



Building a Sustainable and Replicable Approach to Estimating Youth Homelessness Using Linked Administrative Data

SEPTEMBER 2025

CENTER FOR POLICY RESEARCH

Lanae Davis, MPA; Ali McLagan, MA; Kayla Christiani, MA

COLORADO EVALUATION AND ACTION LAB THE UNIVERSITY OF DENVER

Elysia Verson, PhD; Whitney LeBoeuf, PhD;
Chistian Belcher, MPA; Erin Wickerham, MPH



The Center for Policy Research (CPR), in collaboration with the Colorado Evaluation and Action Lab (the Lab), developed an approach to link administrative data from multiple systems to more rigorously count and estimate the prevalence of youth who have experienced homelessness in Colorado. This resource provides a brief overview of the approach to inform state and local agencies on how to use administrative data to generate sustainable and replicable data on youth homelessness.

Data sharing is a powerful tool to improve policies and programs that affect youth experiencing homelessness and housing instability. Data stewards must prioritize creating policies and practices that protect the privacy of individuals while balancing the reporting of accurate reliable data to advance prevention efforts.

Although we focus on the target population of youth experiencing homelessness, this guide is intended for a broad audience including government agencies, community-based organizations, and researchers to support other data reporting and estimation efforts when the ability to link cross-system data is available (e.g., public health and criminal justice populations).

The approach consists of seven steps and two cross-cutting approaches to estimating youth homelessness. These steps are intended to be an iterative process, so that study goals can be expanded or refined based on learnings at each step. This resource describes the iterative steps, as well as how to identify key partners and include youth with lived experience in every step of the process. **Those steps are:**

1. Defining study goals

2. Identifying administrative data systems

3. Prioritizing administrative data systems

4. Selecting identity resolution approaches to link data across systems

5. Conducting analysis

6. Framing and communicating findings

This resource is an abbreviated version of a full community guide that can be found [here](#). Additionally, the full study report provides more details about the analysis and findings and can be found [here](#).

Iterative Approach

FIGURE 1. PROCESS FOR LINKING CROSS-SYSTEMS DATA TO IDENTIFY AND SUPPORT YOUTH EXPERIENCING HOMELESSNESS

Cross Cutting Approaches

When embarking on a process that impacts multiple systems and partners, it is important to build strong support networks as well as engage young people who are experts in the impact of systems involvement and housing instability.

Cultivate champions at every step

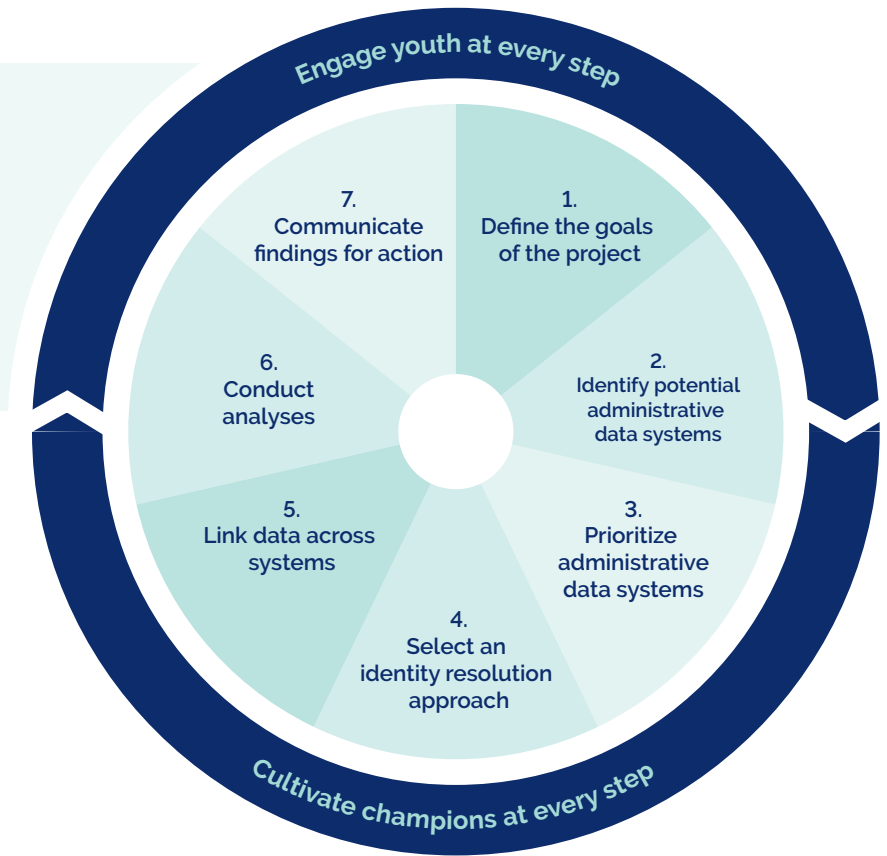
When considering who should be recruited as a potential champion for the project questions like, “who are potential champions for the project,” “who are the stewards of the administrative data that already are or can become champions for this study,” and “what are their needs and opportunities for use of the study results?”

Champions are individuals at data partner agencies who have a vested interest in the project and who can help the research team navigate legal and procedural barriers that come up when accessing, linking, and reporting on analysis from administrative data. Read more about cultivating champions throughout each step of the process in the full community guide section titled Prioritizing Administrative Data Systems.

Engage youth at every step

As youth are engaged throughout the process, it is important to consider how youth with lived experience report they appear in systems, how they define homelessness and housing instability, whether they are equitably represented in the data systems, and how the project will ensure youth with various demographic, social, and geographic variations are represented.

Use youth engagement where they are best and most interested in supporting the process. An in-depth look at how and why engaging youth with lived experience is important to this work is provided in the full community guide section titled Centering the Voices of Lived Experts through an Equity Lens.





Steps

1. Define the goals of the study

Questions that frame this step include: What are the ages, characteristics, and geographic areas of interest for estimating youth homelessness? Ideally, what estimates would be generated (e.g., point in time, annual, known counts, unknown estimates)?

The first step of this approach is to outline the initial vision of how the data will be used, the target population, and the goals for the study. These goals will evolve over the course of the approach.

An important step in counting youth experiencing homelessness is establishing a consistent definition of homelessness that aligns with the study's goals. To establish a working definition, the research team should review agency definitions of homelessness and engage youth with lived experience to define homelessness in a way that reflects their realities. For a more in-depth discussion of how CPR and the Lab established their definition of youth homelessness see the full community guide [here](#).

2. Identify potential administrative data systems

Questions that frame this step include: What are the data systems in those geographic areas that are likely to offer information aligned with the goals of the study? Who provides services to youth experiencing homelessness, and how do they track service delivery? How do youth with lived experience identify themselves in these systems?

After establishing clear goals, the next step is to identify which administrative data systems are most likely to capture meaningful indicators of youth homelessness. Examples of relevant systems include:

- *Homeless Management Information Systems (HMIS):* Tracks service use within the homeless response systems.
- *McKinney-Vento Education Data:* Identifies students experiencing housing instability.
- *Child Welfare Systems (e.g., SACWIS):* May include data on risk factors, placements, and service referrals.
- *Other systems:* Public housing authorities, TANF/SNAP, Medicaid, juvenile justice, and mental health services.

Research teams may use a combination of publicly available data and subject matter experts to begin to rule in and rule out potential data sources for a study. This step requires both technical insight and partnership-building, as systems vary in scope, quality, and readiness for collaboration.

- Leverage existing contacts and partnerships both in governmental and non-governmental agencies. Consider adding data discussions to existing meetings and workgroups to begin the process of identifying and onboarding data partners. This can serve two purposes, one being identifying interested data holders and learning about the data they hold, along with identifying their goals and potential use for the data.
- Capitalize on existing relationships. Strong partnerships with state agency data owners or engaging champions early in this step is crucial to fully explore the data landscape and navigate the legal and technical barriers likely to arise. Youth input at this stage also supports the selection of data sources explored, as youth can share their experiences with various systems to identify where they may have been recorded as experiencing homelessness.

3. Prioritize administrative data systems

Questions that frame this step include: What are the feasibility and cost considerations for each data system? What does each system uniquely contribute to the estimates or the study goals?

With a list of potential systems in hand, the next step is to prioritize those that offer the greatest potential value for analysis, feasibility for access and use, and strength of partnership. Key strategies for prioritization include:

- *Assessing value:* Does the data illuminate key aspects of youth experiencing housing instability or housing status?
- *Evaluating feasibility:* Are data access and linkage technically and legally possible?
- *Fostering collaboration:* Are champions available to support the work long-term?

Key Considerations:

- Use existing agency data sharing agreements, where available.
- Determine if personally identifiable information (PII) will be shared, and if so in what environment.
- Determine informed consent process for accessing and using identifiable data.

The crosscutting approach of cultivating champions is key in this step as the individuals within agencies will advocate internally to help move partnerships forward and resolve barriers. The success of a partnership with data sharing agencies and champions within them requires a commitment to reciprocity, flexibility, and responsiveness.

Initial conversations with data partners should be framed around mutual benefit: researchers should not only request data access but also demonstrate how the study can support agency goals. One-on-one engagement with agency leadership, data privacy officers, and technical staff is critical to shape data governance and sharing frameworks that meet both agency and project needs. Offering flexible engagement options and being transparent about risks and benefits can further strengthen trust with data partners. When partnerships don't show a clear mutual benefit, researchers should be prepared to respectfully move on.

4. Select an identity resolution approach

Questions that frame this step include: What identity resolution tools and services are available to connect individual records across systems? Are there existing data sharing agreements or efforts that can be leveraged (this study used the Linked Information Network of Colorado's data sharing agreements)? Who will link and analyze the data and in what environment?

Explore options for linking and analysis of the data. For example, some states have a data warehouse or other type of system that allows linking in a secure environment. This study used the LINC model, which includes a three-tier legal framework designed to ensure a common, high-level commitment to collaboration among LINC data partners while providing flexibility in lower-level agreements to allow data partners to meet their legal obligations. If that option does not exist, explore working with a research partner to link the data and conduct the analysis. Important considerations are privacy and security of the data. These discussions need to occur before finalizing any partnerships and data sharing agreements are in place. Use existing data share agreements and legal frameworks for data sharing as examples and modify those rather than starting from scratch.

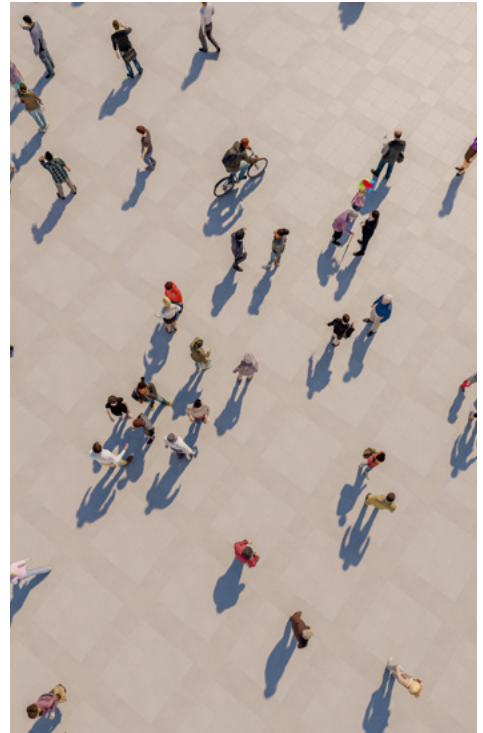
Once data is shared, linking records across systems requires a carefully chosen method to determine whether different data points refer to the same individual. This is a foundational technical decision that affects data quality and privacy.

Common approaches include:

- *Manual matching:* Reviewing records by hand to determine matches.
- *Rule-based matching:* Using exact or deterministic criteria (e.g., full name, birth date).
- *Fuzzy matching:* Allowing for variations due to typos or inconsistent formatting.

The approach selected should balance accuracy, scalability, and the capacity of project partners. Input from system champions and data stewards is essential to navigate these technical decisions. Youth engagement can also help identify common naming or identity challenges experienced by young people in services.

CPR and the Lab used an already established data network called the Linked Information Network of Colorado (LINC) to gain access to potential data systems, select an identity resolution approach, and link data. LINC is a collaborative initiative housed within Colorado's Governor's Office of Information Technology that utilizes a federated data model to link data across state and local agencies on a project-by-project basis. LINC has a dedicated data scientist who specializes in identity resolution, who used probabilistic matching procedures for this project.



5. Link data across systems

Questions that frame this section include: What are the common identifiers that allow for identity resolution across the data from each system? Is there a single system contributing to the population of interest or does each system need to contribute a unique portion of the population? What are the inclusion/exclusion criteria that set the boundaries of the individuals included in the data?

Once an identity resolution strategy is chosen, data must be integrated into a single dataset that is deduplicated, analyzable, and respectful of privacy. This step typically includes:

- Finalizing inclusion and exclusion criteria
- Standardizing identifiers and variable formats
- Executing data linkage protocols
- De-identifying data for analysis
- Conducting match quality reviews and resolving ambiguities

Cross-system linkage often requires negotiation and coordination across agencies with different priorities, timelines, and legal parameters. Champions can help facilitate these discussions.

6. Conduct analysis

Questions that frame this section include: How have the goals of the study evolved? How can the analytic approach be conducted so they are responsive to the decision-making goals of each system contributing data and the broader landscape of preventing or lessening the duration of youth homelessness? What was learned about the quality, strengths, and limitations of each data source? Agencies looking to learn more about youth in their system can ask questions like, where are youth identified for services, where do they access services," and "what systems are youth attached to?"

Conducting analysis using administrative data often requires a series of analytic decisions, each of which must be well documented to ensure replicability. Key decisions may include whether to use demographic data, what geographic level to analyze, whether to produce annual or point-in-time estimates and how to present the data to answer the questions of interest for the agency.

One analytic choice CPR and the Lab made was to estimate the number of youths known to systems and those unknown to systems. Based on interviews with youth and the limited overlap of youth identified in each system, the research team decided to use a multisystem estimation method to estimate the number of youths experiencing homelessness who were not identified in the linked administrative systems for the project. To learn more about this multisystem estimation method, please see the final report linked [here](#).

Other agencies may choose to merge, analyze and report the data to highlight counts of youth, where they access services, demographic characteristics of youth served by geographic area or service provider. Work with the research team and agency partners to develop a plan that aligns the research questions with the analysis and agency goals.

Engaging youth with lived expertise during this phase ensures the resulting dataset reflects the diversity and complexity of youth experiences.

7. Frame and communicate findings for action

Questions that frame this section include: Who is the target audience? What message and messengers will resonate with distinct audiences? How do you balance reporting priorities given competing interests of key partner agencies?

Early in the process of framing the message and communicating the findings, it is vital to share any initial results with primary data partners to garner feedback. This will create an opportunity to demonstrate where the gaps may be and where data may be missing.

Findings should be shared with youth with lived experience, study champions, and organizations with the administrative data to inform the message given to the broader community. This ensures that a broad audience understands the interpretation of the data and the context and implications to the community.



Conclusion

This approach offers communities a practical and replicable roadmap to building locally grounded systems using administrative data to better serve youth experiencing homelessness. The process is not linear—it is intentionally iterative, allowing stakeholders to revisit earlier steps, integrate new learning, and refine methods over time.

Success depends on:

- Starting with shared goals
- Engaging youth meaningfully and consistently
- Building trusted relationships with system champions
- Respecting each partner's data, context, and capacity
- Remaining flexible and adaptive throughout the process

By embedding equity, partnership, and technical rigor throughout the process, communities can create more accurate, actionable, and sustainable estimates—and ultimately drive better outcomes for young people.



This handout was completed with the support of National Prevention Learning Collaborative.