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 datashop@hud.gov for consideration.*

Residential Mobility and Big Data: Assessing the Validity of Consumer Reference Datasets

Alex Ramiller
Taesoo Song
Madeleine Parker
University of California, Berkeley

Karen Chapple
University of California, Berkeley
University of Toronto

Abstract

The increasing availability of privately produced longitudinal Consumer Reference Datasets (CRDs) presents substantial opportunities for housing and urban studies research, permitting the analysis of processes, including residential mobility, migration, and neighborhood change. Despite their growing popularity in academic and policy research, these datasets—which are produced by private companies for sale primarily to commercial interests—are not explicitly designed for research purposes and have not been comprehensively assessed in terms of data quality or representativeness. This article carries out a comparative analysis of the CRDs that two of the most prominent sources of consumer reference data—Data Axle and Infutor Data Solutions—produce for King County, Washington. Comparing these datasets with estimates from the American Community Survey at the county and census tract scales, this article identifies substantial limitations associated with each dataset in terms of population counts,

Abstract, continued

demographic characteristics, distribution across census tracts, and residential mobility rates. It concludes that despite notable advantages, including the ability to provide valuable and novel insights into heretofore unobserved patterns of residential mobility at a range of spatial scales, these datasets contain systematic biases. These biases may lead researchers to underestimate population counts and mobility rates for low-income households, renters, young adults, and people of color, and should, therefore, be used with caution in social, demographic, and policy research.

Introduction

Data for residential mobility research in the United States have traditionally been largely confined to longitudinal surveys, which track mobility outcomes over time but have limited temporal and spatial resolution due to small sample sizes and privacy concerns. Recently, researchers have increasingly employed a new form of data to address these issues—Consumer Reference Data. Data companies produce Consumer Reference Datasets (CRDs) by synthesizing an array of public and private datasets about the characteristics of individuals (that is, tax assessment records, utility bills, change-of-address data), designed for sale to commercial interests. Data Axle (formerly known as "Infogroup," "InfoUSA," and "RefUSA") and Infutor Data Solutions provide two of the most prominent CRDs in the United States. Data Axle and Infutor are private companies with long histories of aggregating, repackaging, and selling consumer marketing data to commercial interests. These datasets have several potential advantages over other longitudinal population datasets; they provide more detailed locational information than is generally available in traditional population data and are more deliberately structured than big "exhaust" data such as social media posts. Despite the growing use of CRDs in housing and demographic research—including the measurement of neighborhood-level population flows, assessing the effects of displacement on locational outcomes, and testing the relationship between housing market changes and population mobility—a systematic assessment of how well these datasets capture the true population composition and mobility patterns needed for evidence-based policymaking has been limited. The lack of validation of CRDs raises concerns about the reliability and equity implications of using these data to inform critical policy decisions around housing, community development, and segregation. Without a clear understanding of which populations and residential moves are being represented or misrepresented in CRDs, researchers and policymakers risk drawing misleading conclusions and perpetuating inequities. This article compares Data Axle and Infutor estimates with American Community Survey (ACS) estimates, analyzing the representativeness of the two CRDs with respect to population counts, demographic composition, and in-migration rates at both the county and tract levels for King County, Washington. This analysis provides a reference for using CRDs in housing research, pointing to limitations in respect to demographic and geographic validity.

Consumer Reference Datasets in Contemporary Research

Recent studies have employed CRDs in a variety of ways to study neighborhood-level population flows, leveraging the availability of longitudinal address-level information for large numbers of individuals. CRDs have been used to calculate neighborhood-level migration rates, determine the destinations of particular households, and assess move volumes between locations (Acolin et al., 2022; Greenlee, 2019; Pan et al., 2020; Song and Chapple, 2024). CRDs also enable the study of the long-term locational effects associated with certain types of residential moves such as evictions (Asquith, 2022; Collinson et al., 2022). These datasets have also been employed to examine the effects of specific housing policies. Diamond, McQuade, and Qian (2019) used Infutor to show that rent control in San Francisco reduces renter mobility and limits residential displacement while also reducing rental housing supply. Several studies have also explored the issue of displacement from public housing, using CRDs to trace the residential outcomes associated with public housing demolitions (Blanco, 2022; Phillips, 2020; Richardson, 2022). Mast (2019) shows that the longitudinal nature of these datasets may be leveraged to uncover the longer-term effects of residential moves, using Infutor to identify multi-year mobility chains generated when the vacancy created by one household's move is filled by another household, and so on. Finally, CRDs have been used to study the relationship between development and displacement, examining the effect of new construction on mobility rates (Asquith, Mast, and Reed, 2019; Chapple et al., 2022; Chapple and Song, 2024; Pennington, 2021).

The growing use of CRDs in social science research has been met with some limited validation. Matching a national dataset of 2008 records from Data Axle (then "InfoUSA") to the U.S. Census Bureau's internal Master Address File (MAF) on geographic fields, Kennel and Li (2009) obtained matches for roughly 85 percent of households nationally, with lower coverage rates for certain housing types such as mobile homes. Given that internal Census Bureau MAF data are not widely available, a more common validation approach compares CRD population counts with census estimates at aggregate spatial scales. Acolin, Decter-Frain, and Hall (2022) align Data Axle records with 2018 ACS estimates at three different scales (county, ZIP Code tabulation area [ZCTA], and census tract), finding that Data Axle household counts in 2018 were within 80 to 120 percent of census estimates in nearly all counties (94 percent) and most ZCTAs (82 percent). Furthermore, they establish that areas with less precise census estimates (larger margins of error), larger shares of young adults, and higher levels of employment were all associated with over- or under-estimates. Diamond, McQuade, and Qian (2019) similarly rely on census comparisons to validate the Infutor dataset, finding that it represents 44 percent of the population in San Francisco as of 1990 and 110 percent of the population as of 2000. The latter overrepresentation is attributed to the lack of recorded deaths in the Infutor dataset, suggesting that their tabulation of the dataset retains households after the point of last observation. This validation also finds that the dataset is most reliable in its representation of individuals between the ages of 30 and 49. Phillips (2020) pursues an alternative validation strategy, leveraging cases of acute residential mobility to demonstrate the value of the Infutor CRD for demographic research, identifying cases in which heightened mobility rates would be anticipated—such as the New Orleans neighborhoods affected by Hurricane Katrina, public housing demolitions and households at imminent risk of homelessness in Chicago, and gentrification in Washington, D.C. Each case exhibits elevated household migration rates relative to baselines, suggesting that CRDs could be used to capture mobility outcomes of acute mobility events. Ramani and Bloom (2021) compare Data Axle household addresses with public U.S. Postal Service change-of-address files to assess the effects of the COVID-19 pandemic on out-migration from cities. Likewise, Chapple et al. (2022) use an applied validation approach, comparing mobility rates for Bay Area households between Data Axle and Equifax credit records, and find similar patterns of mobility across socioeconomic groups within both datasets.

Although these analyses suggest that CRDs are reasonably representative of overall population and mobility dynamics, several questions remain unanswered. CRDs also contain a number of other household characteristics that could be useful in formulating demographic estimates. The current body of CRD analyses is divided on the use of auxiliary household characteristic fields, with some cautiously adopting those characteristics—for example, Chapple et al. (2022) and Greenlee (2019)—whereas others use only mobility information and assess household characteristics based on origin and destination characteristics—for example, Asquith, Mast, and Reed (2019) and Pennington (2021). Although these household characteristics are imputed using proprietary processes, useful insights may be drawn from these demographic characteristic fields. However, the quality of those fields has not yet been systematically assessed.

Data Description

This article examines CRD records between 2015 and 2019 in King County, Washington. King County is one of the most populous counties in the United States and contains geographic, racial, and economic diversity. It includes much of the Seattle metropolitan area. According to ACS 2015–2019 estimates, the county's population is 58 percent White, 26 percent Hispanic, 19 percent Asian, and 7 percent Black, has a high rate of renting (44 percent), and has a relatively mobile population, with an annual average mobility rate of 21.2 percent for adults compared with the U.S. average of 15.4 percent. This relatively more mobile population makes King County a useful case to examine the effectiveness of CRDs in accurately capturing residential mobility.

This article uses 2015–2019 ACS 5-year estimates as a reference point. It is important to acknowledge that ACS data are limited as a representation of the "true" composition of a population. ACS estimates are based on rolling averages of small random samples of the national population and, thus, have margins of error that may be quite large for small populations and geographies (Spielman, Folch, and Nagle, 2014). Nonetheless, given the widespread use of ACS estimates in a variety of population research contexts, this article treats the ACS as the best available approximation of actual population conditions. With a robust sampling and surveying strategy, the ACS serves as a reliable standard against which the authors evaluate the nonrandomly sampled CRD datasets.

For both Data Axle and Infutor, the authors select households observed in King County, Washington, at least once during the 2015–19 period. The sample for both CRDs consists of adults in households for which both demographic and address history information were available. Each dataset is converted into a complete panel structure, with observations for each individual in each year, filling any missing individual-year observations with information from prior years. Moves are identified by observing whether an individual's recorded location changes from one year to the next within the study period.

Analysis

The validity of these CRD datasets is assessed through two sets of analyses. First, this article examines whether the population count and in-migration rates of different demographics groups closely match between ACS and the CRD datasets for King County as a whole. Second, this article uses both descriptive and statistical methods to assess those same relationships at the census tract level.

Demographic Comparison

The authors begin by comparing counts and mobility rates for demographic groups in each CRD with estimates generated from IPUMS tabulations of public-use 2015–2019 ACS microdata (Ruggles et al., 2023; exhibit 1). The authors find that CRDs systematically underestimate both the population counts and relative shares of different demographic categories. Data Axle comes close to approximating the total size of the adult population, identifying 1.9 million adults compared with the roughly 2.2 million estimated in the ACS. By comparison, Infutor captures a yearly average of only 760,000 individual adults per year. Both Data Axle and Infutor substantially overestimate the share of the population that is White. This issue is particularly severe for Infutor, which treats White as a "default" for the category, thus categorizing more than 97 percent of households as White. Both CRDs also overestimate the share of homeowners and older individuals, which likely reflects the data collection techniques employed to construct these datasets, relying on administrative sources (such as property tax records) from which younger households and renters are more likely to be absent. Finally, Data Axle appears to overestimate the share of the population earning low incomes (less than $35,000) and underestimate the share of the population earning relatively higher incomes (greater than $70,000), whereas Infutor underestimates the low-income population and overestimates the higher-income population.

Exhibit 1

Demographic Characteristics of Adults in King County per ACS PUMS, Data Axle, and Infutor (2015–19) (1 of 2)

	Population Count (Share of Population)			In-Migration Rate		
Variable	ACS	Data Axle	Infutor	ACS	Data Axle	Infutor
Asian	396,401 (18.0%)	238,206 (13.2%)	5,689 (1.4%)	17.5%	7.0%	19.0%
Black	136,634 (6.2%)	30,902 (1.7%)	1,027 (0.2%)	20.4%	7.5%	18.7%
Hispanic	212,135 (9.7%)	111,463 (6.2%)	4,371 (1.1%)	19.6%	7.6%	19.4%
Other	141,192 (6.4%)	2,017 (0.1%)	-	21.1%	7.5%	-
White	1,310,029 (59.6%)	1,416,914 (78.7%)	403,418 (97.3%)	16.3%	6.9%	21.8%
Unknown Race	-	119,781	343,462	-	7.5%	31.3%
Own	1,317,032 (61.0%)	1,329,709 (69.3%)	246,457 (85.1%)	8.5%	4.2%	19.2%
Rent	841,209 (39.0%)	589,573 (30.7%)	43,063 (14.9%)	29.8%	14.2%	21.4%

Exhibit 1

Demographic Characteristics of Adults in King County per ACS PUMS, Data Axle, and Infutor (2015–19) (2 of 2)

Variable	Population Count (Share of Population)			In-Migration Rate		
	ACS	Data Axle	Infutor	ACS	Data Axle	Infutor
Unknown Tenure	-	-	468,447	54.6%	-	30.2%
Female	1,095,021 (49.9%)	913,743 (52.1%)	189,401 (50.3%)	16.9%	7.0%	21.4%
Male	1,101,370 (50.1%)	839,594 (47.9%)	187,436 (49.7%)	17.8%	7.0%	21.5%
Unknown Gender	-	165,945	-	-	6.8%	30.7%
< 25	179,744 (10.3%)	60,671 (3.2%)	-	38.6%	20.0%	-
25–44	725,544 (41.5%)	576,134 (30.0%)	104,120 (34.8%)	25.1%	12.0%	20.2%
45–64	558,841 (32.0%)	833,373 (43.4%)	109,652 (36.7%)	9.3%	4.8%	18.5%
65+	284,107 (16.3%)	449,104 (23.4%)	85,201 (28.5%)	6.7%	4.2%	17.0%
Unknown Age	-	-	458,993	-	-	30.9%
Low Income ($0–$34,999)	267,501 (12.2%)	413,457 (21.5%)	35,272 (4.7%)	20.6%	10.0%	19.6%
Moderate ($35,000–$54,999)	214,301 (9.8%)	268,825 (14.0%)	54,861 (7.2%)	19.4%	8.4%	20.8%
Middle ($55,000–$74,999)	167,767 (7.6%)	181,722 (9.5%)	82,700 (10.9%)	18.5%	7.6%	22.8%
High ($75,000+)	1,546,832 (70.4%)	1,055,278 (55.0%)	239,658 (31.6%)	16.4%	5.5%	22.0%
Unknown Income	-	-	345,476	-	-	31.2%
Total Population	2,196,391	1,919,282	757,966	2,138,180	1,691,641	-

- = not available. ACS = American Community Survey.

Notes: Population shares are from individuals for which demographic attributes are identified. All differences in population shares and mobility rates between ACS and Consumer Reference Datasets are statistically significant at the 99.9 percent level.

Sources: Author tabulations of 2019 ACS 5-year estimates; Data Axle; Infutor

The authors also find that each CRD fails to accurately estimate the share of adults that moved between 2015 and 2019. The average annual in-migration rate for all adults according to the ACS was 17.4 percent compared with only 7 percent of Data Axle individuals and 26.2 percent of Infutor individuals. This finding suggests that CRDs are not effective at capturing when households actually move. The authors hypothesize that Data Axle may underestimate mobility rates either because it continues to record individuals in the same location after they have already moved or because those households disappear from the record when they move. This hypothesis is consistent with the manner in which CRD data are collected, which draw from sources such as property tax records that may take a year or more to find updated information and that may take even longer for other households such as renters, for whom such records are not available.

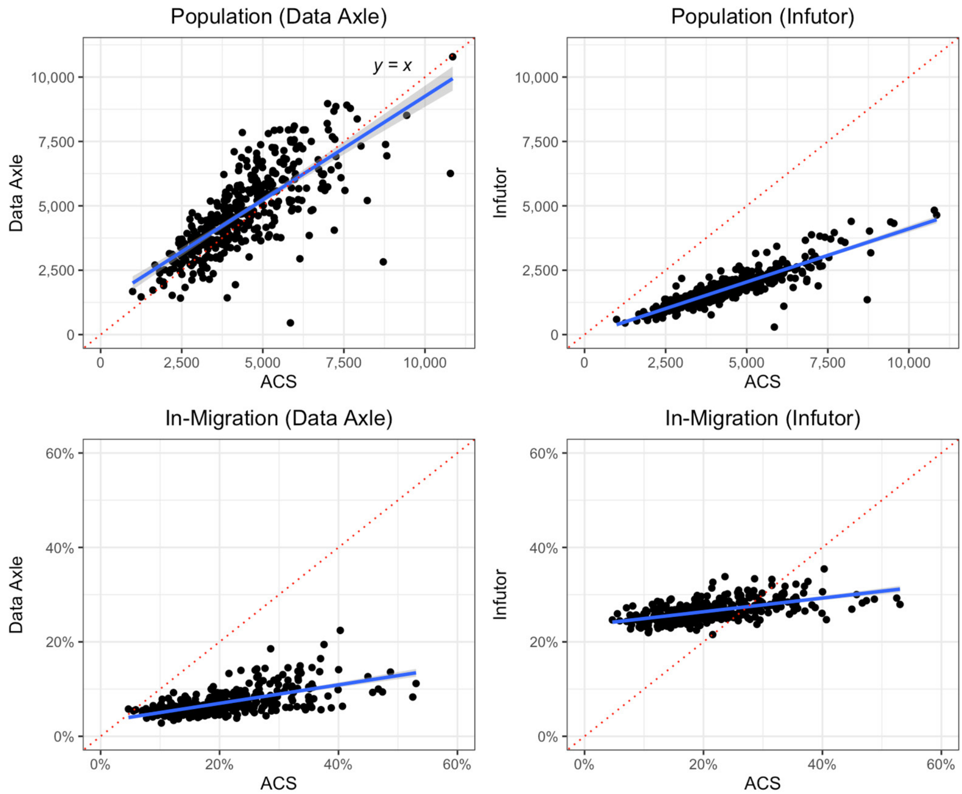
Although Infutor appears to more closely approximate the magnitudes of ACS estimates, Data Axle performs better in capturing the relative differences between demographic groups. For example, Data Axle reflects the substantially higher mobility rate for renters (14.2 percent) than owners (4.2 percent). Although these numbers are roughly one-half the mobility rates reported in the ACS, they reflect the much greater likelihood of mobility for renters. Similarly, Data Axle captures the differential mobility rates across age groups, with a much higher likelihood of mobility among younger households than older households. Data Axle also reflects the relative mobility rates across different income categories, with the highest rates of mobility for low-income households and the lowest for higher-income households. Data Axle shows the weakest relationship to ACS estimates for race. Although it does find lower mobility rates for White and Asian households vis-à-vis Black and Hispanic households, differences between these groups are close to zero. Infutor largely fails to distinguish mobility rates across demographic groups, consistently estimating rates around 20 percent regardless of demographic characteristics. Although the Infutor mobility rate is slightly higher for renters than for homeowners, the difference in these estimates (19.2 versus 21.4 percent) is too small to reflect the actual difference in mobility rates of these groups. Infutor also incorrectly finds the lowest mobility rates associated with low-income households and Black households—the opposite of both ACS and Data Axle.

Neighborhood Comparison

The top panels of exhibit 2 demonstrate the extent to which CRDs represent the census tract-level adult population and in-migration rates reflected in ACS 5-year estimates from 2015 to 2019, with each point representing a single tract within King County. Data Axle roughly approximates census tract populations, with a fitted curve closely aligned with the identity function but with fairly broad dispersion. Although Infutor population counts are clustered more tightly around the fitted curve, they are also far less accurate. Infutor substantially underestimates census tract population across all tracts, although it maintains a linear relationship with true population counts. Both CRDs perform worse with respect to mobility rates, significantly underestimating the share of the population that moved into King County census tracts between 2015 and 2019 (bottom panels of exhibit 2). Data Axle exhibits a positive correlation with ACS in-migration rates but underestimates mobility rates for nearly every census tract. By contrast, although Infutor both under- and over-estimates mobility rates across different census tracts, it is primarily because of a higher baseline mobility rate.

Exhibit 2

Comparison of ACS Population Estimates and Mobility Rate Estimates With 5-Year Population Averages for Data Axle and Infutor



ACS = American Community Survey.
Sources: Author tabulations of 2019 ACS 5-Year estimates; Data Axle; Infutor

The authors use linear regression models with robust standard errors to assess neighborhood characteristics associated with under- or overestimating population counts or mobility.[1] Tract characteristics employed in the model include a control for the ACS population or mobility rate and ACS variables, including median household income, median rent, median home value, populations between 18 to 24 years of age (percent), owner-occupants (percent), one-person households (percent), non-Hispanic White individuals (percent), population density (log), vacant units (percent), foreign born individuals (percent), and housing built after 2010 (percent). The variables were then standardized and mean-centered.

Modeling results suggest that several neighborhood characteristics are significantly related to over- and underestimating neighborhood population and in-migration rates (exhibits 3 and 4). Both CRDs overestimated population in neighborhoods with higher median rents and underestimated

[1] The authors follow Acolin, Decter-Frain, and Hall (2022) in comparing the performance of linear, ridge, and lasso regression models using root mean squared error (RMSE), comparing dependent variable values and predicted values. The authors find that linear regression yields the lowest RMSEs across all models.

population in neighborhoods with a higher share of young adults (aged 18 to 24). However, differences were also present in these relationships between the Data Axle and Infutor datasets. Data Axle underestimated population in higher-income neighborhoods but also overestimated population in neighborhoods with higher ownership rates and home values, whereas Infutor did not. Infutor overestimated population in majority White neighborhoods and neighborhoods with higher shares of one-person households, whereas Data Axle did not.

Exhibit 3

Linear Model Results, Comparing 2015–19 American Community Survey Estimates With 2015–19 Consumer Reference Dataset Averages

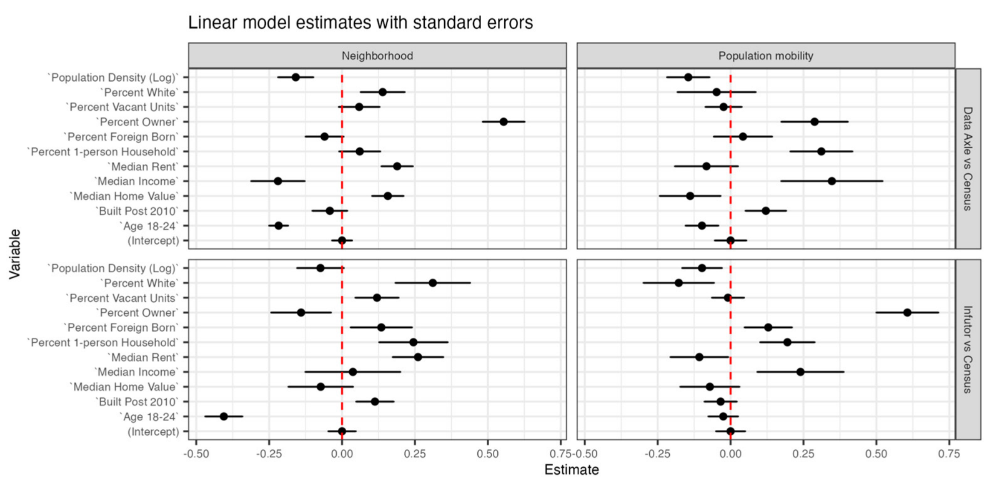
Variable	Population (Data Axle vs. Census)	Population (Infutor vs. Census)	Mobility Rate (Data Axle vs. Census)	Mobility Rate (Infutor vs. Census)
(Intercept)	0.000	0.000	0.000	0.000
Median Income	– 0.220*	0.037	0.347*	0.240
Median Rent	0.189***	0.260**	– 0.083	– 0.107
Median Home Value	0.157**	– 0.073	– 0.139	– 0.071
Age 18–24 (%)	– 0.218***	– 0.405***	– 0.098	– 0.025
Owner (%)	0.554***	– 0.141	0.288*	0.606***
White (%)	0.139	0.311*	– 0.048	– 0.178
1-Person Household (%)	0.061	0.245*	0.311**	0.195*
Population Density (Log)	– 0.159**	– 0.074	– 0.145*	– 0.098
Vacant Units (%)	0.059	0.120	– 0.024	– 0.009
Foreign Born (%)	– 0.060	0.135	0.042	0.129
Built Post 2010 (%)	– 0.042	0.113	0.121	– 0.034
R-squared	0.672	0.371	0.155	0.242

** = p < 0.05. ** = p < 0.01. *** = p < 0.001.*

Sources: Author tabulations of 2019 American Community Survey 5-year estimates; Data Axle; Infutor

Exhibit 4

Linear Model Results, Comparing 2015–19 American Community Survey Estimates With 2015–19 Consumer Reference Dataset Averages



Sources: Author tabulations of 2019 American Community Survey 5-year estimates; Data Axle; Infutor

In terms of mobility rates, the authors observe that neighborhoods with a greater share of homeowners have relatively higher than expected mobility rates across both datasets, as do neighborhoods with a larger share of one-person households and vacancies. On the other hand, a larger share of young adults in a neighborhood predicts underrepresentation of mobility rates for Data Axle, and the share of housing units built in the past decade in a neighborhood predicts underrepresentation for Infutor. In short, in addition to overestimating the volume of homeowners and the population in high-ownership neighborhoods, these datasets also systematically overestimate mobility rates in high-ownership neighborhoods and, in the case of Infutor, overestimates the mobility rates of homeowners. This finding may indicate that CRDs are more likely to capture renters in high-ownership neighborhoods and classify them as homeowners, or it may be indicative of a bias toward homeowners that move more frequently. In any case, given that tenure is a key predictor of residential mobility (Rossi, 1955), these findings raise serious concerns about the ability of CRDs to accurately predict mobility rates at either an individual or neighborhood level.

Conclusion

Both CRDs exhibit clear limitations in terms of external validity, indicating that neither should be treated as equivalent to a complete population census. Each dataset omits or underestimates the share of certain populations—particularly individuals who are lower income, younger, people of color, or renters. Despite the apparent advantages CRDs provide in terms of temporal and spatial detail, they are unlikely to serve as representative alternatives to census microdata. These limitations also pose challenges for the measurement of residential mobility. If CRDs accurately represent only certain neighborhoods and population segments, the calculation of migration rates for any given census tract or demographic group is likely misleading. Given that vulnerable populations such as low-income households, renters, young adults, and people of color are more likely to be systematically undercounted within these datasets, policy research that relies on these datasets risks obscuring those populations and generating erroneous conclusions that could reinforce existing inequalities, whether by underestimating the scale of residential displacement or by undercounting the size of vulnerable populations that could benefit from additional resources. Therefore, the findings suggest that serious reservations should be considered when using these data for studying residential mobility or for producing sociodemographic estimates of an area using raw counts. However, these datasets may still be useful for tracking individual residential moves because they provide a rich longitudinal picture of household locations at multiple points in time. In addition, Data Axle may specifically be useful for comparing the relative differences in mobility rates between demographic groups.

Future research should explore potential strategies to better optimize CRDs for research on residential mobility and demographic change. Notwithstanding the limitations described here, CRDs provide a potentially valuable source of information about mobility patterns, enabling nowcasting of population change and detailed analysis of mobility outcomes in response to acute events (Acolin, Decter-Frain, and Hall, 2022; Phillips, 2020). To address the limitations of these datasets, future research could explore strategies such as population weighting, using existing information from census data regarding demographic characteristics, and population distribution to reweight CRD observations. Without adjustment strategies, any use of either imputed fields or attempts to aggregate these data for demographic research should be approached with an abundance of caution. The authors' conclusions indicate that it is important for policymakers and researchers using CRDs for social, demographic, and policy research to exercise caution and triangulate them with other data sources to account for biases. Greater transparency in the data collection and imputation methods used to construct these datasets is also essential to ensure that these data sources accurately describe actual population characteristics and mobility patterns. Without these safeguards in place, the use of CRDs has the potential to contribute to misguided policy decisions.

Acknowledgments

The authors would like to thank those who provided feedback and guidance on this research, including Arthur Acolin, Andrew Greenlee, and Tim Thomas, as well as audiences at the 2023 UC Berkeley D-Lab Fellows Talk and the 2024 Urban Affairs Association conference.

Authors

Alex Ramiller is a Ph.D. candidate in the Department of City and Regional Planning at the University of California, Berkeley. Taesoo Song is a Ph.D. candidate in the Department of City and Regional Planning at the University of California, Berkeley. Madeleine Parker received her PhD in City and Regional Planning from the University of California, Berkeley. Karen Chapple is director of the School of Cities at the University of Toronto, professor of Geography and Planning at the University of Toronto, and professor emerita of City and Regional Planning at the University of California, Berkeley.

References

Acolin, Arthur, Kyle Crowder, Ari Decter-Frain, Anjum Hajat, and Matt Hall. 2022. "Gentrification, Mobility, and Exposure to Contextual Determinants of Health," *Housing Policy Debate* 33 (1): 194–223. https://doi.org/10.1080/10511482.2022.2099937.

Acolin, Arthur, Ari Decter-Frain, and Matt Hall. 2022. "Small-Area Estimates From Consumer Trace Data," *Demographic Research* 47 (27): 843–882. https://doi.org/10.4054/DemRes.2022.47.27.

Asquith, Brian. 2022. *The Effects of an Ellis Act Eviction on Neighborhood Socioeconomic Status*. Kalamazoo, MI: Upjohn Institute. https://doi.org/10.2139/ssrn.4283770.

Asquith, Brian, Evan Mast, and Davin Reed. 2019. Supply Shock Versus Demand Shock: The Local Effects of New Housing in Low-Income Areas. Working paper 316. Kalamazoo, MI: Upjohn Institute.

Blanco, Hector. 2022. "Pecuniary Effects of Public Housing Demolitions: Evidence From Chicago," *Regional Science and Urban Economics* 98: 103847. https://doi.org/10.1016/j.regsciurbeco.2022.103847.

Chapple, Karen, Julia Greenberg, Jackelyn Hwang, Jae Sik Jeon, Bina Shrimali, and Iris Zhang. 2022. Housing Market Interventions and Residential Mobility in the San Francisco Bay Area. Community Development working paper 2022–01. Federal Reserve Bank of San Francisco. https://ideas.repec.org/p/fip/fedfcw/93849.html.

Chapple, Karen, and Taesoo Song. 2024. "Can New Housing Supply Mitigate Displacement and Exclusion? Evidence from Los Angeles and San Francisco," *Journal of the American Planning Association*: 1–15. https://doi.org/10.1080/01944363.2024.2319293.

Collinson, Robert, John Humphries, Nicholas Mader, Davin Reed, Daniel Tannenbaum, and Winnie van Dijk. 2022. Eviction and Poverty in American Cities. Working paper 2022–24. Human Capital and Economic Opportunity Working Group. https://econpapers.repec.org/paper/hkawpaper/2022-24.htm.

Diamond, Rebecca, Tim McQuade, and Franklin Qian. 2019. "The Effects of Rent Control Expansion on Tenants, Landlords, and Inequality: Evidence from San Francisco," *American Economic Review* 109 (9): 3365–3394. https://doi.org/10.1257/aer.20181289.

Greenlee, Andrew J. 2019. "Assessing the Intersection of Neighborhood Change and Residential Mobility Pathways for the Chicago Metropolitan Area (2006–2015)," *Housing Policy Debate* 29 (1): 186–212. https://doi.org/10.1080/10511482.2018.1476898.

Kennel, Timothy L., and Mei Li. 2009. "Content and Coverage Quality of a Commercial Address List as a National Sampling Frame for Household Surveys." In *Proceedings of the Joint Statistical Meetings*. http://www.asasrms.org/Proceedings/y2009/Files/304010.pdf.

Mast, Evan. 2019. *The Effect of New Luxury Housing on Regional Housing Affordability*. Kalamazoo, MI: Upjohn Institute.

Pan, Haozhi, Si Chen, Yizhao Gao, Brian Deal, and Jinfang Liu. 2020. "An Urban Informatics Approach to Understanding Residential Mobility in Metro Chicago," *Environment and Planning B: Urban Analytics and City Science* 47 (8): 1456–1473. https://doi.org/10.1177/2399808320924437.

Pennington, Kate. 2021. *Does Building New Housing Cause Displacement?: The Supply and Demand Effects of Construction in San Francisco*. Berkeley: University of California. https://doi.org/10.2139/ssrn.3867764.

Phillips, David C. 2020. "Measuring Housing Stability With Consumer Reference Data," *Demography* 57 (4): 1323–1344. https://doi.org/10.1007/s13524-020-00893-5.

Ramani, Arjun, and Nicholas Bloom. 2021. The Donut Effect of Covid-19 on Cities. NBER Working Paper No. 28876. Cambridge, MA: National Bureau of Economic Research. https://doi.org/10.3386/w28876.

Richardson, Edward C. 2022. The Race to Displace: The Long-Run Causal Effect of the 1996 Olympic Legacy Program on Residential Locations in Atlanta, GA. Bachelor's thesis. Harvard College.

Rossi, Peter H. 1955. *Why Families Move: A Study in the Social Psychology of Urban Residential Mobility*. Los Angeles, CA: Free Press.

Ruggles, Steven, Sarah Flood, Matthew Sobek, Danika Brockman, Grace Cooper, Stephanie Richards, and Megan Schouweiler. 2023. "IPUMS USA: Version 13.0 [dataset]." https://doi.org/10.18128/D010.V13.0.

Song, Taesoo, and Karen Chapple. 2024. "Does Gentrification Constrain Housing Markets for Low-Income Households? Evidence From Household Residential Mobility in the New York and San Francisco Metropolitan Areas," *Urban Studies* 61 (13): 2603–2622. https://doi.org/10.1177/00420980241244699.

Spielman, Seth E., David Folch, and Nicholas Nagle. 2014. "Patterns and Causes of Uncertainty in the American Community Survey," *Applied Geography* 46: 147–157. https://doi.org/10.1016/j.apgeog.2013.11.002.