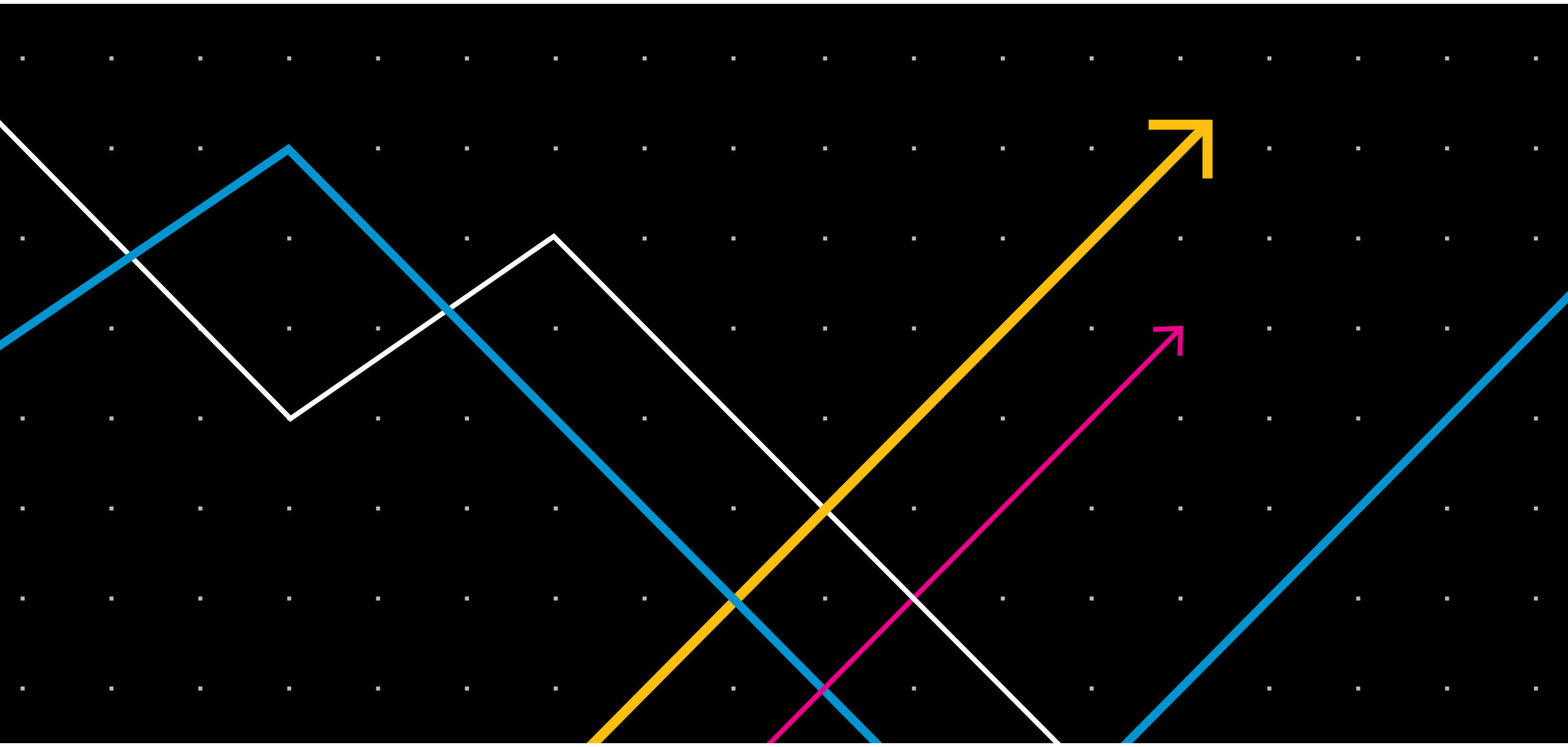


RESEARCH REPORT

Unveiling Disparities in Eviction

A Novel Approach to Demographic Imputation in Cleveland

Alena Stern
URBAN INSTITUTE

Manuel Alcalá Kovalski
URBAN INSTITUTE

Leah Hendey
URBAN INSTITUTE

Elizabeth Burton
URBAN INSTITUTE

Sandra Ambrozy
URBAN INSTITUTE

Carlos Manjarrez
INSTITUTE FOR TECHNOLOGY LAW AND POLICY
GEORGETOWN UNIVERSITY LAW CENTER

January 2025

URBAN INSTITUTE · ELEVATE · THE · DEBATE

ABOUT THE URBAN INSTITUTE

The Urban Institute is a nonprofit research organization that provides data and evidence to help advance upward mobility and equity. We are a trusted source for changemakers who seek to strengthen decisionmaking, create inclusive economic growth, and improve the well-being of families and communities. For more than 50 years, Urban has delivered facts that inspire solutions—and this remains our charge today.

Contents

Acknowledgments

This report was funded by the Salesforce Foundation and the Mellon Foundation. In addition, donations to the Urban Institute from Sandra Ambrozy and a matching gift from the Kresge Foundation along with a gift from the Gerald and Sandra Ambrozy Fund held at the Community Foundation of Southeast Michigan were used to support this project. We are grateful to them and to all our funders, who make it possible for Urban to advance its mission.

The views expressed are those of the authors and should not be attributed to the Urban Institute, its trustees, or its funders. Funders do not determine research findings or the insights and recommendations of Urban experts. Further information on the Urban Institute's funding principles is available at urban.org/fundingprinciples.

We would like to thank our project advisory group for their valuable insights into the civil legal system and knowledge of evictions in Cleveland: Colleen Cotter of The Legal Aid Society of Cleveland, Claudia Coulton of Case Western Reserve University, and Eduardo Gonzalez of the American Academy of Arts & Sciences. A special thanks to Joe Andre, Michael Henderson, and Stephen Steh at the Center on Poverty and Development at Case Western Reserve University for their assistance with validating the imputation results. We would like to thank the Legal Services Corporation for providing the court data used in this research. Thanks to Cory McCartan who provided valuable methodological insight on the BIRDiE methodology. Thank you also to Peter Tatian and Celina Barrios-Millner for technical review, Judah Axelrod for code and methodological review, Rodrigo Garcia for research assistance and Alexandria Dallman and Managed Editing for editing.

Executive Summary

Each year, millions of people engage with the civil legal system for a range of matters, including evictions, traffic violations, small claims, probate, debt collection, divorce, and child custody and support. Civil court case outcomes can have significant impacts on the lives of these individuals, their families, and their communities. Unlike criminal cases—where various agencies and organizations at the federal, state, and local levels collect demographic information about defendants—there is no comparable data collection practice for civil legal matters. State courts are not required to report demographic information to other agencies or authorities; therefore, they overwhelmingly fail to do so. Lack of demographic information on individuals who engage with the civil legal system makes it extremely difficult to identify disparities in who engages with the system and the case outcomes. Consequently, it is difficult for advocates to build an evidence base to support reforms that address disparities.

This research used several imputation methodologies to predict the race/ethnicity of individuals who faced an eviction filing in the Cleveland Municipal Court - Housing Division (Cleveland Housing Court) between January 2016 and June 2023. We tested an extension of the Bayesian Improved Surname Geocoding (BISG) method to address measurement error in census datasets used for imputation as well as the Bayesian Instrumental Regression for Disparity Estimation (BIRDiE) method to improve estimates of racial/ethnic disparities using imputed data. We also tested multiple gender imputation approaches using data from social media profiles and the Social Security Administration. We had the rare opportunity to validate our imputation methodology using eviction records linked to the Case Western Reserve University Center on Poverty and Community Development's Child and Household Integrated Longitudinal Data (CHILD) system. The validation found that our methodology correctly imputed gender for 97.1 percent, and race/ethnicity for 83.8 percent of the linked observations. We believe that our implementation and validation of this range of imputation approaches will advance the field of demographic imputation research.

We then analyzed the imputed data for racial/ethnic and gender disparities and found that Cleveland evictions mirror and add to existing inequities in our society. People of color and women disproportionately face eviction, and people without legal representation are more often evicted and more likely to receive financial judgments. Further, we found widespread disproportionate eviction filings against Black renters. In 90 percent of the City of Cleveland's block groups with 5 or more eviction filings from 2016 to 2023, the share of tenants facing eviction who were Black was greater than the share of renter households that were Black. Although our findings indicate clear racial/ethnic and gender disparities regarding who enters the eviction court system, our analysis of selected eviction case outcomes did not reveal clear racial/ethnic or gender disparities after an individual entered the system. However, data limitations restricted us to analyzing a subset of outcomes, including whether defendants (tenants) or plaintiffs (landlords) had legal representation; whether a writ of eviction was filed or issued, a move was scheduled, or a second cause hearing was held; whether a financial judgment occurred and the amount; and whether wage garnishment occurred. Future analyses should explore additional outcomes of interest, such as causes of eviction, back rent owed, and others.

Our analysis results will help to create an evidence base that will support systemic reforms aimed at reducing the harmful and disparate impacts of the eviction court system. In our discussion, we highlight several potential state and court reforms for addressing eviction filing disparities.

Unveiling Disparities in Eviction

Background

To understand opportunities for addressing systemic racism, we must look at local institutions' habits and their relationships with communities. With the understanding that racism manifests in more ways than the behavior of errant individuals, we explore how local institutions accept, enforce, and perpetuate unequal treatment of residents. Through this inquiry, we open a new set of questions about the patterns and practices of local schools, health care facilities, police forces, and our courts.

The administrative data collected by local institutions to record daily operations and activities offer many opportunities to investigate important public policy questions. School records on performance, attendance, and discipline tell a story about how Black and Latinx students may experience the same institution in dramatically different ways than their white counterparts. Records of police stops and arrests could reveal how law enforcement practices may systematically differ from one neighborhood to the next. Public service utilization data could identify underrepresented communities for targeted outreach to ensure that public program benefits are equitably shared.

Local, state, and national institutions across the country have come to recognize the value of making administrative data available for research. The National Neighborhood Indicators Partnership, a network coordinated by Urban Institute, is an excellent example of how local institutions and community data organizations can form partnerships that leverage administrative data to address critical policy issues, such as childhood lead exposure, housing instability, and more (Coulton et al. 2023; Hendey, Burton, and Pettit 2024). However, civil courts have resisted the call to make administrative data accessible for secondary analysis.

Civil Courts and Data

In 2023, nearly 50 million civil court cases took place in the United States.[1] These cases covered a range of matters, including evictions, traffic violations, small claims, probate, debt collection, divorce, and child custody and support. Similar to other local government functions, resourcing, governance, and administration are decentralized in court systems. Significant variations in local jurisdictions, the ratio of courts and judges to people, court funding, and the training and experience needed to oversee court proceedings, can occur across courts within a state.[2]

Courts' decentralized nature and relative independence have allowed them to avoid the pressure that executive branch agencies have experienced to make their administrative data accessible for public scrutiny. The National Center for State Courts first released recommendations to make data open to the public (along with meaningful guidance and resources) as late as 2020. In 2021, scholars associated with the American Academy of Arts and Sciences released a report decrying the limited access to court data. Additionally, about half of states, including Ohio, do not have unified court systems, which makes it more likely there would be wide variation across local courts in data collection and reporting standards and processes.[3]

The lack of data transparency in court proceedings and judgments severely limits our understanding of how these institutions affect the lives of everyday Americans and how to remedy inequitable outcomes. Courts are the final stop for people to defend themselves against wrongful evictions or predatory lending schemes, where people may have their wages garnished for years when they lose a debt case, or face heavy fines and fees for civil court judgments. Court decisions have a profound impact on the lives of the people involved in each case, and for vulnerable individuals and families, a judgment can push them deeper into poverty. Given the profound impact of the civil court system on economic outcomes for vulnerable communities, decisions about how courts are administered will also have significant ramifications and may exacerbate inequities. This potential for harm was underscored most recently when courts closed during the COVID-19 pandemic. Few people had the literacy, access, or capacity to shift into interacting remotely. Although some changes created more access, such as remote hearings authorization, document automation adoption, and electronic filing expansion, other changes exacerbated the very justice gaps they were meant to address. For example, when some courts shifted to solely remote or virtual court hearings to help parents care for children, reduce travel time, and limit exposure to illness, they did not consider that many communities lacked internet service and devices that would allow people to remotely attend hearings.

Measuring Racial Disparities in Civil Court Data

Research on racial disparities in the civil court system has been limited because courts lack the capacity and standards to collect data on race or ethnicity (O'Hara and Straus 2024, p. 3). Even where courts have the capacity to collect these data, most do not do so, either because data collection is not required, or they lack the will to collect data. In the few cases where courts collect demographic information, they do so only for the subset of individuals who appear in court; additionally, race determination is often based on the court clerk's observation, which is known to be unreliable (O'Hara and Straus 2024). The resulting data gap limits researchers' ability to investigate potential relationships between racial or ethnic characteristics and involvement with or outcomes within the civil court system. However, researchers have developed methods to study racial disparities within the civil court system despite the absence of a data collection standard. One method combines case-level court data with tract- or block-level race and Hispanic ethnicity demographic data. Studies using this method have identified racial disparities, such as Black renters experiencing higher rates of eviction filing and evictions than white renters (Medina et al. 2020). Graetz and colleagues (2023) linked eviction court records to confidential 2006 through 2015 American Community Survey microdata for the year preceding the eviction. The resulting dataset linked households with evictions filed against them with demographic data for the household members threatened with eviction, which allowed researchers to analyze the relationship between civil court involvement and a range of demographic characteristics made available through the American Community Survey. The results showed that, "Non-Hispanic Black renters were the only racial/ethnic group overrepresented in eviction filings and judgments" (Graetz et al. 2023, p. 3).

Some researchers, including researchers at multiple National Neighborhood Indicators Partnership organizations, have matched tenants' eviction case records to other administrative data sources containing demographic information. For example, García-Cobián Richter and colleagues (2019) used this matching method in the Cleveland Eviction Study. By linking eviction records data to public assistance records (Medicaid, Temporary Assistance to Needy Families, and the Supplementary Nutrition Assistance Program) they matched data for nearly 20,000 households with low incomes facing eviction to analyze household demographics and outcomes, and found that 77 percent of the matched tenants identified as Black and 78 percent as female; additionally, 60 percent had children in the household (García-Cobián Richter et al. 2019). They also found that tenants facing eviction resulting in move orders had higher rates of residential mobility, higher homeless shelter use, and higher school absenteeism for the children in those households. DataWorks NC (2023) linked case records to voter registration records containing self-identified race/ethnicity data in Durham County, North Carolina, and found that Black tenants were disproportionately impacted by eviction. Collinson and Reed (2019) began their work at the Furman Center at New York University and matched New York City eviction records to public assistance records from the city's Human Resources Administration.

Lack of access to demographic data from administrative sources or to a secure Federal Statistical Research Data Center to enable data linkage limits researchers' ability to analyze relationships between court involvement and racial or ethnic characteristics. Imputing demographic characteristics onto civil court records using modeling algorithms provides an alternative. The most widely used method for imputing race and ethnicity is Bayesian Improved Surname Geocoding (BISG), developed by the RAND Corporation for the US Department of Health and Human Services.[4] The method is also used by the Equal Employment Opportunity Commission and the Consumer Financial Protection Bureau (Harris 2020, CFPB 2014). BISG combines administrative surname data and residential address data with publicly available census data on surnames by race and ethnicity and with block group racial and ethnic composition in a calibrated Bayesian framework to estimate probabilities by race and ethnicity for each record in a dataset. Government agencies have used BISG-imputed data for numerous racial equity analyses, such as measuring racial and ethnic differences in voluntary disenrollment from Medicare and conducting fair lending analysis of non-mortgage lending products (CFPB 2014; Martino et. al. 2020). A 2009 evaluation of BISG found high accuracy for the four largest racial and ethnicity groups in the US. Concordance statistics for the BISG methodology are 0.94 for Asian/Pacific Islander, 0.93 for Black, 0.94 for Hispanic, and 0.93 for white, where 1.0 is perfect predictiveness (Elliott et al. 2009).

Hepburn, Louis, and Desmond (2020) of the Eviction Lab imputed gender, race, and ethnicity using BISG on 2012 to 2016 eviction data from 39 states to examine disparities in the eviction filing rate, eviction rate, and serial eviction filing rate. They found that Black renters were disproportionately affected by eviction filings and judgments. The authors also imputed gender using R packages *gender* and *generizeR*, and found that evictions were filed against Black and Latinx women more often than Black and Latinx men.

A recent study by O'Hara and Straus (2024) assessed the feasibility of matching individual eviction records from five jurisdictions to self-reported race and ethnicity data from the decennial censuses or the American Community Survey from the US Census Bureau's Federal Statistical Research Data Center and comparing the results of race and ethnicity imputation methods with direct linking. They used a BISG method for imputation but used census tract for the geography to narrow the racial and ethnic probabilities. For records with self-identified race and ethnicity, imputed race matched self-identified race 50 to 83 percent of the time, depending on the jurisdiction. Imputed ethnicity was more accurate, matching self-reported ethnicity 82 to 97 percent, depending on the jurisdiction.

Eviction Process in Cleveland

Eviction is a complicated process to navigate. The process in Cleveland is no exception. According to Urban and colleagues (2019), the process starts with the landlord providing a required "notice to vacate" to the tenant. Legal reasons for initiating this process include nonpayment of rent, lease violations, illegal activity, health and safety codes noncompliance, denying the landlord access after the required 24-hour notice, and the lease term ending. However, landlords may also try to evict tenants for reasons that are not legal, including discrimination and retaliation for reporting health and safety issues. In Cleveland, if the tenant moves out or resolves the lease violations within three days, the process stops, and the eviction does not move forward. If the situation is not resolved within three days, the landlord is permitted to file an eviction complaint and begin the civil court process. After the complaint is filed, a first hearing is scheduled for three weeks after the filing in Cleveland's special jurisdiction court with one judge and a number of magistrates, the Cleveland Housing Court. At this "first cause hearing," the judge determines whether there is a lease violation; the tenant can raise defenses, including that they have come back into compliance with the lease.

At the first hearing, a case could be dismissed, sent to mediation, continued, or a judgment could be issued. In cases where a judgment is issued in favor of the landlord, a "writ of restitution" is granted (a court order that allows a landlord to take possession of a rental property from a tenant) and the landlord can schedule a tenant move out or physical eviction, giving the tenant seven to ten days to vacate the unit. In cases where the landlord is seeking damages from the tenant, including unpaid rent, a second cause hearing is held.

Methodology

This section describes the data sources, data cleaning, imputation methodologies, and validation process.

Data Sources

We entered into a data sharing agreement with the Legal Services Corporation to receive data on eviction cases for the Cleveland Housing Court collected through the Legal Services Corporation's Civil Court Data Initiative (CCDI).[5] The Cleveland Housing Court CCDI data contain 48,399 unique residential and nonresidential eviction filings from 2016 to 2023. However, as described in the data cleaning section, we removed nonresidential eviction filings from our data, resulting in 45,968 filings for analysis. The 2023 data are partial because we received the data transfer in June 2023. The data were separated into multiple files linked by unique case identifiers. We used the following data files for analysis:

- **Filings** contains case filing date, case closing date (if applicable), case status, plaintiff first and last name, defendant first and last name, and plaintiff and defendant representation.
- **Parties** contains plaintiff and defendant addresses used to assign defendants to a census tract.

The CCDI data do not include any demographic information since courts do not collect this data. Our imputation methodology leveraged the defendant name and address fields to predict race/ethnicity and gender.

To supplement the CCDI data, we scraped data for 11,704 cases for 2019 and 2021 from the Cleveland Housing Court's Odyssey Portal. This portal, accessible through the Smart Search tool, provides a public interface that allowed us to retrieve detailed case information, including events associated with each case. To automate data extraction, we used Python's *Selenium* browser automation tool and the *requests_html* package, both of which allowed us to access dynamic web content. We scraped the data by programmatically searching the portal using the CCDI data case numbers. Our searches returned results for 100 percent of the case numbers. We gathered detailed case event comments from the portal, which allowed us to extract other outcomes, such as financial judgment amounts and wage garnishments, through text analysis.

Our approach involved automating several steps for each case. We entered a unique case ID, navigated to the case summary page, and extracted the contents of the "event comments" table, which provided a structured record of all events related to the case, including event names and detailed descriptions that often included relevant financial information and case outcomes.[6]

Data Cleaning

To prepare the defendant names for imputation, we applied a series of steps to ensure data consistency and to specifically focus on individual defendants rather than businesses or organizations. First, we removed entries indicating nonresidential eviction filings, such as businesses, nonprofits, government organizations, and so forth. Using keyword filters, we excluded records containing terms like "inc," "company," "corp," and other indicators of business entities.

We then cleaned names by removing certain phrases and identifiers, such as business names, addresses, and descriptors (e.g., "estate of" or "deceased"). Additionally, we excluded common placeholder names like "John Doe" and any variants of "unknown" to improve uniformity across the dataset. To further reduce formatting inconsistencies, we standardized name suffixes, eliminated digits and special characters, and corrected spacing and hyphenation issues. These adjustments helped ensure more consistent naming patterns across all entries.

Since our original data contained full names in a single column, we developed a parsing method to separate names into first, middle, and last names. This addressed the varying formats present in the data and enabled more precise imputation by consistently aligning name components. Following these steps, we applied our imputation to 45,968 eviction filings.

To identify "prolific evictors" (those who filed 100 or more evictions), we began by cleaning the plaintiff names in the filings data. This involved replacing common misspellings and correcting specific errors in the names of the seven companies or landlords identified by Legal Aid as "prolific" evictors. We then categorized plaintiffs into "company" and "individual" categories using classifications produced by the Case Western Reserve University (CWRU) Poverty Center as part of their Cleveland Eviction Study (García-Cobián Richter et al. 2019). Where a landlord appeared in our data that had not been previously classified by CWRU, we classified the landlord as corporate based on the appearance of keywords: for instance, names including "LLC," "LTD," "L.P.," "corporation," "real estate group," "CMHA," "Cuyahoga County," "Cleveland Housing Network," "church," and "foundation." Entries that did not match these categories and instead featured a first and last name format were labeled as "individual."

Using a string distance with the Jaro-Winkler method, we clustered similar names to create a cleaned and consolidated list of landlord names under the "company" category. We included both company and individual landlords in our analysis of prolific evictors but applied Jaro-Winkler consolidation only to company landlords, because similar company names likely referred to the same entity, whereas similar individual names could represent different people. Based on this cleaned list, we identified the landlords who had filed 100 or more evictions during the study period, from January 2016 to June 2023.

We encountered several limitations in cleaning and categorizing the data and made certain assumptions. Our cleaning of corporate plaintiffs focused on resolving differences in how the same company was represented in the data to enable accurate counts of eviction filings by plaintiff. Despite our efforts to group entities by name, some company groupings were likely missed because of misspellings and clustering method limitations. We were unable to group entities with different names but a common parent entity. Additionally, our keyword filters for identifying businesses may not have been fully comprehensive, resulting in some companies remaining in the individual defendant category. We also attempted to clean individual plaintiff and defendant names to enable accurate counts of eviction filings by individual plaintiffs and support more accurate imputation. Although we manually corrected certain errors that we identified, we were unable to systematically address all name misspellings. Formatting inconsistencies across entries posed another challenge, and it is likely that some inconsistencies in suffix usage were not fully standardized. These limitations taken together with the fact that we were not able to use the Jaro-Winkler method to consolidate individual plaintiff entries means that we likely undercount filings by some individual landlords due to different representations of the same name that we did not have enough information to reconcile. In reviewing the cleaned landlord data, we determined that the counts of filings in these cases were small enough that it likely did not affect our prolific landlord analysis. Finally, we could not rely on addresses, which were often associated with lawyers rather than the landlords themselves, to determine whether two names with similar spellings referred to the same individual.

Court documentation limitations also affected our ability to analyze certain outcomes, such as writs issued, writs filed, second cause hearings, and physical evictions, which we identified by keywords and grouped using the filing ID key. The accuracy of this information depended heavily on the consistency of court docket documentation.

Imputation Methodology

RACE AND ETHNICITY IMPUTATION

To impute race and ethnicity for tenants facing eviction, we tested three different imputation methodologies to determine which produced the best results:

- Bayesian Improved Surname Geocoding (BISG) uses the defendant surname and information about the demographic composition of the block group in which they live to predict the probability of belonging in one of six race/ethnicity groups: American Indian, Asian, Black, Hispanic, other, and white. We used the proportion of renters belonging to each of these six race/ethnicity groups in the block group where the defendant's address was located as the prior probabilities that we fed into the BISG model. For BISG and BISG-Measurement Error (BISG-ME), we also tested using the tract-level racial/ethnic composition as the prior probabilities (shown in table 1 in the BISG Tract and BISG-ME Tract columns) and found that the block group–level probabilities delivered better results. Therefore, for BIRDiE, we only tested the block group prior probabilities.
- BISG-Measurement Error (BISG-ME) builds upon BISG to address two issues of measurement error in census datasets used for BISG that could lead to worse results for smaller population groups. First, undercounts or privacy protections may result in census estimates containing zero counts for a given racial/ethnic group in a geography where members of those groups reside. Second, the Census Bureau data used by BISG for race/ethnicity composition associated with surnames only reports surnames that appear 100 or more times in the population; however, this reportedly covers 90 percent of names. BISG-ME incorporates data from six

southern states where individual self-reported race of registered voters is available for validation to improve the surname coverage (Imai, Olivella, and Rosenman 2022).

- Bayesian Instrumental Regression for Disparity Estimation (BIRDiE) estimates racial disparities using data that have imputed race/ethnicity by using the imputed race/ethnicity probabilities as inputs and producing racial disparity estimates using surnames as an instrumental variable for race. This corrects for the potential correlation between the residuals in imputed race/ethnicity and the outcome of interest and can yield more accurate estimates of racial disparities (McCartan et al. 2024). To the best of our knowledge, BIRDiE has not been used in previous research on evictions. We used this methodology in our disparity analysis discussed below. We also extracted the updated individual-level race/ethnicity probabilities from BIRDiE to compare with the BISG and BISG-ME probabilities.

Each imputation methodology produces a set of race/ethnicity group membership probabilities for each individual in the dataset. We used the probabilities directly in our analyses rather than classifying each individual in a single race/ethnicity group. Using the probabilities for analysis is the established best practice because it produces more accurate results than classifying individuals into a single group prior to analysis (Adjaye-Gbewonyo et al. 2014; Elliott et al. 2008, 2009). Unless specified otherwise, all results in this report were calculated using the probabilities.

Table 1 presents the aggregated racial/ethnic composition of the imputed data for tenants facing eviction filings for each imputation methodology we tested, alongside the racial/ethnic composition of renters in the court jurisdiction for reference. Notably, the BISG-ME method significantly inflated the proportion of individuals assigned to the AIAN, Asian, and other groups to a degree that our project advisers and research team believed was implausible, which was supported by our validation results. This proportion inflation may result from the high margins of error in the American Community Survey estimates used to calculate the prior racial and ethnic probabilities. The inflation issue could be improved in future studies by using decennial census estimates instead. We also noted similarities across the BISG Block Group and BISG Tract results and in the updated probabilities produced by BIRDiE. We decided to use the BISG Block Group probabilities, given the likely greater predictive power of the prior race/ethnicity probabilities for the smaller geography.

TABLE 1

Aggregate Racial/Ethnic Composition of Defendants Facing Eviction Filings in Cleveland Housing Court(2016–23) by Imputation Method

Race/ ethnicity	Court jurisdiction[1]	BISG-ME block group	BISG-ME tract	BISG block group	BISG tract	BIRDiE[2]
AIAN	0.23	17.12	5.14	0.14	0.15	0.15
Asian	2.84	6.87	3.47	0.82	0.70	0.85
Black	51.90	25.45	46.15	64.37	65.88	64.48
Hispanic	10.88	8.07	7.31	7.58	7.25	7.58
Other	4.02	20.36	8.52	2.65	2.60	2.71
White	30.14	22.13	29.41	24.43	23.42	24.23

Source: Authors' calculations using imputed eviction filings from Cleveland Housing Court from January 2016 to June 2023 collected by the Legal Services Corporation Civil Court Data Initiative.

[1]The court jurisdiction proportions use all renters in the court jurisdiction (N = 1,663,089) as the denominator. All other columns use the total number of imputed eviction filings (N = 45,968) as the denominator.

[2]Calculated using BISG Block Group probabilities as input and writ issued variable as outcome. Results were similar across all outcome variables considered.

GENDER IMPUTATION

To impute gender, we implemented a two-step process that leveraged large-scale public data sources: (1) the Genderize API, and (2) the US Social Security Administration baby names data.

- **Genderize API**: The *genderizeR* R package was our primary tool for gender imputation. It uses the Genderize API to predict gender based on name data collected from about one billion social media profiles. This source provided broad, contemporary coverage for gender predictions and allowed us to match a significant portion of the names in our dataset.
- **Baby names data:** For names not matched in the initial Genderize API pass, we used the *gender* R package, which references US Social Security Administration baby names data. This supplementary source helped us assign gender to additional names in our dataset, relying on established historical data to improve coverage.

Through this combined approach, we imputed gender for approximately 93 percent of observations in our dataset. A substantial number of the imputed names were assigned a probability above 95 percent, indicating high confidence in the predictions. The remaining 7.3 percent of observations could not be matched to either source, primarily because unique or less common names were absent from both datasets. For these cases, we used the overall gender probabilities for the court jurisdiction rather than excluding them from the analysis of gender disparities. This gap highlights a limitation in the availability or representativeness of gendered name data.

Data Validation

To validate our imputation methodology, we partnered with the Case Western Reserve University's (CWRU) Center on Poverty and Community Development to compare the imputed race and ethnicity probabilities to the reported race and ethnicity for the same individuals in their Child and Household Integrated Longitudinal Data (CHILD) data system. The CHILD system includes continually updated administrative data from 1989 to the present for more than 750,000 children with data in over 35 administrative systems. Some of the datasets in the CHILD system also include information on the children's parents or legal guardians, including race, ethnicity, and gender.[7] CWRU shared a dataset of 16,637 eviction records from 2013 to 2016 that they had previously linked with the CHILD data system using the name and address fields common to both datasets.[8] We applied our imputation methodology to the records to generate imputed race/ethnicity probabilities, which CWRU compared with the "true" race/ethnicity field in the CHILD system within their secure computing environment, then calculated several aggregated validation metrics that they shared back with us. To calculate the accuracy of our imputation method, we assigned each individual in the dataset to the single race/ethnicity group that had the largest probability (see "Single-Assignment" in table 2).

We also used the validation results to calculate a set of recalibrated probabilities that we validated using the same method just described. First, we fitted a logistic regression model to predict membership in each of our imputed classes using the logit-transformed imputed probabilities as our independent variables, fitting one logistic regression model for each race/ethnicity and gender group. The coefficients generated by these regressions captured the relationship between the predicted probabilities and true class membership. We expected these coefficients to identify the ways in which our imputation model systematically deviated from true class membership, in which case we could use the coefficients to recalibrate the original imputed probabilities to correct for the deviations and thereby improve the probabilities' accuracy. We split the data into two sets to perform this recalibration—a training set used to fit the logistic regression models and a test set used to assess how the recalibration affected imputation accuracy.

The validation results for gender imputation showed that we correctly predicted gender in 97.1 percent of observations, with very similar accuracy rates for men and women (97.3 and 97.0 percent, respectively). The recalibrated probabilities slightly reduced overall accuracy (96.6 percent), mostly because of significantly reduced accuracy for men (91.4 percent) accompanied by a much smaller increase in accuracy for women (98.4 percent). Accordingly, we used the original gender probabilities for our analysis.

Our validation results for race and ethnicity show that we correctly predicted race/ethnicity in 83.8 percent of observations, although the results varied significantly by race and ethnicity group, as shown in figure 1. Notably, our model performed poorly for the smallest race and ethnicity groups (AIAN, Asian, and other) which collectively represent 0.9 percent of total observations. This was likely driven by the need to assign each observation to the single highest probability race/ethnicity group. Since these small groups rarely represent the largest probability share, this "single-assignment" approach is unlikely to classify an individual as a member of these groups. This is underscored in table 2; the various "single-assignment" approaches we tested generally underrepresented the AIAN, Asian, and other race/ethnicity groups. Yet, when we use imputed probabilities directly to calculate population shares, the results overrepresented members of the Asian and other groups.

TABLE 2

Racial/Ethnic Composition of Validation Data by Imputation Approach

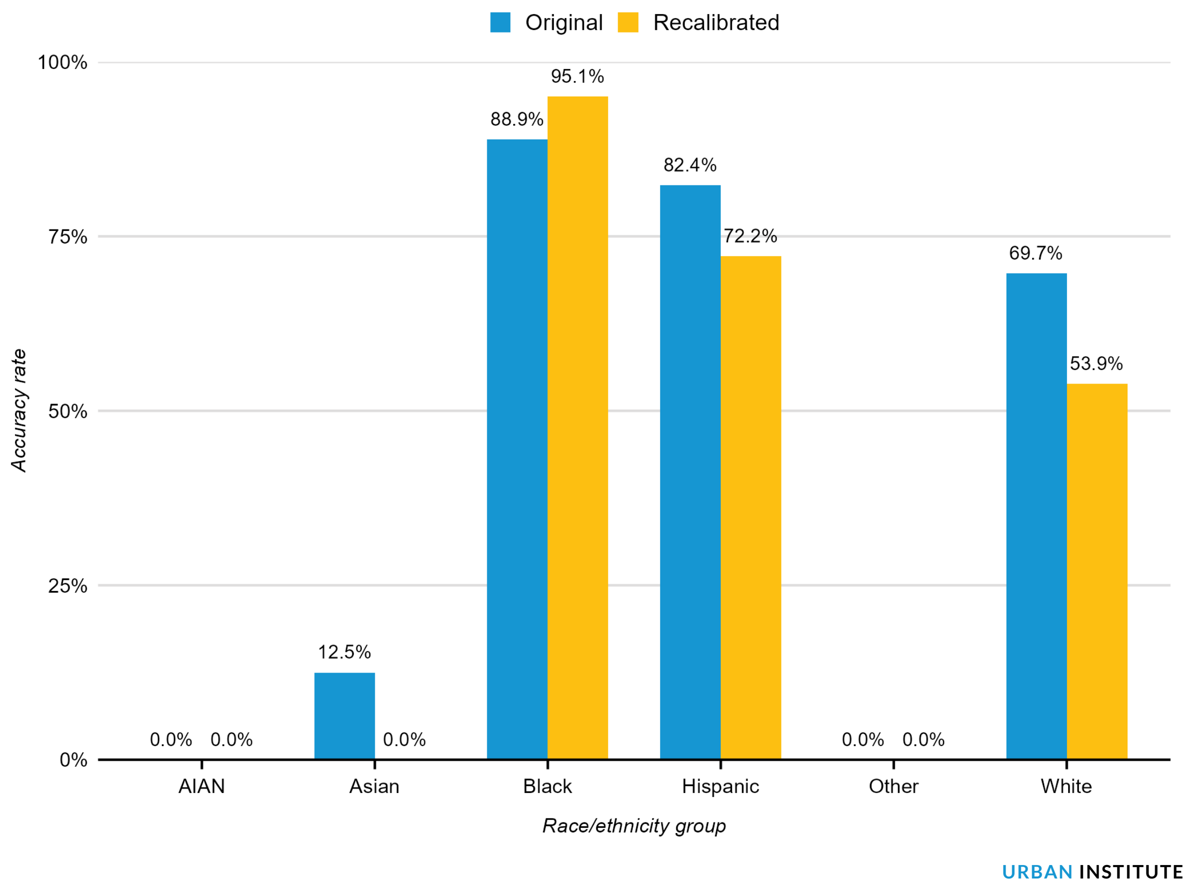
Race	True composition – CWRU data	Imputed – proportions	Imputed – single-assignment	Recalibrated – single-assignment	Block group composition– Single-assignment
White	20.46%	23.28%	22.13%	15.4%	27.51%
Black	73.47%	66.08%	71.15%	79.2%	63.82%
AIAN	0.17%	0.13%	0%	0%	0%
Asian	0.05%	0.43%	0.13%	0%	0.03%
Hispanic	5.09%	7.65%	6.57%	5.36%	8.32%
Other	0.74%	2.44%	0.03%	0%	0%

Source: Authors' calculations based on imputation of 2013-2016 eviction records linked with the Case Western Reserve University CHILD data system.
Note: N = 16,637.

For the single-assignment imputation approaches, we could also calculate the overall accuracy rate by comparing the assigned race/ethnicity group with an individual's true race/ethnicity in the CWRU data, and calculating the proportion of observations for which imputed race/ethnicity matched true race/ethnicity. We found that recalibration slightly increased overall accuracy to 85 percent by improving the accuracy rate for Black individuals from 88.9 to 95 percent; however, that was accompanied by reduced accuracy for other race/ethnicity groups, as shown in figure 1. Table 2 also shows that the recalibrated data perhaps overcorrected for the underrepresentation of Black individuals in the original imputations and pivoted toward significantly overrepresenting the Black population, with 79.2 percent of the recalibrated data classified as Black compared with 73.47 percent of the population. To contextualize our results, we also calculated the CWRU data racial and ethnic composition using a common alternative imputation approach that assigns an individual the most common racial/ethnic group in their block group ("Block Group Composition" in table 2). We found that this approach more substantially overestimated white tenants and underestimated Black tenants, and yielded no observations in smaller racial/ethnic groups. These results highlight the value of employing a probabilistic imputation approach and using the probabilities for imputed data analysis.

FIGURE 1

Accuracy by Race/Ethnicity Group for Original and Recalibrated Imputation Results



Source: Authors' calculations based on imputation of 2013 to 2016 eviction records linked with the Case Western Reserve University CHILD data system.

Unsurprisingly, our accuracy rate for predicting white defendants was highest in primarily white neighborhoods (i.e., population > 60 percent white) and our accuracy rate for predicting Black defendants was highest in primarily Black neighborhoods (i.e., population > 60 percent Black). The accuracy for white and Black predictions is similar in "other neighborhoods" that do not fall into either neighborhood type group.

Data Analysis

Disparity Analyses

We used the BIRDiE methodology to estimate racial and gender disparities for tenants across several eviction court outcomes:

- whether any defendant (tenant) had legal representation
- whether any plaintiff (landlord) had legal representation
- whether a writ of eviction was filed
- whether a writ of eviction was issued
- whether a move was scheduled
- whether there was a second cause hearing

In addition to running a BIRDiE regression analysis to calculate racial disparities, we also directly used the race/ethnicity probabilities to calculate outcomes by race/ethnicity group as a robustness check.

Although the BIRDiE method worked well for outcomes where we had complete information on all involved individuals (such as legal representation or writ issuance for people who received eviction filings), we could not use it to estimate disparities in eviction filing rates because our data only included cases where filings actually occurred, so we did not have a full picture of renters who were never filed against. Without data for both outcomes (filing and non-filing) for the entire population at risk, we could not build the necessary comparisons needed for the BIRDiE model. Therefore, to analyze disparities in filing rates, we aggregated imputed data and compared it to the overall tenant population.

Text Analysis

We performed text pattern matching on the event comments section to extract financial outcomes related to eviction filings. The event comments contained detailed information on case events, which we analyzed to identify judgment amounts and garnishments across cases.

Specifically, we searched for keywords in the comments associated with events titled “Mag Dec Filed, Approved, Confirmed, Judgment Rendered For,” where we could identify the judgment amount and whether it was in favor of the plaintiff or defendant. Additionally, for events containing the term “garnishment,” we parsed the text to determine the garnishment type, which were either personal earnings or “other.”

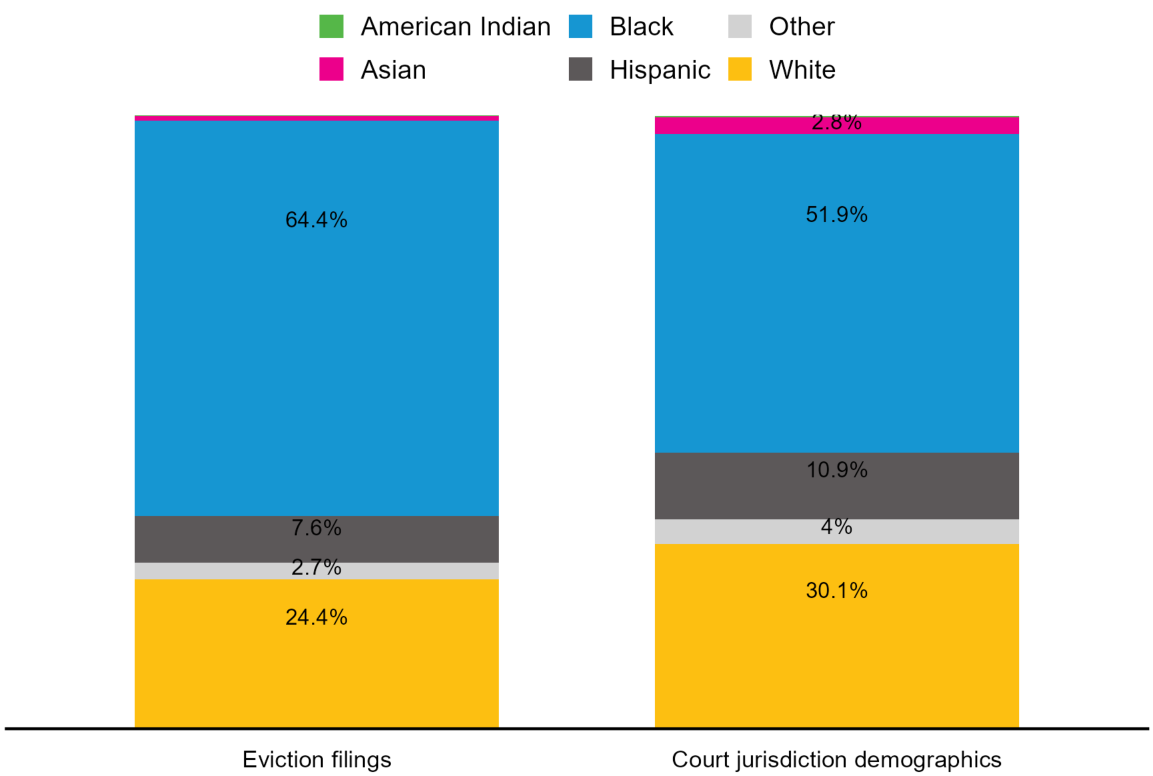
Results

Demographic Composition of Tenants Facing Eviction Filings

To contextualize the demographic composition of tenants whose landlords had filed an eviction complaint against them, we first calculated the potential universe of households. We defined this universe as renter households who live in the Cleveland Housing Court's jurisdiction, which represents the City of Cleveland and Village of Bratenahl. We then calculated the demographic composition of renter households by householder race/ethnicity and gender using data from the 2020 Census Supplemental Demographic and Housing Characteristics File. Relative to the court jurisdiction's demographic composition, we found that Black individuals and women were overrepresented in the population of tenants with eviction filings (figure 2). Further, Black defendants comprised 64.4 percent of the eviction filings between January 2016 and June 2023, which was lower than the share of Black defendants (77 percent) in the Cleveland Eviction Study for 2013–2016 (García-Cobián Richter et al. 2019). Their study matched eviction records primarily with public assistance data; therefore, as the authors noted, they essentially studied households with low incomes. We believe that imputation allowed us to produce a better estimate of the universe of households facing eviction as we were able to use nearly all eviction records. Given the differences in the population studied in the Cleveland Eviction Study, we cannot draw conclusions about trends in eviction by race/ethnicity.

FIGURE 2

Demographics of Tenants Facing Eviction Filings Vs. Court Jurisdiction Population by Race/Ethnicity

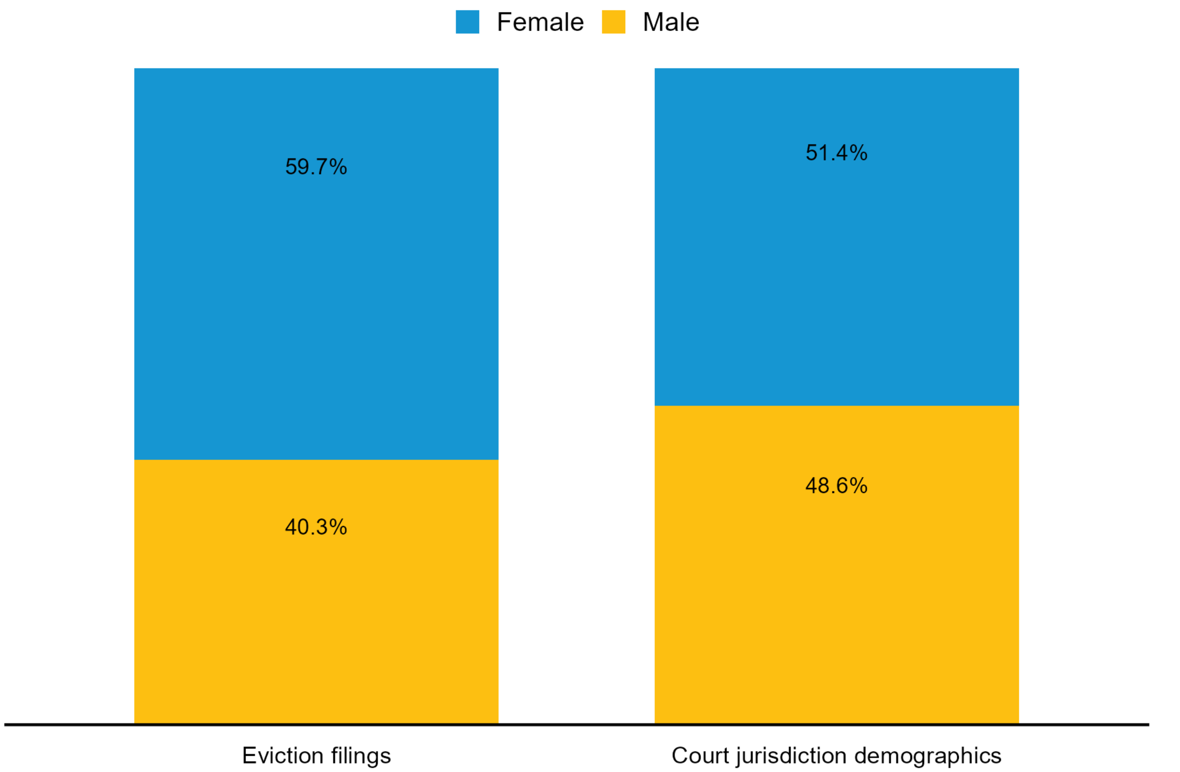


URBAN INSTITUTE

Source: Demographic information on tenants facing eviction calculated using imputed eviction filings from Cleveland Housing Court from January 2016 to June 2023 collected by the Legal Services Corporation Civil Court Data Initiative. Court jurisdiction demographics are calculated using the 2020 Census Supplemental Demographic and Housing Characteristics File.

FIGURE 3

Demographics of Tenants Facing Eviction Filings Vs. Court Jurisdiction Population by Gender



Source: Demographic information on tenants facing eviction calculated using imputed eviction filings from Cleveland Housing Court from January 2016 to June 2023 collected by the Legal Services Corporation Civil Court Data Initiative. Court jurisdiction demographics are calculated using the 2020 Census Supplemental Demographic and Housing Characteristics File.

The data presented in figure 2 and 3 support previous research findings that certain populations are disproportionately affected by eviction. Inspecting the intersection of race/ethnicity and gender, it is clear that Black women are particularly overrepresented, with 62.5 percent of Black defendants imputed as female compared with 57.6 and 53.8 percent of Hispanic defendants and white defendants, respectively.

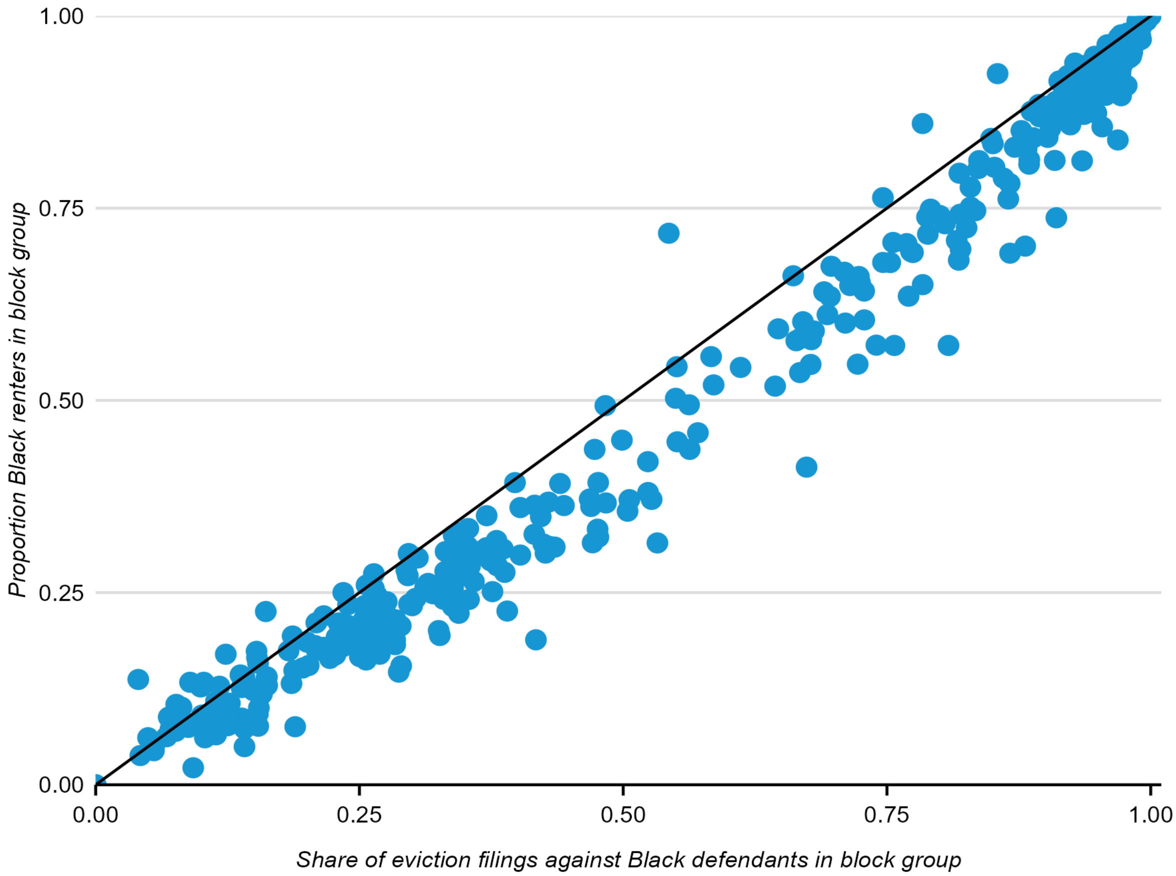
Although we found that Black renters were disproportionally the subject of landlord eviction filings, our analysis of eviction case outcomes did not reveal significant differences across racial/ethnic groups. Proportions of white, Black, and Hispanic individuals facing eviction filings who had any defendant represented, any plaintiff in their case represented, a writ issued, a move scheduled, a physical eviction, or a second cause hearing were similar for both the BIRDiE method and direct calculations using imputed probabilities (see Appendix A for full results). We could not assess whether race or gender disparities existed for the eviction reason, default judgments, amount of rent owed, or for how many months the rent payments were delinquent when landlords chose to file for eviction.

Neighborhood Characteristics of Eviction in Cleveland

We found that Black tenants were disproportionately represented in eviction filings across Cleveland neighborhoods. For 90 percent of Cleveland block groups with 5 or more eviction filings in our data, the share of tenants facing eviction who were Black was greater than the share of all renters who were Black. This pattern holds across tracts with different proportions of Black residents, as shown in figure 4.

FIGURE 4

Black Individuals Overrepresented in Tenants Facing Eviction across Block Groups



URBAN INSTITUTE

Source: Share of eviction filings against Black defendants in block group calculated using imputed eviction filings from Cleveland Housing Court from January 2016 to June 2023 collected by the Legal Services Corporation Civil Court Data Initiative. Proportion of Black renters by block group calculated using the 2020 Census Supplemental Demographic and Housing Characteristics File.

Note: This graph includes the 427 block groups with 5 or more eviction filings during the period of analysis.

The 10 percent of tracts where the share of tenants facing eviction filings who were Black was smaller than the share of tenants who were Black, had higher median incomes and rents on average, as well as smaller proportions of Black residents and renters (table 3).

TABLE 3

Characteristics of Block Groups with Smaller Share Black Tenants Facing Eviction

	Median income	Share Black renters	Gross rent	Share renters
Block groups with smaller share Black tenants facing eviction than Black renters	$52,382	44.8%	$963	45.8%
All other block groups	$41,507	57.0%	$886	57.3%

Source: Demographic information on tenants facing eviction calculated using imputed eviction filings from Cleveland Housing Court from January 2016 to June 2023 collected by the Legal Services Corporation Civil Court Data Initiative. Share of Black renters calculated using the 2020 Census Supplemental Demographic and Housing Characteristics File. Other block group characteristics are calculated using the 2018–22 5-year American Community Survey.

Prolific Evictors in Cleveland

While 95.38 percent of all landlords in our dataset filed 10 or fewer evictions from January 2016 to June 2023, the 0.25 percent of landlords who filed 100 or more evictions accounted for 21.8 percent of all eviction filings. In fact, the 10 landlords with the greatest numbers of evictions in our data represented 15.32 percent of all evictions in our dataset, despite representing only 0.08 percent of unique landlords in our dataset. These prolific evictors are listed in table 4, along with the number of evictions appearing in our dataset and the racial and ethnic composition of individuals who were evicted by each landlord. It is important to note that our extensive data cleaning of landlord names may have missed some cases from a given landlord because of name misspellings in court records, or may have inaccurately grouped landlords with very similar names. In addition to the programmatic data cleaning steps described earlier, we also manually checked unique landlord names that were similar to landlords identified as prolific evictors to improve the accuracy of our counts.[9] It is also important to note that we did not have information on the number or racial/ethnic composition of all tenants renting from each landlord, so we could not determine the share of tenants for which each landlord initiated evictions, or whether the racial/ethnic composition of those tenants was representative of the landlord's tenants overall.[10]

TABLE 4

Ten Most Frequent Evictors and Racial/Ethnic Composition of Tenants Served Evictions, 2016–23

Landlord name	AIAN	Asian	Black	Hispanic	Other	White	# Evictions
CMHA	0.1	0.2	84.8	3.7	2	9.2	3347
K D Management LLC	0.2	8.6	45.7	3.8	5.2	36.5	908
CHN Housing Partners	0.1	0.2	80.5	5.1	2.1	11.9	661
Windsor Realty Management Inc	0.3	5.6	44.4	17.2	3.2	29.3	477
Aiy Properties Inc	0.3	0.5	36.5	13.3	2.3	47.1	374
Millennia Housing Management Inc	0.1	1.3	82.9	1	3.2	11.6	281
Emerald Development Economic Network	0.2	0.1	68.2	3.4	2.9	25.2	258
Independent Management Services of Ohio Inc	0.2	0.3	92.2	0.4	2.5	4.4	254
Neiderst Management Ltd	0.2	3.9	55.4	7	6.3	27.2	245
Clifton Management LLC	0.0	1.4	30.9	7.2	3.7	56.7	241

Source: Authors' calculations using imputed eviction filings from Cleveland Housing Court from January 2016 to June 2023 collected by the Legal Services Corporation Civil Court Data Initiative.
Note: AIAN = American Indian and Alaska Native.

We found that a greater share of the tenants served eviction filings by prolific evictors (100+ evictions) were Black (69 percent) compared with other landlords (63.1 percent), and a smaller share of tenants served eviction filings by prolific evictors were white (20.7 percent) compared with other landlords (25.5 percent). This difference may have been driven by the large numbers of Black tenants served evictions by the Cuyahoga Metropolitan Housing Authority (CMHA); when CMHA was removed from the analysis we found a slightly larger share of the residents evicted by prolific evictors were white (26.4 vs. 25.5 percent for other landlords) and a slightly smaller share of the residents evicted by prolific evictors were Black (61.0 vs. 63.1 percent for other landlords).

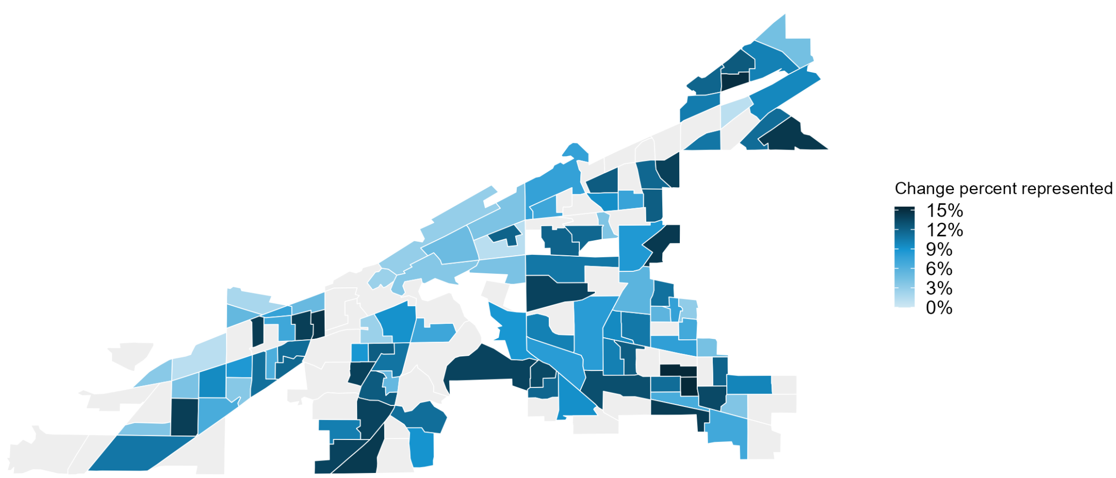
Legal Representation and Right to Counsel Cleveland

Cleveland's right to counsel (RTC-C) law passed in October 2019 and took effect in July 2020. It gives Cleveland residents facing eviction and living at or below the federal poverty line with at least one child the right to free legal help in housing court.[11] Some tenants with low incomes who do not qualify for RTC-C may still qualify to be represented by a lawyer through the Legal Aid Society of Cleveland. Analyzing the cases in our data prior to the passage of RTC-C, from 2016 through June 2020, we found that only 1.7 percent of defendants with eviction filings had any legal representation. After Cleveland's right to counsel passed, 14.3 percent of defendants were represented from July 2020 to 2022. Even in the post RTC-C period, more than 85 percent of defendants in our data did not have legal representation. The presence of legal counsel can have a significant impact on outcomes for tenants facing eviction. For example, only 14 percent of defendants with representation received a writ, compared with 50 percent of unrepresented defendants.

Figure 4 shows the change in the share of defendants with legal representation by census tract after RTC-C. Comparing tracts with low versus high representation (8 percent or more of filings in a tract had representation) before and after RTC-C, we found that tracts with higher representation increases corresponded to areas with a greater share of Black residents, lower median incomes, and a higher share of cost-burdened renter households. This suggests that the RTC-C legislation and outreach efforts effectively target residents and neighborhoods that have more structural barriers to accessing legal representation, such as Black residents and residents with lower incomes.

FIGURE 5

Change in Share of Defendants Legally Represented by Census Tract After RTC-C



URBAN INSTITUTE

Source: Authors' calculations using imputed eviction filings from Cleveland Housing Court from January 2016 to June 2023 collected by the Legal Services Corporation Civil Court Data Initiative. Census tract boundaries from US Census Bureau obtained using tidycensus R package.

Financial Judgments and Wage Garnishments

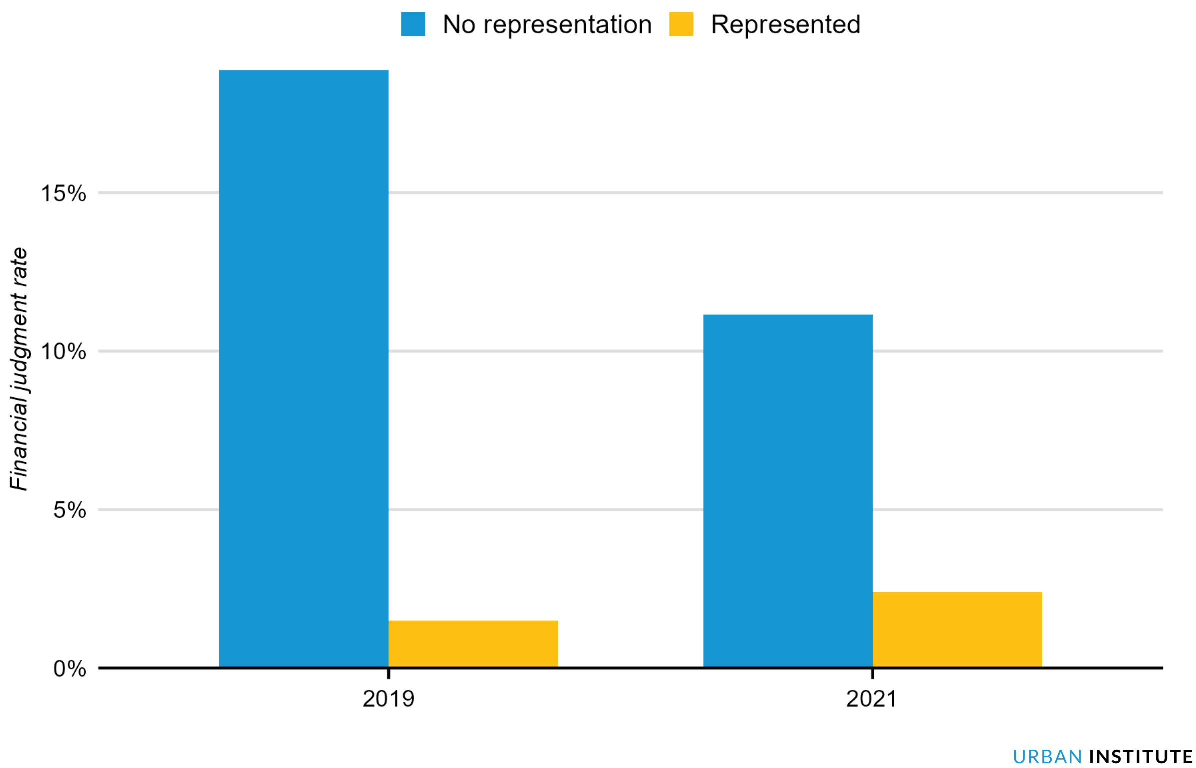
Our web scraping process collected data on 11,704 eviction filings from 2019 and 2021 and identified financial judgments in 1,822 of these cases. Only two of these judgments were rendered against the plaintiff. In both filings and judgments, 2019 saw higher numbers (7,802 filings and 1,449 judgments) than 2021 (3,902 filings and 373 judgments).

We observed a median financial judgment of $2,424.50 across years, with individual judgments ranging from as low as $2 to as high as $120,841. We identified twice as many financial judgments against Black men compared with white men, and three times as many against Black women compared with white women. Despite these differences in absolute numbers, the rate of judgments relative to filings was consistent across these demographic groups.

Legal representation appeared to be a significant factor in judgment outcomes. Defendants without legal representation were far more likely to have a financial judgment rendered against them, with 16.6 percent of unrepresented defendants receiving financial judgments compared with only 2.26 percent of represented defendants. This gap narrowed in 2021, coinciding with the implementation of right to counsel legislation and the pandemic. In 2019, judgments were rendered against 18.9 percent of unrepresented defendants versus 1.49 percent of represented defendants, while in 2021, these rates shifted to 11.1 and 2.41 percent, respectively (see figure 6).

FIGURE 5

Legal Representation Protects against Financial Judgments



Source: Authors' calculations using imputed eviction filings from Cleveland Housing Court from January 2016 to June 2023 collected by the Legal Services Corporation Civil Court Data Initiative and data scraped by authors from Cleveland Housing Court website.

Results were similar across racial groups, with Black and white defendants having comparable judgment rates when unrepresented (17.2 vs. 16.7 percent) and slightly lower rates for Black defendants compared with white defendants when represented (2.3 vs. 3.6 percent). Judgment rates for unrepresented Asian and AIAN defendants were higher (25.3 and 20.0 percent, respectively), but sample sizes were small (24 of 95 cases for Asians, 3 of 15 cases for AIANs). Comparisons for represented Asian and AIAN defendants are not meaningful because of the extremely small sample sizes (5 and 1 cases, respectively), and no significant variation was found across gender or race-gender combinations.

Finally, we identified 272 cases involving wage garnishments. Of these, the majority (225) were categorized as garnishments of personal earnings, while the remaining 47 were classified as "other." All garnishments of personal earnings were limited to amounts up to $3,000, and specific amounts were not disclosed.

Discussion

The assumption that “justice is blind” is a cornerstone of our civil legal system. Unfortunately, the lack of information about which demographic groups are most likely to suffer the potentially life-altering effects of an eviction creates an environment where advocates and practitioners have little or no evidence to challenge this assumption, plaintiffs are not held accountable for racially biased practices, and courts continue to process large volumes of cases and issue judgments without any oversight. In short, we are letting this complex machine run unattended, without understanding or objecting to the disparate impact it is having on large segments of our communities.

Unlike criminal cases, where various federal, state, and local agencies collect demographic information about defendants, civil legal matters have no comparable practice. State courts are not required to collect and report demographic information to other agencies or authorities. Given the financial pressure on state courts, the administrative burden of gathering this type of information, the fact that it is not requested or required by external groups or authorities, and the lack of leadership across courts to overcome these challenges, data on civil litigants’ characteristics is simply not collected.

Fortunately, data science tools are available today that allow us to leverage the limited information that courts do collect about litigants—such as name and address—to explore patterns in who enters the civil court system and the administration of justice. In the case of Cleveland, we have shown that Black tenants and female tenants are more likely to face an eviction filing than might be expected given their representation in the community. Although this is concerning, what is perhaps more significant is that the pattern is widespread and occurs across the city, regardless of local neighborhood composition. For 90 percent of the City of Cleveland’s block groups with 5 or more eviction filings from 2016 to 2023, the share of tenants facing eviction who were Black was greater than the share of renter households that were Black. In other words, it does not matter whether the neighborhood is predominantly Black or predominately white, Black renters are still more likely to face an eviction filing. This type of persistent and widespread outcome is the very definition of systematic bias. Just as importantly, this study reveals how seemingly neutral institutions, like state courts, can perpetuate systematic racial and gender inequities. The imputation methodology used here leverages administrative civil case data that are readily available in state courts today. This same method could be used in many more jurisdictions, with no change to their current procedures or data collection methods. Broadening the analysis to other communities would allow more judges, state court administrators, and advocates to measure and monitor racial and gender disparities in eviction case filings, as well as other civil legal matters that have a significant impact on people's lives, such as debt cases, foreclosures, civil protection orders, and family law matters.

The eviction filing pattern we see in Cleveland matches patterns found in other studies, such as Hepburn, Louis, and Desmond (2020), and patterns that likely occur in other jurisdictions across the country. Unfortunately, given the data limitations outlined above, it is easy to miss these patterns if one focuses solely on the administrative data available from state courts. That is why the process we have employed here, of imputing race and gender of named litigants, holds such promise. For this project, we not only outlined a process for enhancing court records, but we also went a step further to validate the imputations against one of the most sophisticated integrated data systems in the country. After validating and recalibrating the imputation process for both race/ethnicity and gender, we were encouraged by the high accuracy rates, though recognize the need to continue to improve accuracy for smaller racial and ethnic groups. The project demonstrates that, even in communities that have taken significant measures to balance the scales of justice and implemented a right to counsel intervention, disparate filing rates are likely to persist. In the face of such evidence, the question should no longer be whether we should collect and monitor civil litigants' demographic characteristics. The question now becomes, how can we not?

In addition, this work advances the field of demographic imputation research by adding to the knowledge base in several ways. First, it applied the new BIRDiE methodology for estimating racial/ethnic disparities using imputed data, which corrects for the potential correlation between the residuals in imputed race/ethnicity and the outcome of interest and can yield more accurate estimates of racial disparities (McCartan et al. 2024). Additionally, we assessed the effectiveness of using gender data from social media profiles—a relatively new source in imputation studies—to enhance gender classification accuracy. Finally, unlike many imputation studies, we validated our results against ground truth data provided by Case Western Reserve University, offering a rare opportunity to evaluate the accuracy of our imputed values and improve future methodologies.

This research confirms that racial and gender inequities and wealth stripping occur in the Cleveland Housing Court. Specifically, people of color disproportionately face eviction filings and are entering the civil court system at higher rates. People without legal representation are more often evicted and more often have financial judgments. Using this evidence, community leaders may advocate for and policymakers and practitioners may enact the following policy and practice reforms that reduce the entry of tenants into the civil court system and mitigate the potential harms from interacting with eviction processes:

- **Automatic record sealing at time of filing**. An eviction record can have devastating long-term impacts on many aspects of people's lives—it can be a barrier to finding future housing and can lower a person's credit score so that it costs more to borrow money. Ohio law does not give tenants the right to seal their eviction records.[12] Each court decides whether that is possible.

  A state law that requires Ohio courts to automatically seal eviction records at the time of filing would provide optimal consumer protection, especially given our finding of clear racial disparities in who is filed against in housing court. At the very least, a state law that requires courts to automatically seal records when eligibility is reached would place the responsibility on the courts to execute its policies, rather than placing the burden on the tenant.

- **Change eligibility rules to shorten time to seal records**. The Cleveland Housing Court has a process for sealing eviction records, but in cases decided in favor of the landlord, the tenant must wait at least five years before they can file a motion to seal their record.[13] This five-year wait is arbitrary and could be modified to a shorter period (although sealing at the time of filing remains optimal).

- **Restrict third-party access to eviction records**. Third-party entities, including credit card companies, often access court information to add to their database, which may be used by landlords, employers, and others. Sealing a record at the time of filing, as noted before, prevents this access.

- **Increase eviction filing fee.** The current fee for landlords to file a motion to evict is $110 in the Cleveland Housing Court.[14] Gomory and colleagues (2023) with the Eviction Lab found the cost of filing an eviction case effects how often landlords file eviction cases. In some cases, rather than working with tenants whose rent is late to figure out payment plans, landlords use the courts as a collections arm. Higher filing fees generally make filing evictions less economical, which will likely reduce this behavior.

- **Expand eviction diversion programs.** Diversion programs can be developed to intervene at different points in the eviction process, including pre-filing before the court process is initiated, post-filing, or in court. The National Center for State Courts implemented an Eviction Diversion Initiative project and found that across the cohort of participating jurisdictions, "89 percent of the cases that engaged with a diversion program resulted in a settlement agreement or voluntary dismissal, meaning the case was resolved without an eviction judgment against the tenant" (NCSC 2024, p. 3). Philadelphia's Eviction Diversion Program requires pre-court registration by the landlord and a waiting period during which they make a good-faith attempt to resolve the issue before filing. Tenants may qualify to attend mediation with a housing counselor and receive rental assistance in some cases. Dowdall and Goldstein (2023) found that, although mediation itself did not decrease filings, it did increase the likelihood of an agreement being reached between the landlord and tenant, and cases with agreements were less likely to end up with a court filing.

- **Confirm code and habitability standards**. The Cleveland Eviction Study reported nearly one-third of tenants they interviewed had a property conditions issue, a quarter of whom reported lead paint or dust (Urban et al. 2019). In some cases, tenants reported withholding rent, which led to the eviction filing. In other cases, tenants felt the evictions were filed in retaliation for reporting code violations. The Cleveland Housing Court could require that a landlord submit documentation that their building/apartment meets code and habitability standards before filing an eviction complaint.[15] Requiring this documentation from landlords deters landlords from retaliating against tenants for reporting health and safety violations. Additionally, efforts to improve the habitability of rental units outside the eviction process should continue, including enforcing requirements to register rental properties and obtain Lead Safe certification.[16]

In addition to reforms that target court practices and court interactions—given that we found entry into the system is racially disproportionate—advocates should work to reduce filings altogether by targeting efforts to address some of the reasons tenants may not be able to make their payments in full and on time, including income instability and lack of transparency in lease terms and fees:

- **Improve landlord practices.** Reynolds and colleagues (2023) convened a roundtable of rental property owners and developed five principles and associated practices that landlords could adopt that would help reduce the number of evictions and increase tenant stability. They include proactively connecting tenants to resources that allow flexibility in rental terms and payments. The latter issue was raised by Urban and colleagues (2019) in interviews with tenants who reported landlords refusing to accept partial payments or late payments.
- **Eliminate "junk" fees.** These are additional fees that landlords add to rent that lets them get around the restriction of raising rent during a lease. The Cleveland tenants interviewed by Urban and colleagues (2019) were also overwhelmingly cost burdened. Eliminating junk fees would provide transparency; otherwise, a renter may not know about these fees until after they move in and find that their actual costs to rent are much higher than their budget. These wealth

stripping fees may include administrative fees, pet fees, late rent fees, rental application fees, and fees for paying with a credit card.[17] State laws that eliminate these fees allow renters to have more income for their housing. At minimum, fee transparency must be required before a lease is signed.

Our recommendations above focus on reducing entry into the eviction court system as that is where our analysis identifies racial disparities. However, there are numerous other reforms that could improve outcomes and overall fairness for tenants who face eviction, such as expanding the right to counsel program, providing more time for tenants to prepare for their hearings, and requiring landlords in eviction cases related to nonpayment of rent to attach the rent ledger. Further research could explore the impact of these reforms in outcomes for tenants overall and by race/ethnicity and gender.

Although imputation methods such as those used in our analysis can unlock valuable insights in cases where race/ethnicity data are not directly reported, these methods still have limitations. While we validated our imputation method as accurate for 83.8 percent of observations, it performed poorly for smaller Asian, AIAN, and other groups. Relying on imputed data could result in failing to identify disparities affecting these groups. Community leaders can call upon courts to begin direct collection of race, ethnicity, and other demographic information from individuals facing eviction filings to improve data coverage and accuracy. Such data collection efforts should be accompanied by work to build trust between courts and community members—particularly those from communities with a history of data misuse—who may be reluctant to self-report race and ethnicity due to institutional mistrust or concerns of negative impacts on case outcomes (Genthon et al. 2024; Haley et al. 2022). Increasing the amount of self-reported race/ethnicity data and using imputation to fill in data gaps will improve the quality of the evidence that supports efforts to enact the reforms outlined above. Indeed, while our analysis did not find evidence of significant racial/ethnic and gender disparities for case outcomes, the absence of standardized, analysis-ready data limited our ability to analyze numerous other outcomes of interest, such as causes of eviction, back rent owed, and others. Improving civil court data quality and transparency would allow future research to examine a broader set of outcomes and civil case types (e.g., debt cases) to more fully understand potential disparities in the civil court system and the community-level impact. Ongoing investments in systematic analysis of racial and gender disparities in civil court cases is essential to identifying and addressing those disparities and understanding the impact of the reforms discussed above.

These are some of the practice, policy, and court reforms that community advocates can call for and policymakers and practitioners can implement, now better armed with evidence that entry into the current eviction process is biased and strips people of their wealth.

Appendix: Disparity Analysis Results

TABLE A.1

Case Outcomes by Race/Ethnicity Group Calculated Using Probability and BIRDiE Estimates

Group	Value	Outcome	N	Probability estimate	BIRDiE estimate
aian	no	physical_evict	62	94.16	NA
aian	yes	physical_evict	4	5.84	NA
asian	no	physical_evict	361	95.73	NA
asian	yes	physical_evict	16	4.27	NA
black	no	physical_evict	28103	94.98	NA
black	yes	physical_evict	1484	5.02	NA
hisp	no	physical_evict	3331	95.66	NA
hisp	yes	physical_evict	151	4.34	NA
other	no	physical_evict	1157	94.8	NA
other	yes	physical_evict	63	5.2	NA
white	no	physical_evict	10631	94.7	NA
white	yes	physical_evict	594	5.3	NA
aian	no	writ_issued	29	49.46	46.37
aian	yes	writ_issued	29	50.54	53.63
asian	no	writ_issued	177	53.2	53.21
asian	yes	writ_issued	156	46.8	46.79
black	no	writ_issued	13664	51.65	52.35
black	yes	writ_issued	12789	48.35	47.65
hisp	no	writ_issued	1669	53.91	55.7
hisp	yes	writ_issued	1427	46.09	44.3
other	no	writ_issued	551	50.83	49.29
other	yes	writ_issued	533	49.17	50.71
white	no	writ_issued	5021	50.51	48.28
white	yes	writ_issued	4919	49.49	51.72
aian	no	any_defendant_represented	61	93.9	66.87
aian	yes	any_defendant_represented	4	6.1	33.13
asian	no	any_defendant_represented	365	96.7	89.47
asian	yes	any_defendant_represented	12	3.3	10.53
black	no	any_defendant_represented	27922	94.37	94.67
black	yes	any_defendant_represented	1666	5.63	5.33
hisp	no	any_defendant_represented	3279	94.17	93.97
hisp	yes	any_defendant_represented	203	5.83	6.03
other	no	any_defendant_represented	1156	94.74	84.08
other	yes	any_defendant_represented	64	5.26	15.92

Group	Value	Outcome	N	Probability estimate	BIRDiE estimate
white	no	any_defendant_represented	10643	94.82	95.68
white	yes	any_defendant_represented	582	5.18	4.32
aian	no	any_plaintiff_represented	23	35.83	56.31
aian	yes	any_plaintiff_represented	42	64.17	43.69
asian	no	any_plaintiff_represented	65	17.34	21
asian	yes	any_plaintiff_represented	312	82.66	79
black	no	any_plaintiff_represented	9225	31.18	30.24
black	yes	any_plaintiff_represented	20363	68.82	69.76
hisp	no	any_plaintiff_represented	1118	32.11	32.05
hisp	yes	any_plaintiff_represented	2364	67.89	67.95
other	no	any_plaintiff_represented	346	28.33	32.46
other	yes	any_plaintiff_represented	874	71.67	67.54
white	no	any_plaintiff_represented	3342	29.77	31.55
white	yes	any_plaintiff_represented	7883	70.23	68.45
aian	no	second_cause_hearing	14	24.29	35.32
aian	yes	second_cause_hearing	44	75.71	64.68
asian	no	second_cause_hearing	77	23.19	25.47
asian	yes	second_cause_hearing	254	76.81	74.53
black	no	second_cause_hearing	6800	25.7	25.04
black	yes	second_cause_hearing	19655	74.3	74.96
hisp	no	second_cause_hearing	740	23.9	23.76
hisp	yes	second_cause_hearing	2355	76.1	76.24
other	no	second_cause_hearing	292	27.02	37.57
other	yes	second_cause_hearing	790	72.98	62.43
white	no	second_cause_hearing	2686	27.06	27.55
white	yes	second_cause_hearing	7240	72.94	72.45
aian	no	move_scheduled	27	47.38	37.43
aian	yes	move_scheduled	30	52.62	62.57
asian	no	move_scheduled	170	51.09	49.25
asian	yes	move_scheduled	163	48.91	50.75
black	no	move_scheduled	13035	49.23	36.87
black	yes	move_scheduled	13444	50.77	63.13
hisp	no	move_scheduled	1581	50.97	39.84
hisp	yes	move_scheduled	1521	49.03	60.16
other	no	move_scheduled	525	48.4	44.79
other	yes	move_scheduled	560	51.6	55.21
white	no	move_scheduled	4770	47.92	34.44
white	yes	move_scheduled	183	52.08	65.56

Source: Authors' calculations using imputed eviction filings from Cleveland Housing Court from January 2016 to June 2023 collected by the Legal Services Corporation Civil Court Data Initiative.

Notes

[1] "The Court Statistics Project," National Center for State Courts, accessed January 3, 2025, www.courtstatistics.org.

[2] "State Court Organization Data," Court Statistics Project, accessed January 3, 2025, https://www.ncsc.org/sco.

[3] "How Many States Are 'Unified'? There Is No Definitive Answer Because There Is No Definitive Definition." OCLC, February 24, 2021, https://ncsc.contentdm.oclc.org/digital/api/collection/ctadmin/id/2502/download#:~:text=The%20result%20is%20that%20today,many%20are%20legally%20described%2C%20defined.

[4] "RAND Bayesian Improved Surname Geocoding," RAND Health Care, accessed January 3, 2025, https://www.rand.org/health-care/tools-methods/bisg.html.

[5] CCDI collects and cleans civil court records from over 1,200 counties in more than 30 US states and territories. CCDI has collected over thirty million case records filed since 2016 through a combination of web scraping from civil court case lookup websites and direct data sharing with civil courts. For more information, see "Methodology," Legal Services Corporation Civil Court Data Initiative, accessed January 3, 2025, https://civilcourtdata.lsc.gov/about/methodology.

[6] To avoid overwhelming the website and ensure that our scraping efforts did not disrupt the Cleveland Housing Court's Odyssey Portal, we incorporated time delays between requests in our code. By adding these delays we limited the frequency of requests, allowing us to scrape data at a controlled rate of approximately three cases per minute. This cautious approach respected the portal's traffic capacity and reduced the risk of our IP address being blocked. However, this method limited efficiency because it slowed the scraping rate. It took about a month of continuous, uninterrupted scraping to gather all cases from 2019 and 2021, which highlights the challenges of scraping large datasets from public portals without a public-facing API.

[7] CWRU derived the race/ethnicity and gender data used for validation from two distinct sources, birth records and public assistance records. According to CWRU, the public assistance records they have access to are a mix of voluntary self-identified and observer-identified demographic details. Data collection guidelines require that race/ethnicity and gender be recorded (among other factors), and allows for observer identification, which effectively allows collection without self-identification. CWRU reports that data providers state that the collection is primarily self-report, but there is no mechanism to verify in which cases, or to what degree overall, the data represent clients' or observers' reports. This is less of an issue with electronic records, but the confusion over self-report versus observer-report persists when collecting data from historic records.

The rbirth records CWRU has access to are similar in that self-identification is voluntary and observer identification is allowed, but the collection requirement differs. Hospitals and birthing centers collect birth data and submit it to the State of Ohio and various Federal agencies. These records are sometimes missing race/ethnicity and do not explicitly collect parents' gender. CWRU used birth data from the Ohio Department of Health to validate the imputation methodology reported here. This should not be considered an endorsement from the Ohio Department of Health for this study or our conclusions.

[8] These records are a subset of the records produced by the CWRU team and used in García-Cobián Richter and colleagues 2019 study, which started with approximately 42,000 case filings from 2013 to 2016 and linked 19,748 eviction filing cases with the CHILD data system, each represented by one head of household.

[9] Manual data cleaning steps at this phase included (1) grouping "chn housing partners," "cleveland housing network," and "cleveland housing network inc" under "chn housing partners" because web searches for all three names led to the CHN Housing Partners site; (2) grouping "vesta corporation" and "vesta management corp" under "vesta corporation" because both names led to the Vesta Corporation website; and (3) ungrouping "k d

management" and "p k management" because we did not have sufficient evidence that these refer to the same company.

10 HUD's Picture of Subsidized Households for 2022 for the city of Cleveland shows that 88 percent of households living in public housing are Black, and 84 percent of households with Housing Choice Vouchers are Black.

11 "Right To Counsel – Cleveland Launches, Providing Free Legal Help to Low-Income Tenants Facing Eviction," Legal Aid Society of Cleveland, July 1, 2020, https://lasclev.org/07012020-2/.

12 "How Can I Have My Eviction Record Sealed?" Legal Aid Society of Cleveland, accessed November 13, 2024. https://lasclev.org/sealing-an-eviction-record/.

13 "How Can I Have My Eviction Record Sealed?" Legal Aid Society of Cleveland.

14 "Eviction Filing Checklist", City of Cleveland Housing Court, accessed January 3, 2025, https://www.clevelandhousingcourt.org/sites/default/files/Final%20eviction-filing-checklist-(updated-as-of-9_27_21).pdf.

15 Victoria Bourret and Hannah York, "NLIHC State and Local Tenant Protection Series: A Primer on Renters' Rights. Habitability Protections: Two Case Studies," National Low Income Housing Coalition, accessed January 3, 2025, https://nlihc.org/sites/default/files/2024-11/Habitability_Protections_Case_Study.pdf.

16 "Getting and Keeping Rentals in Good Condition through Tougher Housing Codes," Cleveland City Council, February 12, 2024, https://www.clevelandcitycouncil.org/resources/news-media/getting-and-keeping-rentals-good-condition-through-tougher-housing-codes.

"Lead Safe Certification: Important Update," City of Cleveland, October 11, 2024, https://www.clevelandohio.gov/city-hall/departments/building-housing/divisions/records-administration/lead-safe-certification.

17 For examples of tenants affected by junk fees, see "NLIHC State and Local Tenant Protection Series: A Primer on Renters' Rights," The Legal Aid Society of Cleveland, August 22, 2024, https://lasclev.org/08222024-2/.

References

Adjaye-Gbewonyo, Dzifa, Robert A. Bednarczyk, Robert L. Davis, and Saad B. Omer. 2014. "Using the Bayesian Improved Surname Geocoding Method (BISG) to Create a Working Classification of Race and Ethnicity in a Diverse Managed Care Population: A Validation Study." *Health Services Research* 49 (1): 268–83. https://doi.org/10.1111/1475-6773.12089.

CFPB, 2014. *Using Publicly Available Information to Proxy for Unidentified Race And Ethnicity: A Methodology and Assessment*. Washington, DC. Consumer Financial Protection Bureau.

Collinson, Robert, and Davin Reed. 2019. *The Effects of Evictions on Low-Income Households*. South Bend, IN: University of Notre Dame Department of Economics.

Coulton, Claudia, Francisca García-Cobián Richter, Youngmin Cho, Jiho Park, Jeesoo Jeon, and Robert L. Fischer. 2023. "Making the Case for Lead Safe Housing: Downstream Effects of Lead Exposure on Outcomes for Children and Youth." *Health & Place*, *84*, 103118. https://www.sciencedirect.com/science/article/abs/pii/S1353829223001557?via%3Dihub.

DataWorks NC. 2023. *Tenant Demographics and Eviction Filings in Durham County*. Durham, NC: DataWorks NC.

Dowdall, Emily, and Ira Goldstein. 2023. *Eviction Diversion in Philadelphia: Evaluation of Efforts to Reduce Eviction Filings in Two Program Phases.* Philadelphia, PA: City of Philadelphia Reinvestment Fund.

Elliott, Marc N., Allen Fremont, Peter A. Morrison, Philip Pantoja, and Nicole Lurie. 2008. "A New Method for Estimating Race/Ethnicity and Associated Disparities Where Administrative Records Lack Self-Reported Race/Ethnicity." *Health Services Research* 43: 1722–36. https://doi.org/10.1111/j.1475-6773.2008.00854.x.

Elliott, Marc N., Peter A. Morrison, Allen Fremont, Daniel F. McCaffrey, Philip Pantoja, and Nicole Lurie. 2009. "Using the Census Bureau's Surname List to Improve Estimates of Race/Ethnicity and Associated Disparities." *Health Services and Outcomes Research Methodology* 9 (2): 69–83. https://doi.org/10.1007/s10742-009-0047-1.

García-Cobián Richter, Francisca, April Hirsh Urban, Claudia Coulton, Stephen Steh, and Tsui Chan. 2019. *The Cleveland Eviction Study: Downstream Paths of Evictions into Homelessness and Loss of Human Capital*. Cleveland, OH: Case Western Reserve University Center on Urban Poverty and Community Development.

Genthon, Kathryn, Andrea Miller, and Miriam Hamilton. 2024. *Race and Ethnicity Data as a Tool for Equal Justice: Practical Considerations for State Courts*. Williamsburg, Virginia: Court Statistics Project.

Gomory, Henry, Douglas S. Massey, James R. Hendrickson, and Mathew Desmond. 2023. "The Racially Disparate Influence of Filing Fees on Eviction Rates." *Housing Policy Debate* 33 (6): 1463–83.

Graetz, Nick, Carl Gershenson, Peter Hepburn, Sonya R. Porte, Danielle H. Sandler, and Mathew Desmond. 2023. "A Comprehensive Demographic Profile of the US Evicted Population." *PNAS* 120 (41): e2305860120.

Haley, Jennifer M, Lisa Dubay, Bowen Garrett, Clara Alvarez Caraveo, Ilyse Schuman, Katy Johnson, Jason Hammersla, James Klein, Jay Bhatt, David Rabinowitz, Heather Nelson, and Becca Depoy. 2022. *Collection of Race and Ethnicity Data for Use by Health Plans to Advance Health Equity*. Washington, DC. Urban Institute.

Harris, Ada. *Using Bayesian Improved Surname Geocoding (BISG) to Classify Race and Ethnicity in Administrative Employment Data by Industry: A Validation Study*. Presented at Joint Statistical Meetings, August 6, 2020.

Hendey, Leah, Elizabeth Burton, and Kathryn L. S. Pettit. "Commentary: Improving Housing Policy with Neighborhood Data." 2024. *Cityscape: A Journal of Policy Development and Research* 26 (1): 183–91.

Hepburn, Peter, Renee Louis, and Mathew Desmond. 2020. "Racial and Gender Disparities Among Evicted Americans." *Sociological Science* 7: 649–62. https://doi.org/10.15195/v7.a27.

Imai, K., Santiago Olivella, and Evan T. R. Rosenman. 2022. "Addressing Census Data Problems in Race Imputation via Fully Bayesian Improved Surname Geocoding and Name Supplements." *Science Advances* 8 (49): 1–10. https://doi.org/10.1126/sciadv.adc9824.

Lens, Michael C., Kyle Nelson, Ashley Gromis, and Yiwen Kuai. 2020. "The Neighborhood Context of Eviction in Southern California." *City & Community* 19 (4): 912–32. https://doi.org/10.1111/cico.12487.

Martino, Steven C., Megan Mathews, Cheryl L. Damberg, Joshua S. Mallett, Nate Orr, Judy H. Ng, Denis Agniel, Loida Tamayo, and Marc N. Elliott. 2021. "Voluntary Disenrollment from Medicare Advantage Plans Is Higher among Racial and Ethnic Minorities." *Health Services Research* S1 (55): 54–55. https://doi.org/10.1097/mlr.0000000000001574.

McCartan, Cory, Robin Fisher, Jacob Goldin, Daniel E. Ho, and Kosuke Imai. 2024. "Estimating Racial Disparities When Race is Not Observed." NBER Working Paper No. w32373. Cambridge, MA: National Bureau of Economic Research.

Medina, Richard M., Kara Byrne, Simon Brewer, and Emily A. Nicolosi. 2020. "Housing Inequalities: Eviction Patterns in Salt Lake County, Utah." *Cities* 104: 102804. https://doi.org/10.1016/j.cities.2020.102804.

NCSC (National Center for State Courts). 2024. *Reimagining Housing Court: A Framework for Court-Based Eviction Diversion*. Eviction Diversion Initiative Interim Report. US and International: NCSC.

O'Hara, Amy, and Stephanie Straus. 2024. *Linking Demographic Characteristics to Eviction Records: A Preliminary Analysis*. Washington, DC: Massive Data Institute, McCourt School of Public Policy, Georgetown University.

Reynolds, Kathryn, Katie Fallon, Owen Noble, Abby Boshart, Lee D. Evans, and Andrew Jakabovics. 2023. *Preventing and Mitigating Evictions After the COVID-19 Crisis*. Washington, DC: Urban Institute.

Urban, April Hirsh, Aleksandra Tyler, Francisca García-Cobián Richter, Claudia Coulton, and Tsui Chan. 2019. *The Cleveland Eviction Study: Observations in Eviction Court and the Stories of People Facing Eviction*. Cleveland, OH: Case Western Reserve University.

About the Authors

Alena Stern is the chief data scientist at the Urban Institute, leading Urban's data science team. She is also a member of the Racial Equity Analytics Lab within the Race and Equity Division. Her research focuses on creating data science tools and informing policy solutions to advance equity and inclusion.

Manuel Alcalá Kovalski is a data scientist at the Urban Institute's Office of Technology and Data Science, where he applies diverse data science methodologies—such as machine learning, web scraping, and spatial analysis—to drive equitable policy research. He co-leads projects predicting neighborhood change, developing interactive dashboards for public agencies, and analyzing judicial disparities through imputation techniques. His past work includes NLP-based extraction from federal documents, equity modeling for autonomous ride-share fleets, and dashboard development for local governments tracking federal recovery spending.

Leah Hendey is a principal research associate in the Housing and Communities Division at the Urban Institute. She also codirects the National Neighborhood Indicators Partnership.

Elizabeth Burton is a former research analyst in the Housing and Communities Division at the Urban Institute. Her research interests include affordable housing, with a focus on eviction prevention and decommodified housing, and neighborhood change. At Urban, she provided technical assistance to Promise Neighborhood grantees and helped manage the National Neighborhood Indicators Partnership.

Sandra Ambrozy is a nonresident fellow in Urban's Center on Nonprofits and Philanthropy and a senior fellow at the Full Frame Initiative. Her work advances a fair civil justice system for all that is coordinated, accountable, and person-centered.

Carlos Manjarrez is a senior fellow at the Justice Lab at the Institute for Technology Law and Policy within the Georgetown University Law Center. He advises and supports the Civil Legal Data Commons Project. He previously served as the chief data officer for the Legal Services Corporation (LSC) where he was the founding director of the Office of Data Governance and Analysis, conducted LSC's first national survey of legal needs, and established the Civil Court Data Initiative. Prior to LSC, he served as the inaugural director of the Office of Planning Research and Evaluation at the Institute of Museum and

Library Services. Dr. Manjarrez currently serves as an advisor to the National Center for Access to Justice. Carlos holds a BA in sociology and Latino studies from the University of Michigan and a PhD in urban and regional planning from the University of Maryland.

Statement of Independence

The Urban Institute strives to meet the highest standards of integrity and quality in its research and analyses and in the evidence-based policy recommendations offered by its researchers and experts. We believe that operating consistent with the values of independence, rigor, and transparency is essential to maintaining those standards. As an organization, the Urban Institute does not take positions on issues, but it does empower and support its experts in sharing their own evidence-based views and policy recommendations that have been shaped by scholarship. Funders do not determine our research findings or the insights and recommendations of our experts. Urban scholars and experts are expected to be objective and follow the evidence wherever it may lead.

URBAN INSTITUTE · ELEVATE · THE · DEBATE

500 L'Enfant Plaza SW
Washington, DC 20024

www.urban.org