3

EXPANDING MOBILITY:
The Power of Linked Administrative Data for Multi-Gen Analysis

AISP

Acknowledgments

"Expanding Mobility: The Power of Linked Administrative Data for Multi-Gen Analysis" was created by Actionable Intelligence for Social Policy (AISP) at the University of Pennsylvania. The report is a part of our series on administrative data use for economic mobility research and analysis. It was developed based on findings from our 2020 national survey and landscape analysis of data integration efforts across the US, interviews with AISP Network sites and domain experts, and from more than a decade of working with and learning alongside our Network and Learning Community Sites.

AISP is supported by grant funding from the Bill & Melinda Gates Foundation and the Annie E. Casey Foundation. The findings and conclusions contained within are those of the authors and do not necessarily reflect positions or policies of the funders.

Many thanks to everyone who made it possible.

AISP Contributors
Lead authors:
Della Jenkins and
Emily Berkowitz

TC Burnett
Dennis Culhane, PhD
Amy Hawn Nelson, PhD
Matthew Katz
Kristen Smith
Sharon Zanti

We would also like to recognize the following experts and reviewers who contributed their expertise to the report: Katie Barghaus, Katie Breslin, Claudia Coulton, Stephanie Cuccaro-Alamin, Jessica Cunningham, John Fantuzzo, Regan Foust, Robert Goerge, Whitney LeBoeuf, Barb Lucenko, Deleena Patton, Emily Putnam-Hornstein, Heather Rouse, and Emily Wiegand.

Suggested Citation: Jenkins, D., Berkowitz, E., Burnett, TC, Culhane, D., Hawn Nelson, A., Katz, M., Smith, K., & Zanti, S. (2021). Expanding Mobility: The Power of Linked Administrative Data for Multi-Gen Analysis. Actionable Intelligence for Social Policy. University of Pennsylvania.

Introduction: Data Across the Ecosystem

Decades of research and our own life experiences tell us that children do not develop in isolation; they are impacted by a complex, nested ecosystem and the resources available within this environment.[1] Our ecosystem includes our school, neighborhood, community, and even our broader society. We are most immediately impacted by those most proximal to us: our family and household.[2] The ecosystem lens comes from ecological systems theory, and research grounded in this theory concludes that a parent or caregiver's involvement in a child's environment, as well as a caregiver's own well-being, influences a child's developmental outcomes. Increasingly, governments, researchers, and service providers are recognizing the ecosystem's impact on children and taking a whole-family or multi-generational (referred to hereafter as "multi-gen") approach to programs and policy.

Programmatically, this means designing services for both children and caregivers that mutually reinforce family well-being, the classic examples being the Head Start and Early Head Start programs. At a policy level, a multi-gen approach broadens conceptions of risk and resiliency from individuals to households and families, and considers how these conceptions interact with the systems that contribute to poverty or expand economic mobility.[3]

But how do we know if multi-gen policy approaches are working? How do we understand the impact of an intervention on more than one person in a household, or on members of a family over generations? What do these words—"family" and "household"—even mean, and how do we study them when most datasets capture information on only one person?

In this paper, we will discuss how integrated administrative data can help. Drawing on input from experts and our decade-plus experience working with cross-sector administrative data, we will explore the types of multi-gen research questions that become answerable when data are linked across individuals in a family or household. We'll examine the benefits and challenges of different data sources for enabling linkages, and share examples from integrated data systems (IDS) across the US that have used multi-gen analysis to drive action. Finally, we'll discuss opportunities for investment in data infrastructure that could make multi-gen analysis more routine and dynamic, thus transforming our capacity to understand human development and mobility across generations.

This report is the third in the three-part "Expanding Mobility" series exploring the use of integrated data and IDS to deepen understanding of economic mobility. To learn more about how we can take a holistic approach to understanding and expanding mobility, read our first report, "Expanding Mobility: The Power of Linked Administrative Data and Integrated Data Systems" here.

Linking Administrative Data for Multi-Gen Analysis

Despite what we know about ecosystems theory and the ways in which multi-gen approaches can promote well-being and mobility, the vast majority of programs are still designed to serve children and adults separately. Few serve both, limiting the datasets that are readily available for multi-gen analysis.

Data integration requires that data are linked at the individual level based on common data fields. These might include personal identifiers, such as name, birth date, social security number, or a common encrypted "unique ID."

By linking individuals in one dataset to household or family members in other datasets, we can:

- Observe multiple interactions that household members have with programs and systems
- Evaluate the spillover effects of an intervention (or the cumulative effects of several interventions) that one household member receives on other household members
- Predict how household characteristics impact individual outcomes in order to better design interventions
- Coordinate across programs serving various household members to maximize impact

In summary, using linked administrative data for multi-gen analyses allows us to see a clearer picture of an individual in their unique ecosystem at a particular point in time. Importantly, with longitudinal administrative data, we may also look across multiple points in time to better understand both individual and household trajectories. With sufficient years of records, data linkage can illuminate patterns of mobility throughout and across the life course, and across generations. This is important because poverty is pernicious and persistent. Research consistently confirms that children born into poor families are more likely to be poor as adults than their peers whose families have greater financial resources.[4] To put it simply, money begets opportunity and resources. The resulting "path dependency" of socioeconomic status is further compounded by structural barriers to mobility within one's community and immediate environment, such as school and neighborhood segregation.[5] There is an existing literature that relies upon the use of linked administrative data to describe individual trajectories using a "life course" analysis. Multi-gen linkage enables us to take this work one step further, capturing intergenerational impacts on multiple people over time.[6]

With linked, longitudinal multi-gen administrative data, we can:

- Observe the sequence of touchpoints that individuals in a family have with programs and systems across generations
- Evaluate the ripple/spillover effects of an intervention in one generation on future generations
- Predict how the characteristics of prior generations impact individual outcomes in order to better design interventions

While multi-gen linkage is a developing field with significant challenges and limitations, these high-impact approaches demonstrate the value of linking individuals to their family and household members.

Families and Households in Administrative Data

One challenge in asking multi-gen research questions is simply operationalizing terms—what is a family? What is a household? "Family" and "household" are often used interchangeably, despite key differences: "households can contain multiple families, and families can exist across multiple households." [7] Further, most government definitions of "family" and "household" do not capture changing, complex formations, thereby limiting the applicability of information collected by public systems. In order to move beyond hegemonic and outdated definitions, those examining program impacts need a more expansive view of what it means to be a family. But in practice, operationalizing complex, dynamic families and households using existing administrative data is rife with tradeoffs. As a result, definitional questions about what makes a family or household are central when considering potential data sources, linkage procedures, and analytic methods for multi-gen research.

What are complex or dynamic families and households?

What social scientists now refer to as the "postmodern family" reflects trends that began at the end of World War II with the changing role of women in the labor force. These trends, combined with other social phenomena such as increasing rates of divorce and cohabitation, and the extension of marriage rights to LGBTQIA+ couples, necessitate a more dynamic definition of "family" than has traditionally been used. Two-parent households have been declining steadily since the 1960s, and, today, "blended" households (those with a stepparent, stepsibling, or half- sibling present, according to the Census Bureau) are increasingly common.[8]

At the same time, groups made vulnerable through historical policy and (in)action have long participated in household formations that do not reflect nuclear or heteronormative frameworks. Such groups include households headed by Black women (mothers, grandmothers, aunts, aunties, etc.) and members of the queer community[9] among others. To that point, the term "household," rather than "family," may be considered more inclusive of the dynamic structures of caregivers and kinship present in the home, and for the nonbiological relationships that can be essential to development and resiliency.[10] It also enables the capture of changing market dynamics that influence the makeup of households. For example, increasing housing costs coupled with stagnant wages have forced more families to "double up" (multiple households living under the same roof). [11]

The term "households" is limited in its own way. It may exclude family members who are part of the family unit but residing elsewhere (e.g., a young adult away at college, or a parent who is incarcerated). It becomes less robust when living arrangements are highly dynamic (e.g., when families are experiencing changes due to custody arrangements, or moves due to homelessness or housing instability), or in situations where significant caregiving is provided by someone or somewhere outside the home.

To the extent possible, those interested in multi-gen outcomes should start from an evidence-based framework about the relationships that matter most for those outcomes, and then construct linked datasets and data models based on those relationships. Ample literature exists (much of it drawing on linked administrative data) exploring which multi-gen variables are predictive of child outcomes. For example, in Philadelphia, researchers linked data across a number of child- and parent-serving agencies to explore the risk and resiliency factors (e.g., parent education, lead exposure) that most affect the educational well-being of vulnerable children.[12] They later used these same factors to map cumulative risk across the city and target new investments in early education to those areas with the greatest need.[13]

In practice, multi-gen data models can be limited by definitions used in data collection protocols, data access opportunities, and data quality considerations. While these data may capture some important relationships, they could also miss others altogether[14] (e.g., capturing people who reside at an address together, but not other family members who play key caregiving roles). We may only be able to see relationships beyond primary caregivers for certain populations (e.g., people receiving public assistance) and only for a specific period of time (e.g., the month when eligibility was certified). In other words, we may have to adjust our data model or limit the population we study based on what is answerable with the data we have.

Importantly, this also means we may need to consider our data linkage approach each time we perform multi-gen analysis. Because data limitations will affect the definition of household that is possible today, which in turn will affect results, we should remain cognizant of what isn't captured or knowable through our research. Incorporating discussion of these limitations into findings is important even as we strive to do better over time by seeking new ways to capture nuanced definitions of families and households.

Considerations for Family and Household Linkage and Multi-Gen Analysis

If there is no simple or standard approach that applies to all families, all households, or all research questions, what are our options, and how can we as researchers and policymakers begin to approach family and household linkages with existing administrative datasets?

Below, we explore several potential definitions of family and household and the administrative datasets that can be used to construct linkages based on those definitions. Each definition has benefits and challenges that should be considered prior to use. In addition, the quality and availability of potential data sources that correspond to each definition will vary by jurisdiction; data collection and data access standards are often inconsistent across states and counties. As a result, the information in this chart is intended to be a summary of common elements to jumpstart your thinking, rather than a definitive guide.

FAMILY/HOUSEHOLD DEFINITION	BENEFITS + CHALLENGES OF DEFINITION	POTENTIAL DATA SOURCES
Parent/caregiver and child/ dependent	***Benefit:*** The significance of the parental/caregiver relationship for child outcomes is well documented. ***Challenge:*** This definition of family is limited and does not capture other significant relationships in a household, such as grandparents, siblings, stepparents, etc.	Birth records Adoption records Income tax records School enrollment Child care subsidy
Shared multi-gen service receipt (caregiver and child dyads)	***Benefit:*** Like birth and tax records, multi-gen program data capture proximal caregiving relationships based on service receipt. ***Challenge:*** Data capture only the primary recipient dyad (most often a mother and child).	Home visiting Child welfare services Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) Head Start
Shared accommodations, or household roster	***Benefit:*** This definition has the potential to capture more members of complex families and important nonbiological relationships. ***Challenge:*** It does not consider anyone who is not currently sharing accommodations (e.g., an incarcerated parent, a sibling away at college).	Supplemental Nutrition Assistance Program Temporary Assistance for Needy Families Assisted housing Medicaid
Shared address	***Benefit:*** Shared address data has the potential for connecting more members of complex families and important nonbiological relationships. It may also be used to validate family or household matches found in other datasets. ***Challenge:*** It may be hard to distinguish which relationships at a shared address are important/relevant to analysis, given that multiple families may reside at one address (e.g., in a large apartment building with multiple apartments or among families who are doubled up).	Any administrative data source that has reliable address data

Key Data Quality, Access, and Use Considerations

Just as each of the definitions above presents benefits and challenges that must be weighed in the context of the research questions, each of the datasets carries particular considerations regarding quality and access.

Parent/child
(birth, adoption, income tax, school enrollment, and child care subsidy records)

Key considerations:

- Birth records provide a straightforward link between parents and biological children.[15] Of course, depending on your research question, family of birth may be less relevant than caregiving/custody at a later point in time. Tax records are a strong alternative because they provide information on dependent children, biological or not.[16]

- Because neither birth nor tax records are reliant on service usage or benefit receipt, they are more representative than most administrative datasets, have generally good data quality, and can be matched across large populations.[17]

- At the same time, birth and tax records have notable limitations. Birth records will not contain information on immigrant children and noncitizens, and tax records do not capture nonfilers, who are more likely to be very poor families (even if they qualify for EITC[18]), people who are insecurely housed or homeless,[19] and those without a social security number or formal documentation.[20]

- While between 85 and 93% of fathers are captured in birth records,[21] that number is often lower among vulnerable populations.[22] Likewise, tax records capture dependency status, which may or may not indicate a householding parent and does not fully capture complex caregiving relationships.

- It is also important to note that birth and adoption records are sometimes confidential, with limited access for research, and in some states, unavailable beyond very specific operational uses.[23] Income tax records are also highly restricted at the federal level,[24] so this avenue may be less feasible to pursue because of time or resource constraints.

- On the other hand, the information on parents and caregivers from school enrollment or child care subsidy records is hyper-local. When these data are reported up to the state, identifiable parent information is often not included. For more on navigating data access considerations at the federal, state, and local level, see AISP's *Introduction to Data Sharing and Integration*.

Shared multi-gen service receipt, caregiver and child dyads (home visiting, child welfare, WIC, and Head Start records)

Key considerations:

- A major benefit of using service receipt or program data, particularly from explicitly multi-gen programs like home visiting, Head Start, and WIC, is that nonmaternal and even nonbiological sibling relationships may be captured.[25]

- At the same time, programs like WIC will often capture only the role of a maternal parent because mothers receive the benefit, regardless of whether she is the only parent present in the infant's life.

- Information captured in child welfare records will differ depending on the result of an investigation and response.[26] In other words, the level of system involvement will affect the amount and quality of information captured. For example, the information captured when a child is removed from the home is different than when a family remains unified.

Shared accommodations, or household roster (public benefits, assisted housing, and Medicaid)

Key considerations:

- One major limitation of household rosters is that they are capturing only families in means-tested programs, and therefore are "only as expansive as state and federal programs are." [27]

- It is also important to remember that each social program that captures household roster defines the concept differently. For example, a SNAP household includes "everyone who lives together and purchases and prepares meals together." [28]

- Alternatively, Medicaid and other public health insurance programs utilize a more specific formula where a household is defined as a tax filer, legally married spouse, and any dependents claimed as such during tax filing.[29]

Shared address (any administrative data source that collects reliable address data)

Key considerations:

- Shared address data are helpful for cross-validation but are not optimal as the basis of multi-gen linkages in most places, because apartment buildings will contain multiple households or families that cannot be distinguished without unit numbers.

- It can also be difficult to determine the length of time that individuals were living together—especially among more mobile groups.[30] For that reason, when constructing multi-gen linkages based on shared address data, it is important to consider whether housing instability or residential mobility is particularly high for certain subgroups, and the implications this may have for match rates—either between data sources or between members of the same family/household.[31]

- The quality of address data varies widely, depending on the source system and how data are collected. Still, shared address data can be beneficial in facilitating spatial analysis on mobility. Learn more in *Expanding Mobility: The Power of Linked Data for Spatial Analysis.*

Opportunities and Challenges in Multi-Gen Matching and Validation

Each of the datasets described above offers a glimpse into the multi-gen relationships in a child's ecosystem. But because administrative data are not collected for the purpose of research, these glimpses are often incomplete.[32] By linking information across multiple datasets, we begin to see additional points of connection and opportunities to expand the multi-gen unit of analysis.[33] For example, more people living in the home may be identified, including additional caregivers who don't reside there full time but influence child outcomes, like half-siblings or grandparents. Further, incorporating more datasets allows matches to be cross-validated in order to increase accuracy and quality of findings. Of course, more data can also mean more complications.

Matching generally involves the use of probabilistic record linkage, a statistical approach that matches records based on the likelihood that they represent the same person, using a combination of identifiers, such as address, name, and date of birth.[34] For example, Jane Doe may be receiving SNAP, which also captures information about children and other adults in the household with whom she shares meals. An algorithm can be used to match Doe's information (e.g., name, date of birth) to data from another program, like child welfare, in order to identify other people in Doe's life with whom she does not currently share meals. Together, these data sources provide a wider view of the household network—verified using multiple matches—than one dataset could alone.

Linkage is more challenging when working with complex families, who often experience more changes in household formation.[35] In particular, changing names can make it difficult to link child and parent, or even to link the same person across datasets. These changes may occur when a child moves in with different family members due to a change in parental capacity (e.g., hospitalization, incarceration), or at the time of marriage, divorce, or other family structure shifts that lead to the changing or dropping of last name(s).

Algorithms and machine-learning (ML) tools can also help reconcile multiple identifiers, but have their own limitations. The overrepresentation of White, European names in datasets on which ML tools are trained can reinforce biases and lead to lower match rates for nonwhite, culturally diverse families and communities. Hyphenated or double last names, ubiquitous in Latinx communities and increasingly common in general, can be particularly challenging for matching algorithms.[36] It is essential that bias in match rates be assessed, documented, and mitigated in order to ensure that analysis is both accurate and ethical. For more on this topic and how to address racial equity across the data lifecycle, see AISP's *Toolkit for Centering Racial Equity throughout Data Integration*.

Integrated Data Systems and Multi-Gen Work in Action

Given the promise and complexity of multi-gen linkage, communities that have already invested in shared data capacity are the ones leading the way. Across the United States, state and local governments are striving to build more routine, streamlined sharing and linkage of data for research, evaluation, resource allocation, and service delivery. Those that have succeeded in building integrated data systems (also sometimes known as data collaboratives, data hubs, etc.) are well on their way toward being able to answer multi-gen policy and practice questions with linked data. Their governance structures provide a venue and roadmap for collaborative deliberation about ethical data use and the definitional and methodological considerations outlined above. In addition, their legal frameworks for sharing, technical infrastructure for linkage, and interdisciplinary approach allow them to transform siloed data into actionable insights.

The following examples of multi-gen work in action are drawn from our national network of IDS sites, which are exploring a wide range of questions and linkage approaches. Each work in action demonstrates a different use case, and they are arranged in order of increasing methodological and technical complexity.

KENTUCKY STATISTICS	
PROJECT & AUTHORS	Kentucky Head Start Report and State Dashboard Dashboard | Technical Report Kentucky Center for Statistics (KYSTATS) December 2019
PURPOSE OF ANALYSIS	The Kentucky Head Start Report and State Dashboard was developed to provide information on the number of children served by Head Start grantees and whether children served by Head Start are kindergarten-ready. Cross-agency data integration allows the dashboard to also capture information on family characteristics and needs in other domains, such as SNAP or WIC receipt.
DEFINITION OF FAMILY/ HOUSEHOLD	Shared multi-gen service receipt (caregiver and child dyads)
DATA SOURCES	The dashboard is powered by data from the Kentucky Longitudinal Data System and visualizes key multi-gen information from: • Kentucky Department of Education (KDE) • Head Start (includes SNAP and WIC receipt)

FINDINGS & RESULTS	Lawmakers, community members, and other stakeholders are now able to see overlap between children in Head Start programs and caregivers receiving public and food assistance in order to better understand the landscape of early education and family well-being in their state. Dashboard data are viewable at the county level and provide state averages for comparison.
KEY LEARNINGS	KYSTATS began as an education-focused statewide longitudinal data system, but quickly demonstrated the value of reporting and visualizing multi-service involvement. Their effort, which began in 2009, has since expanded the scope of their data access to include birth to five data and workforce programs. This increase in cross-sector sharing has enabled researchers to begin utilizing the family unit as a level of analysis in research and evaluation. Kentucky's public-facing multi-gen Head Start dashboard has also helped to improve accessibility and use of evidence among policymakers, service providers, and community members. The detailed technical report also provides the transparency necessary for building and maintaining public trust.

IOWA'S INTEGRATED DATA SYSTEM FOR DECISION-MAKING

PROJECT & AUTHORS	Substance Use among Iowa Families: An Intergenerational Mixed Method Approach for Informing Policy and Practice Cassandra Dorius, Shawn Dorius, Heather Rouse, Elizabeth Richey, Elizabeth Talbert, Kelsey Van Selous, and Darien Bahe January 2020
PURPOSE OF ANALYSIS	As part of a CDC-funded grant to study the impact of the opioid crisis on families, researchers leveraged their early childhood integrated data system (Iowa's Integrated Data System for Decision-Making) to better understand characteristics and trajectories of families in home visiting programs with and without histories of substance abuse.
DEFINITION OF FAMILY/ HOUSEHOLD	Shared multi-gen service receipt (caregiver and child dyads)
LINKAGE & DATA SOURCES	Using probabilistic and deterministic matching, home visiting records and birth records were linked at the child level for any family who participated in services during 2017. Additionally, birth records provided the population-level sample for comparisons. • Home visiting records • Birth records
FINDINGS & RESULTS	Authors found that children born into families with substance abuse (SA) histories experience significantly more risks evident at birth than children whose families do not have SA experiences. These same families are also less likely to complete the home visiting program and more likely to lose custody of children over the course of the program. Findings suggest several key opportunities to better coordinate services and resources to support children in home visiting programs whose families have SA histories.
KEY LEARNINGS	Researchers in Iowa drew on a prior environmental scan of state administrative data that had identified sources of information on families with substance use histories. Since two of those data sources—birth records and home visiting records—could already be accessed through existing data sharing agreements with the state's early childhood IDS, the project was able to progress quickly and efficiently. The project also served as an important demonstration of the feasibility of expanding their data infrastructure beyond early childhood to support the state's multi-gen research agenda. Administrative data linkage and analysis was performed as part of a larger mixed methods approach to understanding the experience of families with substance use histories. Qualitative data from ethnographic interviews with families and survey data sources proved helpful for filling in gaps where administrative data were insufficient.

CHILDREN'S DATA NETWORK, CALIFORNIA	
PROJECT & AUTHORS	An Examination of Child Protective Service (CPS) Involvement among Children Born to Mothers in Foster Care Andrea Lane Eastman and Emily Putnam-Hornstein February 2019
PURPOSE OF ANALYSIS	Recognizing the "early and concentrated risk of CPS involvement among children born to mothers in foster care," Children's Data Network researchers sought to identify all children born to mothers in care between the years 2009 and 2012 and explore subsequent CPS involvement among those children. Analysis of "two-generation CPS involvement" among this birth cohort aimed to evaluate how maternal experiences in foster care were related to children's later outcomes.
DEFINITION OF FAMILY/ HOUSEHOLD	Parent/caregiver and child/dependent
LINKAGE & DATA SOURCES	Researchers probabilistically matched birth and CPS records in order to identify all mothers in foster care on or after conception from 2009 to 2012, and children "were followed prospectively using linked records to identify CPS involvement occurring during the first three years of life." • California Child Protective Service records • Birth records
FINDINGS & RESULTS	Authors found that 53% of children born to mothers in care were reported to CPS for maltreatment. To deepen understanding of this trend, the analysis identified subpopulations from which three distinct "classes" of mother-child dyads emerged. These classes varied by stability of the mother's care environment, mother's age at time of birth, and mental health conditions. Notable differences among mothers' experiences in care and among their mental health conditions, as the authors state, underscore "the importance of providing services that fit the needs of [mother-child] dyads." Tailored support for both mom and child during early years could improve two-generation outcomes but can be challenging to determine without linkage capacity.
KEY LEARNINGS	This study was the first to develop profiles of two-generation involvement among mother-child dyads in the foster care system at the time of birth. By looking at these dyads using cross-sector data, researchers were able to deepen their understanding of the variation in risk for future child welfare involvement of children born to mothers already in foster care, which has notable policy implications. These and other linkages provided a foundation for an ongoing record reconciliation project in California that generates unique identifiers for the population, in order to streamline integration across datasets while maintaining privacy and security. Read more about the record reconciliation project here. Informed by this project and others, researchers at Children's Data Network and their collaborators are leading the exploration of methods to grow multi-gen analyses using integrated administrative data. Read their methods-focused paper from January 2021 here.

CHILD HOUSEHOLD INTEGRATED LONGITUDINAL DATA SYSTEM, CUYAHOGA COUNTY, OHIO	
PROJECT & AUTHORS	An Integrated Data System Lens into Evictions and Their Effects Francisca García-Cobián Richter, Claudia Coulton, April Urban, and Stephen Steh March 2020
PURPOSE OF ANALYSIS	By linking eviction records and public assistance (PA) data, researchers in Cleveland were able to examine the disruptive effects of eviction on low-income households. Both adults and children experience the deleterious effects of housing loss and instability, and this study surfaced the extent of those impacts as well as how they differ among particular subgroups (i.e., public vs. private housing tenants).
DEFINITION OF FAMILY/ HOUSEHOLD	Shared accommodations, or household roster

LINKAGE & DATA SOURCES	The head of household's name and address on an eviction filing record was linked to monthly PA records from across several programs. Demographic and identifying information from these records was then used for analysis, as well as to construct a dataset of children of defendants, which is not usually captured in eviction records. Finally, these two-gen records were linked to homeless shelter data, school records, lead testing data, and spatial information about neighborhood quality. • Cleveland Municipal Court, Housing Division, eviction filing records • Public assistance data from Medicaid, TANF, and SNAP • Cuyahoga County Office of Homeless Services (HMIS, homeless shelter use) • Cleveland Metropolitan School District attendance records • Ohio Department of Health lead testing data • 2013 area deprivation index (census tract–level information indicating neighborhood quality)
FINDINGS & RESULTS	By accounting for several potentially confounding factors in their longitudinal comparative analysis, researchers were able to demonstrate that eviction filings and, to an even greater extent, eviction orders have a detrimental impact on housing stability. Impacts were greater and more prolonged for tenants in public housing than for those in private housing. Using attendance and lead testing data, authors also found that children in households experiencing eviction filings had higher rates of chronic absenteeism than their peers, particularly in middle through high school. Young children in households experiencing eviction were also more likely to have elevated lead levels, but were less likely to receive timely testing and treatment. This finding—made possible by multi-gen linkage—is particularly troubling, as it suggests that evictions and resulting episodes of housing instability may be compounding disadvantage for an already vulnerable population.
KEY LEARNINGS	The authors' creative linkage approach across a large number of datasets provided more information than ever before on the multi-gen effects of eviction. This integrated data lens into evictions surfaces new insights about how filings against adults and subsequent instability negatively impact health and education outcomes for children. However, as the authors note, reliance on PA records for multi-gen linkage constrains the analysis to only low-income households accessing benefits.

Family/Household Dynamics Over Time: Emerging Methods

The examples of multi-gen data linkage and research described above are exciting because they each situate children and families in an ecosystem by linking them to proximal household members at a given point in time. In doing so, they teach us valuable lessons about the mechanisms behind risk, resilience, health, and mobility. They also establish a baseline against which to measure evidence-based practice changes that improve outcomes.

Of course, in reality, families, households, and ecosystems are dynamic, which prompts us to ask, how do family formations change over time? And how do these changes impact our trajectory? Longitudinal administrative data hold the promise of one day enabling researchers to answer these questions too, using multi-gen linkage for key points in time comparisons (e.g., pre- and post-divorce, incarceration, or child welfare involvement).

At Chapin Hall at the University of Chicago (hereafter referred to as Chapin Hall), researchers leveraged the Integrated Database on Child and Family Programs in Illinois—a state-level IDS—to capture complex and vulnerable families across multiple datasets.[37] A "family network" unit of analysis was developed using a base population of SNAP and child welfare cases. Individuals who were in multiple cases across systems or over time were used to identify and connect family units through decades of data; the number of people identified as part of each longitudinal network was substantially greater than typical single point-in-time approaches to linkage. Authors call these longitudinal family networks "supercases" and visualize the process as follows:

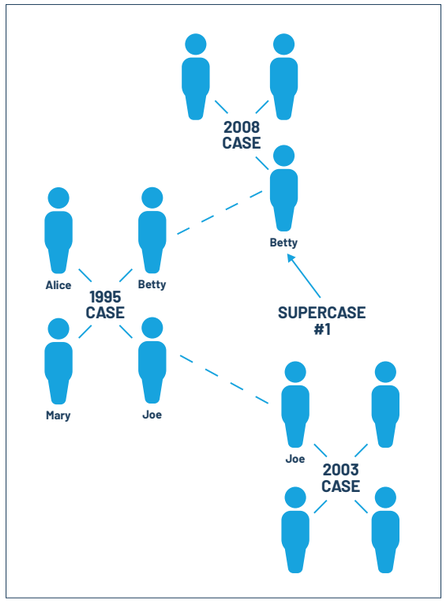

Image description: Betty was part of a SNAP household between 2007 and 2008. Looking back at all public assistance data since 1989, researchers also found that Betty was a child in 1995 on a case with her mother (Alice) and siblings (Joe and Mary). They also found that Betty's brother Joe had a SNAP case in the early 2000s with his family. Since Betty and Joe had been previously connected by the 1995 case, researchers were able to group all three cases into a family network.

Adapted with permission from Goerge, R. & Wiegand, E. (2019). Understanding vulnerable families in multiple service systems. The Russell Sage Foundation Journal of the Social Sciences, 5(2), 86-104.

Next, individuals were linked to participant data from five programs that represent Illinois' costliest services—mental health treatment, substance abuse treatment, juvenile incarceration, adult incarceration, and child welfare—to determine the prevalence of involvement by family networks with one or more of these systems. Findings showed that 23% of family networks received services in two or more of these areas, therefore experiencing a greater number of more severe problems. Further, these 23% of families accounted for 86% of service funding used for the entire sample population.

Beyond the stated findings, the process of mapping networks also generates opportunities to incorporate dynamic changes in family networks across time. The authors have noted that, in order to refine and improve multi-gen research, "temporal factors" should be incorporated, including measures of the depth or duration of service receipt among an individual, family, or household and the order or sequence of system involvement. Administrative data could also be linked to census data and surveys in order to capture richer information about family composition and multi-gen experiences across the life course.[38]

The complex data models and advanced linkage procedures necessary to capture family change over time make this work purely aspirational for most communities today. Nevertheless, it is the logical next horizon for multi-gen linkage and holds the potential to transform our understanding of multi-gen mobility.

So, how do we get there?

Recommendations to Improve Linked Data Access and Use for Multi-Gen Research

The following recommendations for agencies, researcher partners, and policymakers would increase the ease, frequency, and quality of multi-gen linkage and analysis.

- **Work to clarify definitions across programs, services, and agencies, and document within metadata.** Fiscal and legislative mandates often dictate how family and household are defined, which causes inconsistencies in terminology and limits data linkage opportunities. Clear metadata are a best practice and helpful starting point. When possible, data sharing partners should work to harmonize definitions and clarify understanding of terms in order to better collaborate. As part of this process, partners should consider how ideologies and judgments are embedded in definitions of family and caution against harmful misconceptions.

- **Expand data collection to better capture parental and caregiver relationships.** A necessary reliance on datasets that primarily capture the mother-child dyad has led to a dearth of research that explores other biological or caregiver relationships, particularly paternal or father figures. While birth records are effective in identifying the vast majority of fathers, much public benefits information disproportionately captures information on mom and not dad.[39] More expansive and representative datasets should be used when possible.

- **Routinize data quality and validity testing in order to build evidence-based multi-gen data models.** It bears repeating that a robust literature exists on the indicators and characteristics that contribute to resiliency during childhood and greater mobility in adulthood. This existing—and growing!—evidence base should be consulted when developing research questions and utilized to test the validity of the data models being employed. Doing so ensures a strong theoretical backing to findings and contributes to greater success in the use of research results for policy and practice. Firm but flexible standards around relevant data elements as well as effective models and tests could help routinize the practice to add rigor to the field.

- **Supplement administrative data with other sources that provide greater context on family and household dynamics over time.** A key challenge for multi-gen research is capturing information across longer time periods as well as being able to follow parents, caregivers, and children across generations. As the use of linked administrative data becomes more common and routine, the intricacy of these analyses will grow. To capture dynamic movement across generations and changes in complex family structures/relationships over time, researchers may look to **national panel studies** (e.g., Panel Study of Income Dynamics, Medical Expenditure Panel Survey, Survey of Income and Program Participation) to fill in the gaps. **Local panel studies** may also be available for use at the local level (e.g., Robin Hood Poverty

Tracker in New York City from Columbia University, Annie E. Casey Foundation Making Connections Initiative panel survey on residential mobility and neighborhood change across 10 major US cities). These sources often monitor long-term trends related to mobility and well-being for multiple people in the same home. Many data integration efforts have the capacity to link administrative data with panel studies, as long as the use of data is approved and the data owner's security requirements are met. To streamline this process, future panel studies may wish to consider crafting consent forms to explicitly allow for approved integrations with administrative records.

- In other cases, it may be possible to identify and utilize hyper-local administrative data sources, like **registration forms** or **intake surveys administered by programs,** in order to improve understanding of family dynamics, especially sibling relationships (which are typically included for programmatic purposes). This approach comes with clear challenges; in addition to being sensitive information, forms and intake surveys often aren't digitized. As a result, researchers and analysts will need to develop practical, ethical, and legal processes to transform and protect this information. They should also take great care to use information only in safe and appropriate contexts, such as internal program evaluations.

- **Blend administrative data analysis with qualitative data on family and household experiences.** Another way to source more information on the household or family ecosystem that impacts a child is through new qualitative data collection (e.g., phone calls, surveys administered by programs). This process allows the voice of study populations to be incorporated more directly and gives researchers more flexibility in how variables are defined. Whenever new primary data collection is undertaken, participants must be asked for consent to share or integrate that information with existing administrative data. Because primary data collection is notably time- and resource-intensive, it is best used for small or modest studies. But we do recommend mixed methods approaches as a path toward more comprehensive, culturally nourishing, and even self-reported definitions of families and households.

- **Increase secure researcher and public agency access to confidential, population-level records, like birth records and tax records, for multi-gen studies of economic mobility.** Large and comprehensive administrative datasets are challenging to access, despite their clear benefit for research and policy improvement. Barriers may vary depending on an effort's specific context (for example, as noted earlier, use of birth records are more restricted in some jurisdictions than others, regardless of whether state agencies want to share the information), but, typically, they involve a combination of burdensome access procedures, data security concerns, and cost.[40] (We address the first and third concerns in the bullet below.) One way access can be advanced in situations where data owners are particularly worried about data sensitivity is by utilizing emerging privacy preserving technology. Recent advancements have made it easier than ever before to safely allow users to access and link sensitive datasets for approved research and analytics purposes. Advocates are promoting these techniques as part of the call for a new National Data Service within the Executive Branch[41] following the recommendations of the 2017 US Commission on Evidence-Based Policymaking,[42] and subsequent Foundations for Evidence-Based Policymaking Act of 2018.[43] While such an endeavor is likely to take time, movement toward less burdensome and more democratic federal data access procedures is encouraging. Meanwhile, privacy-preserving techniques and secure remote access procedures have already begun to be adopted by several leading states and their data partners. However, significant resources will need to be dedicated to ensure that more state and local governments are able to follow suit.

- Most importantly, we need to **invest in the future of multi-gen data infrastructure.** Multi-gen linkages require sustained investments in both human and technical capacity. We recommend that both public and philanthropic stakeholders with an interest in promoting multi-gen mobility support existing IDS efforts as they work to modernize their infrastructure and build increasingly complex multi-gen data models. These efforts have already laid the groundwork with trusted data partners and are uniquely positioned to use administrative data to expand our understanding of the family and household ecosystem. However, sustained funding is needed to move beyond individual projects toward more routine linkage, and to grapple with more complex and dynamic family formations. As the work in actions and examples throughout this brief demonstrate, enormous opportunities exist to build innovative, new multi-gen approaches on solid foundations.

Notes

[1] Urie Bronfenbrenner's Ecological Systems Theory views children as developing within a complex set of relationships at various levels (micro, meso, exo, macro, chrono) in their environment.

[2] Likewise, we are deeply influenced by our immediate environment and what we're exposed to as a result of our immediate surroundings. This idea is explored in more detail in Expanding Mobility: The Power of Linked Administrative Data for Spatial Analysis. Specifically, we consider historical (i.e., past events and policies that impact the present), built (i.e., physical space, resources, and exposures within a community), and social (i.e., people and relationships that share the built space) environments, and their role in helping or hindering a child's mobility in adulthood.

[3] For more on 2Gen and multi-gen programs and policies, explore the Aspen Institute resources.

[4] Ratcliffe, C. (2015). Child Poverty and Adult Success. Urban Institute. https://www.urban.org/sites/default/files/publication/65766/2000369-Child-Poverty-and-Adult-Success.pdf

[5] Sampson, R. J. (2016). Individual and community economic mobility in the Great Recession era: The spatial foundations of persistent inequality. In A. Brown, D. Buchholz, D. Davis, & A. Gonzalez (Eds.), Economic mobility: Research & ideas on strengthening families, communities, & the economy (pp. 35–42). Federal Reserve Bank of St. Louis & the Board of Governors of the Federal Reserve System.

[6] Zaretzky, K. and Flatau, P. (2013). The cost of homelessness and the net benefit of homelessness programs: a national study, AHURI Final Report No. 218. Melbourne: Australian Housing and Urban Research Institute.

Felitti, V. J., Anda, R. F., Nordenberg, D., Williamson, D. F., Spitz, A. M., Edwards, V., Koss, M. P., & Marks, J. S. (1998). Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults. *American Journal of Preventative Medicine*, 14(4), 245-258.

Kuhn, R., & Culhane, D. P. (1998). Applying Cluster Analysis to Test a Typology of Homelessness by Pattern of Shelter Utilization: Results from the Analysis of Administrative Data. *American Journal of Community Psychology*, 26(2), 207-232.

[7] Cuccaro-Alamin, S., Eastman, A. L., Foust, R., McCroskey, J., Nghiem, H. T., & Putnam-Hornstein, E. (2021). Strategies for constructing household and family units with linked administrative records. Children and Youth Services Review, 120. https://doi.org/10.1016/j.childyouth.2020.105706

[8] Pew Research Center. (2015). Parenting in America: Outlook, worries, aspirations are strongly linked to financial situation. https://www.pewresearch.org/social-trends/2015/12/17/parenting-in-america/

[9] Collins, P. H. (2005). Black women and motherhood. In S. Hardy & C. Wiedmer (Eds.), Motherhood and space, chap. 9. Palgrave Macmillan. https://doi.org/10.1007/978-1-137-12103-5_9; Bailey, M. M. (2009). Performance as Intervention: Ballroom Culture and the Politics of HIV/AIDS in Detroit. Souls, 11(3), 253–274. https://doi.org/10.1080/10999940903088226

[10] Goerge, R. & Wiegand, E. (2019). Understanding vulnerable families in multiple service systems. The Russell Sage Foundation Journal of the Social Sciences, 5(2), 86–104. https://doi.org/10.7758/RSF.2019.5.2.05

[11] Bush, H. & Shinn, M. (2017). Families' experiences of doubling up after homelessness. Cityscape, 19(3), 331–356.

[12] Rouse, H., Fantuzzo, J., & LeBoeuf, W. (2011). Comprehensive challenges for the well being of young children: A population-based study of publicly monitored risks in a large urban center. Child & Youth Care Forum, 40, 281–302. https://doi.org/10.1007/s10566-010-9138-y

[13] LeBoeuf, W., Barghaus, K., Henderson, C., Coe, K., Fantuzzo, J., & Moore, J. (2017). The use of integrated data to inform quality pre-k expansion in Philadelphia. Research Briefs, 1. https://repository.upenn.edu/pennchild_briefs/1

[14] Jutte, D. P., Roos, L. L., & Brownwell, M. D. (2011). Administrative record linkage as a tool for public health research. Annual Review of Public Health, 32, 91–108. https://doi.org/10.1146/annurev-publhealth-031210-100700

[15] Lee, H., Warren, A., & Lakhpreet. G. (2015). Cheaper, faster, better: Are state administrative data the answer? The mother and infant home visiting program evaluation: Strong start second annual report. Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. https://www.mdrc.org/sites/default/files/MIHOPE-StrongStart-2yr_2015.pdf

[16] Card, D., Chetty, R., Feldstein, M., & Saez, E. (2010). Expanding access to administrative data for research in the United States [White paper]. National Science Foundation. https://eml.berkeley.edu/~saez/card-chetty-feldstein-saezNSF10dataaccess.pdf

[17] Card, D., Chetty, R., Feldstein, M., & Saez, E. (2010). Expanding access to administrative data for research in the United States [White paper]. National Science Foundation. https://eml.berkeley.edu/~saez/card-chetty-feldstein-saezNSF10dataaccess.pdf

[18] Internal Revenue Service. (2021, January 19). About EITC. Earned Income Tax Credit & Other Refundable Credits. https://www.eitc.irs.gov/eitc-central/about-eitc/about-eitc

[19] Slemrod, J. (2016). Caveats to the research use of tax-return administrative data. National Tax Journal, 69(4), 1002-1020. https://doi.org/10.17310/ntj.2016.4.13

[20] McClure, D., Santos, R., & Kooragayala, S. (2017). Administrative records in the 2020 US Census: Civil Rights considerations and opportunities. Urban Institute. https://www.urban.org/sites/default/files/publication/90446/census_ar_report.pdf

[21] Putnam-Hornstein, E. & Needell, B. (2011). Predictors of child protective service contact between birth and age five: An examination of California's 2002 birth cohort. Children and Youth Service Review, 33(8), 1337–1344. https://doi.org/10.1016/j.childyouth.2011.04.006

[22] Cuccaro-Alamin, S., Eastman, A. L., Foust, R., McCroskey, J., Nghiem, H. T., & Putnam-Hornstein, E. (2021). Strategies for constructing household and family units with linked administrative records. Children and Youth Services Review, 120. https://doi.org/10.1016/j.childyouth.2020.105706

[23] Kim, S. Y., Ahuja, S., Stampfel, C., & Williamson, D. (2015). Are birth certificate and hospital discharge linkages performed in 52 jurisdictions in the United States? Maternal and Child Health Journal, 19(12), 2615-2620. https://doi.org/10.1007/s10995-015-1780-4; Lee, H., Warren, A., & Lakhpreet. G. (2015). Cheaper, faster, better: Are state administrative data the answer? The mother and infant home visiting program evaluation: Strong start second annual report. Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. https://www.mdrc.org/sites/default/files/MIHOPE-StrongStart-2yr_2015.pdf

[24] Card, D., Chetty, R., Feldstein, M., & Saez, E. (2010). Expanding access to administrative data for research in the United States [White paper]. National Science Foundation. https://eml.berkeley.edu/~saez/card-chetty-feldstein-saezNSF10dataaccess.pdf

[25] Lin, V., Shaw, S., & Maxwell, K. (2019). Administrative data sources to address early care and education policy-relevant research questions. Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. https://www.childtrends.org/wp-content/uploads/2019/11/ccadac_admin_data_sources_ChildTrends_December2019.pdf

[26] Wiegand, E. R. & Goerge R. M. (2019). Using and linking administrative datasets for family self-sufficiency research. Washington, DC: Family Self-Sufficiency and Stability Research Consortium.

[27] O'Hara, A., Young, I., & Amman, L. (2020, August 6). Who is missing from administrative data? Georgetown University McCourt School of Public Policy. https://mccourt.georgetown.edu/news/who-is-missing-from-administrative-data/

[28] United States Department of Agriculture. (2021, January 1). SNAP Eligibility. USDA Food and Nutrition Service. https://www.fns.usda.gov/snap/recipient/eligibility

[29] United States Centers for Medicare and Medicaid Services. (n.d.). How to count income and household members: Who to include in your household. Healthcare.gov. https://www.healthcare.gov/income-and-household-information/household-size/

[30] E. Wiegand, personal communication, May 18, 2021.

[31] Jenkins, S. P., Lynn, P., Jäckle, A., & Sala, E. (2006). The feasibility of linking household survey and administrative record data: New evidence for Britain. International Journal of Social Research Methodology, 11(1), 29–43. https://doi.org/10.1080/13645570701401602

[32] O'Hara, A., Young, I., & Amman, L. (2020, August 6). Who is missing from administrative data? Georgetown University McCourt School of Public Policy. https://mccourt.georgetown.edu/news/who-is-missing-from-administrative-data/

[33] Wiegand, E. & Goerge, R. (2019). Administrative data quality varies between programs, affects use of data. Chapin Hall at the University of Chicago. https://www.chapinhall.org/research/administrative-data-quality-varies-between-programs-affects-use-of-data/

[34] Actionable Intelligence for Social Policy, MetroLab Network, National Neighborhood Indicators Partnership. (2018). Technology for civic data integration. https://metrolabnetwork.org/wp-content/uploads/2018/09/Technology-for-Civic-Data-Integration.pdf

[35] Doyle, P. & Long, S. E. (1988). The impact of unit analysis on measures of serial multiple program participation [White paper]. United States Census Bureau. https://www.census.gov/content/dam/Census/library/working-papers/1988/demo/SIPP-WP-48.pdf

[36] Christen, P. (2006, December). A comparison of personal name matching: Techniques and practical issues. In Sixth IEEE International Conference on Data Mining-Workshops (ICDMW'06) (pp. 290–294). IEEE; Ansolabehere, S. & Hersh, E. D. (2017). ADGN: An algorithm for record linkage using address, date of birth, gender, and name. Statistics and Public Policy, 4(1), 1–10, http://doi.org/10.1080/2330443X.2017.1389620

[37] Goerge, R. & Wiegand, E. (2019). Understanding vulnerable families in multiple service systems. The Russell Sage Foundation Journal of the Social Sciences, 5(2), 86–104. https://doi.org/10.7758/RSF.2019.5.2.05

[38] National Academies of Sciences, Engineering, and Medicine. (2016). Using linked census, survey, and administrative data to assess the longer-term effects of policy: Proceedings of a workshop—in brief. The National Academies Press. https://doi.org/10.17226/23583; O'Hara, A., Shattuck, R. M., & Goerge, R. M. (2017). Linking federal surveys with administrative data to improve research on families. The ANNALS of the American Academy of Political and Social Science, 669(1), 63–74. https://doi.org/10.1177/0002716216678391

[39] Jutte, D. P., Roos, L. L., & Brownwell, M. D. (2011). Administrative record linkage as a tool for public health research. Annual Review of Public Health, 32, 91–108. https://doi.org/10.1146/annurev-publhealth-031210-100700

[40] Goerge, R. (2018). Barriers to accessing state data and approaches to addressing them. The Annals of the American Academy of Political and Social Science, 675. http://doi.org/10.1177/0002716217741257

[41] Data Coalition. (2020, July 29). An open letter to Congress on the rapid development of a national secure data service. Data Coalition. https://www.datacoalition.org/an-open-letter-to-congress-on-the-rapid-development-of-a-national-secure-data-service/

[42] Office of Planning, Research and Evaluation. (n.d.) Commission on evidence-based policymaking. Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. https://www.acf.hhs.gov/opre/project/commission-evidence-based-policymaking-cep

[43] Foundations for Evidence-Based Policymaking Act of 2018, Publ. L. No. 115-435 (2019). https://www.congress.gov/bill/115th-congress/house-bill/4174

THIS REPORT is the third in AISP's 3-part "Expanding Mobility" series, which explores how linked administrative data can be used to deepen understanding of economic mobility. The two complementary reports, linked below, focus on the use of integrated administrative data for mobility research and multi-gen analysis.

- *Expanding Mobility: The Power of Linked Administrative Data and Integrated Data Systems*
- *Expanding Mobility: The Power of Linked Administrative Data for Spatial Analysis*

To learn more about building and scaling IDS, visit the AISP website and view our Quality Framework.


AISP