to underestimate the extent of the problem.

Administrative data on homelessness collected by government agencies depends on the definition of homelessness used by each agency. Counts based on the Department of Education's definition -which applies to students in doubled-up arrangements- are consistently and remarkably higher than counts obtained using the HUD definition (Pergamit et al., 2013). It is estimated that one million students lived doubled up in the 2018-2019 school year (U.S. Department of Education, 2020). Exclusion of certain types of unstable housing states, such as street-dwelling, and doubling up, may also be a product of the challenges in locating and identifying youth in these precarious conditions for inclusion in the population (Sullivan, 2023).

Sullivan (2023) pointed out that differences in definitions of homelessness can lead to significant differences in the ranking of prevalence estimates across communities with direct implications on the funding distribution for homeless services. His analysis found that communities whose rates of homelessness were high under the Department of Education's definition were different from those with high rates based on the narrower HUD definition. Relative to the latter, the former tended to be more rural, have higher poverty rates and a larger share of students identified as Black or Hispanic.

Studies conducted in California (Burns et al., 2021), Louisiana (Meltzer et al., 2019), Minnesota (Lowell & Hanratty, 2022), and Washington state (Building Changes, 2019) also provide evidence that the number of youth facing homelessness is underestimated when using HUD's definition compared to the number identified using the McKinney-Vento definition. According to the above studies, most students experiencing homelessness were doubled up or living temporarily in hotels/motels, and thus did not meet HUD's definition of homelessness.

Acknowledging the limitations of surveys and administrative data to fully assess the extent of homelessness, some researchers have leveraged integrated administrative data systems for this purpose. Fischer and his colleagues (2019) note that by connecting information across multiple service delivery agencies, integrated data systems provide significant value for needs assessment, program planning, policy decision making, and collective impact evaluation across a range of social issues. The following examples from King County, WA, and Mecklenburg County, NC, demonstrate successful implementations of integrated data systems for more accurate estimations of individuals facing homelessness.

Case 1: Using integrated data systems to measure homelessness in King County, WA.

Recognizing the limitations of PIT counts and individual data systems, King County invested in data infrastructure to link data from Homeless Management Information System (HMIS), Health Care for the Homeless Network (HCHN), and the Behavioral Health and Recovery Division (BHRD) since 2018. This integrated data system was used to estimate the population size of homelessness in 2020. Washington State has an "opt-in" HMIS consent policy, allowing individuals to decide if they want to share personal identifiers.

According to the King County's Department of Community and Human Services (2021), there were 40,800 people in 2020 experiencing homelessness at some point in the year, with HCHN and BHRD contributing substantially (7,300 people) to the count of people facing homelessness beyond what was reported in HMIS. The integrated data system in King County improved the accuracy of estimating the number of individuals experiencing homelessness, enhanced the identification of demographic characteristics of those affected, provided a more accurate assessment of housing needs, and better equipped policymakers and program implementers to develop effective responses (King County's Department of Community and Human Services, 2021).

Case 2: Using integrated data systems to measure child and youth homelessness in Mecklenburg County, NC.

The 2014-2015 Charlotte-Mecklenburg Family Homelessness Snapshot Report (Clark et al., 2017) by the University of North Carolina at Charlotte and the Urban Institute studied children (0-17 years old) and youth (18-24 years old) experiencing housing crises or homelessness using linked data from HMIS, Charlotte-Mecklenburg Schools (CMS), and Mecklenburg County Department of Social Services (DSS). The report examined connections and gaps between students in shelters, those using McKinney-Vento resources, and those receiving DSS services such as Food and Nutrition Services (FNS), Child Protective Services (CPS), and Foster Care. The Child & Youth Homelessness Integrated Data Report (Anderson et al., 2020) found that between August 1, 2016, and July 31, 2017, 2,936 children and youth used HMIS services, and 4,114 students were identified for McKinney-Vento services in CMS. In total, 6,558 children and youth used either HMIS or McKinney-Vento services, with only 492 (7.5%) seen in both administrative lists. Anderson and her colleagues (2020) found that students experiencing homelessness and/or housing instability may not be identified as McKinney-Vento for a variety of reasons such as lack of knowledge about the program; lack of self-report; or failure among staff to identify students. It is estimated that a large proportion of McKinney-Vento students are identified when they need transportation to school. Anderson and her colleagues (2020) indicated that the integrated data system helped characterize child and youth homelessness and service utilization patterns across systems, identifying gaps and opportunities for accessing safety net services. It also highlighted the importance of linking housing solutions with other service sectors, such as education and health, to effectively address youth homelessness and improve service coordination.

The two cases mentioned above show how integrated data systems are being utilized in communities to improve estimates of the homeless population and enhance service delivery. Integrated data systems provide a more comprehensive view of homelessness, helping to identify system interactions and service gaps, and enhance coordination among providers.

Furthermore, integrated data systems serve as inputs for multiple system estimation methods (MSE) to estimate the full size of the population of individuals facing or at risk of homelessness, including those not appearing in administrative data (sometimes referred to as ‘hidden’). Bird and King (2018) expounded on the workings of MSE and its broad applicability in addressing social issues for estimating hidden population across diverse domains. For instance, researchers have explored its use in estimating unreported human trafficking victims, addressing homelessness and drug use, assessing casualties in wars, and even monitoring wildlife populations in specific regions. Estimating the share of populations in distress that do not engage with government systems and thus are not seen in administrative data is crucial to inform policies aimed at addressing their specific needs (Bird & King, 2018).

We conclude this section noting that the relationship between data and better programs to address homelessness is bidirectional. High quality, integrated data can inform planning and policies for better services. At the same time, these improved services may be more suited to engage with and support youth facing homelessness, which would improve the quality and coverage of administrative data, thus completing a virtuous cycle. An example of this positive reinforcement is transpired in the Built for Zero approach to end homelessness. Led by the nonprofit Community Solutions, Built for Zero aims to 'measurably and equitably end homelessness' where ending homelessness refers to achieving Functional Zero. According to Evans and Baker (2021), functional zero serves as a benchmark and methodology used by communities to effectively address homelessness. It ensures that the number of people experiencing homelessness never exceeds the community's capacity to move people into permanent housing. Recognizing the importance of data to interrupt homelessness more effectively, Community Solutions has designed a data system -the By-Name List - to implement Built for Zero. The By-Name Data or List is a high-quality, real-time, comprehensive, individual-level data system on people facing homelessness in a community (Community Solutions, 2024). Community Solutions provides a scorecard to guide the development of this data system as well as case studies of communities that have effectively implemented Built for Zero.

2.2. Systematic review of the literature on using Integrated Data System to count youth facing homelessness

2.2.1. Background

As addressed in the previous section, accurately estimating the homeless population poses a persistent challenge due to the intrinsically volatile and transient conditions of this population. Traditionally, methods such as the PIT count have been widely used in the United States and other countries to capture a snapshot of the homeless population at a specific moment. However, these methods have significant limitations, as they tend to undercount individuals who are not institutionalized or connected to formal support systems. This has led to an underestimation of the homeless population and an incomplete understanding of their characteristics and needs. In response to these limitations, researchers have explored alternative and complementary approaches to improve the accuracy of estimates, such as tapping on integrated data systems and in some cases applying Multiple Systems Estimation to the integrated data. Multiple Systems Estimation (MSE) is a statistical methodology used to estimate the total number of individuals in a population, based on a sample of that population appearing on one or more lists. It is also known as capture-recapture estimation, given its applications in ecology. The integration of data systems and the use of advanced statistical methods like MSE offer significant potential for program planning, policy decision-making, and evaluating the impact of policies on the homeless population.

Therefore, this systematic review focuses on identifying studies that use integrated data systems to estimate the homeless population. We assess the extent to which integrated data systems have been able to provide additional value to the counting of homeless individuals and the methods employed to derive estimates.

2.2.2. Method

Search strategy

The search terms for this review were based on those used in previous relevant reviews (Embleton et al., 2016; Morton et al., 2018; Morton et al., 2020), with additional criteria referring to youth homelessness. The search was conducted in January 2024 in the following databases: Web of Science, PsycInfo, PubMed, and Scopus. There were no limitations used regarding publication language or date.

The search strategy was performed by an iterative process using multiple combinations of keywords in five groups. The following terms were included: (homeless* OR “unstabl* hous*” OR unsheltered OR “vulnerable hous*” OR doubled-up) AND (youth OR adolescen* OR “emerging adulthood” OR young OR teen* OR juvenile OR unaccompanied OR child OR minor OR student) AND (shelter OR Street OR "juvenile justice" OR “foster care” OR "unsheltered location" OR "emergency shelter" OR "transitional housing" OR "rapid rehousing" OR "permanent supportive housing" OR "coordinated intake") AND (“integrated data” OR “administrative data” OR estimat* OR “multiple systems estimation” OR “linked data” OR “data integration system”) AND (prevalence OR count OR “total population” OR incidence).

The search was limited to titles, abstracts, keywords, and scientific articles. Table 1 details search terms by category.

Category	Search terms
Population	homeless* OR “unstabl* hous*” OR unsheltered OR “vulnerable hous*” OR doubled-up
AND	
Population	youth OR adolescen* OR “emerging adulthood” OR young OR teen* OR juvenile OR unaccompanied OR child OR minor OR student
AND	
Context	shelter OR Street OR "juvenile justice" OR “foster care” OR "unsheltered location" OR "emergency shelter" OR "transitional housing" OR "rapid rehousing" OR "permanent supportive housing" OR "coordinated intake"
AND	
Input	“integrated data” OR “administrative data” OR estimat* OR “multiple systems estimation” OR “linked data” OR “data integration system”
AND	
Output	prevalence OR count OR “total population” OR incidence

Table 1. *Search strategy categories*

Note: “”* following the search term instructs the database to search for anything containing the root of the search term; for example, homeless, homelessness, etc.

Eligibility

The inclusion criteria for studies were: 1) published in peer-reviewed journals; 2) leverage integrated data systems; 3) concerned with improving the count of youth experiencing homelessness.

To identify studies that met the eligibility criteria, two study team members independently screened the articles by title and abstract. The full texts of all articles that met the potential eligibility criteria were assessed separately by the team members. Discrepancies in the decisions of the reviewers were resolved by discussion and consensus with an additional reviewer.

Data extraction

A trained research assistant extracted data from all studies selected for inclusion. Data extracted included authors, country, year of publication, definition of homelessness, data sources (lists of sources from which data were obtained), sample characteristics (description of the population, total number of homeless people identified, and other relevant variables), results of homelessness estimates, and a contextualization of the results. The information extracted from the studies was discussed with the principal investigators of the project who verified this data. Discrepancies that arose were resolved by discussion.

2.2.3. Results

Screening findings

Using the search terms described in the method section, the initial search of the four databases (PsycInfo, Web of Science, Scopus, and PubMed) identified 346 potentially relevant articles. After removing duplicates, a total of 173 articles remained. These articles were reviewed by title and abstract. A total of 156 articles were excluded because they did not meet one or more of the review inclusion criteria. The remaining 17 articles were reviewed in full text. Of these, 13 were excluded because they did not meet the eligibility criteria based on further review. The reasons were as follows: they did not use integrated administrative data (n = 9); they used integrated administrative data but not to obtain a population estimation (n = 4). Finally, 4 studies identified through databases and meeting the inclusion criteria were included in the systematic review. In Figure 2, the flow diagram displays the results of identification, screening, and inclusion at each stage.

Qualitative findings

In this section, we focus on the qualitative findings from the identified studies. Table 2 shows the key characteristics of these results. The first characteristic is related to the country and year of development of the studies. The four studies included in the systematic review (Bezerra et al., 2011; Coumans et al., 2017; Dutton & Jadidzadeh, 2019; Evangelist & Shaefer, 2020) were conducted in different countries: Brazil, Netherlands, United States, and Canada. This fact will allow us to examine how administrative data are used in different realities and socio-cultural contexts for estimating the homeless population. The articles were published from 2011 (Bezerra et al., 2011) through 2020 (Evangelist & Shaefer, 2020).

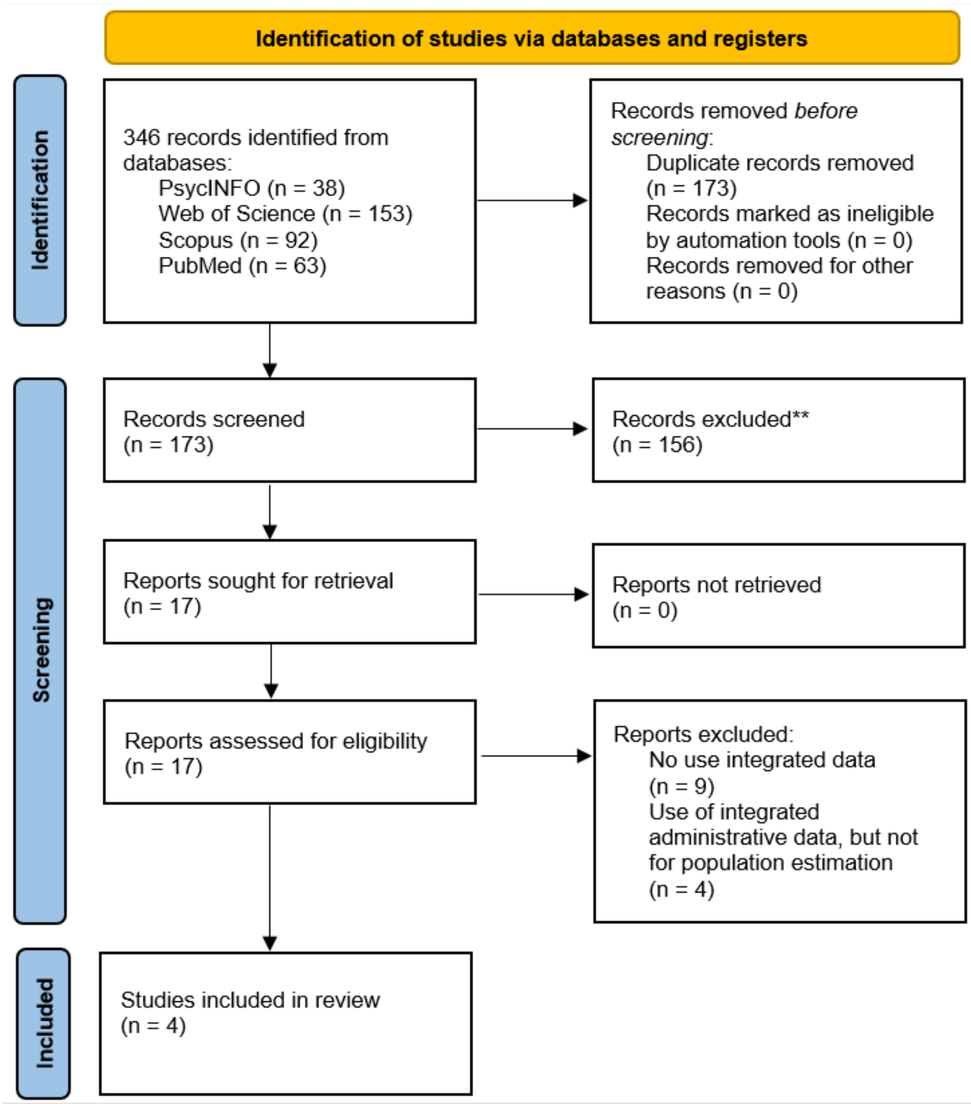


Figure 2. *Flow diagram with identification, screening, and inclusion results at each stage of the systematic review process.*

The studies established a definition of homelessness, which guides the specific type of population on which they will make an estimate. There is a great variability in these definitions depending on the studies. One study focuses on street children (SC) defined as those unsupervised children and adolescents who spend much of their days or nights on the streets (Bezerra et al., 2011). In the study by Coumans et al. (2017), they specify that in the Netherlands, there was no established definition of homelessness that was widely accepted. Consequently, they relied on the definition established by Wolf et al. (2002) from a large-scale homelessness study. In this study they focused on roofless people according to Wolf et al. (2002), that is, people who did not have permanent accommodation on the reference date, distinguishing between the following categories of: people sleeping outdoors or in open public spaces; people sleeping indoors in transitional shelters for homeless people; people sleeping indoors in the homes of friends, acquaintances or relatives. In the Evangelist and Shaefer (2020) study, they indicated that homelessness represents a dynamic state that contributes to other inequalities, and they focused explicitly on families, distinguishing between those doubling up and those facing literal homelessness. Finally, Dutton and Jadidzadeh (2019) focused on people using emergency shelters. They defined homelessness as referring to the situation in which people do not have a safe and stable place to reside, leading them to rely on emergency shelters to meet their basic shelter needs. The authors state that these shelters provide accommodation for a significant and varied proportion of people facing homelessness.

Regarding the population estimated in the studies included in the systematic review, we can observe significant heterogeneity. Two of the studies focused on estimating the population size of adults experiencing homelessness (Coumans et al., 2017; Dutton & Jadidzadeh, 2019). Coumans et al. (2017) focused on an age range between 18 and 65 years, noting that Dutch homelessness policy assigns a separate track for minors and individuals above 65 years old typically transition from living on the streets to residing in facilities (residential services), such as homeless hostels or nursing homes. The other study covered late adolescence and adulthood, from age 16 to 75 years (Dutton & Jadidzadeh, 2019). Another study focused on estimating the population of children and adolescents from kindergarten to twelfth grade (Evangelist & Shaefer, 2020). And finally, the study by Bezerra et al. (2011) focused on boys and girls but did not specify the exact age range of the study population.

To estimate the homeless population, studies employed a variety of lists. There were three studies that used three lists (Bezerra et al., 2011; Coumans et al., 2017; Dutton & Jadidzadeh, 2019), while one study used one list across multiple years (Evangelist & Shaefer, 2020). Among the studies that used three lists, Bezerra et al. (2011) was the only one that combined survey and administrative data. Two lists included data collected from street surveys (on two different days of 1 week) with a third official list obtained from the municipal secretary of citizenship and social assistance and from the public prosecutor office in two different cities (Maceió and Arapiraca). The rest of the studies used lists with data from various organizations, institutions, and services. Some of these lists provided data on people registered in shelters, as in Dutton and Jadidzadeh (2019) (Calgary Homeless Foundation (CHF) and Toronto's Shelter, Support & Housing Administration (SSHA) lists). Other lists used recorded information on people who received income support but did not have a permanent residence (Coumans et al., 2017; WWB list) or who were registered as homeless with the National Alcohol and Drug Information System (Coumans et al., 2017; LADIS list).

The only study that used one list across 15 years was Evangelist and Shaefer (2020), who used administrative data on public school student population through the Michigan Department of Education. In addition, this study used county-level data on socioeconomic characteristics, housing market conditions, availability of housing assistance, and housing and social hardship indicators from various official government agencies.

Of the four studies included in the systematic review, two used multiple system estimation (Bezerra et al., 2011; Coumans et al., 2017). Dutton and Jadidzadeh (2019) leverage integrated administrative data on shelter use to obtain shelter-to-population count ratios as estimates of shelter entry incidence. Evangelist and Shaefer (2020) used administrative public-school student level data integrated over time from kindergarten through twelfth grade to estimate cumulative risk models of homelessness over the school age period.

Characteristics of studies		***N* (%)**
Country of study	Brazil	1 (25)
	Netherlands	1 (25)
	United States	1 (25)
	Canada	1 (25)
Publication year	2010-2015	2 (50)
	2016-2020	2 (50)
Definition of homelessness	Street children	1 (25)
	Roofless	1 (25)
	Family homelessness	1 (25)
	Shelter use	1 (25)
Estimated population	Children	1 (25)
	Children and adolescents	1 (25)
	Adults	2 (50)
Number of lists	3 lists	3 (75)
	1 list, 15 years	1 (25)
Estimation approach	Multiple System Estimation	2 (50)
	Others	2 (50)

Table 2. *Characteristics of included studies*

Estimation results by approach

The studies included in the systematic review employed different approaches using administrative data to estimate the size of the homeless population. Studies 1 and 2 (Bezzera et al., 2011; Coumans et al., 2017) used multiple system estimation. Results are summarized in tables 3a and 3b.

The study by Bezerra et al. (2011) aimed to estimate the number of street children and adolescents, as well as their characteristics, in the Brazilian cities of Maceió and

Arapiraca. The previous estimate by official records was 565 and 157, respectively. The authors apply MSE to integrated street survey and administrative data. The latter come from the municipal secretary of citizenship and social assistance of Maceió and from the public prosecutor office in Arapiraca. They estimated 5,225 children in Maceió and 1,191 children in Arapiraca. Thus, in Maceio the ratio of unlisted to listed was 4.2, while in Arapiraca the ratio of unlisted to listed was 2.7. Their results estimated that for every young person observed in survey or government data, there are between seven to eight people unobserved who spend most of their day in the streets. Regarding the population characteristics, the study results indicated that these were similar in both cities in most cases. The majority of the estimated population were males (71.4% in Maceió and 71.8% in Arapiraca) and were born in the same city where they lived (73.4% in Maceió and 73.5% in Arapiraca). In terms of education, a little over half had dropped out of school (56.6% in Maceió and 50.3% in Arapiraca), but the majority still maintained contact with their families (85.5% in Maceió and 89.6% in Arapiraca).

The second study that used the MSE method is by Coumans et al. (2017). This study aims to provide a national estimate of the homeless population in the Netherlands as an indicator of social exclusion. Importantly, the authors assert that The Netherlands is the first country to rely on MSE estimation (since 2009) to produce official statistics on homelessness. For this purpose, three registers were taken into account, along with characteristics such as sex, age, place of residence, and origin. The results indicated that in 2009, the total estimated number of homeless people was 17,767, of which only 5,169 were registered in any of the three lists used in the study. Therefore, they estimated a total of 12,589 homeless people who were not included in any of the registers. The ratio of unlisted to listed was 2.4. Regarding sex, the results showed that men are overrepresented compared to women. People aged between 30 and 49 are also overrepresented compared to the younger and older age categories. As for the non-Western origin category, it was more than three times higher than that of the total Dutch population.

The other two studies in the review used different approaches from MSE (Evangelist & Shaefer, 2020; Dutton & Jadidzadeh, 2019). Evangelist and Shaefer (2020) estimated the cumulative risk faced by students over time, from kindergarten to tenth grade. The study used event history analysis to calculate the initial and cumulative probabilities of student homelessness for the school years 2002-2003 to 2016-2017 by grade and race, evaluating both general and specific categories such as literal homelessness and doubling up. The study's results showed that between kindergarten and twelfth grade, nearly 1 in 10 students (9.5%) experienced homelessness at some point while in school, with 3.2% of students experiencing literal homelessness and 7.4% living in shared housing. Additionally, Black students were three times more likely to experience literal homelessness than White students. Fixed-effects models by county indicated that rental costs, forced housing relocations, and the opioid epidemic were associated with a higher number of homeless students.

Lastly, Dutton and Jadidzadeh (2019) combined administrative data on shelter stays in Calgary (2008-2014) and Toronto (2011-2015) with publicly available metropolitan census population estimates from Statistics Canada to calculate the incidence of shelter use per 1000 person-years in these two cities. The study also employed the k-means clustering method to classify shelter users based on the transience, periodicity, or chronicity of their homelessness situation. The results indicated that Calgary showed a higher overall risk of shelter use than Toronto, with 3.58 versus 1.18 per 1000 person-years. However, the results also indicated a higher likelihood of being a chronic shelter user in Toronto (0.09) compared to Calgary (0.06). Regarding age range, the risk of shelter use was nine times higher for individuals aged 16 to 20 in Calgary than in Toronto. In terms of macroeconomic factors, in Calgary, the employment rate was correlated with the incidence of homelessness ($r = 0.88$), whereas in Toronto, there was no such correlation ($r = -0.28$). The authors concluded that homelessness in Canada is a population-wide phenomenon that is significantly influenced by local macroeconomic factors.

Multiple Systems Estimation					
Study number, authors and country	**Defining Homelessness**	**Data sources**	**Sample characteristics**	**Results**	**Contextualizing results**
[1] Bezerra et al. (2011) Brazil	Children and adolescents who live or spend part of their day or night on the street.	Three lists: - *List A* : street survey - *List B* : street survey - *List C* : official lists from the municipal secretary of citizenship and social assistance in Maceió and from the public prosecutor office in Arapiraca	Children and adolescents living in the streets of Maceió and Arapiraca Maceió: - List A: N = 225 - List B: N = 266 - List C: N = 565 Total of uniquely identified individuals: N = 1,005 Arapiraca: - List A: N = 106 - List B: N = 92 - List C: N = 157 Total of uniquely identified individuals: N = 319	The estimated number of SC: Maceió: N = 5,225 Ratio unlisted to listed: 4.2 Arapiraca: N = 1,191 Ratio unlisted to listed: 2.7	Maceió and Arapiraca, the largest cities in Alagoas, a state with poor sociodemographic indicators, have a high prevalence of children and youth in the streets, especially Maceió, which significantly surpasses comparable cities due to factors such as poverty, infant mortality, and internal migration. Children and adolescents who live on the streets constitute an elusive population, with bad experiences with official agencies, who suffer abuse and violence on a regular basis. Hence the difference in SCs detected between the lists and the estimate made in the study.
[2] Coumans et al. (2017) Netherlands	Individuals who had no permanent accommodation (NPA) on the reference date. Three categories: - People who sleep outdoors - People who spend the night indoors in transient accommodation - People who sleep indoors in the home of friends, acquaintances or relatives	Three lists: - Shelter list: data from the basic system of municipal administration - WWB list: individuals between 18 and 65 years old who received income support (WWB), but did not have a permanent residence - Ladis list: individuals identified as homeless in the National Alcohol and Drug Information System	People with NPA between 18 and 65 years old in the Netherlands - Shelter list: N = 1548 - WWB list: N = 3492 - Ladis list: N = 883 Total of uniquely identified individuals: N = 5,169	Estimated number of homeless: 2009: N = 17,767 Ratio unlisted to listed: 2.4	The study's estimate is in line with a previous study on homelessness population counts in the Netherlands (De Bruin et al., 2003). Differences between the count and the estimate may be due to the elusive nature of this population or the lack of interaction with the agencies that conduct the homeless count.

Table 3a. Studies included in the systematic review that use Multiple Systems Estimation methods.

Other Methods					
Study number, authors and country	**Defining Homelessness**	**Data sources**	**Sample characteristics**	**Method**	**Results**
[3] Evangelist and Shaefer (2020) United States	Homelessness represents a dynamic state that contributes to other social inequalities. This study is focused explicitly on family homelessness. They distinguish doubling up and literal homelessness.	15 years of administrative data from the State of Michigan Department of Education integrated at the student level County-level Socio-economic data: - PEP: Population Estimates Program - SAIPE: Small Area Income and Poverty Estimates - BLS LAUS: Bureau of Labor Statistics Local Area Unemployment Statistics - HUD: US Department of Housing and Urban Development - ACS: American Community Survey - FRBNY: Federal Reserve Bank of New York - CDC: Centers for Disease Control and Prevention	Students from kindergarten through twelfth grade who are part of the public schools of the State of Michigan.	The study used life table methods to calculate the initial and cumulative probabilities of student homelessness for school years 2002-2003 to 2016-2017 by grade and race, assessing overall and specific categories like literal homelessness and doubling up.	Between kindergarten and twelfth grade, close to 1 out of 10 students (9.5 percent) were homeless at some point, with 3.2 percent of students experiencing literal homelessness and 7.4 percent having been doubled up.
[4] Dutton and Jadidzadeh (2019) Canada	People using emergency shelters. Homelessness refers to the situation in which people do not have a safe and stable place to reside, leading them to rely on emergency shelters to meet their basic shelter needs.	Three lists: - Calgary Homeless Foundation (CHF): central authority that collects statistics from local shelters, including basic demographics of those who use the shelter each day. - Toronto's Shelter, Support & Housing Administration (SSHA): service manager for housing and homelessness, including emergency shelter, social housing, and a range of housing stability options. - Statistics Canada's Canadian Socioeconomic Database (CANSIM): demographic and economic data providing the age and gender breakdown of the population and the employment rate.	Individuals using programs for single adults in emergency shelters, excluding families and youth-specific programs, aged 16 to 75 years old. - CHF list: *N* = 24,760 - SSHA list: *N* = 27,019	The study analyzes administrative data from shelter stays in Calgary (2008-2014) and Toronto (2011-2015). It combines these with population estimates to calculate the incidence of shelter use per 1,000 person-years in these cities.	Incidence of shelter use (per 1000 person-years): *Calgary* : 3.58 *Toronto* : 1.18 Incidence of shelter use by group (per 1000 person-years): *Calgary* (transitional 3.18; episodic 0.33; chronic 0.06) *Toronto* (transitional 0.99; episodic 0.10; chronic 0.09) Incidence of shelter use by age: *Calgary* : the highest incidence (4.50; 21 to 25 age group), the lowest incidence (0.49; 71 to 75 age group) *Toronto* : The highest incidence (1.71; 26 to 30 age group), the lowest incidence (0.28; 16 to 20 age group)

Table 3b. Studies included in the systematic review that used methods other than Multiple Systems Estimation.

2.2.4. Discussion

This systematic review has focused on studies published in peer-reviewed journals that have used integrated data systems to improve the count of homeless individuals. The aim was to understand the different sources that other studies use to implement integrated data systems and how these estimate the homeless population.

The two studies that used the MSE method indicated that the estimate of the homeless population was substantially higher than previous estimates using other methods. Becerra et al. (2011)'s results place the ratio of listed to unlisted youth at 4.2 for the city of Maceió and 2.7 for Arapiraca. They contextualize these results by indicating that these cities are located in the state of Alagoas, which has some of the worst sociodemographic indices in all of Brazil, with street children and adolescents being a significant problem. Therefore, sociodemographic indicators could help understand why the number of street children is so large, especially in Maceió, which significantly surpasses comparable cities due to factors such as poverty, infant mortality, and internal migration. The results obtained in this study indicated that the estimated number of street children is substantially higher than what is recorded in official lists. The MSE method is suitable to improve the count of a population that may have had negative experiences with official agencies and thus, be counted in their data. More reliable estimates can help improve the health and socioeconomic situation of street children in Maceió and Arapiraca.

Regarding the study by Coumans et al. (2017), the population facing homelessness is estimated to be 17,767, with a ratio of unlisted to listed of 2.4. Their estimation is consistent with a previous study that used the extrapolation method and estimated 15,200 people in the Netherlands (De Bruin et al., 2003). However, there were two earlier studies that also used the extrapolation method and obtained a higher estimate of homeless people (Heydendael & Brouwers, 1989, Van Der Zwet et al., 1990). Coumans et al. (2017) report that these differences may be due to the fact that those studies were conducted 30 years ago, so the time frame was completely different, as well as the definition of homeless people used, and the methods employed. The authors conclude that the differences between the register counts and MSE estimates may be due to the elusive nature of this population or the lack of interaction with the agencies conducting the homeless count.

In the study by Evangelist and Shaefer (2020), the authors highlight that analyzing the cumulative risk of students over the years allowed them to demonstrate how experiencing homelessness is a common occurrence, especially among minority students. Evangelist & Shaefer (2020) stated that when observing homelessness over an extended period, dynamic markers such as income poverty experience, food stamp usage, and incarceration are much more common than shown in cross-sectional studies (Enns et al. 2019; Grieger & Danziger, 2011).

Prior to Dutton and Jadidzadeh (2019)'s study there had been no estimates of the incidence of emergency shelters in Canadian cities. The results of this study indicated that these figures can differ drastically between large municipalities, attributing the differences in risks partly to population-level factors. The results showed that the risk of becoming homeless was three times higher in Calgary than in Toronto. This is true despite Calgary having a higher median income than Toronto, a smaller immigrant population, and lower market-based housing prices, all factors identified as individual-level correlates of homelessness (Statistics Canada 2016). Dutton and Jadidzadeh (2019) believe that their results provide a portrait of the homelessness problem beyond its individual-level correlates and consider that the risks could be partly attributed to population-level factors such as employment rates, the ratio of social assistance to rental prices, the extent of public benefits, and trends toward deinstitutionalization in mental health care.

The systematic review screened out some articles that aimed to estimate country-level population sizes of people facing homelessness. These were not included in the review as they did not meet the established inclusion criterion that the studies use integrated data systems. Some studies employed the MSE method but without using integrated data systems. For instance, to estimate the outdoor homeless population in Toronto, Brent (2007) applied MSE to data collected from street interviews. According to the author, this street-based approach is suitable for urban environments where homeless individuals are more likely to be seen. Brent (2007) considers that administrative lists result from selective and self-selective referral forces, whereas data collected on the street are independent of such biases. In the study by Stark et al. (2017), several estimation methods were used to determine the size of the homeless adolescent population in seven cities in Cambodia. On one hand, they conducted a complete count of homeless adolescents aged 13 to 17, and on the other hand, they used the MSE technique to perform a statistical estimation of completeness based on the overlap of the counts. Additionally, they conducted street interviews. Stark et al. (2017) concludes that using innovative methods will make it more feasible to collect data on this hard-to-reach population. Two other studies, however, did not use the MSE method. The first of these studies, by Kidd and Scrimenti (2004), conducted a homeless count through a single point-in-time survey. The study by Vameghi et al. (2019) used four methods (direct count, indirect count, wisdom of the crowd, and unique object multiplier) to estimate the size of the homeless child population but did not use an integrated data system.

There are other studies in the literature that were not found in the systematic review but that are important to highlight (Beata & Snijders, 2002; Feir & Akee, 2018; Fisher et al., 1994; Metraux et al., 2001). Three of these studies did not use integrated data system (Beata & Snijders, 2002; Feir & Akee, 2018; Metraux et al., 2001). Metraux et al (2021) analyzed HMIS data in nine US jurisdictions to produce estimates of annual prevalence counts and rates relative to the jurisdiction's population. Beata and Snijders (2002) estimated the homeless population in Budapest using two methods: snowball sampling and capture-recapture (MSE) methods. Feir & Akee (2018) use register and census data along with migration estimates to infer a more complete count of First Nations Canadians experiencing institutionalization and homelessness. We highlight their careful attention to how marginalization impacts individuals' representation in the data, Similar to our approach, they provide estimates stratified by age groups and sex assigned at birth. Finally, the study "Estimating numbers of homeless and homeless mentally ill people in northeast Westminster by using capture-recapture analysis" by Fisher et al. (1994) does make use of linked data from hospitals, homeless service agencies, and other agencies. Their estimates imply that for every person found in administrative lists there are about two people who were unaccounted for (1,640 observed; 3,293 estimated). However, our search strategy did not identify this study because it did not focus on a young population. And while the study relied on linked administrative data and the MSE method, their references to data and methods didn't include such terms; rather, they used terms such as 'routinely collected data', 'statutory or voluntary agencies', and 'capture-recapture.'

Limitations

This systematic review has some limitations. First, by focusing on populations of youth, our search strategy may fail to identify relevant articles as exemplified in the Fisher et al. (1994) article. Note however, that our search strategy did not necessarily require the keyword 'multiple systems estimation' to be included nor did it prevent the inclusion of studies with the term 'capture-recapture'). The search strategy was defined as a logical condition made up of sub conditions linked by the Boolean AND operator. One of such sub conditions is ("integrated data" OR "administrative data" OR estimat* OR "multiple systems estimation" OR "linked data" OR "data integration system"). While we did not explicitly include the term capture-recapture in this condition, the inclusion of OR estimat* relaxed the requirement for any of the other keywords to appear in the study.

Second, the included studies showed great heterogeneity in terms of the definition and conceptualization of homelessness, as well as the range of the estimated population. Third, this review is limited because it includes only peer-reviewed articles, which implies the exclusion of works not published in scientific journals, such as dissertations, reports from organizations, and gray literature.

Conclusions

This systematic review has demonstrated that the use of integrated data systems can significantly improve the estimation of the homeless population. The studies included in this review have utilized various sources and methodologies to implement these systems, which has allowed for more precise and detailed estimates.

The reviewed studies have shown that integrated data systems can be used to provide more accurate estimates of the homeless population relative to what is recorded on any one data system alone. This is particularly true when administrative agencies provide services based on different definitions of homelessness, so that lists have little overlap with each other. The use of integrated data systems also allows for the identification of sociodemographic factors that contribute to homelessness. For example, in the study by Becerra et al. (2011), it was highlighted how sociodemographic indicators can help understand the high prevalence of street children in certain cities. Furthermore, in our systematic review, we observed significant differences in the estimation of homeless populations across various countries. Notably, some countries exhibited a larger discrepancy between initial counts and subsequent estimates compared to others. This suggests that certain countries have more effective systems for collecting and identifying data on homeless populations.

Integrated data systems allow for the centralization of information from different sources, making information management for service provision more efficient (Fischer et al., 2019). Additionally, they can provide more reliable estimates of the extent of homelessness to inform social policy.

3. Analytical Approach

Consistent with Objective 1, we study the potential of integrated administrative data systems (IDS) to estimate a more comprehensive community-level count of youth experiencing homelessness that may provide valuable planning information for a more effective Continuum of Care. For this purpose, we leverage a community IDS, the Child Household Integrated Longitudinal Data (CHILD) system, spanning over 35 linked data systems from social service agencies in Cuyahoga County, Ohio, and focus on youth aged 13 to 25 years old, in compliance with HUD’s request.

The first part of the analysis provides an assessment of the administrative data held in CHILD, which includes the Homeless Management Information System (HMIS), and newly acquired administrative data to explore whether and how housing instability is tracked in the data. For this, we draw from the knowledge of key staff from service agencies, analyze the administrative data, and validate our findings with community stakeholders (section 3.1).

Following the assessment of existing and newly acquired data, we use it to develop a baseline registry of youth who were identified in records of homeless services in HMIS or administrative school data after careful deduplication. This baseline registry is then extended using a novel Address List method that leverages CHILD, our integrated data system. The Address List method flags administrative records of social services agencies where the address field is indicative of youth facing homelessness. Through this process, additional youth are identified leading to an extended registry of our target population (section 3.2). We proceed to estimate the number of youth not seen in the registry via Multiple Systems Estimation methods and we contextualize our results with benchmarks derived from survey and census data, as well as community knowledge (section 3.3).

Since the definition of homelessness varies across administrative systems and homelessness is a time-dependent state rather than an individual permanent trait, we are careful in describing the population whose size we aim to estimate. Youth in the registry have either been in contact with a system that provides homeless services, or they have provided an address to another service provider that is associated with homelessness or housing instability when interacting with other systems such as food support programs. This included when an address field was explicitly populated as homeless, or used government office, health care and other service providers, and hotel/motel/other temporary private shelter addresses. Thus, we seek to estimate **the annual prevalence of youth who experienced levels of housing instability similar to those in the registry, but who may or may not have interacted with the agencies represented in our data during that year.**

The registry includes those that have interacted with such agencies. To estimate the annual prevalence of unstably housed youth including those who have not interacted with agencies, we apply Multiple Systems Estimation (MSE) methods introduced in section 3.2.

3.1. Assessment and assembly of linked administrative data

3.1.1. Key informant interviews on data systems held in CHILD

The Center systematically receives and integrates administrative data from partners with which it has standing data sharing agreements to maintain the individual-level integrated data system, CHILD. This has allowed for conducting analyses to inform public policy and local program planning to address poverty, its causes, and its impact on communities and their residents. The CHILD data are stored within a secure research environment that meets the highest standards of physical, administrative, and technical controls. The linkage and analysis of data are allowed under the appropriate data use agreements and IRB approvals.

A core component of CHILD is data from the Homeless Management Information System (HMIS), which is vital to identifying a minimum count estimate of homeless youth. Though the Center has conducted numerous studies over time that have drawn on HMIS and other social service records, it had not fully explored the degree to which they can contribute to an understanding of youth homelessness. To do so, we reviewed available data documentation from our partners to identify data mechanisms that may indicate homelessness. We then conducted interviews with these partners to 1) evaluate whether the candidate data items were of sufficient reliability and quality, and 2) assess additional relevant data that have not been provided to us as part of standard extracts. These interviews were transcribed and thematically analyzed to provide additional insight into administrative data collection practices and challenges.

Nine primary group interviews with 31 organizational representatives were completed. Follow-up conversations were scheduled when further clarity was needed or to include additional personnel. Interview participants were part of the following agencies or organizations:

- County Office of Homeless Services
- County Department of Children and Family Services
- Public school districts covering the central city and surrounding suburbs (16 districts)
- County housing authority
- County juvenile court
- Regional food bank
- A local homeless service provider who manages both the by-name listing and coordinated entry for the Continuum of Care
- County Job and Family Services - which administers statewide food assistance benefits

In advance of the interviews, participants were provided with a summary and interpretation of the data received by the research team. Participants were asked to confirm the accuracy of the data and to contribute to the interpretation presented by researchers. Interviews were conducted as semi-structured interviews, allowing participants significant latitude to direct the conversation to issues which were particularly relevant to their organization, but with an aim to address a set of core concerns across agencies. Specific questions regarding the units of analysis, types of records, and characteristics recorded sought a nuanced understanding of the administrative data. Participants were also asked to describe the process and frequency of updating address data in their administrative data system. Answers to questions related to the collection, frequency, validity, and purpose of entering the address afforded deeper understanding of programmatic data collection and administrative use.

Finally, participants were asked about the reliability, universality, and representativeness of their data system in measuring youth experiencing homelessness. Agencies were asked pointedly to rate their system in terms of reliability and to describe if and how their data system captures information on vulnerable subpopulations of youth, including those identifying as LGBTQIA+, pregnant, and parenting youth. Interviews were transcribed and thematically analyzed by two members of the research team. Thematic analyses were combined and reviewed by the full research team for accuracy. In the initial interviews the research team also inquired about youth involvement in the programming provided by participants' agencies, including their engagement in a feedback process or commenting on policies and decisions that impact them.

Interview finding 1: *Administrative program data differ widely.* All key informants noted practical nuances of data collection practices at their respective organizations. These were closely aligned with processes that facilitated or hindered staff ability to provide needed services. Understandably, providers have specific goals for data collection, typically centered on funding, legal reporting requirements and improving service delivery for clients. One example is food assistance benefit data, where addresses are overwritten frequently to reflect current residence, ensuring uninterrupted assistance. These updates happen at least every six months, when benefits eligibility is redetermined. The organizations that administer these services do not have a need to maintain historical records of address since the current one is all that's needed to verify eligibility and mail necessary correspondence. In fact, maintaining historical records of addresses could unnecessarily increase data management costs for the organizations.

Some key informant agencies have flags for homelessness among the population they serve but expressed difficulty in standardized data entry and understanding of homeless experience by front-line staff. Multiple agencies indicated that more detailed information about homelessness would be in case notes, though case notes aren't necessarily filled out in a consistent manner. Moreover, there is difficulty determining stability in housing, as a mailing address may not change as frequently as the place in which they are residing. Once a youth is connected with CoC services, their address and demographic information is tracked in the data system of the provider. Multiple informants noted that specific addresses, namely those associated with known shelter locations or main social service provider offices, were considered indicators of housing instability. As noted by one informant

> "Sometimes people will leave [address] blank if they're at a shelter. But when the worker interviews them, and they find out they're in a shelter, they'll put that address in there as their residence address. They usually do have a different mailing address, and they may have it go to their brothers or something else, somebody a friend of theirs. Some shelters I think offer mailboxes, but not all."

Another informant reported

> "Sometimes when someone's homeless and not in a shelter, we'll actually use our address as their address. It's not a great solution, but we have to put something in there. So you may see some data that says that they live [at address]. That almost certainly means that they're homeless. So it's rare that we do that, and especially during the pandemic, it's very impractical, the lobby is closed to the public. But that's another kind of factor that will be in there."

Interview finding 2: Program experience is critical to accurate data interpretation. Administrative and client serving staff provided critical practice-based knowledge not captured in system data. One example of this is how data are captured in HMIS. Addresses of youth who enter the Continuum of Care (CoC) at Coordinated Entry are entered for this service only, which is typically one day (or less) in duration. Further, HMIS data elements for youth are not required to be captured by Coordinated Entry, leading to known gaps in the data. One key informant stated

> "Data quality issues come into play because of those variables [client perception and report], in my opinion, and also Coordinated Intake or Coordinated Entry across the country, they're not required to collect youth specific data elements. And so we have RHY providers, runaway and homeless youth providers, that are required to enter into HMIS. And it's this really robust data set. But if we could somehow, from a universality perspective, right, require coordinated entry to also collect those same elements. I think, you know, HUD could…do a better job of looking at the population."

For child welfare involved youth, homelessness may be experienced as part of their family of origin or as unaccompanied individuals. Data entered into child welfare systems records homeless status at time of intake and address as where youth were removed. Key informants with child welfare expertise noted the importance of knowing homeless status for case workers and staff supporting families. As evidence of commitment to this value, administrators at child welfare agencies voluntarily coordinated a pilot project with the local office of homeless services, wherein they manually checked the HMIS system records of youth and families experiencing homelessness. Without this pilot, access to HMIS records were restricted and child welfare providers resorted to using a flag of homelessness at entry, address data, and lists of youth who were "AWOL" from placements as indicators of homelessness. Key informants noted the complexity of youth experience and noted that some youth considered "AWOL" from placements were staying with family or friends who were not approved by the child welfare department but may be a stable living arrangement.

Multiple key informants noted changes to their data systems and transitions of staff during and following the COVID-19 pandemic. These changes led to loss of programmatic, institutional, and data system knowledge needed to contextualize youth experience of homelessness within their respective programs or service systems.

Interview finding 3: Defining youth homelessness is complicated, even in homeless-serving programs. Identification of youth homelessness is inconsistent and complicated by program specific administrative processes and willingness of clients to disclose. Programs do not transactionally keep the homelessness information but may do so depending on the preference of the specific organization or worker. While some programs have flags for homelessness or use specific guidelines for determining homelessness experience, data may be overwritten or not frequently updated. Each program had particular needs and perspective on youth homelessness, and therefore unique methods for gathering additional and rich data regarding youth homelessness. This presents a challenge when attempting to adapt a systematic data mining approach across nuanced administrative systems. Unaccompanied youth under age 18 are particularly challenging to track, as address is not captured in HMIS data unless they are part of a family.

> "I know that we've heard from specifically [the local school district], that there are some kids that they know of that just don't seem to be accounted for in some of our point in time counts. And I think some of that comes from just sort of the definition of homelessness that they're using. HMIS is really for those that are literally homeless as much as we can [know]. And then, you know, they're saying, well, they're just living with family because they have nowhere else to go. But you know, we're considering them homeless. So I think you have some of that sort of mismatch in definition that can lead to that. But I do think that there are kids that just don't hit a shelter that provides or that puts their data into HMIS. Or they hit some other resource that we're just not tracking in our database."

Interview finding 4: Centering youth is critical. Key informants desired to identify and help youth who are perceived as falling through the cracks, noting restrictive definitions of homelessness and lack of tailored services as access barriers. They identified the need for youth specific processes such as distinct staff and programs for youth diversion and prevention, as youth may not want or need emergency shelter.

Provider staff named agencies and programs that exemplify cross agency coordination. Despite the deep appreciation and critical role of service coordination in the continuum of care, staff acknowledged the role of funding in facilitating and inhibiting such projects. Some key informants identified the desire to link data systems across organizations to improve service coordination and youth experience. Doing so was also noted as critical to increasing awareness of vulnerable subpopulations of those experiencing homelessness, such as LGBTQIA+ identified and pregnant/parenting youth.

Key informants identified substantial pandemic-related impacts on the ability to center youth voice. In person interactions were minimized so focus groups, advisory boards, and planned youth specific intake initiatives were postponed or moved to remote access. Intake services shifted to phone based. Despite these limitations and shifts to program delivery, providers found prevention funding increased and allowed for expanded services.

3.1.2. New data acquisition

The second phase of the work identified other potential data sources held by community partners about youth experiencing homelessness to develop a fuller identification and understanding of homeless youth. The Center pursued new data use agreements with multiple partners for the purpose of this study. Sources were chosen based upon their presumed utility for identifying youth experiencing homeless and included:

- State education data: each school district maintains a listing of students in their enrolled population who meet the definition of homelessness under the McKinney-Vento Act, so they can provide specialized services to them. These data reflect high-risk younger youth (age 13-18).
    - McKinney-Vento Act data were obtained from the Cleveland Metropolitan School District and the State of Ohio Department of Education for the entire state[2].
- Food assistance
    - The Greater Cleveland Food Bank is a multi-county hunger relief organization providing access to, and connection with, a range of food-service programming across six counties. Data reflect persons who received services from the food bank during the study period.
    - SNAP benefits are administered through the state Office of Job and Family Services and provide food assistance to individuals and families based upon specific income-related eligibility criteria. SNAP records are at the case level which may be an individual of family group where the individual members are identified.
- By-name listing: this source is a cumulative listing of homeless young adults aged 18-24 who have been engaged through street-based outreach and referral efforts in Cuyahoga County. This listing was developed by partners following the Youth Count survey effort in 2015 to track their involvement with this segment of the youth population experiencing homelessness. This listing was used to manage outreach and engagement efforts by service providers with broad youth outreach. Note that this By-name listing is not related to the Built for Zero model.

Each of these sources required the development of a data-use agreement allowing individually identifiable data to be transferred to the research team for linkage and analysis. This process required the assistance of a university supported legal team but is a familiar part of university-based research.

---

[2] Data from the state were linked for analysis using the Statewide Student Identifier (SSID). The SSID is maintained within the CHILD data system through data use agreements with sixteen school districts. Only those youth whose SSID existed within the CHILD system were included in this analysis.

3.2. An IDS-based registry of youth facing homelessness: CCUYR

The Child-Household Integrated Longitudinal Data (CHILD) System is one of the oldest and most comprehensive integrated county-level data systems in the country (Kitzmiller, 2013). CHILD holds administrative records from over 35 administrative systems including Cuyahoga County Health and Human Services, area school districts, the Homeless Management Information System, the Ohio Department of Job and Family Services and other local agencies. Integrated data systems serve to break down agency silos and can be a unique and powerful resource for planning, monitoring, evaluation, and action to improve services for families.

Maintaining CHILD requires routinely integrating agency administrative data that has been pre-processed and standardized to ensure high quality linking. A similar process was implemented with newly acquired data specific to this project. The integration process relies on deterministic and probabilistic data linkage methods (Coulton et al., 2017) using Soundex phoneticization, Edit-Distance, and Jaro-Winkler algorithms to compensate for multiple error types in the administrative data, including typographical errors, transcription errors, or misspellings. A sequential automated and manual process determines new linkages for individuals with and without previous records in CHILD.

Variables used for matching include individual and parent names, birth date, and sex assigned at birth. When available, identifiers like home addresses are also considered for matching. We draw race and sex at birth from the earliest administrative records pertaining to an individual that routinely document these variables. In most cases, this means that sex comes from birth certificates and race comes from public assistance records (Medicaid, Supplemental Nutrition Assistance Program – SNAP, or Temporary Assistance for Needy Families - TANF)[3]. However, we recognize that race and gender are social constructs and may change over time. When the linking process is completed, each individual has a unique identifier that allows us to track their records across agencies and over the study period. This process is used to create a registry of youth facing homelessness.

3.2.1. The Address List Method

For the current analysis, we created a registry of youth identified in our linked administrative data as having experienced homelessness or housing instability in Cuyahoga County during the calendar years 2017, 2018 and 2019. We refer to this condition as “facing homelessness.” For each year, we focused on youth between their

[3] About 87% of our study population have an identifier related to public assistance programs.

13th and 25th birthday. More specifically, for the 2017 count, we select all individuals born at any time between 1/1/1992 through 12/31/2004, as the time period between their 13th and 25th birthday that would have included at least one day in 2017. Likewise, the birth date ranges for 2018 and 2019 are 1/1/1993 – 12/31/2005 and 1/1/1994 – 12/31/2006, respectively.

The Cuyahoga County Unhoused Youth Registry (CCUYR) leverages the CHILD system, starting with all youth in CHILD that fit the age criteria within the study period, and identifying youth with homeless experiences based on the following sources of data:

- Homeless Services: Continuum of Care Homeless Service data come from two sources that are closely related: the Homeless Management Information System (HMIS) and the By-name Listing (BNL).
- School-Based Homeless Services: School data on students experiencing homelessness is collected in compliance with the McKinney-Vento Act. The Center has data use agreements with the Cleveland Metropolitan School District (CMSD) and 15 other school districts in the county for inclusion of their data into CHILD. This represents about half of the school districts in the county. For this project, we obtained Ohio Department of Education and Workforce (ODE) data that allowed us to identify students attending any of the 16 school districts in CHILD with records in McKinney-Vento data.

We identify additional youth facing homelessness by comparing the address field of administrative data related to social services and other systems not directly related to homelessness against a list of addresses associated with homelessness. **We call this the Address List method.** Address fields were chosen using locations listed in the Street Card[4] or Relink.org[5] (a list of housing resources provided by non-profit organizations). We included addresses of transitional housing, addiction rehabilitation centers, or government agency addresses that are used when an individual interacts with such agencies and lacks a fixed home address. Additionally, since some administrative records may explicitly state homelessness in the address field, those records were included. The address list method was developed based on community knowledge drawn from interviews with agency partner informants (section 3.1), which pointed to the multiple ways in which homelessness could be flagged in their administrative data.

We applied the address list method to search address fields in Supplemental Nutrition Assistance Program (SNAP) records for 2017 – 2019, Greater Cleveland Food Bank data for 2019, child welfare (DCFS), and juvenile court (JC) filings data for 2017 to 2019. After

[4] Produced and distributed by the Northeast Ohio Coalition for the Homeless (NEOCH). Available at https://www.neoch.org/street-card

[5] https://needs.relink.org/services/homeless-shelter-shelter-crisis?cid=61

completing this address search, we obtained counts of youth 13-25 years of age that have at least one of their records associated with homelessness or housing instability each year.

The unique count of youth with homeless records by year is 6,060 in 2017, 6,198 in 2018, and 5,896 in 2019. The total number of unique youth 13-25 years of age included in the CCUYR across all three years is 11,652. Most youth appear in records for just one year (58%), but 28% of youth appear in records in two years and 14% in all three years, indicating longer periods of housing instability.

Table 4 provides a detailed account of the number of youth contributed by the address list method beyond the population accounted for via HMIS and McKinney Vento school data. *Across all three years, the address list method taps into our integrated data system to add between 15% and 24% of youth to the registry that had not been accounted for in administrative data associated with homeless services.* It is important to note that most of the additional records come from SNAP, a program that has a large population coverage relative to the other administrative data used. SNAP records are a valuable source of information on housing instability when processed with the address list method because they provide monthly addresses for program participants. In comparison, food bank administrative data in our community records the address of people at the time of their most recent interaction with the food pantry, overwriting the previous value in the address field. Thus, we lose the ability to identify instances of homelessness for food pantry users prior to their most recent interaction. Given this restriction in the administrative data, we limit the use of Food Bank data to the most recent year, 2019.

The design of the address list to identify instances of homelessness in administrative data was done in close consultation with our data partners. For reporting purposes, we classify the types of addresses associated with homelessness in Cuyahoga County broadly into (1) government office addresses, (2) health care and other service providers, (3) an indication of homeless or homeless services address, and (4) hotel/motel/other temporary private shelter. Across all three years, records where a government office address was entered instead of a home address are the ones that most contribute to the registry. For instance, in 2017 there were 1,453 young people identified by the Address List method that had not been identified in HMIS, the By-name listing, or school homeless service data. Of those, 52% had an administrative record with a government address recorded in the home address field. This figure is 57% and 47% for the years 2018 and 2019, respectively. However, it merits noting that the final list of addresses selected will reflect the unique characteristics of each community, and the alternatives to shelter made available to people experiencing homelessness.

	2017		2018		2019	
Total Annual Registry Count	6,060		6,198		5,896	
Flagged with address list method	**All**	**New additions**	**All**	**New additions**	**All**	**New additions**
SNAP	1,833	1,389	1,690	1,271	1,012	665
Juvenile Court	25	21	19	17	23	17
DCFS	108	43	86	28	80	31
Food Bank	0	0	0	0	241	194
Total	1,966	1,453	1,795	1,316	1,356	907
As share of Annual Registry Count	32%	24%	29%	21%	23%	15%
Share contributed by address type						
Government Office		52%		55%		47%
Health Care Provider		28%		25%		14%
Homeless, Homeless services		12%		15%		27%
Hotel, mobile home shelter		8%		6%		12%
Total		100%		100%		100%

Table 4. Counts of youth at risk of homelessness identified by the Address List method in selected data sources in CHILD.

3.2.2. Characteristics of youth identified as facing homelessness in the CCUYR

Once the registry is complete, the integrated data system upon which it was built provides rich information on youth facing or at risk of homelessness that can be used to guide local programs and policies. Table 5 provides a summary of only a few characteristics available through CHILD, spanning early childhood to young adulthood. We note the elevated levels of lead exposure that youth in the registry present, with more than half of youth ever tested presenting elevated levels of lead in their blood. Also noteworthy is that more than 60% of youth had been enrolled in SNAP in the year prior to the experience of homelessness or housing instability documented in the registry. While the association between food and housing insecurity is not surprising[6], it suggests that cross-system partnerships between SNAP and housing stability programs might play a role in preventing youth homelessness.

[6] Although not directly comparable, the Department of Job and Family Services reports that 15% of the population in Cuyahoga County were enrolled in SNAP as of July 2024. https://data.ohio.gov/wps/portal/gov/data/view/snap-population-metrics

	2017	2018	2019
Total count	6,060	6,198	5,896
Youth with Birth Certificates in CHILD	4,084	4,141	3,848
Youth with Birth Certificates (%)	67.4	66.8	65.3
Among those with a Birth Certificate			
Low Birth Weight (%)	12.7	13.6	13.5
Premature Birth (%)	13.1	13.3	14.1
Adequate Prenatal Care (%)	62.6	61.7	58.3
No Prenatal Care (%)	4.8	4.4	3.9
Mother's age (mean)	23.8	24.0	24.0
Mother's age (median)	22	23	23
Mother has HS diploma (%)	50.3	52.2	52.5
Previous year programs			
SNAP enrollment	66.2	64.7	62.2
TANF Cash receipt	10.6	9.4	8.8
HMIS any service	27.9	30.5	33.1
HMIS shelter	7.4	8.2	7.9
Youth with Lead Test	4,037	4,099	3,887
Youth with Lead Test %	66.8	66.1	65.9
Among tested			
Elevated blood lead level (%)	64.7	61.5	56.3
Youth aged 13 to 18	3,121	3,372	3,521
Previous year events			
Child welfare report	20.5	23.0	23.8
Placement	4.1	4.3	5.8
Juvenile court filing	7.9	7.2	4.9

Table 5. Select characteristics of youth in CCUY Registry derived from the CHILD integrated data system.

3.3. Multiple Systems Estimation (MSE)

Multiple Systems Estimation (MSE) is a statistical methodology used to estimate the total number of individuals in a population, based on a sample of that population appearing on one or more lists. It is also known as capture-recapture estimation, given its applications in ecology. The hidden figure, a term used to denote individuals in the population that do not appear on any of the observed lists, is estimated using the patterns of list overlaps under modeling assumptions that in recent years have been relaxed to align closer to reality.

3.3.1. A textbook example of MSE would be one in which the total number of children in a population, *Nc*, is estimated with two independent random samples (lists) of children drawn in two consecutive days, as seen in figure 3. In this example, *Nc* = 14. The count of children in list 1 ($nc1 = 6$), list 2 ($nc2 = 5$), and the count of children appearing in both lists ($nc12 = 2$) can be used to estimate the count of children not appearing on either list: the hidden figure.

This is because, *if lists are independent random samples of the population and the population is closed*, the prevalence of List 1 in the population should approach the prevalence of List 1 within List 2, for large enough sample sizes.

$$\frac{nc1}{Nc} = \frac{6}{14} \sim \frac{nc12}{nc2} = \frac{2}{5}$$

Thus, *if the samples overlap* – that is, if *nc12 > 0,* we can estimate the total number of children in the population as $\widehat{Nc} = \frac{nc1}{nc12} nc2$

In our example, the estimated number of children is $\widehat{Nc} = \frac{nc1}{nc12} nc2 = \left(\frac{6}{2}\right) 5 = 15$

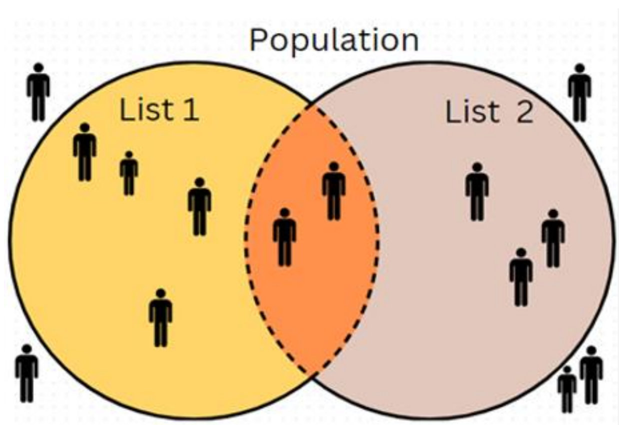


Figure 3. *Illustration of Multiple Systems Estimation with two lists.*

While standard MSE methods account for more than two lists, current advances allow for the presence of non-overlapping lists and introduce modeling considerations to account for the fact that lists are not independent random samples of the population (Chan et al., 2021). Far from random samples, lists are often derived from administrative records by government or non-profit organizations. However, methodological advances have made MSE a valuable tool for estimating populations that are vulnerable and difficult to quantify, such as victims of human trafficking or terrorism and state violence, for instance.

The input for MSE can be represented in a table of count data for each combination of list appearances observed. With the two lists in our example, we have $2^2 - 1 = 3$ possible combinations or histories as seen in table 6. For instance, (1,0) is a list history representing appearance in list 1 but not list 2, and nc(1,0) (nc(1,0) = 4 in our example) is the count of children with list history (1,0). In our example, (1,1) is the list history representing appearance in both lists, and so on. The number of children not appearing on either list (= 5 in our example) is not observed.

List 1	List 2	Number of children	List history description
1	0	nc(1,0) = 4	# children seen in list 1 but not 2
0	1	nc(0,1) = 3	# children seen in list 2 but not 1
1	1	nc(1,1) = 2	# children seen in list 1 AND 2
0	0	nc(0,0) = 5	Hidden figure: # children not seen in either list
		nc = 14	

Table 6. List histories and counts for the textbook example of MSE.

In that sense, integrated data systems are key to implementing MSE since they allow deriving list history counts based on individual-level data linked across multiple lists of data.

We use MSE to estimate the number of youth experiencing housing instability in the county based on their appearance in an integrated data system of administrative records, following recommendations in the literature regarding the use of administrative data.

3.3.2. Addressing the violation of standard MSE modeling assumptions

Jones et al. (2014) show that a main assumption of MSE -independence of lists- is violated when agency administrative lists are connected by a direct referral system between agencies. Direct referrals between agencies can be a powerful tool to address the needs of people. However, for estimation purposes, referral-related lists imply complex models with high-order correlation parameters that may not be estimable due to having more parameters than count data. Thus, while we have access to eight administrative lists, we combine them into four lists, where correlation is likely between each pair due to agency coordination and referrals. HMIS and BNL form one list (HMIS); CMSD and ODE is another list (CODE); SNAP and Food Bank data form a third list (SNAP/FB), and child welfare and juvenile court (DCFS/JC) make up the fourth and last list.

This method also requires the target population to be closed. We know that youth facing homelessness or housing instability can enter or exit the county at any point during our study period. However, by estimating yearly counts rather than counts over longer time periods, we reduce the instances of in and out mobility of this population.

Another important consideration is the violation of the MSE assumption of homogeneity. MSE requires that individuals have equal probability of appearing on a list (homogeneity), or that heterogeneity in the probability of appearing in a list is parametrized in the model. However, qualitative research and community knowledge derived from Data Chats with youth who had experienced homelessness[7] suggest that young people's reliance on shelter depends on their sense of safety in shelter relative to other alternatives, their social networks, and the stigma they experience when reaching out to homeless service providers (Richter et al., 2023). When embedded in a society that discriminates by race or gender identity, we recognize that these features will influence their likelihood of appearing on the aforementioned lists. Thus, we follow recommendations by Lum et al. (2013) and estimate counts by year-race-sex-age group strata to reduce heterogeneity.

Even though LGBTQ youth are overrepresented in the population experiencing homelessness (see Rice et al., 2013), most administrative data systems do not provide reliable information on gender identity. In addition, we recognize that racial categories have no biological basis and have been found to perpetuate bias (see Compton et al., 2023). However, since discrimination continues to play a role in access to housing and wellbeing across multiple domains, we consider it important to account for it in our stratified analysis. We classify youth into two categories: white and non-white, based on administrative records. Finally, we create two age categories to distinguish school-aged youth from older youth who will face different options for system engagement.

3.3.3. Analysis of CCUYR by strata

Figure 4 presents yearly counts of the CCUYR broken down by sex assigned at birth, age and racial identification groups as discussed in the previous section. Notably, across these subgroups, there are three to four times more youth identified as non-white than those identified as white. More specifically, in every year, between 76% and 79% of youth in the registry are non-white identified. This fact reflects large racial disparities in housing

[7] https://cwru-dsci.org/?page_id=70

stability and opportunity. Absent those disparities, we would expect to see closer to an even distribution of youth identified as white versus non-white in the registry, given that 47% of youth aged 13-25 are non-white identified in Cuyahoga County[8].

Figure 4 also shows that the time trend in the yearly count of youth facing homelessness varies by age group. Between 2017 and 2019, there is a slight increase in the yearly count of school aged youth (13-18) facing homelessness, while for older aged youth (19 to 25 years of age), the count is clearly lower in 2019 relative to the previous years. However, differences in trends are difficult to interpret in the registry, as they might be influenced by changes in agency policies and procedures, ranging from eligibility criteria to staff training and capacity.

Similarly, differences in the count of youth across age groups is at least partly data dependent. McKinney Vento data is available for school aged youth, but not for the older group. Regarding sex differences, among school-aged youth, men and women are evenly represented in the data, but fewer men than women in the older age group are recorded in the youth registry.

Figure 4. *Counts of youth with records indicating homelessness by year and strata according to the Cuyahoga County Unstably Housed Youth Registry.*

---

[8] *IPUMS USA, University of Minnesota, www.ipums.org*

3.3.4. MSE Estimation of youth facing homelessness

We estimate MSE models following the methods derived by Chan et al. (2021) to address issues due to sparse data and non-overlapping lists. Their associated R program, SparseMSE, is used to test a large number of model specifications, screen for the existence and identification of model parameters, and ultimately produce estimates for the strata that meet both criteria. In order to estimate these models, we structure the data as counts of youth associated with each possible list history. The counts corresponding to all list histories are modeled as a Poisson Loglinear Model with parameters indexed by the possible list histories.

Given our four combined administrative lists, a list history can be any of 15 possible combinations of patterns describing whether a person appears or not on these lists. For example, for a given year, some youth only appear in HMIS and no other lists. We label this list history as HMIS. Another group appears in the By-name listing and school data but not on any other lists. This list history is labeled as HMIS+CODE, etc. Youth in the registry are associated with one of the 15 list histories possibly observed with four lists ($2^4 - 1 = 15$). We aim to estimate the total population of youth facing homelessness or housing instability that includes those not seen in any of the four lists, sometimes called the 'hidden figure.'

Table 7 shows the total count of youth that appear in each of all possible list histories or combinations by year. Not surprisingly, most youth are seen in homeless service data, whether it be HMIS or CODE (McKinney-Vento) lists and not in other system data. However, the number of youth seen in any two data sets ranges from below 10 (CODE+JUV/DCFS) to almost 300 in the case of HMIS+SNAP. When estimating the total population count, we do so by age-race id-sex at birth strata, in which case the counts by list histories will be much smaller.

List history	CODE	HMIS	JUV/ DCFS	SNAP/ FB	2017 youth count	2018 youth count	2019 youth count
In one list only							
CODE	1	0	0	0	1,882	1,993	2,099
HMIS	0	1	0	0	2,081	2,246	2,299
JUV/DCFS	0	0	1	0	64	45	49
SNAP/FB	0	0	0	1	1,389	1,271	871
In exactly two lists							
CODE+JUV/DCFS	1	0	1	0	<10	<10	<10
CODE+SNAP/FB	1	0	0	1	66	48	52
HMIS+CODE	1	1	0	0	175	199	198
HMIS+JUV/DCFS	0	1	1	0	11	10	11
HMIS+SNAP/FB	0	1	0	1	294	280	246
SNAP/FB+JUV/DCFS	0	0	1	1	13	11	10
In exactly three lists							
CODE+SNAP/FB+JUV/DCFS	1	0	1	1	<10	<10	<10
HMIS+CODE+JUV/DCFS	1	1	1	0	<10	<10	<10
HMIS+CODE+SNAP/FB	1	1	0	1	42	57	29
HMIS+SNAP/FB+JUV/DCFS	0	1	1	1	17	12	<10
In all four lists							
HMIS+CODE+SNAP/FB+JUV/DCFS	1	1	1	1	<10	<10	<10
Not on any list (hidden figure)	0	0	0	0	?	?	?

Table 7. *Yearly counts of Cuyahoga County youth (ages 13 -25) with records in administrative data indicating homelessness by list history. There are 15 possible list histories given the four groups of administrative data used in the study: Homeless Management Information System and the By Name List (HMIS), McKinney-Vento Cleveland Metropolitan School District and Ohio Department of Education (CODE), Supplemental Nutrition Assistance Program (SNAP) and Greater Cleveland Food Bank (SNAP/FB), juvenile court filings and child welfare - DCFS (JUV/DCFS).*

Results of the model estimation for each stratum are presented in table 8. We follow Chan et al. (2021)'s stepwise method for optimal model selection. All selected models include an intercept and three main effect parameters underscoring the relevance of HMIS, school and food assistance data.

The first stratum refers to young people between the ages of 13 to 18 identified (in administrative data) as nonwhite and female who appear in the 2017 CCUYR. The registry counts 1,137 individuals in this group. Given their list histories, the model estimates that there are 2,875 additional young people in this stratum that are unlisted, making up a total of 4,012 youth. The upper and lower bounds of this total (MSE) estimate are provided, along with the ratio of unlisted to listed youth. We can interpret this ratio as stating that for every young person identified by administrative lists as experiencing housing instability, there are 2.5 more under similar homelessness circumstances that have not been identified by such lists. The estimates are consistent for this stratum over the following two years, with ratios of 2.7 and 2.6 for 2018 and 2019 respectively.

Across all strata, we see that the ratios of unlisted to listed youth range from 1.5 to 3.2, suggesting that in some cases, there might be up to three times more youth facing homelessness unseen in administrative data relative to those that are identified. Out of the 24 strata, there are two groups for which the method fails to produce reliable estimates. In both cases, these models correspond to young males identified as white and in the older age group, where school data is less able to capture their housing instability. This is not surprising given the estimation issues presented by sparse data.

According to the American Community Survey (ACS) 5-year 2017-2021 extract, Cuyahoga County is estimated to have 204,175 young people aged 13 to 25, 22% (45,869) of which have incomes that fall below the poverty line[9]. It is important to note that the Census Bureau's population estimates are based on a random stratified and weighted sample of the population. Contrastingly, the CCUYR includes individuals served by some administrative agencies in the County but misses individuals not engaged with those social service agencies. Furthermore, CCUYR provides a count of individuals found to have homeless records in administrative data over the course of an entire year (annual prevalence), which is likely larger than the point-prevalence, or the count of this population on any given day.

With that understanding, we cannot benchmark the CCUYR counts or the MSE estimates to the ACS estimate, but we can cautiously compare them to subsequently gauge the plausibility of our estimates. The yearly counts of youth with administrative records signaling homelessness hovered around 6,000, which represents about 3% of the size of the County youth population aged 13-25 and 13% of the size of the youth population falling below the federal poverty level. As a reference, Metraux et al. (2001) find that the 1998 prevalence of *sheltered* homelessness across 9 jurisdictions in the United States represented between 1.3% to 10.2% of the population in poverty.

Our estimated unlisted-to-listed ratio is consistent with the studies analyzed in the systematic review that used MSE. Notably, Coumans et al. (2017) uses high quality national registries for the Netherlands and an approach similar to our Address List method. Their population estimates reflect an unlisted-to-listed ratio of 2.4, in line with our estimates.

[9] *IPUMS USA, University of Minnesota, www.ipums.org.*

Sex at birth	Age group	List Year	Racial Id group	Count youth listed	MSE estimate - total	MSE lower bound	MSE upper bound	Unlisted estimate (MSE - listed)	Estimate unlisted-to-listed ratio
F	13-18	2017	nonwhite	1,137	4,012	3,355	4,864	2,875	2.5
M	13-18	2017	nonwhite	1,222	4,868	4,019	5,974	3,646	3
F	13-18	2017	white	373	1,298	967	1,813	925	2.5
M	13-18	2017	white	383	1,615	1,176	2,298	1,232	3.2
F	19-25	2017	nonwhite	1,189	3,496	3,069	4,020	2,307	2
M	19-25	2017	nonwhite	1,076	4,014	3,398	4,794	2,938	2.7
F	19-25	2017	white	314	1,042	796	1,415	728	2.3
M	19-25	2017	white	355	882*	713*	1,131*	527*	1.5*
F	13-18	2018	nonwhite	1,247	4,596	3,856	5,546	3,349	2.7
M	13-18	2018	nonwhite	1,277	4,550	3,844	5,451	3,273	2.6
F	13-18	2018	white	411	1,662	1,222	2,339	1,251	3
M	13-18	2018	white	423	1,498	1,142	2,031	1,075	2.5
F	19-25	2018	nonwhite	1,248	3,439	3,036	3,933	2,191	1.8
M	19-25	2018	nonwhite	1,003	3,729	3,142	4,476	2,726	2.7
F	19-25	2018	white	297	1,151	839	1,643	854	2.9
M	19-25	2018	white	272	715	569	933	443	1.6
F	13-18	2019	nonwhite	1,366	4,973	4,267	5,852	3,607	2.6
M	13-18	2019	nonwhite	1,352	5,359	4,471	6,500	4,007	3
F	13-18	2019	white	370	1,224	941	1,646	854	2.3
M	13-18	2019	white	406	1,704	1,248	2,406	1,298	3.2
F	19-25	2019	nonwhite	1,095	2,722	2,380	3,154	1,627	1.5
M	19-25	2019	nonwhite	823	2,034	1,740	2,422	1,211	1.5
F	19-25	2019	white	248	935	674	1,354	687	2.8
M	19-25	2019	white	203	618*	458*	878*	415*	2*

Table 8. *MSE Estimated counts of youth facing homelessness or housing instability by strata. The asterisk following two of the strata estimates (*) correspond to estimates that failed the model identifiability test.*

4. Summary and Recommendations

Enumerating the number of youth experiencing homelessness is crucial for effective policy-making and resource allocation. Understanding the prevalence of youth homelessness allows for targeted intervention and appropriate youth support to transition to stability. However, several factors complicate accurate estimation. The dynamic nature of homelessness, data collection challenges, and lack of resources at social service agencies may hinder reliable enumeration.

Furthermore, varying definitions of homelessness across agencies lead to systematic undercounting of the population. Age-specific nuance, such as the developmentally appropriate tasks of youth individuation and limited engagement with formal support systems, further complicate accurate counts. Youth experiencing homelessness often rely on peer relationships and trusted adults rather than shelters, which are less utilized due to restrictive policies and limited access. Socially marginalized youth face additional hurdles, engaging less with services and thus absent from administrative data counts.

Ideally, to get a true sense of the magnitude and scope of youth homelessness in a community, a data system would need to be documenting a person's housing at regular and frequent intervals as housing instability results in a good deal of fluctuation in one's address over time. This is not the case with administrative data. These data are often collected as a part of service delivery and, as such, regular and frequent documentation of details pertinent to housing stability or instability may not be captured at multiple time points, instead these details may only exist in the data at a single point in time.

Importantly, the ability of administrative data to identify youth facing homelessness may be impacted by larger systemic issues. First, administrative data systems are a product of the society in which they are created and thereby reflect the biases, namely systemic racism, that are present in communities. Communities of color have historically been over-represented in many administrative data systems due to systemic racism and a history of over-surveillance. Second, administrative data are principally created for program delivery and are subject to decision making at multiple levels, including what data to collect and in what format. These collective decisions may not make the data ideal for answering questions at the system level without considerable investment in understanding the underlying data.

With those caveats acknowledged, we set out to analyze the potential for integrated administrative data systems to enhance accuracy, providing a more comprehensive understanding of the prevalence of youth experiencing homelessness and housing instability.

ADMINISTRATIVE DATA AND INTEGRATION HOLD MUCH PROMISE FOR IMPROVING UNDERSTANDING OF YOUTH EXPERIENCING HOMELESSNESS

The methods of this study sought to determine the value of utilizing collective knowledge among community partner organizations in regard to youth homelessness. Our findings show the incremental value of a series of tactics on this front for communities that use HMIS in the Continuum of Care. See Figure 5. By integrating HMIS data with student records from schools, estimates of the total number of youth experiencing homelessness were increased by approximately 75%, for example from 2,820 to 4,871 in 2018. Next by using address-level information across systems, total estimates were increased by 20-30%. Lastly, using MSE methods across several lists, estimates were dramatically increased (by over 200%).

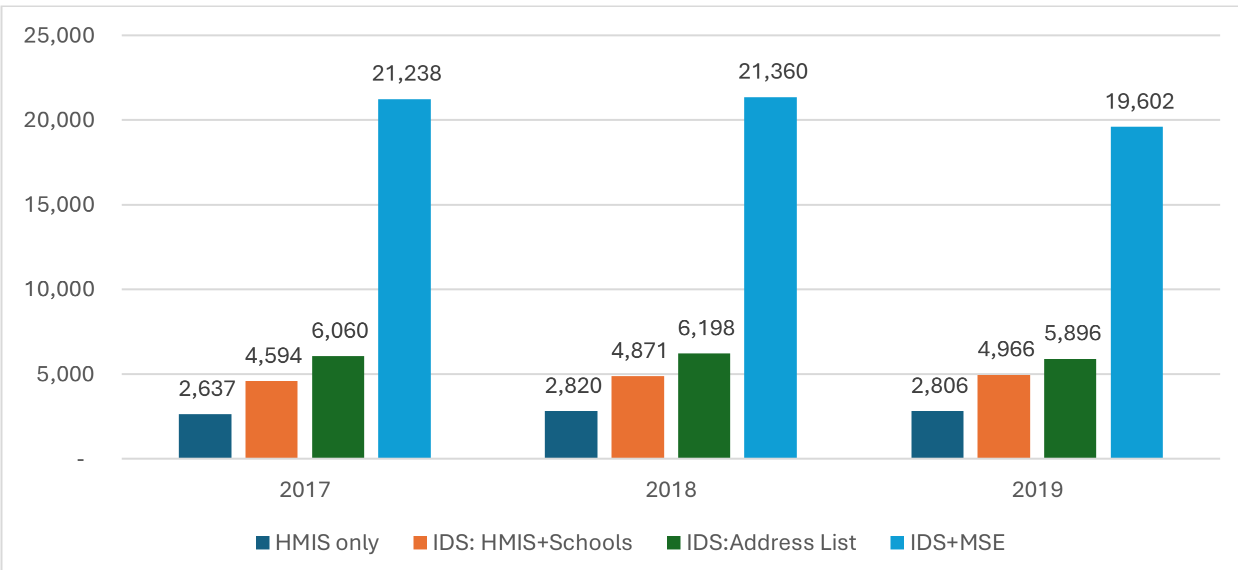


Figure 5. *Estimates of youth experiencing homelessness by year in Cuyahoga County by Different Data Integration Strategies*.

When expanding on data in HMIS, by far, the largest value comes from the inclusion of school records. This source alone nearly doubles the estimates of youth experiencing homelessness by identifying school-age youth not represented in other systems. The addition of address-listing methodologies further improves these estimates by 20-230% by identifying youth known to other systems. Lastly, MSE provides an estimate of the full population of youth experiencing homelessness by leveraging the value of these multiple lists to estimate counts of youth who are unconnected to system partners.

ENGAGING COMMUNITY PARTNERS TO UNDERSTAND ADMINISTRATIVE RECORDS ON YOUTH HOMELESSNESS

Each data source considered for this analysis was designed for unique purposes relevant to the administration of programming for different groups of clients. These data were not collected for the purposes of conducting research and as such, they are limited in important ways that cannot be ignored. For example, the unit of analysis within data systems are unique and designed around the delivery of service. Therefore, youth living within families may not always have distinct records in the data as a "head of household" or an "assistance group" may be preferable based on the service being delivered (i.e. food pantry usage, public assistance, public housing, etc.) routinely list the head of household, other members of the household are captured but not always in as routine a manner).

Collaborating with data partners was essential to gain a better understanding of the data, whether we had ample experience working with it (such as HMIS) or were unfamiliar with it (such as McKinney-Vento school data). In both cases, data partners provided invaluable support.

Insights from partners and youth with experiential knowledge suggested that marginalized youth facing homelessness may be less likely to seek formal services than non-marginalized youth, leading to pattern of missingness in the data that is not random. Missingness may reflect safety concerns or distrust in the system by some of the most vulnerable young people who experience homelessness. This has implications for the way in which we implemented our estimation models, segmented in a way to reduce the heterogeneity within groups.

DEFINITIONS AND DOCUMENTATION ACROSS PROVIDERS

Through interviews with key stakeholders, including homeless service and social safety net providers, it became apparent that data collection, quality, and retention, directly impact the ability to track youth experiencing homelessness in our community. Programs typically use their administrative data systems to provide services and report outcomes. Staff are trained to collect information and data control measures internal to the organization support quality aligned with reporting and program-specific goals. Paper forms and data systems may not have fields to collect information key to homelessness, such as type or length of current residence. Addresses of local shelters or transitional housing locations may not be allowable at the program level and youth may provide address information of a family member or friend instead. If shelter or transitional housing addresses are allowable by programs, these data fields may be updated frequently (weekly or monthly, such as each time a youth frequents a food pantry). But some data systems rarely maintain historical data due to cost and technical requirements of storing it. As such, an experience of homelessness would be lost as soon as the address field was updated. For example, a youth who experiences three weeks of homelessness or housing instability in June but returned to housing stability in July would not reflect this unstable state in an August data draw.

PREVALENCE-BASED COUNTS

This study used a proprietary integrated longitudinal data system that included individually identifiable information from more than thirty-five administrative data sets within our county. These data provide valuable contextual information about the population of individuals included in the analysis and facilitated linkage across data acquired for the study. Youth were included in the study through a homelessness flag in the administrative data related to homeless services (HMIS, By-name listing, CMSD, ODE), or through the Address List method, which used address field records in other data (SNAP, food bank, child welfare or juvenile court) as an indicator of homelessness or housing instability. This approach estimated annual prevalence counts of 6,060 youth in 2017, 6,198 in 2019, and 5,896 in 2019.

Youth from all data sources were combined to create the CCUYR, the analytic data set used to conduct MSE and generate population prevalence estimates. As noted in the previous analysis section, 24 MSE strata were created to address the violation of the MSE assumption of homogeneity. For each year we had eight strata formed dividing the population into groups by age, racial identification, and sex assigned at birth (2x2x2 = 8). Across all strata, ratios between 1.5 and 3.2 were produced by the MSE, indicating that there may be up to three times more youth facing homelessness than are documented in the current administrative data systems.

BENEFICIAL ADMINISTRATIVE DATA SYSTEMS

Access to HMIS and McKinney-Vento education data are critical components to tracking youth experiencing homelessness and housing instability. Continuums of Care and school districts are consistently funded and have the technical assistance necessary to collect and maintain data systems that produce valid and reliable information on youth. Lists maintained by homeless outreach providers (such as the By-name list used in this analysis) are critical to documenting youth who may not interact with other systems of care. For data that don't contain discrete homelessness variables, using addresses that indicate housing instability are useful to identifying youth who receive public assistance, interact with child welfare, or juvenile legal systems and may not disclose homelessness. With support of our community data partners, we were able to identify additional youth facing homelessness by comparing the address field of their administrative records (not directly related to homeless services) against a list of addresses associated with homelessness.

RECOMMENDATIONS

Access to individually identifiable data on youth and young adults necessitates data use agreements, high levels of data security, and strong community collaboration. Having a central partner to facilitate these agreements and navigate relationships among the collaborative team can improve efficiency.

The collaborative process to integrate community knowledge into data analysis is continuous. Sharing findings and seeking feedback from community data partners and people represented and not represented in the data is essential to promote a reinforcing cycle of better data and better programs to address homelessness and advance equity in housing stability for young people.

Administrative data held by organizational partners contain substantial value for understanding youth experiencing homelessness in a region. Communities beginning down the path of data integration are advised to prioritize school records as the first source to consider expanding on the value of HMIS. Address list techniques also provide a meaningful addition once data are available and integrated across multiple partners. Finally, multiple systems estimation offers a novel approach to better assessing the vulnerable population of youth likely to be facing issues of homelessness and housing instability.

5. References

Anderson, B. E., Williams Clark, A., Hill, D., & and LaCaria, C. (2020). Child & youth homelessness in Charlotte-Mecklenburg integrated data report (2016 – 2017). UNC Charlotte Urban Institute. https://mecklenburghousingdata.org/child-youth-homelessness-report/

Anthony, E. R., & Fischer, R. L. (2016). Surveying homeless and unstably housed youth: methodological considerations when estimating the prevalence and characteristics of the population (research note). Families in Society, 97(4), 330–335. https://doi.org/10.1606/1044-3894.2016.97.40

Auerswald, C. L., & Adams, S. (2018). Counting all homeless youth today so we may no longer need to tomorrow. *Journal of Adolescent Health*, *62*(1), 1-2.

Beata, D., & Snijders, T. A. (2002). Estimating the size of the homeless population in Budapest, Hungary. *Quality and Quantity*, *36*, 291-303. https://doi.org/10.1023/A:1016080606287

Bezerra, K. F., Gurgel, R. Q., Ilozue, C., & Castaneda, D. N. (2011). Estimating the number of street children and adolescents in two cities of Brazil using capture–recapture. *Journal of Pediatrics and Child Health*, *47*(8), 524-529. https://doi.org/10.1111/j.1440-1754.2011.02015.x

Bird, S. M. & King, R. (2018). Multiple systems estimation (or capture-recapture estimation) to inform public policy. *Annual Review of Statistics and Its Application*. https://doi.org/10.1146/annurev-statistics-031017-100641

Brent, B. (2007). A repeated observation approach for estimating the street homeless population. *Evaluation Review*, *31*(2), 166-199. https://doi.org/10.1177/0193841X06296947

Brown, A. (2020) The changing categories the U.S. census has used to measure race. Pew Research Center Publication. https://www.pewresearch.org/short-reads/2020/02/25/the-changing-categories-the-u-s-has-used-to-measure-race/

Building Changes. (2019). Students experiencing homelessness in Washington's K-12 public schools: trends, characteristics and academic outcomes, 2015-2018. https://buildingchanges.org/resources/students-experiencing-homelessness-in-washington-s-k-12-public-schools-trends-characteristics-and-academic-outcomes-2015-2018-2/

Burns, D., Espinoza, D., Ondrasek, N., & Yang, M. (2021). Students experiencing homelessness: the conditions and outcomes of homelessness among California students. https://learningpolicyinstitute.org/product/students-experiencing-homelessness-report

Chan, L., Silverman, B. W., & Vincent K. (2021). Multiple systems estimation for sparse capture data: Inferential challenges when there are nonoverlapping lists. *Journal of the American Statistical Association*, *116*(535), 1297-1306. https://doi.org/10.1080/01621459.2019.1708748

Clark, A. W., Lane, J. T., & Gaines, A. M. (2017). Charlotte-Mecklenburg family homelessness snapshot 2014-2015. University of North Carolina at Charlotte Urban Institute. https://www.neighborhoodindicators.org/sites/default/files/publications/Family%20Homeless_3_17.pdf

Community Solutions (2024) By-Name Data. A pillar of the Built for Zero methodology. https://community.solutions/quality-by-name-data/

Coumans, A. M., Cruyff, M. J. L. F., Van der Heijden, P. G., Wolf, J. R. L. M., & Schmeets, H. J. S. I. R. (2017). Estimating homelessness in the Netherlands using a capture-recapture approach. *Social Indicators Research*, *130*, 189-212. https://doi.org/10.1007/s11205-015-1171-7

Cowan, C. D., Breakey, W. R., & Fischer, P. J. (1988). The methodology of counting the homeless. *Homelessness, Health, and Human Needs*, 12-20.

Cutuli, J. J., Treglia, D., & Herbers, J. E. (2020). Adolescent Homelessness and Associated Features: Prevalence and Risk Across Eight States. *Child Psychiatry & Human Development*, *51*(1), 48–58. https://doi.org/10.1007/s10578-019-00909-1

D'Onise, K., Wang, Y., & McDermott, R. (2007). The importance of numbers: Using capture-recapture to make the homeless count in Adelaide. *Australian Journal of Primary Health, 13*(1), 89–96. https://doi.org/10.1071/PY07012

De Bruin, D., Meijerman, C., Verbraeck, H., Braam, R., Leenders, F., & Van de Wijngaart, G. (2003). Dwelling in the 21st century: An explorative study for mental illness among and nuisance from roof- and homeless people in the Netherlands. *Utrecht: Centrum voor Verslavingsonderzoek.*

Dutton, D. J., & Jadidzadeh, A. (2019). The incidence of homelessness in canada is a population-level phenomenon: a comparison of Toronto and Calgary shelter use over time. *Canadian Studies in Population*, *46*, 161-171.https://doi.org/10.1007/s42650-019-00013-8

Embleton, L., Lee, H., Gunn, J., Ayuku, D., & Braitstein, P. (2016). Causes of child and youth homelessness in developed and developing countries: A systematic review and meta-analysis. *JAMA Pediatrics*, *170*(5), 435-444. https://doi.org/10.1001/jamapediatrics.2016.0156

Enns, P. K., Yi, Y., Comfort, M., Goldman, A. W., Lee, H., Muller, C., ... & Wildeman, C. (2019). What percentage of Americans have ever had a family member incarcerated?: Evidence from the family history of incarceration survey (FamHIS). *Socius, 5*, 2378023119829332. https://doi.org/10.1177/2378023119829332

Evangelist, M., & Shaefer, H. L. (2020). No place called home: Student homelessness and structural correlates. *Social Service Review*, *94*(1), 4-35.https://doi.org/10.1086/707569

Evans, J., & Baker, T. (2021). Breaking through the epistemic impasse: Ending homelessness with the invention of 'functional zero'in the Anglo-American world. *Futures*, *129*, 102730.

Feir, D. L., & Akee, R. (2018). Estimating Institutionalization and Homelessness for Status First Nations in Canada. *International Indigenous Policy Journal, 9*(4), 1-28. https://www.jstor.org/stable/48767548

Fischer, R. L., Richter, F. G.-C., Anthony, E., Lalich, N., & Coulton, C. (2019). Leveraging administrative data to better serve children and families. Public Administration Review, 79(5), 675–683. https://doi.org/10.1111/puar.13047

Fisher, N., Turner, S. W., Pugh, R., & Taylor, C. (1994). Estimating numbers of homeless mentally ill people in north east Westminster by using capture-recapture analysis. *British Medical Journal, 308*, 27–30. https://doi.org/10.1136/bmj.308.6920.27

Glasser, I., Hirsch, E., & Chan, A. (2013). Ethnographic study of the group quarters population in the 2010 census: homeless populations. https://www.census.gov/content/dam/Census/library/working-papers/2013/adrm/ssm2013-14.pdf

Grieger, L. D., & Danziger, S. H. (2011). Who receives food stamps during adulthood? Analyzing repeatable events with incomplete event histories. *Demography, 48*(4), 1601-1614. https://doi.org/10.1007/s13524-011-0056-x

Heydendael, P. H. J. M., & Brouwers, H. G. (1989). *Mensen in de marge in soorten en maten. [People on the fringes in all kinds and measures]. Tijdschrift voor Sociale Gezondheidszorg, 8, 4–8.*

Hopper, K., Shinn, M., Laska, E., Meisner, M., & Wanderling, J. (2008). Estimating numbers of unsheltered homeless people through plant-capture and postcount survey methods. *American Journal of Public Health*, *98*(8), 1438-1442. https://doi.org/10.2105/AJPH.2005.083600

Jones, H. E., Hickman, M., Welton, N. J., De Angelis, D., Harris, R. J., & Ades, A. E. (2014). Recapture or precapture? Fallibility of standard capture-recapture methods in the presence of referrals between sources. American journal of epidemiology, 179(11), 1383–1393. https://doi.org/10.1093/aje/kwu056

Kidd, S. A., & Scrimenti, K. (2004). Evaluating child and youth homelessness. *Evaluation Review*, *28*(4), 325-341. https://doi.org/10.1177/0193841X04264820

King County's Department of Community and Human Services. (2021). Integrating data to better measure homelessness DCHS data insights series. https://kingcounty.gov/~/media/depts/community-human-services/department/documents/KC_DCHS_Cross_Systems_Homelessness_Analysis_Brief_12_16_2021_FINAL.ashx?la=en

Kitzmiller, E. M. (2013). IDS Case Study: Case Western Reserve University. Actionable Intelligence for Social Policy. University of Pennsylvania. https://www.datanetwork.org/wp-content/uploads/2017/02/Cuyahoga-County_CaseStudy.pdf

Lowell, W., & Hanratty, M. (2022). Who counts? Educational disadvantage among children identified as homeless and implications for the systems that serve them. Social Service Review, 96(4), 581–616. https://doi.org/10.1086/722003

Lum, K., Price, M. E., & Banks, D. (2013). Applications of Multiple Systems Estimation in Human Rights Research. The American Statistician, 67(4), 191–200. https://doi.org/10.1080/00031305.2013.821093

Mast, B.D. (2020). Measuring homelessness and resources to combat homelessness with PIT and HIC data. *Cityscape: A Journal of Policy Development and Research 22*:1. https://www.jstor.org/stable/26915494

Meltzer, A., Quintero, D., & Valant, J. (2019). Better serving the needs of America's homeless students. Brookings. https://www.brookings.edu/articles/better-serving-the-needs-of-americas-homeless-students/

Metraux, S., Culhane, D., Raphael, S., White, M., Pearson, C., Hirsch, E., ... & Cleghorn, J. S. (2001). Assessing homeless population size through the use of emergency and transitional shelter services in 1998: Results from the analysis of administrative data from nine US jurisdictions. *Public Health Reports*. https://doi.org/10.1093/phr/116.4.344

Metraux, S., Manjelievskaia, J., Treglia, D., Hoffman, R., Culhane, D. P., & Ku, B. S. (2016). Posthumously assessing a homeless population: services use and characteristics. Psychiatric Services, 67(12), 1334–1339. https://doi.org/10.1176/appi.ps.201500388

Meyer, B. D., Wu, D., Mooers, V., & Medalia, C. (2021). The use and misuse of income data and extreme poverty in the United States. *Journal of Labor Economics, 39*(S1), S5-S58.

Meyer, B. D., Wyse, A., Grunwaldt, A., Medalia, C., & Wu, D. (2021). Learning about homelessness using linked survey and administrative Data. https://bfi.uchicago.edu/wp-content/uploads/2021/06/BFI_WP_2021-65.pdf

Morton, M. H., Dworsky, A., Matjasko, J. L., Curry, S. R., Schlueter, D., Chávez, R., & Farrell, A. F. (2018). Prevalence and Correlates of Youth Homelessness in the United States. *Journal of Adolescent Health*, *62*(1), 14–21. https://doi.org/10.1016/j.jadohealth.2017.10.006

Morton, M. H., Kugley, S., Epstein, R., & Farrell, A. (2020). Interventions for youth homelessness: A systematic review of effectiveness studies. *Children and Youth Services Review*, *116*, 105096. https://doi.org/10.1016/j.childyouth.2020.105096

Morton, M., Dworsky, A., Samuels, G. M., & Patel, S. (2018). Voices of youth count comprehensive report: Youth homelessness in America.

National Network for Youth. (n.d.). Homeless and runaway youth in the juvenile justice system. https://www.juvjustice.org/sites/default/files/resource-files/Homeless%20and%20Runaway%20Youth_0.pdf

Pergamit, M., Cunninghan, M. K., Burt, M. R., Lee, P., Howell, B., & Bertumen, K. D. (2013). Counting Homeless Youth | Urban Institute. https://www.urban.org/research/publication/counting-homeless-youth

Richter, F. G.-C., Coury, N., Nelson, E. 2023. Integrating Community Knowledge into Data Analytics. Online Resource. https://cwru-dsci.org/

Schneider, M., Brisson, D., & Burnes, D. (2016). Do we really know how many are homeless?: An analysis of the point-in-time homelessness count. *Families in Society, 97*(4), 321-329. https://doi.org/10.1606/1044-3894.2016.97.39

Smith, C., & Castañeda-Tinoco, E. (2019). Improving homeless point-in-time counts: uncovering the marginally housed. Social Currents, 6(2), 91–104. https://doi.org/10.1177/2329496518812451

Snow, D. A., Baker, S. G., Anderson, L., & Martin, M. (1986). The myth of pervasive mental illness among the homeless. *Social Problems, 33*(5), 407-423. https://doi.org/10.2307/800659

Sullivan, A. A. (2023). What does it mean to be homeless? How definitions affect homelessness policy. *Urban Affairs Review, 59*(3), 728-758. https://doi.org/10.1177/10780874221095185

Stark, L., Rubenstein, B. L., Pak, K., Taing, R., Yu, G., Kosal, S., & Roberts, L. (2017). Estimating the size of the homeless adolescent population across seven cities in Cambodia. *BMC Medical Research Methodology*, *17*, 1-8. https://doi.org/10.1186/s12874-017-0293-9

U.S. Department of Education. (2020). ED Facts school Data status. US Department of Education. https://www2.ed.gov/about/inits/ed/edfacts/data-files/school-status-data.html

U.S. Department of Housing and Urban Development Office of Community Planning and Development. (2007). The annual homeless assessment report to congress. https://www.huduser.gov/portal/sites/default/files/pdf/ahar.pdf

U.S. Government Accountability Office. (2021). The challenges in counting and serving homeless populations. https://www.gao.gov/blog/challenges-counting-and-serving-homeless-populations

Vameghi, M., Roshanfekr, P., Ali, D., Noroozi, M., Madani, S., McFarland, W., & Mirzazadeh, A. (2019). Population size estimates of street children in Iran: synthesis of multiple methods. *Journal of Urban Health*, 96, 549-557. https://doi.org/10.1007/s11524-019-00354-4

Van der Zwet, G. P., Van der Meijden, R. R., & Burgers, L. (1990). *Dak-en thuislozen. Aantallen, opvang en gemeentelijk beleid; een inventariserend onderzoek. [Roof-and homeless people. Numbers, shelter and municipally policy].* The Hague: Vereniging voor Nederlandse Gemeenten (VNG; Council for Dutch Municipalities), afdeling Sociaal-Geografisch en Bestuurskundig Onderzoek (SGBO; department of Social-Geographic and Science of Public Administration).

Whitbeck, L., Lazorita, M. W., Crawford, D., & Hautala, D. (2016, April). Data collection study final report. Accessed from https://www.acf.hhs.gov/sites/default/files/documents/fysb/data_collection_study_final_report_street_outreach_program.pdf

Wolf, J., Zwikker, M., Nicholas, H., Van Bakel, H., Reinking, D., & Van Leiden, I. (2002). *Op achterstand. Een onderzoek naar mensen in de marge van Den Haag. [About deprivation. A study into marginal people in The Hague].* Utrecht: Trimbos Institute.

6. Appendix

6.1. Qualitative Data Collection Instrument

Key Informant Interviews

Semi-Structured interview questions for administrative data system leaders/data managers to get a better understanding of if and how youth homelessness might be documented in their systems

Thank you for your willingness to participate in this interview. Please answer these questions as best as you can, there are no 'right' answers. If you have any questions or are not sure what I'm asking, please stop me and ask. At any point during the interview, you can skip a question for any reason. You are also free to stop participating at any time for any reason.

Do you have any questions before we get started? (Answer questions if any). Okay, great. Let's get started.

[Provide each respondent with a short-written summary of data we currently have in-hand and our interpretation of what it contains]

1. Can you confirm what's described here? Is this accurate? Anything we're missing? Can you provide additional detail for us on the unit of analysis, frequency of update, types of records and characteristics recorded, years included, how to interpret missing data fields etc…

2. Does your data system record the home address of individual persons? If so, at what frequency?

3. How is a person's race recorded in the system? Is it self-reported?

4. If the person is homeless at the time of recording and/or they are experiencing housing instability of some kind how would that be documented, if at all?

5. Would a shelter ever be entered as an address for this person?

6. Given the frequency with which your data system is updated, how would a period of homelessness or housing instability be documented? (would it be a spell, a point in time, is there always a date attached?)

7. On a scale of 1 to 10 with 10 being the best, how would you characterize your data system's documentation of youth homelessness in terms of
    a. Reliability? (the degree to which you are confident in its accuracy)
    b. Universality? (is the data collected in a consistent way for all, or is it more haphazard?)

8. Does your data system document the extent to which individuals might be included in sub-populations that we are interested in learning more about, such as youth identifying as LGBTQ, pregnant youth, parenting youth? If so, how is that data documented and at what intervals?

9. Do you convene youth for the purposes of getting their insight and feedback at your agency? If so, would it be possible for us to engage with this group to learn more about their perceptions of youth homelessness?

6.2. Homeless Services Data Intake Process: Perspectives from the Data Chat with Youth who faced homelessness

As part of a project to Integrate Community Knowledge into Data Analytics (Richter et al., 2023), we partnered with the Northeast Ohio Coalition for the Homeless to host a Data Chat with youth who had experienced homelessness in the past year.

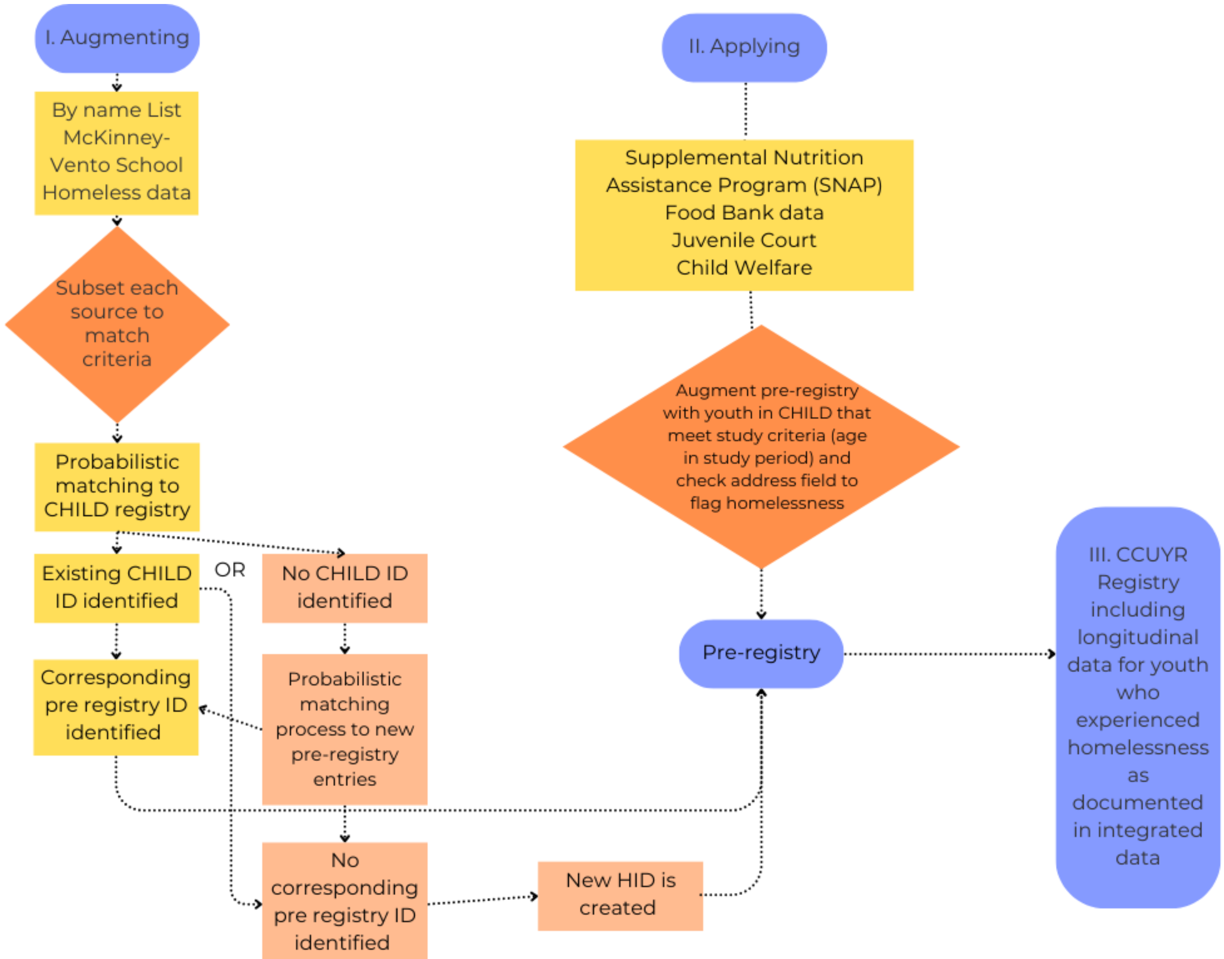
The aim was to enhance our understanding of the meta data and data for homeless services, drawing from the experiential knowledge acquired by young people who engaged in the data intake process for HMIS.

The main points below suggest that the quality and comprehensiveness of data will improve as the system becomes more capable of meeting the needs of youth. The full analysis can be accessed here https://cwru-dsci.org/?page_id=70

Themes	Youth voices
Data intake process varies by location, tied to capacity and resources	They didn't offer me no assistance, no programs, nothing. I was just in there clueless.
	At XX I didn’t talk to nobody. I even asked them for three days and they said I didn’t have a caseworker. I almost left and went back to my family. At intake, they just threw me to the wolves. At YY it was way different. I couldn’t even go to my room without going through intake. I met my social worker, I met the facilitator, I met the security guard.
Data intake experience varies case by case	I think the questions were fine. They got the information they needed. And they kind of made it seem like they cared about what was going on with me.
	The data intake, first going into the shelter, is useful because they ask you what all you need on there, then you tell them, but I just wish that the people there, or the workers who work at the shelter, at least should reach out to programs and help them apply for assistance that they need to get out of the shelter.
	Just the same questions just want to check them out and that's the end of it. There's no words of hope...
Data intake could be geared to connect people with needed services	I wish what happened more often ... for coordinated intake, I wish like once we find out there's an issue, ... like you need counseling? Let's get you a referral right then and there. We used to have our therapists no longer there. And ... they would call us and they'll help us when we get housed rejected.
Data intake could help triage people into different spaces for safety and better service	For shelters, I would compartmentalize, like mental health, young people, people with kids, I would have different buildings. Even if it is the same building, I would have different spots. Like this building
Ideas about using shelters as opportunities to provide training to youth so they can build a stable future.	same with, how you get into college, and then figure out like, what level you should start at. I feel like that should be inside shelters too...

Table A1. Summary of insights provided by youth Data Chat participants on their experiences with the data intake process for homeless services.

6.3. Using Integrated Data to Develop the Cuyahoga County Unhoused Youth Registry (CCUYR)


I. Augmenting
By name List McKinney-Vento School Homeless data
Subset each source to match criteria
Probabilistic matching to CHILD registry
Existing CHILD ID identified
OR
No CHILD ID identified
Corresponding pre registry ID identified
Probabilistic matching process to new pre-registry entries
No corresponding pre registry ID identified
New HID is created
II. Applying
Supplemental Nutrition Assistance Program (SNAP)
Food Bank data
Juvenile Court
Child Welfare
Augment pre-registry with youth in CHILD that meet study criteria (age in study period) and check address field to flag homelessness
Pre-registry
III. CCUYR Registry including longitudinal data for youth who experienced homelessness as documented in integrated data