Evaluating Approaches to Linking Evictions Records:

Assessing the Feasibility of Research with Integrated Data

J. J. Cutuli, PhD; Mary Joan McDuffie, MA;[2] Erin Nescott, MS[3]

1. Senior Research Scientist, Nemours Children's Health

2. Policy Scientist, Center for Community Research & Service, Joseph R. Biden, Jr. School of Public Policy and Administration, University of Delaware

3. Associate Policy Scientist, Center for Community Research & Service, Joseph R. Biden, Jr. School of Public Policy and Administration, University of Delaware

Acknowledgements

Abstract

Objectives: This study investigates different approaches to integrating evictions data with Medicaid and homeless shelter utilization records at the individual level for the state of Delaware. We especially focus on evaluating the feasibility of creating an integrated dataset focused on children and adolescents through different approaches to matching. **Methods:** We attempt to link existing statewide records on evictions, Medicaid, and shelter from 2017-2019. We first compare direct match and probabilistic match approaches to linking evictions and Medicaid records, and then incorporate shelter records. Finally, we consider a limited set of characteristics relevant to potential future public health research among children who experienced eviction, had a shelter stay, and were enrolled in Medicaid. **Results:** Direct matching resulted in a lower match (14%) rate than probabilistic matching (22%) of eviction records to Medicaid data. Homeless shelter records had a high match rate to Medicaid records, even when using a direct match (75%). A sizeable subset of children (n=216) were linked across the three data sources, though this was from a small percentage of cases in the evictions data. Among this subset of children, most (71%) were enrolled in Medicaid in all three years considered by this study and Black children were greatly overrepresented (75%). **Conclusions:** Integrating evictions records with other health and human service data involves a number of challenges. Probabilistic matching yielded a considerably higher number of matches after manual review, resulting in a possible study sample of children who have experienced eviction, a homeless shelter stay, and were enrolled in Medicaid. Strategies to increase the match rate for eviction records through using records from other, more universal services may be necessary for investigations that require more comprehensive coverage of the population.

Introduction

This study considers the feasibility of matching evictions records to Medicaid and shelter use administrative records at the individual level for families with children and youth. Integrated datasets are an important tool for researchers in public health, healthcare delivery, and varied health and human service arenas. However, linking records requires that a set of relatively unique individual-level identifier variables be present across sources, commonly first name, last name, date of birth and, sometimes, social security number. This presents a challenge for researchers interested in linking evictions records with others: evictions data often only include the first and last name of the lease holder(s), the property address, and a small set of other details related to the eviction process. We investigate data match-rates between sources for children with different experiences of forced housing disruption (homeless shelter stay, eviction without a homeless shelter stay, and neither). Our goal is to quantify the match rate using different linkage approaches.

The Importance of Research on Child, Youth, and Family Eviction and Homelessness

Relatively little is reliably known about families who experience eviction, though studies of families and youth who experience homelessness suggest considerable risk for becoming disconnected from services and for poor outcomes. This underscores the importance of developing reliable methods to further research on family eviction. Crises that involve residential mobility challenge families to stay connected or reconnect to health care,[1] early childhood programs,[2] school,[3] and a host of other formal and informal community supports.[4] Homelessness and eviction are already viewed as prevalent issues though the economic consequences of the COVID-19 pandemic threaten great increases. Over 300,000 minor children and teens stay in U.S. homeless shelters each year,[5] and about 500 specifically in Delaware.[6] Children and minor teens in shelter have higher rates of chronic disease, acute disease, and emergency department utilization.[7–9]

Much less is known about children in families who are evicted, including reliable estimates of their number.[10,11] Analyses linking evictions to census information suggest risk for very low birth weight and infant mortality,[12] though other pediatric health outcomes are uninvestigated. Eviction or a move to shelter disproportionately involve families in deep poverty, women, and those from racial minority backgrounds.[11,13] These disruptions may disconnect children from primary healthcare providers while forcing parents to prioritize competing basic needs. Interagency and inter-system collaboration may help maintain healthcare connections during crises, but there is essentially no rigorous evidence for children who experience eviction and homelessness.

Past Attempts at Linking Evictions Records With Other Data

Few published studies have attempted to integrate evictions records with other health and human service records to investigate impacts on children, though integrated data approaches that include shelter and other housing intervention records are much more common.[14–16] Focusing on linking methodology, Collinson and Reed provide an early and most-detailed description of their linking procedure.[17] This group linked housing court records on eviction filings (containing only address, date, and first and last name) to a large database containing recipients of multiple public benefits in New York City (containing first and last name, address, and dates benefits were

received), but only for cases where benefits receipt predated the eviction record. They cleaned the identifiers (e.g., removing non-numeric characters and aliases from names in the evictions data) and geocoded addresses to the parcel level to account for multiunit properties at the cost of precision. After processing and blocking the data, they engaged a matching algorithm that prioritized exact matches before probabilistically linking non-exact matches based on best-available name and geocode combinations past a predefined minimum threshold. This resulted in matching 40% of evictions records to benefits records, though this match rate increases to 57% when benefits records are allowed to occur before or after the eviction record date. Though a 40% match rate seems low and would likely contribute to bias, the authors note that a majority of actual eviction cases (filings that proceeded to eviction) are contained in the 40% of filings that matched.

A more recent study by Schwartz and colleagues linked New York City eviction filings to New York State Medicaid records.[18] The reported linking procedure is less detailed. The study team first geocoded each eviction filing to the parcel level and then identified patients as evicted if their most recent Medicaid enrollment information (as of the eviction date) matched the unit number and building of an eviction. Name was not considered in the match. This yielded 6,922 Medicaid recipients identified as having been evicted out of 20,810 total evictions, though it is not clear how many Medicaid recipients shared a household. At best, the match rate for evictions from this approach was about 33.3%.

Neither of the above studies attempted to assign eviction records to children. This presents an additional challenge because children are not listed on eviction records. Instead, children must be linked to parents/caregivers using other data sources and the parents/caregivers must then be matched to the eviction records. This also implies two—instead of one—dynamic indicators: addresses can change as people move and family structures can change through separation or loss. Richter and colleagues linked eviction filing records to a benefits database containing Medicaid and other program records (analogous to Collinson & Reed[17]), though their method ultimately extends evicted-status to children connected to adults in the benefits data.[19] In an attempt to contain the possibility of address and family structure records being out of date, they limited consideration to benefits data that occurred within six months prior to each eviction record. This group geocoded addresses in both datasets to the census block level, and used Soundex (phonetic indexing) to match names and addresses. This resulted in a match rate of 47.4% of eviction filings.

Finally, though not focused on evictions, another effort attempted linkage on similar adult identifiers (guardian first and last name, address) found in education records from a school district, connecting data on adults to infer the presence of risk in the lives of children.[20] The goal was to match guardians from the education data to an existing integrated data system that contained all hospital-based healthcare contacts, various population-wide public-records datasets that contain, at least, names and addresses (e.g., voter registration records), and other sources. This effort standardized addresses using a geocoding service. This process successfully linked 67% of guardians in the education records through an exact/deterministic match, and another 14% matched with an acceptable level of confidence through probabilistic matching (81% match rate, overall).

The Current Study

This study aims to describe the degree to which a linking protocol integrates data from Medicaid records, shelter utilization data, and court records ordering eviction at the household level. Past research matched between 33.3% and 47.4% of eviction filings at an individual level. We seek to answer a fundamental question: What is the match rate when trying to link eviction and Medicaid records using approaches that rely on name and address?

Methods

We attempted to link records at the individual level from three different datasets to reflect the number of children who were evicted, those who stayed in a homeless shelter, and, for future analyses, have access to indicators of health and health care utilization using Medicaid records. This study considers all records from 2017 through 2019. All linkages were performed at the Center for Community Research & Service (CCRS) by policy analysts with individual business associate agreements with the state of Delaware granting permission to use the data. The project was reviewed and approved by the Delaware Division of Medicaid and Medical Assistance and the Institutional Review Boards of the University of Delaware and Nemours Children's Health. All analyses were done on CCRS's encrypted drive to safeguard the data and individual data has not been released.

Eviction records

The Justice of the Peace court system maintains a publicly-available database of eviction filings in Delaware. Records included dates of filings and all listed defendants (max: 9 defendants per filing). Available identifier variables were first name, last name, sex, and address. Addresses were cleaned (e.g., removed non-residential addresses) and geocoded to the address level using ARCGIS (92% of addresses were successfully geocoded), yielding longitude and latitude coordinates. We removed duplicate and repeated eviction filings (e.g., when additional filings contained the same defendant at the same address).

Homeless Shelter Records

Individual-level administrative records describe stays in homeless shelters for individuals and families in Delaware. These data come from the Community Management Information System (CMIS) which functions as the Homeless Management Information System for Delaware. Records contain a family code that allowed us to cluster individuals into family units. CMIS is managed by Housing Alliance Delaware.[21]

Medicaid Records

Analyses consider Medicaid Client data. These data were approved for use and made available through the Medicaid research partnership between the Delaware Division of Medicaid and Medical Assistance, the University of Delaware (UD) Center for Community Research and Service, and the UD College of Health Science. Data Client addresses were geocoded to the address level using ARCGIS (89% of addresses were successfully geocoded). Medicaid data also permit individuals to be clustered into families, allowing us to identify any minor children based on Medicaid household codes. We applied a logic based on ages to demarcate children (minors under 18 years old) and guardians (e.g., those more than 15 years older than each minor) for cases linked to eviction records. Medicaid eligibility in Delaware is based on a combination of

factors including household income and composition, and health status subject to income limits.[22]

Linkage Procedures

We used Link Plus[23] to link Medicaid, eviction, and homeless shelter records, first using direct matches on the complete identifiers (defendant first name, last name, sex, geocoded address). Direct matching requires the data to match exactly for records to be linked. We repeated the match starting with the direct-match cases and then adding additional matches using a probabilistic matching method on the same set of identifiers, using only the first three letters of the first and of the last names Probabilistic matching permit differences in field values to accommodate errors in data entry, alternate spelling of names, nicknames, or nonstandard abbreviations, for example. All matching involved a final step of manual review. CCRS staff reviewed each match to ensure it was a true match without any apparent logical errors. All analyses were completed by CCRS staff and we report linkage results for Medicaid and eviction records in aggregate form. We then attempted to link homeless shelter records with Medicaid records using a more-standard set of identifiers (first name, last name, birth date, and sex) and then integrating all three datasets.

Results

Linkage rate details for matches between eviction and Medicaid records are provided in Table 1. Overall, about 14% of defendants with evictions filings matched to the Medicaid records using direct matching. This rate increases to 22% when using probabilistic matching and subsequent manual review of matches. This represents nearly 2,500 additional matches.

We linked homeless shelter data with Medicaid data by first name, last name, birth date, and sex via direct matching. There were 6,368 adult clients who had a shelter stay; 4,746 or 75% were matched to Medicaid records.

Table 1. Match Rates by Approach Linking Evictions to Medicaid Records.

Year	Eviction filings Unique Persons N	Direct Match Evictions-Medicaid		Probabilistic Match Evictions-Medicaid	
		n	%	n	%
2017	11,930	1,225	10%	1,749	15%
2018	11,924	1,570	13%	2,578	22%
2019	11,005	1,259	11%	2,310	21%
All Years	31,151	4,337	14%	6,821	22%

The linkage was then made between the Medicaid-Homeless data set and the Evictions-Medicaid data set by the Medicaid ID. First considering results from a direct linkage approach, there were 100 individuals who were present in all three datasets (2% of the linked Evictions-Medicaid data set). The household identifier was used to extract anyone living in the household of the eviction defendant. This resulted in 270 full benefit covered Medicaid clients who were in a household that had evictions claim and someone who experienced homelessness.

The same records were matched using a probabilistic linkage approach: There were 168 individuals who were present in all three datasets (2.46% of the linked Evictions-Medicaid data

set), from 167 unique households. We then used the household indicator to identify other members living in the household as the eviction defendant. This resulted in 440 Medicaid clients who were in a household that had eviction claims and lived with or was someone who stayed in a homeless shelter. About half of these individuals were children and teens under 19 years old (See Table 2). Considering just the children, nearly three-quarters (74%) received Medicaid for at least one month in each of the three years of data considered in this study (2017-2019), the vast majority (71%) were Black, non-Hispanic race and ethnicity, and a slight majority (53%) were male.

Table 2. Individuals Indicated by All Three Data Sources by Age Group

Age Group	N	%
Birth - 5	84	19%
6 - 10	65	15%
11 - 18	67	15%
19 - 30	78	18%
31 - 50	110	25%
51+	< 35	< 10%
Missing	< 10	< 2%
Total	440	

Discussion

Different linkage techniques, different types of data, and differences in available identifiers have sizeable implications for match rates when integrating records across sources. This study was interested in determining the match rates when linking records from evictions, Medicaid, and shelter records across Delaware. Our goal was to evaluate the feasibility of future integrated data approaches to testing associations between evictions, stays in family shelter, and pediatric health care. Compared to deterministic matching techniques alone, probabilistic matching of eviction and Medicaid records resulted in a higher number of matches that remained after manual review. Though, overall match rates were low. Meanwhile, the match rate between shelter and Medicaid records was high, even when using a more conservative direct matching approach.

Integrating Eviction, Medicaid, and Homeless Shelter Records

Our attempts to link eviction and Medicaid records yielded match rates of 14% for a direct match approach and 22% for a probabilistic match approach. It is not surprising that probabilistic matching matched a higher number of cases since probabilistic linkage allows for differences in the fields that identify cases, such as misspellings or other data entry errors, nicknames and abbreviations, or other differences. However, only locating 22% of evictions cases in the Medicaid data results in a match rate that is below what has been typically found in other studies (albeit few in number), which generally match between 33-50% of evictions filings.[17–19]

There are several possible explanations for the lower match rate. First, past research used datasets containing enrollment in multiple benefit programs that included Medicaid as well as others (e.g., SNAP, TANF). It is possible that some individuals enrolled in these other programs

and their information was not available to our match in Delaware, potentially lowering our match rate compared to other localities where this information was available. Furthermore, the reliance on Medicaid records in Delaware also means that persons who do not qualify could not be successfully matched (e.g., those without low income, without a child under 18 in the household and not blind, disabled or with a disabled household member, or over age 65). This is less a concern for the current study interested in matching families, and multiple other studies on evictions in Delaware suggest that families and those with low income appear to be disproportionally subjected to eviction filings.[21,24] However, it might further help explain the relatively lower match rate since benefit programs such as SNAP (available in past research involving match rates) allows single adults to participate. Without these additional records, the match rate in Delaware would underperform as it will fail to link evicted persons who are not enrolled in Medicaid (e.g., single adults).

Building on this theme, the match rate will incrementally improve as increasingly universal datasets are included in the match. This is evident in past research linking guardians in school records to health, human service and other publicly available records and using nearly the same set of identifiers (first and last name, address) and similar linkage procedures. This yielded a very high match rate (81%.[20]) In that study, health records were more comprehensive as they included all hospital-based contacts (e.g., those with private insurance or no insurance) as well as records not tied to program eligibility or service delivery (such as voter registration records). Higher match rates are possible with more comprehensive datasets.

An additional consideration is how the current study differed from past research in representing addresses. We took a relatively more-precise approach of geocoding x,y coordinates, which also would contribute to a more conservative match compared to other studies that geocoded to larger geographic units (e.g., parcel or census block[17–19]). This decision has implications for the likely sensitivity (true positive rate) and specificity (true negative rate) that would result from the match. Furthermore, actual match would fail if the Medicaid address data were not accurate (e.g., not current) as of the time of the eviction.

We achieved robust linkage (75%) between homeless shelter and Medicaid records, even when using the more conservative direct matching approach. This high match rate may be an example of the considerations described above: Shelter services frequently involve case management services, which can provide a direct pathway to Medicaid enrollment (and inclusion in the Medicaid data) for families in shelter. Also, shelter records contain identifiers that are more commonly used: first name, last name, and date of birth. This avoids the complications and potential pitfalls of different approaches to matching on address.

Despite the low-match rate for evictions and Medicaid data, we were able to construct a sizeable study group of 216 children who appeared in all three datasets. Importantly, 71% of these children were enrolled in Medicaid for at least a portion of each year considered in this study. This suggests an adequate opportunity to test hypotheses on the pathway from eviction to shelter for families, and consider patterns of health and healthcare contained in Medicaid claims and other data. Consistent with past research,[10] children from Black race groups were drastically overrepresented (75%) in this group, also suggesting the possibility of important programs of research focused on addressing inequity.

Limitations

This study has important limitations. First, it is primarily an analysis of record linkage and was not thoughtfully designed to rigorously test any other hypothesis or describe any other substantive phenomenon. We urge caution in reading more into this study than originally intended.

Second, our approach does not permit a definitive evaluation of the sensitivity (true positive rate) and specificity (true negative rate) of different linkage methods. Simulation-type studies, for example, allow investigators to manipulate particular aspects of the data to determine sensitivity and specificity of each method under different conditions while the true solution is known. Instead, we explored match rates using different approaches to real-world data where the true solutions are not known with certainty. While the probabilistic method produced a higher number of matches, and these matches were subjected to manual review for confirmation, we cannot make any strong claims about the true positive rate, and less so about the true negative rate.

In addition, our match rates can only be as complete as the datasets that we use to inform them. For example, some private shelters and public hotel/motel voucher programs did not contribute to Delaware's CMIS and, therefore, would not be included in the homelessness data considered here. This coverage gap leaves open the possibility that people we identified as evicted could have also received homeless services that were unrecorded in CMIS. Administrative data systems that are more comprehensive would help ensure that the situations and needs of important subsets of families are included in analyses designed to inform how health and human service systems can better operate. This administrative data match also misses people who were homeless and did not use homeless services, as well as households that were evicted from housing without going through a formal, court-based eviction process.

Future research using integrated data may want to attend to the sequence of these records (e.g., if Medicaid enrollment precedes eviction or a shelter stay, or a shelter stay preceded eviction). We did not attempt to sequence records during the 3-year study period.

Public Health Implications

Integrated health, housing, and other human service data is a powerful tool for public health surveillance, practice, and policy research. However, evictions data present challenges to usual approaches to integration. Nevertheless, our findings suggest that integrated evictions, Medicaid, and homeless shelter data can produce a sizeable study group that may be appropriate for some research questions. The children identified in this integrated dataset were enrolled in Medicaid during multiple years, suggesting possible studies of health and healthcare utilization among this group. Black children were also drastically overrepresented, suggesting the need for and utility of studies focused on structural inequity.

Tangibly, the current results affirm that researchers can construct a sizeable study group through linking administrative data sources from Medicaid, shelter utilization and, to a degree, evictions records to interrogate questions about health and healthcare utilization among children with and without these experiences, such as whether forced residential displacement contributes to higher rates of chronic disease or differences in either/both routine and non-routine healthcare. Incorporating additional datasets, especially from universal programs (e.g., education), would

likely further improve the match while allowing additional tests of whether and how eviction and homelessness interfere with developmental outcomes.

Dr. Cutuli may be contacted at JJ.Cutuli@Nemours.org.

References

1. Council on Community Pediatrics. (2013, June). Providing care for children and adolescents facing homelessness and housing insecurity. *Pediatrics*, *131*(6), 1206–1210. PubMed https://doi.org/10.1542/peds.2013-0645

2. Cutuli, J. J., & Willard, J. (2019). Building early links for learning: Connections to promote resilience for young children in family homeless shelters. *Zero to Three*, *39*(4), 43–50.

3. Herbers, J. E., & Cutuli, J. J. (in press). Homelessness. In M. H. Bornstein & P. E. Shah (Eds.), *Developmental behavioral pediatrions and developmental science*.

4. Cutuli, J. J., & Herbers, J. E. (2014). Promoting resilience for children who experience family homelessness: Opportunities to encourage developmental competence. *Cityscape (Washington, D.C.)*, *16*(1), 113–140.

5. U.S. Department of Housing and Urban Development. (2020). *The 2018 annual homeless assessment report (AHAR) to Congress*. https://www.huduser.gov/portal/sites/default/files/pdf/2018-AHAR-Part-2.pdf

6. Metraux, S., Solge, J., Mqangi, O., & Culhane, D. P. (2021). *An overview of family homelessness in Delaware: A report to Housing Alliance Delaware*. https://static1.squarespace.com/static/59ca9d72268b96cb977e74fd/t/61520b359e8d026e5e1 2c9e9/1632766774483/UD-CCRS+Family+Homelessness+in+DE+-+Final+Report+v2.pdf

7. Beharry, M. S., & Christensen, R. (2020, April). Homelessness in pediatric populations: Strategies for prevention, assistance, and advocacy. *Pediatric Clinics of North America*, *67*(2), 357–372. https://doi.org/10.1016/j.pcl.2019.12.007 PubMed

8. Cutuli, J. J., Herbers, J. E., Lafavor, T. L., Ahumada, S. M., Masten, A. S., & Oberg, C. N. (2014, May). Asthma and adaptive functioning among homeless kindergarten-aged children in emergency housing. *Journal of Health Care for the Poor and Underserved*, *25*(2), 717–730. https://doi.org/10.1353/hpu.2014.0099 PubMed

9. Grant, R., Gracy, D., Goldsmith, G., Shapiro, A., & Redlener, I. E. (2013, December). Twenty-five years of child and family homelessness: Where are we now? *American Journal of Public Health*, *103*(2, Suppl 2), e1–e10. PubMed https://doi.org/10.2105/AJPH.2013.301618

10. Desmond, M., An, W., Winkler, R., & Ferriss, T. (2013). Evicting children. *Social Forces*, *92*(1), 303–327. https://doi.org/10.1093/sf/sot047

11. Lundberg, I., & Donnelly, L. (2019, February). A research note on the prevalence of housing eviction among children born in US Cities. *Demography*, *56*(1), 391–404. PubMed https://doi.org/10.1007/s13524-018-0735-y

12. Hazekamp, C., Yousuf, S., Day, K., Daly, M. K., & Sheehan, K. (2020, October). Eviction and pediatric health outcomes in Chicago. *Journal of Community Health*, *45*(5), 891–899. https://doi.org/10.1007/s10900-020-00806-y PubMed

13. Desmond, M. (2012). Eviction and the reproduction of urban poverty. *American Journal of Sociology*, *118*(1), 88–133. https://doi.org/10.1086/666082

14. Brumley, B., Fantuzzo, J., Perlman, S., & Zager, M. L. (2015, January 1). The unique relations between early homelessness and educational well-being: An empirical test of the continuum of risk hypothesis. *Children and Youth Services Review*, *48*, 31–37. PubMed https://doi.org/10.1016/j.childyouth.2014.11.012

15. Cutuli, J. J., & Herbers, J. E. (2019, September). Housing interventions and the chronic and acute risks of family homelessness: Experimental evidence for education. *Child Development*, *90*(5), 1664–1683. https://doi.org/10.1111/cdev.13041 PubMed

16. Palmer, A. R., Piescher, K., Berry, D., Dupuis, D., Heinz-Amborn, B., & Masten, A. S. (2023, January). Homelessness and child protection involvement: Temporal links and risks to student attendance and school mobility. *Child Abuse & Neglect*, *135*, 105972. PubMed https://doi.org/10.1016/j.chiabu.2022.105972

17. Collinson, R., & Reed, D. (2018). The effects of evictions on low-income households. *Unpublished Manuscript.[Google Scholar]*, 1-82.

18. Schwartz, G. L., Feldman, J. M., Wang, S. S., & Glied, S. A. (2022, February). Eviction, healthcare utilization, and disenrollment among new york city medicaid patients. *American Journal of Preventive Medicine*, *62*(2), 157–164. PubMed https://doi.org/10.1016/j.amepre.2021.07.018

19. Richter, F. G.-C., Coulton, C., Urban, A., & Steh, S. (2021). An integrated data system lens into evictions and their effects. *Housing Policy Debate*, *31*(3-5), 762–784. https://doi.org/10.1080/10511482.2021.1879201

20. Cutuli, J. J., Torres Suarez, S., Truchil, A., Yost, T., & Green, C. (under review). Strategies to better identify student homelessness using data in an urban school district.

21. Metraux, S., Mwangi, O., & McGuire, J. (2022, August 31). Prior evictions among people experiencing homelessness in Delaware. *Delaware Journal of Public Health*, *8*(3), 34–38. PubMed https://doi.org/10.32481/djph.2022.08.009

22. DMMA. (nd). *Medicaid*. Retrieved from https://dhss.delaware.gov/dhss/dmma/medicaid.html

23. Centers for Disease Control and Prevention. (n.d.). *Link Plus*. Retrieved from https://www.cdc.gov/cancer/npcr/tools/registryplus/lp.htm

24. Metraux, S., & Guterbock, A. (2020). Eviction and Legal Representation in Delaware-An Overview.