Data Shop

Data Shop, a department of Cityscape, presents short articles or notes on the uses of data in housing and urban research. Through this department, the Office of Policy Development and Research introduces readers to new and overlooked data sources and to improved techniques in using well-known data. The emphasis is on sources and methods that analysts can use in their own work. Researchers often run into knotty data problems involving data interpretation or manipulation that must be solved before a project can proceed, but they seldom get to focus in detail on the solutions to such problems. If you have an idea for an applied, data-centric note of no more than 3,000 words, please send a one-paragraph abstract to chalita.d.brandly@hud.gov for consideration.

HUD Administrative Data Now Linked to National Longitudinal Study of Adolescent to Adult Health

Atticus Jaramillo
University of North Carolina at Chapel Hill

Michael D. Webb
Community Science

Jon M. Hussey
University of North Carolina at Chapel Hill

Abstract

This article describes a new dataset that links U.S. Department of Housing and Urban Development (HUD) administrative data (1995–2017) to survey and biomarker data from the National Longitudinal Study of Adolescent to Adult Health (Add Health).[1] *Add Health is a nationally-representative cohort study that aims to broaden understanding of how overlapping life contexts—such as school, family, and*

[1] Add Health is directed by Robert A. Hummer and funded by the National Institute on Aging cooperative agreements U01 AG071448 (Hummer) and U01AG071450 (Aiello and Hummer) at the University of North Carolina at Chapel Hill. Waves I-V data are from the Add Health Program Project, grant P01 HD31921 (Harris) from *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD), with cooperative funding from 23 other federal agencies and foundations. Add Health was designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill.

Abstract, continued

peer context—influence the outcomes of adolescents as they age into adulthood, with a primary focus on health and well-being. Add Health has tracked these outcomes through five waves of data collection between 1995 and 2018. The linked HUD-Add Health dataset identifies 1,159 Add Health respondents who received federal rent assistance between 1995 and 2017 and provides information about their stay in federally assisted housing (for example, type of assistance received, years assisted). This article describes how this dataset was created and outlines analytic considerations for researchers.

Introduction

Housing is an important social determinant of health, with numerous studies showing that low-income households often experience housing stressors that degrade their health, such as eviction, rent unaffordability, and foreclosure (Alidoust and Huang, 2021; Singh et al., 2019). This general finding has sparked significant interest in the potential for housing policy interventions to promote positive health outcomes among low-income households (Bourdeaux et al., 2020; Fenelon et al., 2018, 2017; Simon et al., 2017; Slopen et al., 2018). Recently, HUD has aimed to advance this body of research by supporting the linkage of HUD administrative data to nationally representative health studies, such as the National Health Interview Survey (Golden and Mirel, 2021; Lloyd et al., 2017). Such linked datasets provide researchers a unique opportunity to explore best practices for promoting positive health outcomes among families who receive federal housing assistance.

In 2018, HUD's Office of Policy Development and Research awarded a grant to the Center for Urban and Regional Studies (CURS) and the Carolina Population Center (CPC) to link HUD administrative data (1995–2017) to the National Longitudinal Study of Adolescent to Adult Health (Add Health). The research team used probabilistic linkage methods to search 70.6 million HUD administrative records for Add Health study respondents. Through both statistical and manual review, the research team ultimately identified 1,159 Add Health respondents who lived in HUD-assisted housing at some point between 1995 and 2017.

Because Add Health is a longitudinal study, the linked dataset provides a unique opportunity to explore the relationship between HUD rent assistance and health outcomes across the life course. The dataset is also complex, however, and has caveats that researchers should account for when conducting their analyses. This article highlights some of these caveats and provides practical guidance about how to address them by describing (1) the underlying data sources of the linked dataset, (2) the methods used to link these sources, and (3) analytic considerations researchers must account for when using the data.

Data Sources

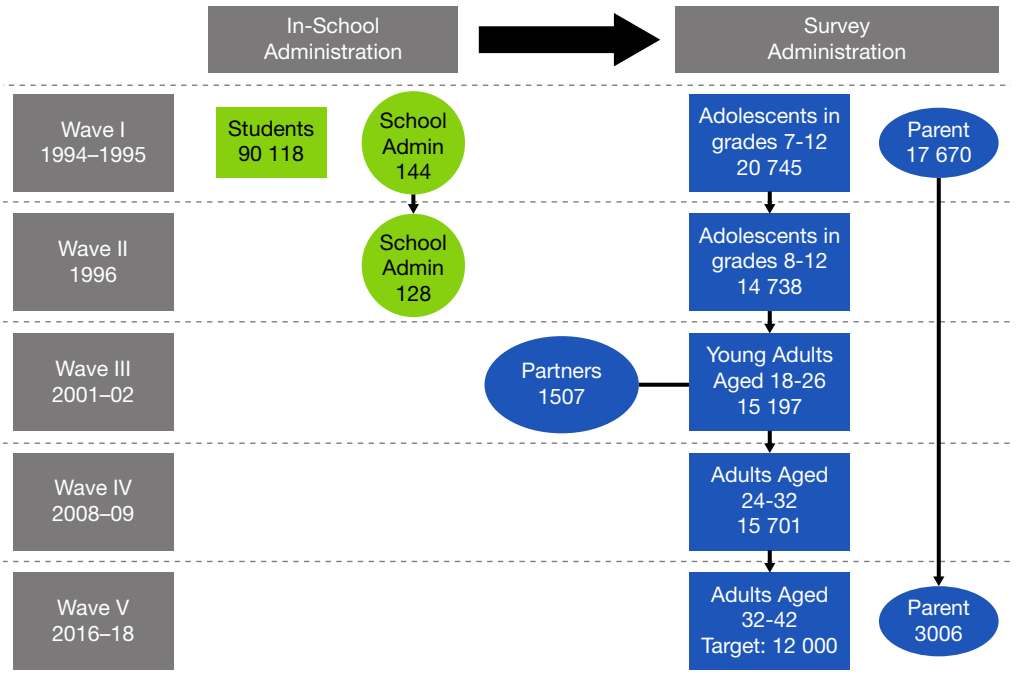
Add Health

Add Health is a cohort study that has tracked a nationally representative sample of U.S. adolescents as they age into adulthood (Harris et al., 2019). The Add Health study began in 1994–95 with the initial goal of understanding how multiple life contexts influence health and health behaviors during adolescence. In subsequent years, the study's goals expanded to include how adolescent and early adult experiences influence health and well-being outcomes in mid-adulthood.

To shed light on these topics, Add Health collects both survey data and biomarker data, such as blood samples, blood pressure, and body mass index (BMI). Add Health researchers have also linked a variety of geographic "contextual" data to respondents' data files, such as Census data on neighborhood characteristics and county-level health data. At Wave I of the Add Health study (1994–95), researchers used a multistage, stratified, school-based, cluster-sampling design to select an initial cohort of 20,745 adolescents in grades 7 to 12 who attended 132 different schools across 80 communities (Harris et al., 2019). Add Health researchers have continued to survey members of the Wave I cohort as they age into adulthood through four additional Waves of data collection: Wave II, 1996; Wave III, 2001–02; Wave IV, 2008–09; and Wave V, 2016–18 (see exhibit 1).

Exhibit 1

Add Health Study Design



Add Health = National Longitudinal Study of Adolescent to Adult Health
Source: Harris et al., 2019

HUD Administrative Data

HUD helps low-income renters and their families access affordable homes that meet basic livability standards through a variety of housing assistance programs (McCarty, Perl, and Jones, 2014). The largest programs include "Section 8" Housing Choice Vouchers (HCVs), public housing, and multifamily housing programs. Local entities that administer these programs regularly collect administrative data on program participants and their household members (Lloyd et al, 2017). Data are typically collected at two times: when a tenant initially moves into a unit and, at minimum, annually thereafter. Public housing agencies (PHAs)who participate in the Moving to Work (MTW) demonstration, however, often collect data less frequently in an effort to achieve cost efficiencies, which is a key goal of the demonstration (Webb, Frescoln, and Rohe, 2016). The data collected on program participants include a broad range of topics, such as date of data collection, Social Security numbers, household size, and income.

Since the mid-1990s, data on program participants have been uploaded to different centralized data repositories; each data upload is referred to as a "transaction record" (Lloyd et al., 2017). Prior to the early 2000s, PHAs submitted data on HCV and public housing recipients to the Multifamily Tenant Characteristics System (MTCS). After the early 2000s, PHAs submitted data to the Public and Indian Housing Information Center (PIC), which was recently updated into the Information Management System of the Public and Indian Housing Information Center (IMS/PIC). For multifamily program participants, data is uploaded to the Tenant Rental Assistance Certification System (TRACS). Data uploaded to these repositories are substantively identical across programs. To complete the HUD-Add Health data linkage, data from MTCS, PIC, and TRACS were used, as the Add Health study (1995–2018) spanned the time periods when each of these data systems was used by HUD.

Linkage Methods and Results

Research staff at the University of North Carolina's CPC used probabilistic linkage methods to search HUD administrative records (1995–2017) for Add Health respondents. The linkage identified 1,159 Add Health respondents who had received HUD rent assistance at any point between 1995 and 2017. Some of these respondents were living with a parent when they received HUD assistance (as they have transaction records before their 18th birthday), others received assistance as an adult, and some received assistance as both a child and adult. The final linked dataset includes 8,587 HUD transaction records on 1,159 unique Add Health respondents. This section describes the methods used to create the linked dataset, the structure of the linked dataset, and how it can be accessed.

Methods

CPC research staff used probabilistic linkage methods to search 70.6 million individual HUD residents (members) and household-level records to locate the Add Health respondents. Probabilistic linkage methods use multiple partial identifiers (for example, name, date of birth) to determine the probability that two people listed in disparate data sources are the same person (Winkler, 2015). To ensure valid "links," probabilistic linkage methods assign stronger "agreement

weights" to partial identifiers that are particularly unique in both data sources. For example, an uncommon last name will have more discriminatory power—that is, the ability to distinguish between two people—than a common last name and, therefore, will be assigned a more substantial agreement weight (Winkler, 2015). After assigning those weights, the overall match or linkage score is calculated as the sum of agreement weights for the full set of partial identifiers. A low score indicates a low degree of agreement across partial identifiers—thus an unlikely match—whereas a high score indicates the opposite. If the linkage score is high enough, it is assumed that partial identifiers belong to a single person.

Probabilistic methods differ from deterministic matching, which identifies matches based on exact matches of partial identifiers (for example, Social Security number + name). The research team opted not to use deterministic methods because they are difficult to implement in longitudinal studies as partial identifiers can change over time; for example, last names may change after a person marries. Further, Add Health does not collect information on reliable person identifiers (for example, Social Security numbers). Thus, the probabilistic linkage was most appropriate given the context of the Add Health study.

For this linkage, CPC research staff used Centers for Disease Control and Prevention's Link Plus (version 2.0)[2] software and eight partial identifiers: (1) first name; (2) middle name; (3) last name; (4) month of birth, (5) day of birth, (6) year of birth; (7) sex; and (8) 12-digit Federal Information Processing Standard (FIPS) code for the county of residence.[3] Prior to the linkage, the research team determined which Add Health and HUD administrative records were deemed "linkage eligible" because they were free of apparent data errors. Add Health respondents were considered linkage eligible if they had complete data for all partial identifiers. HUD administrative records were considered linkage eligible if they *did not* have the following data errors: (1) records sharing unique member IDs but having different birth dates and (2) unique member IDs that were associated with different household IDs within the same calendar year and form type. After determining linkage eligibility, the research team searched the HUD records for the 7 years in which the Add Health study was actively surveying participants. Those 7 years are significant because they are the source of the current name and 12-digit FIPS code of Add Health respondents that are used in the match; thus, the linkage implemented the most recent address—as summarized by the 12-digit FIPS code—to conduct the link. Next, the research team used unique HUD household head and member identification numbers from these matches to search for additional HUD records for these Add Health participants across the full set of 23 annual HUD administrative files (1995–2017). Rationales for these pre-linkage steps and general linkage steps are outlined in further detail in an expanded technical report (Jaramillo et al., 2022).

All Link Plus-recommended matches were subjected to two stages of review. First, all Link Plus-recommended matches were programmatically evaluated using SAS, which was previously known as the Statistical Analysis System. Recommended matches were rejected if the comparison of both first and last names yielded values exceeded 300 (as evaluated by SAS's string comparison function (i.e., "COMPGED"), which indicated the names were highly dissimilar. Further, matches were

[2] Link Plus has been successfully applied to link a variety of databases, such as hospital discharge data (Bigback et al., 2015) and vital records (Zhang et al., 2012).

[3] See Jaramillo et al., 2022 for more details.

rejected if the date of birth did not match satisfactorily; digit transposition (for example, 9/12/1977 versus 9/21/1977) or differences of one were tolerated (for example, 9/12/1977 versus 9/13/1977). Finally, matches were rejected if respondent sex disagreed. Second, all remaining records underwent manual review. Through these two stages of review, the CPC research staff rejected 142 recommended matches based on linking variables and approved 1,159 Link Plus-recommended matches. The main limitations of these steps are that (1) the decision of COMPGED thresholds and date of birth rejections were ultimately subjective and (2) rejecting matches based on sex agreement assumes HUD administrative records are accurate, which may not be the case. Due to these limitations, some accurate matches may have been wrongly rejected. Ultimately, however, the research team opted for these steps to ensure the most conservative match possible.

Variables and File Structure

The linked dataset on the 1,159 Add Health respondents is a single file that includes two general pieces of information. First, unique Add Health respondent identifiers, known as AIDs, are provided for each respondent. Researchers can use these identifiers to merge the linked dataset to all other Add Health data files, including data from each Add Health survey wave, geographic contextual data, and biomarker data. Second, the file provides key variables from household and person-level HUD transaction records (N=8,587) for the Add Health respondents.

The linked dataset is organized in a "hierarchical" or "long" format, with each row corresponding to a unique transaction record. Each row also includes several data points about that transaction record, including: (1) the year and financial quarter for the record; (2) the type of administrative form the record is based on; (3) the type of housing assistance the respondent was receiving (for example, public housing, Housing Choice Voucher); (4) the record type (for example, annual income certification, move-out certification, etc.); (5) total household members; and (6) an episode demarcation flag. A full codebook and detailed documentation for these variables are available on the Add Health user website.[4]

The episode demarcation flag is a particularly important variable. This variable allows researchers to determine whether HUD records depict a single, continuous housing assistance "episode" or two distinct episodes. An episode is a single, sustained period of time in which a linked sample member was living in a HUD-assisted unit. To determine episodes, the research team used standards that were developed to complete a similar linkage of HUD administrative data and National Center for Health Statistics (NCHS) data (Lloyd et al., 2017). These criteria are different for PHAs that participate in the MTW demonstration versus those that do not, as MTW PHAs are subject to different data reporting standards. The specific criteria are:

- Non-MTW records: if a break in transaction records was 425 days or more, those records were to represent two distinct episodes. If the break was less than 425 days, those records were considered to represent a single episode. The 425-day period is the standard because non-MTW sites must complete recertifications every 425 days (1 year plus 60 days of leeway).

[4] Link: https://addhealth.cpc.unc.edu/documentation/codebooks/

- MTW records: if a break in transaction records was 790 days or more, those records were considered to represent two distinct episodes. If the break was less than 790 days, those records were considered to represent a single episode. The 790-day period is used as the threshold because most MTW sites complete recertifications every 790 days.

Subsample Characteristics

Exhibit 2 provides unweighted descriptive statistics on the subsample of Add Health respondents in the linked dataset, but Add Health also includes sampling weights that researchers can use to compare weighted results and assess link quality. The HUD subsample is primarily composed of females (73 percent) and non-Hispanic African-Americans (59 percent). Slightly under one-half of the subsample (42 percent) first began receiving assistance as adolescents (18 or younger), and all remaining members first began receiving assistance as adults. Most HUD subsample members received HCV assistance (42 percent). A sizable proportion of respondents, however, also received some form of place-based assistance, including public housing assistance (11 percent), project-based Section 8 (18 percent), or multifamily assistance (3 percent). With these points in mind, one important insight provided by exhibit 2 is that the racial composition of Add Health respondents in the linked sample differs from the overall Add Health cohort and the overall population of HUD-assisted adults (for example, overrepresentation of women and Black or African-American households). Thus, statistical estimates from this sample may be biased toward or against certain racial and ethnic minority groups.

It is notable that a relatively large proportion of households received assistance through two or more programs, which may be attributable to two potential explanations. First, some respondents may have, in fact, received two forms of assistance. Second, over time HUD has implemented a variety of demonstration programs and policy reforms that can alter the type of assistance that households receive. For example, the HOPE VI program transitioned many households from public housing to HCV assistance, and the Rental Assistance Demonstration (RAD) allows PHAs to transition public housing units to project-based Section 8 or project-based HCV. Unfortunately, due to data confidentially concerns, the research team was unable to identify whether respondents receiving housing assistance were served by a PHA participating in such a demonstration.

Exhibit 2

Linked Subsample Characteristics Versus All HUD Householders (1 of 2)

	Linked Sample		Add Health		HUD Householders
Demographics	**N**	**%**	**N**	**%**	**%**
Female	850	73	10,480	51	70
Male	309	27	10,263	49	30
Non-Hispanic White	323	28	10,455	50	46
Non-Hispanic Black or African-American	681	59	4,320	21	45
American					
Other Race	155	13	2,424	12	8
Hispanic or Latino (all races)	76	16	3,525	17	18

Exhibit 2

Linked Subsample Characteristics Versus All HUD Householders (2 of 2)

	Linked Sample		Add Health		HUD Householders
Assistance Characteristics	**N**	**%**	**N**	**%**	**%**
Assisted as Adolescent (18 or under)	489	42	–	–	–
Assisted as Adult (18 or older)	670	58	–	–	–
Housing Choice Voucher	488	42	–	–	–
Project-Based Section 8	206	18	–	–	–
Public Housing	126	11	–	–	–
Multifamily	31	3	–	–	–
Assisted by Two or More Programs	308	27	–	–	–
Total Observations	**1,159**	**100**	**20,745**	**100**	**–**

Add Health = National Longitudinal Study of Adolescent to Adult Health

Notes: Linked sample characteristics based on Wave I (1994–95) Add Health Survey Data. Information for all HUD households only includes head-of-households and is obtained from Eggers' (2020) analysis of 2017 HUD administrative data. Tabulations for the overall Add Health sample may not perfectly match the total sample size (N=20,745) due to missing data. Percentages may not add to 100 due to rounding.

Sources: Add Health; Eggers (2020)

Access Options

Although the linked HUD-Add Health dataset excludes all personal identifiers, and the data have undergone deductive or statistical disclosure risk assessment, they are considered restricted-access data and will not be released as public-use files. Consequently, researchers interested in obtaining access to the linked data set and any other restricted-use Add Health data must apply for a restricted-use contract using the CPC Data Portal. To learn more about how to obtain a restricted-use contract, see the CPC Data Portal website.[5]

Analytic Considerations

This section highlights four analytic considerations that researchers should be aware of when using the linked dataset, including: (1) survey bias and sampling weights; (2) HUD data quality; (3) sequencing of HUD data collection and Add Health survey collection; and (4) geographic specificity. The main purpose of this discussion is to raise awareness of these considerations rather than provide specific analytic guidance, as researchers are encouraged to think carefully about how to address these considerations based on the goals of their analysis: there is no "one-size-fits-all" solution. To this end, the authors strongly encourage researchers to consult the articles and resources referenced in this section before using the dataset.

Survey Bias and Weights

Researchers should consider if three forms of survey bias may confound their analysis when using the linked HUD-Add Health dataset, including: (1) sampling bias; (2) attrition bias; and (3) missing data bias, which arises due to question non-response. With respect to sampling bias, it is

[5] Link: https://data.cpc.unc.edu/

crucially important to researchers to remember that the respondents in the linked Add Health dataset differ from the overall Add Health cohort and the overall population of HUD-assisted adults; thus, researchers must think carefully about the potential for such differences to bias their analysis (see exhibit 2). A common issue affecting school-based samples is potential bias due to the exclusion of school dropouts. High school dropouts during the 2 years of Wave I (1993–95) were eligible for selection into this study (Harris et al., 2019), however. Further, Udry and Chantala (2003) found that no problematic dropout bias existed on a host of key outcomes. It is also important to note that some school dropouts are still included in the study, as they may have dropped out after the first survey wave.

A related sampling consideration is whether researchers should use sampling weights, which adjust statistical estimates based on Add Health's sampling strategy. Add Health recommends that researchers use sampling weights if they wish to generalize their results to the overall U.S. population (Chen and Harris, 2020). Thus, weights should be used if the HUD-Add Health linked dataset is used to create a variable that, for example, aims to understand whether adolescent receipt of HUD assistance is a risk factor for health outcomes among the overall population of U.S. adolescents. Using weights on the subsample of Add Health respondents using the linked dataset, however, is more complex, as researchers may need to adjust these sampling weights given subpopulation characteristics. Researchers should consult Chen and Harris (2020) for more information about the Add Health sampling weights and how they can potentially be adjusted for sub-sample population analysis.

Like all longitudinal surveys, attrition is a potential issue. Harris et al. (2019) summarizes previous evaluations of attrition bias across Add Health study waves and finds that attrition is lower among "female, younger, higher socioeconomic status, urban, native-born, and White respondents at Wave III and IV" (Harris et al., 2019: 1415b). Although there are clear attrition patterns, attrition bias is "minimal" (Harris, 2019). That noted, a caveat of the Add Health data is that respondents may participate in some, but not all, waves of the study; for example, respondents may participate in Wave I and IV, but not Wave III. This issue can complicate the longitudinal analysis of the Add Health data and should be considered by researchers.

Finally, data missingness is also a potential source of bias, as respondents may not answer all survey questions. In such instances, it is important for researchers to determine whether there are systematic differences between respondents with missing and complete data, as systematic differences between these groups can introduce bias (Bhaskaran and Smeeth, 2014). Because the Add Health study collects a vast body of data, there is no single study that explores the issue of data missingness among respondents. Numerous articles provide practical insight into how researchers can assess and address bias due to missing data, however, and we encourage researchers to use these resources.

HUD Data Quality

The U.S. Government Accountability Office (GAO) and HUD's Office of the Inspector General (OIG) have identified data quality issues associated with MTCS, PIC, and TRACS data used to complete this linkage (Lloyd et al., 2017). Some major errors of the MTCS data, as identified by GAO and

OIG reports, included missing addresses, names, or Social Security numbers. Those omissions are considered "fatal" because it is virtually impossible to accurately identify a person without those variables, and it affects an average of 7 percent of data fields. Other reports have further found that HUD did not have sufficient data collection or management controls in place during the early 2000s. Specifically, at that time, the PIC system did not ensure that Social Security information was accurately collected, although this information was not used in this linkage. Not only did the system allow for the submission of incomplete or inaccurate Social Security numbers, but it also did not require the use of a common format (for example, 123-45-6789 versus 123456789).

These data-quality limitations are important because they may have prevented accurate cross-identification of people who resided in HUD-assisted housing *and* participated in the Add Health study. The risk is noteworthy because Add Health participants were adolescents during roughly the same period of time when HUD data were arguably the least reliable. That noted, HUD undertook several actions to improve data quality during the 2000s, and those efforts have greatly improved the accuracy and completeness of HUD data. Nonetheless, it is important for researchers to keep in mind that the linked dataset is based on administrative records subject to data quality issues.

Sequencing

The sequencing of when HUD administrative records were collected versus when the Add Health interview took place may be important to some researchers, particularly when researchers aim to draw conclusions about the linked subsample life circumstances while living in HUD-assisted housing. Each Add Health data record includes an interview date—that is, when that data was collected. Researchers can use this interview date in tandem with HUD transaction record dates to determine whether a given linked sample member was living in HUD-assisted housing at the time of the Add Health survey. The research team consulted with HUD to determine a general guideline for determining "concurrence" and decided on the following standard: if an Add Health respondent had a transaction record for the same year and quarter when they responded to an Add Health survey, it could be assumed that those respondents were living in HUD-assisted housing at the time of their Add Health interview.

Geographic Specificity

Add Health collects extremely sensitive data on respondents, such as their exposure to sexual abuse, use of illegal drugs, and even their genetic information. Consequently, Add Health maintains strict data confidentiality and deductive disclosure protocols to ensure that Add Health respondents cannot be identified based on their survey data. One such protocol is to anonymize the specific communities and states where respondents live: researchers can only determine the Census region (for example, Midwest, South) where respondents live. To address this issue, Add Health provides a variety of geographic contextual data (for example, neighborhood census data, county health data) with pseudo-GEOIDs and geographical relationship files, the latter of which can help researchers identify whether respondents live in a common neighborhood or metropolitan area.

Conclusion

The linked HUD-Add Health dataset provides a unique opportunity to explore the health and well-being outcomes of a subsample of low-income adolescents (N=1,159), now adults, who received federal housing assistance between 1995 and 2018. Researchers should account for a variety of analytic considerations when using this dataset, however, including (1) survey bias and sampling weights; (2) HUD data quality issues; (3) the sequencing of HUD data collection and Add Health survey collection; and (4) geographic specificity. This article provides a high-level overview of these analytic considerations, but we encourage researchers to consult the resources cited in this article to gain a better understanding of how to use the linked HUD-Add Health dataset appropriately. The authors also encourage researchers to consult the Add Health user website to obtain further information about how to access and analyze this dataset.

Acknowledgments

The authors would like to thank HUD's Office of Policy Development and Research (PD&R) for awarding our research team a collaborative research grant to complete this project. We also thank several key staff at PD&R for overseeing this project and providing valuable input, including Jagruti Rekhi, Kinnard Wright, and Veronica Helms. Finally, we thank the CPC Add Health research team for their expert assistance with this project.

Authors

Atticus Jaramillo is a Visiting Assistant Professor at the University of North Carolina at Chapel Hill. Michael Webb is Managing Associate at Community Science. Jon Hussey is Assistant Professor at the University of North Carolina at Chapel Hill.

References

Alidoust, Sara, and Wei Huang. 2021. "A Decade of Research on Housing and Health: A Systematic Literature Review," *Reviews on Environmental Health*. https://doi.org/10.1515/reveh-2021-0121

Bhaskaran, Krishnan, and Liam Smeeth. 2014. "What is the Difference Between Missing Completely at Random and Missing at Random?" *International Journal of Epidemiology* 43 (4): 1336–1339.

Bigback, Kristyn M., Megan Hoopes, Jenine Dankovchik, Elizabeth Knaster, Victoria Warren-Mears, Sujata Joshi, and Thomas Weiser. 2015. "Using Record Linkage to Improve Race Data Quality for American Indians and Alaska Natives in Two Pacific Northwest State Hospital Discharge Databases," *Health Services Research* 50: 1390–1402.

Boudreaux, M., A. Fenelon, N. Slopen, and S.J. Newman. 2020. "Association of Childhood Asthma with Federal Rental Assistance," *JAMA Pediatrics* 174 (6): 592–598.

Chen, Ping, and Kathleen Mullan Harris. 2020. *Guidelines for Analyzing Add Health Data*. Chapel Hill: Carolina Population Center, University of North Carolina at Chapel Hill.

Eggers, Frederick J., Econometrica Inc., SP Group LLC. 2020. *Characteristics of HUD-Assisted Renters and Their Units in 2017*. U.S. Department of Housing and Urban Development, Office of Policy Development and Research. Washington DC: Government Publishing Office.

Fenelon, Andrew, Patrick Mayne, Alan E. Simon, Lauren M Rossen, Veronica Helms, Patricia Lloyd, Jon Sperling, and Barry L. Steffen. 2017. "Housing Assistance Programs and Adult Health in the United States," *American Journal of Public Health* 107 (4): 571–578.

Fenelon, Andrew, Natalie Slopen, Michel Boudreaux, and Sandra J. Newman. 2018. "The Impact of Housing Assistance on the Mental Health of Children in the United States," *Journal of Health and Social Behavior* 59 (3): 447–463.

Golden, Cordell, and Lisa B. Mirel. 2021. "Enhancement of Health Surveys with Data Linkage." In *Administrative Records for Survey Methodology*. John Wiley & Sons, Inc.: 271–296.

Harris, Kathleen Mullan, Carolyn Tucker Halpern, Eric A. Whitsel, Jon M. Hussey, Ley A. Killeya-Jones, Joyce Tabor, and Sarah C. Dean. 2019. "Cohort Profile: The National Longitudinal Study of Adolescent to Adult Health (Add Health)," *International Journal of Epidemiology* 48 (5): 1415–1415k.

Jaramillo, Atticus, Jon M. Hussey, Michael D. Webb, Tim Monbureau, and Maria Marrufo. 2022. *Linkage of National Longitudinal Study of Adolescent to Adult Health Data to 1995 to 2017*. Washington, DC: U.S. Department of Housing and Urban Development.

Lloyd, Patricia C., Veronica E. Helms, Alan E. Simon, Cordell Golden, James Brittain, Eileen Call, Lisa B. Mirel, Barry L. Steffen, Jon Sperling, Elizabeth C. Rudd, Jennifer D. Parker, and Carol S. Star. 2017. "Linkage of 1999–2012 National Health Interview Survey and National Health and Nutrition Examination Survey Data to U.S. Department of Housing and Urban Development Administrative Records," *Vital and Health Statistics* (60):1–40.

McCarty, Maggie, Libby Perl, and Katie Jones. 2014, April. *Overview of Federal Housing Assistance Programs and Policy*. Congressional Research Service, Library of Congress.

Simon, Alan E., Andrew Fenelon, Veronica Helms, Patricia C. Lloyd, and Lauren M. Rossen. 2017. "HUD Housing Assistance Associated with Lower Uninsurance Rates and Unmet Medical Need," *Health Affairs* 36 (6): 1016–1023.

Singh, Ankur, Lyrian Daniel, Emma Baker, and Rebecca Bentley. 2019. "Housing Disadvantage and Poor Mental Health: A Systematic Review," *American Journal of Preventive Medicine* 57 (2): 262–272.

Slopen, Natalie, Andrew Fenelon, Sandra Newman, and Michel Boudreaux. 2018. "Housing Assistance and Child Health: A Systematic Review," *Pediatrics* 141 (6).

Udry, J. Richard, and Kim Chantala. 2003. "Missing School Dropouts in Surveys Does Not Bias Risk Estimates," *Social Science Research* 32 (2): 294–311.

Webb, Michael D., Kirstin P. Frescoln, and William M. Rohe. 2016. "Innovation in U.S. Public Housing: A Critique of the Moving to Work Demonstration," *International Journal of Housing Policy* 16 (1): 111–124.

Winkler, William E. 2015. "Probabilistic Linkage." In *Methodological Developments in Data Linkage*. John Wiley & Sons, Ltd.: 8–35.

Zhang, Yujia, Bruce Cohen, Maurizio Macaluso, Zi Zhang, Tonji Durant, and Angela Nannini. 2012. "Probabilistic Linkage of Assisted Reproductive Technology Information with Vital Records, Massachusetts 1997–2000," *Maternal and Child Health Journal* 16 (8): 1703–1708.