When a City Isn't a City: Aggregating Data From the Picture of Subsidized Households to the Municipal Scale for Research Purposes

Will B. Payne
Lauren E. Nolan
Eric Seymour
Rutgers University

Abstract

The authors have developed a primarily automated process to take the "city"-level dataset from the Picture of Subsidized Households (PSH), which corresponds to the U.S. Census Bureau's Populated Place Areas, and reassemble it at the scale of a state's municipalities. Municipalities are the relevant scale of governance for many critical issues that have outsize local and regional impacts on housing affordability and residential segregation, like zoning and rent control. This article and accompanying R code (https://github.com/willbpayne/NJSOARH/) outlines the reasons that transforming the spatial scale of PSH data may be necessary, and the steps the authors took to synthesize it into the county subdivision level in New Jersey. This effort was done to allow U.S. Department of Housing and Urban Development (HUD) data users to adapt this process to their needs and better understand the correspondence between municipal-level policies and housing goals and outcomes.

Introduction

HUD releases the Picture of Subsidized Households (PSH) every year, providing detailed unit counts and household demographics for a variety of subsidized housing programs to researchers and practitioners (Taghavi, 2008). Scholarship using the PSH is extensive, including work aimed at understanding affordable housing needs and program assessment (see Bailey et al., 2016; Greenlee, 2019; Metzger, 2014; Silverman, Patterson, and Wang, 2020) and assessing risk and opportunities for residents in subsidized households (see Chakraborty et al., 2021; Gabbe and Pierce, 2020; Gabbe, Pierce, and Oxlaj, 2020; Lens, 2014). The PSH is available at various scales of spatial

aggregation, including summaries by state, Core Based Statistical Area (CBSA), public housing agency, project, ZIP Code, census tract, and city. It is fairly straightforward to link most of these scales to other datasets, both tabular and spatial, but the "city" level is significantly more complex, since it contains a mix of incorporated, unincorporated, and statistical entities. As such, this dataset needs significant processing to be useful to researchers interested in the municipal scale.

The city level in the PSH is of limited use in understanding the effects of housing policy if it cannot be translated to the municipality, the most granular and responsive level of governance for most housing issues such as rent control (Ambrosius et al., 2015) and land-use restrictions, most notably zoning (Pendall, 2000; Stacy et al., 2023). As part of the New Jersey State of Affordable Rental Housing (NJSOARH) research project, the authors developed a method to assemble statewide datasets of HUD-subsidized unit counts and demographic data at the municipal (county subdivision) level using data from the city, census tract, and project level of the PSH. The remainder of this article outlines this method, its benefits over alternative approaches, and the applicability of this process to other states that are, like New Jersey, entirely composed of incorporated local governments.

Data and Methods

Problem Overview

At first glance, the city-level table in the PSH appears to correspond to municipalities, which are easily joined by the GEOID field to spatial boundary files and additional datasets, but on closer examination, many geographic entities in the dataset do not follow this pattern. First, Census Designated Places (CDPs) with subsidized units are included in the dataset (172 of them in New Jersey's 2022 data). Although spatial boundary files for CDPs can be used to visualize data, CDPs have no legal or political authority, and their geographies are not always intuitive for residents. Second, since CDPs in fully incorporated states like New Jersey are entirely contained in one or more municipalities with governing authority, there are also entities (87 in New Jersey in 2022) with the format "Remainder of X Township," consisting of the portions of each township not covered by any CDPs. Although their actual shapes vary, for purposes of illustration these can be thought of as "Swiss cheese" geometries consisting of a township with one or more CDPs removed. These remainders have GEOIDs of 3499999 in New Jersey (the state code of 34 and 99999 to designate an unknown entity), making them impossible to match to spatial boundary files without the processing steps outlined below.

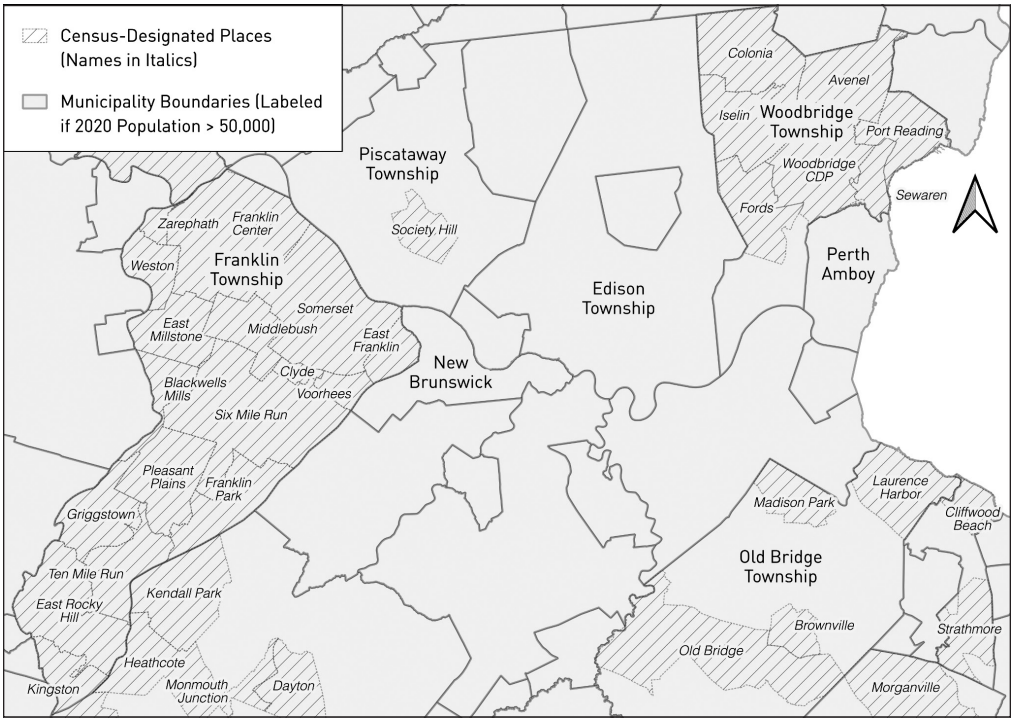
Finally, even for rows that correspond to a single municipality, many records (117 in New Jersey in 2022) have missing GEOIDs, which also appear as 3499999. From comparing these results across states, these 99999 results appear for any towns or townships in a state, because HUD uses Populated Place Area geographies for this report, not minor civil divisions (MCDs) or county subdivisions. This usage is more appropriate for states that were surveyed and settled using the Public Land Survey System (PLSS) after 1785, in which survey township and range divisions do not necessarily correspond to administrative geographies, and unincorporated county land is widespread. However, in states like New Jersey and other Northeastern states, whose political divisions precede the PLSS, towns and townships are fully functioning minor civil divisions, and being able to accurately enumerate HUD-subsidized housing totals in these areas is crucial to understand local governance and supply of affordable housing.

Problem Scope

Exhibit 1 illuminates the problem by presenting a map showing the arrangement of CDP and municipal boundaries in the vicinity of the authors' institution, Rutgers University, New Brunswick. Starting on the western portion of the map, Franklin Township (with a population of 68,364 in the 2020 census) in Somerset County, New Jersey, contains 17 different CDPs and no areas not covered by a CDP. The township's neighbor across the Raritan River in Middlesex County, Piscataway Township (population 60,804) has one CDP ("Society Hill") and one "Remainder of Piscataway Township" row. Nearby Woodbridge Township (population 103,639) has five different CDPs and one "Remainder of Woodbridge Township" row. For these three townships, all of which are incorporated local government units that have significant administrative authority over local land use, 0 units of HUD-subsidized units would be counted if using the PSH city-level file as downloaded from HUD's website. After joining these constituent rows into full municipalities, Franklin has 664 units in the 2022 PSH, Piscataway has 164, and Woodbridge has 1,231.

Exhibit 1

A Map of Portions of Middlesex and Somerset Counties in New Jersey, Showing Municipal Boundaries and Census-Designated Places



Source: Authors

Statewide, 40,408 units, or 23.9 percent of all units in the state of New Jersey, are unjoinable, meaning that if the PSH city data were not cleaned with the process described below, these units would be missing. These problems affect every HUD program in the New Jersey dataset for 2022

(see exhibit 2 for a breakdown by program type and kind of error). Setting aside the moderate rehabilitation program (which had only 205 units statewide) and the Rent Supplement and Rental Assistance Payment (RAP) programs (which had no units), the HUD program least affected by this issue is public housing, at only 12.4 percent of total units unjoinable. This result is intuitive, since public housing buildings are more likely to be in incorporated cities or boroughs, which do not contain CDPs or remainders. Both the Project Based Section 8 and Housing Choice Voucher (HCV) programs, which are less clustered in larger municipalities, had unjoinable rates close to the statewide average of 23.9 percent, with vouchers having the highest unjoinable rate of the three largest HUD programs at 22,144 units unjoinable statewide, or 25.7 percent of all HCV units. Although it is a smaller program, with only 3,766 units statewide, Section 202 Project Rental Assistance Contract (PRAC) has the highest unjoinable rate at 60.6 percent (2,280 units). Section 202, which provides assistance to affordable housing for the elderly, has a particularly dispersed pattern in New Jersey, with many facilities in suburban areas affected by these joinability issues.

The overall unjoinable rate of 23.9 percent is particularly dangerous for unprepared data users: enough to throw off statewide analyses, but not necessarily high enough that the error from a simple tabular join on the "code" field would be immediately obvious. Around three-fourths of the total units in the state would still be accounted for, and some municipalities would be entirely unaffected by the error, including larger cities such as Newark, Trenton, and Camden. Nevertheless, if a researcher was trying to understand the impact of vouchers on suburban housing affordability and residential segregation, places like the Franklin, Piscataway, and Woodbridge Townships discussed above would disappear completely from view, resulting in an incomplete picture of assisted housing, biased toward incorporated cities. Whereas it remains possible to study the neighborhood characteristics of HCV recipients using tract-level data, researchers may be interested, for instance, in studying the effect of municipal source of income antidiscrimination laws on utilization (Freeman, 2012; Freeman and Li, 2014; Tighe, Hatch, and Mead, 2017).

Exhibit 2

Breakdown of Unjoinable Units for New Jersey in 2022 by Program and Error Type

Program	Total Units	Joinable		Unjoinable							
		Units	%	A. CDPs		B. Remainders		C. Names Without GEOID		Total (A + B + C)	
				Units	%	Units	%	Units	%	Units	%
Housing Choice Vouchers	86,286	64,142	74.3	6,284	7.3	7,035	8.2	8,825	10.2	22,144	25.7
Project Based Section 8	48,695	37,409	76.8	2,732	5.6	2,018	4.1	6,536	13.4	11,286	23.2
Public Housing	27,545	24,120	87.6	467	1.7	536	1.9	2,422	8.8	3,425	12.4
202/PRAC	3,765	1,485	39.4	705	18.7	621	16.5	954	25.3	2,280	60.6
S236/BMIR	1,574	869	55.2	0	0.0	196	12.5	509	32.3	705	44.8
811/PRAC	961	421	43.8	148	15.4	194	20.2	198	20.6	540	56.2
Moderate Rehabilitation	205	182	88.8	8	3.9	0	0.0	15	7.3	23	11.2
Total	169,031	128,623	76.1	10,346	6.1	10,599	6.3	19,463	11.5	40,408	23.9

BMIR = Below Market Interest Rate. CDP = Census Designated Places. PRAC =Project Rental Assistance Contract.
Notes: Statewide, 874 units had locations listed as "Missing" in the 2022 PSH, 802 of them (91.8 percent of the total) in the Housing Choice Voucher program. There were no units of "RentSup/RAP" reported for the year.
Source: HUD Picture of Subsidized Households 2022

Methods

To resolve this complexity and aggregate PSH data to municipalities, the authors carry out a series of spatial and tabular operations. The R code that the authors used is available for adaptation and reuse on GitHub (https://github.com/willbpayne/NJSOARH/), but this section summarizes the key points. The first step (section 1 in the code) involves reading in the city-level PSH dataset for a given year, filtering it to the state of interest, and removing extraneous columns. The next step (section 2 in the code) takes all the rows that are split up by program type in the PSH and merges them into a single row for each geographic entity. In doing so, all demographic data for each program are preserved by appending prefixes to each column name for each program. For example, "pct_occupied" from a public housing row (program code of 2) would be renamed to "Pg2_pct_occupied."

Every row now has columns for every demographic data field for each program type (460 columns, up from 74 when each row only represented one program). Doing this process nationally all at once could lead to performance issues here and during the spatial joins, so this process is best done state-by-state. Addressing a single state at a time also allows for closer scrutiny of any potential edge cases and state-specific quirks, such as the fact that a municipality type can mean contradictory things in different states. In Pennsylvania and New Jersey, for example, boroughs are fully incorporated, sovereign local governments, but in Connecticut a borough is an incorporated section of a surrounding town and has no autonomous authority.

The next phase of the code sorts the different rows into those that already represent entire municipalities and those that need further processing and aggregation. As outlined above, four kinds of rows are in the dataset at this point: (1) municipalities (boroughs and cities, in the New Jersey example) that can be used as given, (2) CDPs that need to be combined with (3) remainders of townships and/or other CDPs to aggregate into (4) townships, some of which already exist as rows if they contain no CDPs, but none of which have accurate GEOIDs at the Populated Place Area level. Section 3 uses patterns in the "name" and "code" fields to divide the dataset into these different subsets for further processing. Section 4 uses these categories to generate the summary statistics of unjoinable units by program and error type seen in exhibit 2. Section 5 uses census boundary files for both CDPs and county subdivisions to generate a crosswalk allowing units assigned to CDPs to be ascribed to the correct municipality. Section 6 sets aside any rows that already represent full municipalities for incorporation later. Section 7 assigns municipality GEOIDs to CDPs and remainders based on the CDP crosswalk and name matches. Section 8 allows for manual adjustments to be made to GEOIDs for rows that are still ambiguous between multiple townships (see below). Section 9 merges all the demographic data of any constituent rows that are being combined into one municipality. Finally, section 10 performs final cleanup, merges all rows by municipal GEOID, and outputs the finished file as a CSV.

Limitations of the Automated Crosswalk

Several steps in this crosswalk process can require manual attention from the analyst, since disambiguating the location of HUD-subsidized units at the city level is not necessarily possible from spatial and tabular joins alone without incorporating additional datasets. The two biggest potential problems are caused by CDPs that cross municipal boundaries and townships with the

same name in multiple counties (for example, there are five different Washington Townships in New Jersey), which make it impossible to accurately assign units to a single municipality without bringing in additional data.

The authors' code aims to address these problems by checking how many CDPs intersect multiple county subdivisions and saving those as a separate dataframe. In New Jersey in 2022, eight CDPs intersect multiple county subdivisions, but only five of those contain PSH units for the year. For townships with duplicated names, if only one contains CDPs, disambiguation of the "Remainder of X Township" row is straightforward (and handled by the automated workflow), but that can still leave multiple distinct municipalities with merged data in the "X Township" row. Both situations require triangulation with the tract and project level datasets to determine within which municipality the units in question fall; the code sets "manual aggregation" flags for these areas to ensure that they are handled outside the automatic workflow.

Alternative Approaches

Depending on a researcher's questions and geographies of interest, there may be alternative ways to arrive at a similar outcome, but they are limited in scope. County-level data can be helpful in illuminating broad spatial trends across a state (Bailey et al., 2016), but in a state like New Jersey, many counties span urban, suburban, and rural land uses, making the utility of direct comparison limited. In some areas, public housing authorities overlap with municipal borders, and the PSH is available at the housing authority level, but many areas in New Jersey and other states examined here either lack a dedicated housing authority or share one across part or all of a county. Many analyses of PSH data use spatial scales that can be readily compared to demographic and economic data, like the census tract or ZIP Code Tabulation Area (ZCTA) to examine questions like the neighborhood characteristics of units supported by different HUD programs (Lens and Reina, 2016; Reina and Aiken, 2022), but these scales are not meaningful in the administration and regulation of housing policy and land use.

Aggregating from the census tract level of the PSH into municipalities would seem to sidestep these issues, but tracts often overlap multiple municipalities, since they are only guaranteed to nest perfectly within counties. Also, demographic data are suppressed for rows that contain fewer than 11 subsidized households or fewer than 50 percent of households reporting data, so in some cases a municipality would have usable demographic data, but its constituent tracts would not. For reference, New Jersey has 564 municipalities in 2023, but the state contains 2,010 different census tracts (per the 2010 tract boundaries used for the 2022 PSH). This demographic data suppression also makes working from the PSH's project file (one row per development) less useful, since smaller projects in municipalities with more than 11 units would have data obscured at the project level but included at the city level. A bigger issue with using the project-level data, which scholars have used to answer questions about housing affordability for multifamily properties (Hamidi, Ewing, and Renne, 2016), is that the data are entirely unable to provide information on HCV recipients, because individual HCV households are not identified by location for privacy reasons. In New Jersey in 2022, HCV units represented over one-half of the total HUD-subsidized housing units in the state, so neglecting to include them in analysis would result in an incomplete view of the assisted housing landscape.

Discussion

While this process was developed for research in New Jersey, preliminary tests indicate that it could be helpful in other states that have similar issues with PSH data ambiguity at the city level. Exhibit 3 shows the breakdown of unjoinable units for the nine Northeastern states that are, like New Jersey, fully incorporated. The first column contains all HUD-subsidized units in the PSH at the city level for 2022 with any specific location given within the state. The next two columns contain the absolute number and proportion of these units that are already joinable without following the process outlined above. The next six columns contain the number and proportion of units in each state that are unjoinable for each of the following reasons: (A) they are contained within CDPs, (B) they are contained in "Remainder of X Township" rows, and (C) they have 99999 in their GEOID codes. The final two columns show the total number and proportion of unjoinable units for each state.

Exhibit 3

Summary of 2022 Picture of Subsidized Households Units and Their Ability To Be Joined at the Municipal Scale Across a Number of Fully Incorporated Northeastern States

State	Total Units	Joinable		Unjoinable							
		Units	%	A. CDPs		B. Remainders		C. Names Without GEOID		Total (A+B+C)	
				Units	%	Units	%	Units	%	Units	%
New York	594,317	504,148	85.0	67,949	11.5	16,907	2.8	4,301	0.7	89,157	15.0
Pennsylvania	222,354	183,518	82.7	11,780	5.3	13,901	6.3	12,654	5.7	38,335	17.3
Massachusetts	196,945	160,806	82.0	22,100	11.3	5,764	2.9	7,468	3.8	35,332	18.0
New Jersey	169,905	128,623	76.1	10,346	6.1	10,599	6.3	19,463	11.5	40,408	23.9
Connecticut	84,093	62,870	75.0	14,152	16.9	3,751	4.5	3,001	3.6	20,904	25.0
New Hampshire	22,177	15,483	70.1	4,094	18.5	1,263	5.7	1,248	5.7	6,605	29.9
Rhode Island	38,585	26,641	69.2	2,739	7.1	2,346	6.1	6,790	17.6	11,875	30.8
Maine	27,109	17,056	63.4	7,251	27.0	1,025	3.8	1,559	5.8	9,835	36.6
Vermont	13,245	7,353	55.8	3,424	26.0	1,257	9.5	1,133	8.6	5,814	44.2
Total	1,368,730	1,106,498	81.1	143,835	10.5	56,813	4.2	57,617	4.2	258,265	18.9

CDP = Census Designated Places.

Note: Across the nine states, 3,967 units had locations listed as "Missing" in the 2022 PSH.

Source: HUD Picture of Subsidized Households 2022

As these final columns show, far from being an outlier, New Jersey is close to the middle of the distribution within Northeastern states as far as the proportion of unjoinable units. Certain regional trends are clear in the results. The states with relatively urban population distributions tend to have more units in incorporated cities that are already joinable without following this process, including New York (15.0 percent unjoinable), Pennsylvania (17.3 percent), Massachusetts (18.0 percent), New Jersey (23.9 percent), and Connecticut (25.0 percent). The less densely populated New England states of New Hampshire (29.9 percent), Rhode Island (30.8 percent), Maine (36.6 percent), and Vermont (44.2 percent) have higher percentages of unjoinable HUD-subsidized units. Combined, these nine states contain more than a quarter (27 percent) of all subsidized units nationally, including the majority (74 percent) of S236/Below Market Interest Rate (BMIR) units, 37 percent of all moderate rehabilitation program units, and 36 percent of all public housing units per the 2022 PSH.

Researchers and practitioners in any of these states who are interested in studying questions of municipal governance and its impact on housing affordability could benefit from this open-source data workflow and the discussion above of its strengths and weaknesses. This workflow could also be adapted to many other states that include large sections of unincorporated county land by aggregating any units that fall outside of incorporated municipalities at the census county division (CCD) level and up from there to the county level, because CCDs are statistical fictions created by the Census Bureau with no administrative authority; the authors intend to pursue this in future work.

Acknowledgments

Support for this research came from a Robert Wood Johnson Foundation grant funding the NJSOARH project at the Ralph W. Voorhees Center for Civic Engagement at the Edward J. Bloustein School of Planning and Public Policy at Rutgers University, New Brunswick. The authors would also like to thank Kathe Newman and Shiloh Deitz of the NJSOARH research team for their helpful suggestions on this project. Analyses and mapping were conducted using the following open-source software: R/RStudio (including the sf and tidyverse libraries) and QGIS.

Authors

Will B. Payne is an assistant professor in the Edward J. Bloustein School of Planning and Public Policy at Rutgers University. Lauren Nolan is a Ph.D. candidate in the Edward J. Bloustein School of Planning and Public Policy at Rutgers University. Eric Seymour is an assistant professor in the Edward J. Bloustein School of Planning and Public Policy at Rutgers University.

References

Ambrosius, Joshua D., John I. Gilderbloom, William J. Steele, Wesley L. Meares, and Dennis Keating. 2015. "Forty Years of Rent Control: Reexamining New Jersey's Moderate Local Policies after the Great Recession," *Cities* 49: 121–133. https://doi.org/10.1016/j.cities.2015.08.001.

Bailey, Kathryn T., John T. Cook, Stephanie Ettinger de Cuba, Patrick H. Casey, Mariana Chilton, Sharon M. Coleman, Diana Becker Cutts, Timothy C. Heeren, Ruth Rose-Jacobs, Maureen M. Black, and Deborah A. Frank. 2016. "Development of an Index of Subsidized Housing Availability and Its Relationship to Housing Insecurity," *Housing Policy Debate* 26 (1): 172–87. https://doi.org/10.1080/10511482.2015.1015042.

Chakraborty, Jayajit, Ashley A. McAfee, Timothy W. Collins, and Sara E. Grineski. 2021. "Exposure to Hurricane Harvey Flooding for Subsidized Housing Residents of Harris County, Texas," *Natural Hazards* 106 (3): 2185–2205. https://doi.org/10.1007/s11069-021-04536-9.

Freeman, Lance. 2012. "The Impact of Source of Income Laws on Voucher Utilization," *Housing Policy Debate* 22: 297–318. https://doi.org/10.1080/10511482.2011.648210.

Freeman, Lance, and Yunjing Li. 2014. "Do Source of Income Anti-discrimination Laws Facilitate Access to Less Disadvantaged Neighborhoods?" *Housing Studies* 29 (1): 88–107. https://doi.org/10.1080/02673037.2013.824559.

Gabbe, C.J., and Gregory Pierce. 2020. "Extreme Heat Vulnerability of Subsidized Housing Residents in California," *Housing Policy Debate* 30 (5): 843–60. https://doi.org/10.1080/10511482.2020.1768574.

Gabbe, C.J., Gregory Pierce, and Efren Oxlaj. 2020. "Subsidized Households and Wildfire Hazards in California," *Environmental Management* 66 (5): 873–83. https://doi.org/10.1007/s00267-020-01340-2.

Greenlee, A. 2019. "Redefining Rental Housing Choice in the Housing Choice Voucher Program." In *The Routledge Handbook of Housing Policy and Planning* (1st ed.), edited by Katrin B. Anacker, Mai Thi Nguyen, and David P. Varady. New York: Routledge. https://doi.org/10.4324/9781315642338.

Hamidi, Shima, Reid Ewing, and John Renne. 2016. "How Affordable Is HUD Affordable Housing?" *Housing Policy Debate* 26 (3): 437–455. https://doi.org/10.1080/10511482.2015.1123753.

Lens, Michael C. 2014. "The Impact of Housing Vouchers on Crime in U.S. Cities and Suburbs," *Urban Studies* 51 (6): 1274–1289. https://doi.org/10.1177/0042098013497407.

Lens, Michael C., and Vincent Reina. 2016. "Preserving Neighborhood Opportunity: Where Federal Housing Subsidies Expire," *Housing Policy Debate* 26 (4–5): 714–732. https://doi.org/10.1080/10511482.2016.1195422.

Metzger, Molly W. 2014. "The Reconcentration of Poverty: Patterns of Housing Voucher Use, 2000 to 2008," *Housing Policy Debate* 24 (3): 544–67. https://doi.org/10.1080/10511482.2013.876437.

Pendall, Rolf. 2000. "Local Land Use Regulation and the Chain of Exclusion," *Journal of the American Planning Association* 66 (2): 125–142. https://doi.org/10.1080/01944360008976094.

Reina, Vincent J., and Claudia Aiken. 2022. "Moving to Opportunity, or Aging in Place? The Changing Profile of Low Income and Subsidized Households and Where They Live," *Urban Affairs Review* 58 (2): 454–492. https://doi.org/10.1177/1078087420969895.

Silverman, Robert Mark, Kelly L. Patterson, and Chihuangji Wang. 2020. "Questioning Stereotypes about U.S. Site-Based Subsidized Housing," *International Journal of Housing Markets and Analysis* 14 (3): 613–31. https://doi.org/10.1108/IJHMA-05-2020-0057.

Stacy, Christina, Chris Davis, Yonah Slifkin Freemark, Lydia Lo, Graham MacDonald, Vivian Zheng, and Rolf Pendall. 2023. "Land-use reforms and housing costs: Does allowing for increased density lead to greater affordability?" *Urban Studies*. https://doi.org/10.1177/00420980231159500.

Taghavi, Lydia B. 2008. "HUD-Assisted Housing 101: Using 'A Picture of Subsidized Households: 2000,'" *Cityscape* 10 (1): 211–220. https://www.jstor.org/stable/20868649.

Tighe, J. Rosie, Megan E. Hatch, and Joseph Mead. 2017. "Source of Income Discrimination and Fair Housing Policy," *Journal of Planning Literature* 32 (1): 3–15. https://doi.org/10.1177/0885412216670603.