URBAN INSTITUTE

Opening the "Black Box" of Tenant Screening

Analyzing Data Matches in Court Data

Judah Axelrod, Brendan Chen, Katie Fallon, and Sonia Torres Rodríguez

March 2025

The rise of new data gathering and management technologies has led to an increase in companies aggregating and selling compiled datasets—which may include details such as people's income, credit score, and eviction and criminal histories—to landlords to inform their rental decisions. Research shows that landlords overwhelmingly rely on these tenant screening reports. But given the proprietary nature of the datasets, there is a lack of transparency in how they are assembled and managed, which can hide poor data matching and management practices and result in applicants being incorrectly denied housing. This brief explores two components of tenant screening databases—eviction filings and criminal records—to highlight the challenges of accurately matching records without unique identifiers. We analyze the limits of data matching and provide avenues for improved practices and greater protections for renters.

What Are Tenant Screening Reports?

Tenant screening is the process by which landlords receive and use information about applicants to make leasing decisions. In recent years, new data gathering technologies have allowed more organizations to assemble background information on prospective tenants, which can include credit and income details, eviction filings, criminal records, and rental history (CFPB 2022).

A vast majority of landlords—almost 90 percent—use tenant screening reports to select tenants (Roman and Travis 2004; Shiffer-Sebba 2021).[1] After submitting an application, potential renters are usually required to pay a fee to have their report shared with the landlord. Specific reports vary in what they present to the landlord. Some companies list all related records for an individual (see figure 1), while other companies generate risk scores or recommendations, such as a score from 1 to 10 or a green checkmark or a red "X" (see figure 2). An analysis by Consumer Reports found that the eight most prominent tenant screening companies include an algorithmically generated score or a recommendation to accept or reject an applicant.

FIGURE 1

Example of Tenant Screening Report

Evictions Report

WARNING! Not every State or County provides their EVICTION RECORDS to the National Eviction Databases. TenantTracks uses the most comprehensive database of eviction records available but you should be aware there is no database available with 100% of all records that can be searched for tenant screening purposes.

Case 1: HHCCCC23423

DEFENDANT: **JOE DOE**	PLAINTIFF: **JOANNA DOE**
Address: 10, None New Britain, CT - 6082	Court Description: Housing Session - Hartford Filing Type Description: Forcible Entry/detainer Filling Date: 10/12/2012
Amount: $-	Release Date: -

Suits, Liens & Judgments Report

WARNING! Not every State or County provides their EVICTION and CIVIL COURT RECORDS to the National Eviction Databases. TenantTracks uses the most comprehensive database of eviction and civil records available but you should be aware there is no database available with 100% of all records that can be searched for tenant screening purposes.

Case 1: HHCCCC23423

DEFENDANT: **JOE DOE**	PLAINTIFF: **JOANNA DOE**
Address: 10, None New Britain, CT - 6082	Court Description: Housing Session - Hartford Filing Type Description: Forcible Entry/detainer Filling Date: 10/12/2012
Amount: $-	Release Date: -

Office of Foreign Assets Control Report

Best name	**Listing**
Joe Doe	Birth Date: 1981-04-18 Address: 831 Aaaa Bbbbb Ct 48206 Source: oig.hhs.leie Notes1: General: NURSING PROFESSION Specialty: NURSE/NURSES AIDE Notes2: Exclusion Type: 1128a2

TenantTracks Performance Report

RECORD 1. Last activity: 12/12/2010

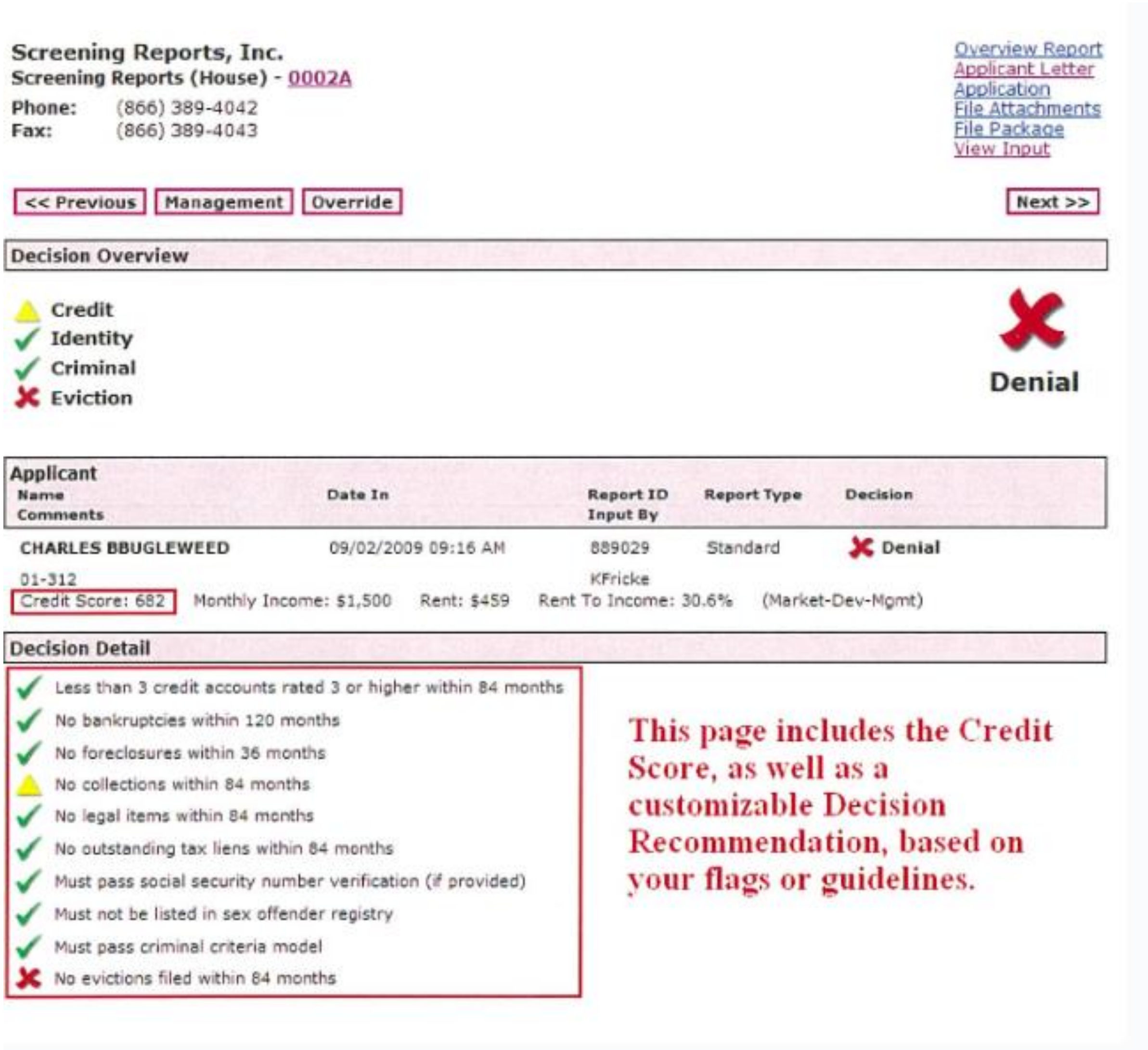
Name: Joe Doe

Employers: AJAX	DOB: XX/XX/1990	Home Phone: 234-567-890	Work Phone: 234-567-890
Address: 10 None St, New Britain, CT - 06050	Cash for keys: 12/12/2010	Notice to quit: 12/12/2010	Summons & Compaints: 12/12/2010
Paid rent late: 12/12/2010	Made partial payment: 12/12/2010	No payment/no legal action: 12/12/2010	Lease violation: 12/12/2010

Source: Tenant Tracks sample screening report. "Tenant Tracks (Optimum 10)," OpenTSS, accessed March 11, 2025, https://open-tss.net/en/companies/15.

FIGURE 2

Example of Tenant Screening Report with Risk Score



Screening Reports, Inc.
Screening Reports (House) - 0002A
Phone: (866) 389-4042
Fax: (866) 389-4043

Overview Report
Applicant Letter
Application
File Attachments
File Package
View Input

<< Previous | Management | Override | Next >>

Decision Overview

Credit
Identity
Criminal
Eviction

Denial

Applicant Name / Comments	Date In	Report ID / Input By	Report Type	Decision
CHARLES BBUGLEWEED / 01-312	09/02/2009 09:16 AM	889029 / KFricke	Standard	Denial

Credit Score: 682 Monthly Income: $1,500 Rent: $459 Rent To Income: 30.6% (Market-Dev-Mgmt)

Decision Detail

- Less than 3 credit accounts rated 3 or higher within 84 months
- No bankruptcies within 120 months
- No foreclosures within 36 months
- No collections within 84 months
- No legal items within 84 months
- No outstanding tax liens within 84 months
- Must pass social security number verification (if provided)
- Must not be listed in sex offender registry
- Must pass criminal criteria model
- No evictions filed within 84 months

This page includes the Credit Score, as well as a customizable Decision Recommendation, based on your flags or guidelines.

Source: BetterNOI sample screening report. "BetterNOI (Screening Reports)," OpenTSS, accessed March 11, 2025, https://open-tss.net/en/companies/35.

What Is the Impact of Tenant Screening Reports?

Landlords use tenant screening reports in varying ways, but studies show they often defer to the reports, especially when the reports include automated recommendations or scores (So 2022; TechEquity 2024). As a result, a single negative item, such as an eviction filing or an arrest record, can bar people from housing in the private market for years, regardless of the outcome of the filing or the accuracy of the data (see box 1) (Eisenberg and Brantley 2024; Housing Justice Center 2021; Shiffer-Sebba 2021; So 2022).[2] Those with negative reports are exposed to high costs of repeated application fees and multiple housing denials, which can cause housing instability and homelessness (Couloute

2018; Duke and Park 2019; Housing Justice Center 2021; Lewis et al. 2019; Reosti 2021; Stanley-Becker 2022; Weiss 2018).[3] This is particularly true in areas with tight housing markets where rental housing is already scarce.

Furthermore, these reports have critical racial equity implications, given that Black, Latinx, and Indigenous people are overrepresented in core data sources used in the reports. A 2020 study found that Black people made up 19.9 percent of all adult renters in sampled counties, but 32.7 percent of all eviction filing defendants (Lake and Tupper 2021).[4] Similarly, Black, Latinx, and Indigenous people are overrepresented in arrest and criminal record data because of racist policing and incarceration practices (Barak, Leighton, and Flavin 2007; Nellis 2021; Pinard 2013).[5]

Asian, Black, and Latinx people are more at risk of receiving negative tenant screening reports because of the inherent challenges with data matching (Stanley-Becker 2022). Given lower levels of diversity in surnames, people in these groups are more likely to be incorrectly matched with other people, and as a result, to be associated with negative history.[6] They are more likely to be in disqualifying data sources and also at higher risk of being incorrectly associated with an eviction filing or criminal record. They therefore face comparatively high risk of denials in the housing market.

BOX 1

The Challenge of Having Unified and Consistent Court Records

The complex court system in the United States makes data collection and standardization at the state and national levels challenging. Only a few states have standardized and aggregated court records for criminal data and eviction filings.[a] This is in part because court records are highly localized, conforming to the requirements, data capabilities, and data-entry processes of local courts.

Court systems vary in how records are maintained and how publicly accessible they are.[b] While some courts facilitate web-based data sharing, others do not. And some states like California, for example, seal eviction records at the point of filing and unseal them depending on the outcome.[c] As a result, there is substantial variation in who and what can be captured.

Court records also vary in the detail and accuracy of data. While most court systems include names and geographic locations, given the variation in information systems, the ability to include components such as a middle name, a hyphen into a last name, and geographic specificity varies. For example, some records include the full address while others include only the county, many records do not include details like the date of birth, and almost none of the records include a unique identifier (e.g., social security number). Accuracy of the data also varies based on local staffing capacity. Many court records include misspellings or inaccurate details in the following two ways:

- **Inaccurate source data:** Court records contain data-entry mistakes, including misspelled names. Incorrect or incomplete names make matching data and accurately identifying individuals with eviction filings difficult.
- **Incomplete outcome data:** Outcomes of court filings are often omitted, vague, or misleading.[d] A study analyzing 3.6 million administrative eviction court records from 12 states found that, on average, 22 percent of eviction records contained ambiguous information on how the case was resolved, and cases with more complex disputes (e.g., cases with multiple tenants and lawyers) were more likely to contain inaccuracies.[e]

[a] See, for example, "Examples of Statewide Criminal Justice Data Repositories," Ohio Criminal Sentencing Commission, updated February 2023, https://www.supremecourt.ohio.gov/docs/Boards/Sentencing/committees/uniformSentEntry/RepositoryExamplesFAQ.pdf.
[b] See US Government Accountability Office, *Evictions: National Data Are Limited and Challenging to Collect* (Washington, DC: GAO, 2024), https://www.gao.gov/products/gao-24-106637; and US Department of Housing and Urban Development, *Report to Congress on the Feasibility of Creating a National Evictions Database* (Washington, DC: HUD, Office of Policy Development and Research, 2021), https://www.huduser.gov/portal/publications/Eviction-Database-Feasibility-Report-to-Congress-2021.html.
[c] Lauren Fung, Isabella Remor, Katie Fallon, and Nyla Holland, "Masking the Scarlet 'E': A Study on California's Attempt to Mask Eviction Records through AB 2819" (Washington, DC: Urban Institute, 2023), https://www.urban.org/research/publication/masking-scarlet-e.
[d] Rudy Kleysteuber, "Tenant Screening Thirty Years Later: A Statutory Proposal to Protect Public Records," *The Yale Law Journal* 116 (2007): 1344–88, https://www.yalelawjournal.org/pdf/539_3abmt4eo.pdf.
[e] Adam Porton, Ashley Gromis, and Matthew Desmond, "Inaccuracies in Eviction Records: Implications for Renter and Researchers," *Housing Policy Debate* 31, no. 3–5 (2020): 377–94, https://doi.org/10.1080/10511482.2020.1748084.

Current Tenant Screening Policy and Regulation

Current regulation of tenant screening practices and companies is quite limited. Tenant screening companies operate like "black boxes," with little insight into their data collection and management practices that undoubtedly affect the accuracy of the reports and the scope of their impact (Bhatia 2020; Desmond and Bell 2015; Desmond and Shollenberger 2015; Pasley, Oostrom-Shah, and Sirota 2021; Reosti 2020).

Federal agencies have released some guidance to help minimize harm caused by the screening reports, including instructions on tenant rights from the Consumer Financial Protection Bureau, Federal Trade Commission, US Department of Housing and Urban Development, among other agencies.[7] However, existing regulation is limited and primarily relies on the Fair Housing Act, which protects people from discrimination in the sale, rental, and financing of housing; or the Fair Credit Reporting Act, which requires that companies adopt procedures to ensure reasonable accuracy of their reports.

For example, a lawsuit was filed against CoreLogic stating that the company should be liable under the Fair Housing Act on a disparate impact claim, because it produced a background report disqualifying a tenant applicant based on a prior arrest that did not result in a conviction.[8] Similarly, a lawsuit was filed against RealPage alleging that the company violated the Fair Credit Reporting Act, which caused consumers with similar names to be incorrectly associated with criminal records.[9]

While these two regulations provide some oversight, tenant reporting companies may evade their requirements by adding a disclaimer. Likewise, the burden is currently on the renters to identify inaccuracies and sue the companies. While applicants can obtain a copy of their tenant screening reports for free and correct or supplement any incomplete or incorrect information, they are often difficult to obtain. Moreover, prospective tenants are often too busy searching for housing to file a lawsuit (TechEquity 2024).

To augment the federal guidance, some states and cities have limited the look-back period (the number of years since the most recent record) or consideration of certain records. For example, fair chance laws aim to limit adverse actions landlords can take against applicants based on criminal history, and Philadelphia's Renters' Access Act aims to reduce landlords' ability to use specific eviction records when screening tenants.[10] Yet, monitoring and enforcement is limited for these policies as well, and there are few mechanisms—outside renter-initiated lawsuits—to hold various companies accountable. The tenant screening companies are largely left unregulated because of the decentralized nature of tenant screening companies and landlord decisionmaking, limited resources for monitoring and enforcement, and reliance on tenants to report lack of adherence.

Given the outsized role tenant screening reports play in housing access, evidence is needed to identify pathways to mitigate their harmful impacts. We address this by opening part of the "black box" of tenant screening and assessing the implications of various data management practices for two types of data: eviction filings and criminal history. We analyze the challenges companies face when trying to link unique individuals, given the existing limitations in court data. We highlight the ways current data capture and management practices can cause harm by incorrectly linking individuals with incorrect and inaccurate records, which may limit their access to housing. Finally, we conclude with actionable steps to help key stakeholders encourage more accurate and equitable screening reports.

Data and Methods

Below is a summary of how we used exploratory interviews and data to reverse engineer the data aggregation process tenant screening companies likely follow. We also discuss the quality and quantity of information we obtained, our approaches to standardizing and harmonizing records at the individual level, and the model input data to tenant screening we created, as well as the assumptions and scenarios that arose throughout our analysis.

Exploratory Interviews

We interviewed housing advocates, people who work with court data, and researchers focused on tenant screening to understand data contained within the tenant screening reports, challenges stakeholders have with the reports, and knowledge gaps surrounding the reports.

Following are key challenges that informed much of our analysis:

- **Access:** Gaining access to tenant screening reports and data included in them is challenging. Screening companies do not share data or their data management practices, which makes regulation, monitoring, and enforcement difficult.
- **Data capture:** A lack of transparency in how companies capture and manage data creates challenges for understanding the representativeness, completeness, and accuracy of the records collected. Some court systems, for example, seal eviction filings until the defendant is found guilty, while others make filings public immediately. Advocates note that less data

sharing helps to prevent adverse consequences for tenants. However, it is unclear how tenant screening companies account for these differences in data availability.

- **Name matching:** Companies often rely on first and last names only, with some geographic identifiers, to match court records. Although federal guidance warns against the use of name-only matching, given that court records lack unique identifiers, it is unclear how tenant screening companies are matching and using these records.
- **Record updates and accuracy:** Companies do not provide information on how they remove or review records that have been sealed, expunged, or updated (CFPB 2022). Renters have reported experiencing challenges when trying to address data errors in their reports generated by a single company, as well as seeing the same errors appear across records generated by multiple companies, which suggest that many companies may be using inaccurate data and/or making consistent errors in their data matching and management systems. For example, AppFolio settled with the Federal Trade Commission for allegedly failing to check whether court records purchased from a third party and used in tenant screening reports had been sealed.[11]

Data Access

To ensure the accuracy and completeness of screening records, we used data from Pennsylvania. The state has a unified court system and has also instituted clean slate laws, which "seal older, more minor, or nonconviction criminal records," at least in principle.[12] We chose Pennsylvania because its data would allow us to conduct additional research on the impact of those policies. We worked with the Administrative Office of Pennsylvania Courts (AOPC) to obtain desired case- and docket-level data on criminal and landlord-tenant case records for the period of 2014–24.[13]

The details included in the criminal data and the landlord-tenant data varied substantially. The fields in the criminal records were more expansive and more likely to be complete than in the eviction records (see table A.1). This variation enabled us to better understand how dataset quality and completeness affect the feasibility and reliability of matches.

Methods

For our analysis, we cleaned, de-duplicated, and matched criminal and landlord-tenant records using a range of probabilistic (or "fuzzy") matching techniques on fields such as name and geography to examine how various decisions and assumptions affect racial bias in outcomes, including racially disproportionate representation in the datasets and prevalence of low confidence matches.

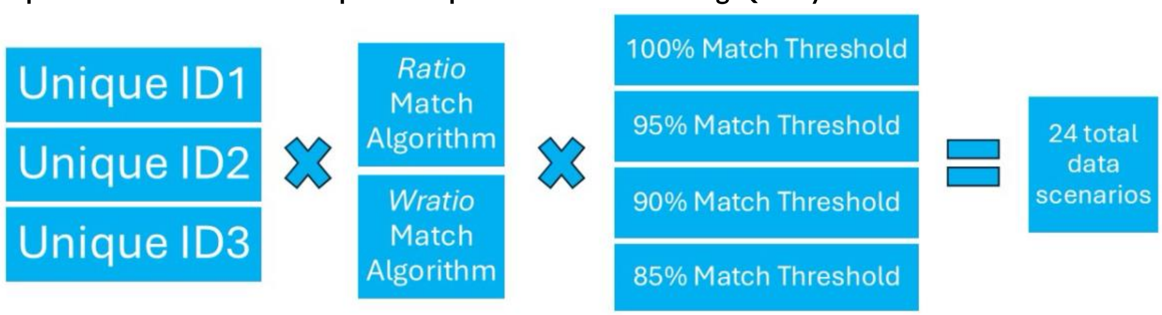
We created 24 data scenarios for which we make differing assumptions about how to assign case records to a particular individual and how to match those individuals across criminal and eviction databases. These scenarios were designed to highlight the extent to which the total number of people—and their corresponding records—varies based on subjective data assumptions. We used the following three dimensions to measure the quality of the data matches:

1. “Unique ID” refers to how different case records are attributed to unique individuals.
2. Match algorithm is how we “score” the similarity of two individuals when combining the criminal and landlord-tenant records.
3. Match threshold is the numeric cutoff point at which we determine an algorithm’s score to be a match.

Figure 3 and the section below describe how the 24 data scenarios were created and explain what each of the three unique identifiers, two different match algorithms, and four different match thresholds mean.

In general, these scenarios explore how different ways of applying the three dimensions affect the overall dataset. For example, the strictest scenario (unique ID1, *ratio* match algorithm, and 100 percent threshold) illustrates what a tenant screening database would look like if the data assumptions were set as rigorously as possible. Each scenario becomes progressively less strict, ending with the most lenient scenario (unique ID3, *wratio* match algorithm, and 85 percent threshold), which shows the effect of relaxed assumptions. We explore variations among each of these scenarios to demonstrate how the thresholds affect how people are captured in tenant screening reports.

FIGURE 3
Unique Data Scenarios to Compare Unique-Individual Matching Quality across Datasets



Source: Authors created the image.
Notes: The image visualizes our methodological approach to exploring different unique identifiers, match algorithms, and thresholds, and their impact on who is captured in the harmonized dataset. The unique identifiers range from most conservative to least conservative matching, where “Unique ID1” requires the same name and nonmissing date of birth, “Unique ID2” requires the same name and a missing date of birth that is considered unique only if the zip codes are the same, and “Unique ID3” requires name-only matching. The probabilistic matching algorithms include a *ratio* score and a composite, more lenient *wratio* score. The various thresholds, ranging from 100 to 85 percent, require different levels of text similarity when identifying a match between the criminal and landlord-tenant datasets.

Unique Identifiers of Individuals in the Data

The data we received from AOPC were much cleaner than the courts’ raw data, but we still performed a number of cleaning steps, including dropping corporate defendants; splitting multi-defendant dockets into individual records; removing extraneous text from names; and organizing offenses, grades, and dispositions into meaningful categories.

Court records, even within unified court systems, do not have established identifiers for determining what constitutes a unique individual or defendant. Dockets do not clearly flag individuals who have been found guilty of multiple offenses. For example, there may be three records indicating that "John Smith" was convicted of a crime, which can mean one John Smith who was convicted three times or three John Smiths who were each convicted of a crime. While name-only matching practices across different types of data in the tenant screening industry have received national attention and scrutiny,[14] matching records *within* the same data source to a unique individual is a similarly subjective process that can have enormous downstream repercussions. The more lenient the de-duplication process, the more records are attributed to the same individual.

Because the court records did not have unique identifiers, to test the sensitivity of the unique identifier used on the resulting dataset, we created our own unique IDs under the following three assumptions:

1. **Unique ID1 (strict case):** Any records containing the *same first/last name AND either same date of birth OR same docket number* are considered to pertain to the same individual. All others are assumed to be different individuals.
2. **Unique ID2 (medium case):** Any records matching the conditions for unique ID1 plus any records that have the *same name*, *same zip code*, *AND a missing date of birth* are considered to pertain to the same individual. All others are assumed to be different individuals. Notably, while 99.4 percent of criminal records have a date of birth, only 22.4 percent of landlord-tenant records do (see table A.1 for fields and missingness).
3. **Unique ID3 (lenient case):** Any records matching the conditions for unique ID1 or ID2 plus any records that have the *same name* are considered to pertain to the same individual. All others are assumed to be different individuals.

Matching Algorithms

To assess how different matching scenarios affect outcomes, we implemented a record linkage procedure that harmonized (or combined) the criminal and landlord-tenant records.

After we aggregated the case data to the individual level based on the unique identifiers above (e.g., ID1, ID2, and ID3), we implemented a "blocking" step to limit the possible candidate matches to those that share the same zip code, first initial, and first three letters of the last name. This is a common step in the record linkage literature that vastly reduces the computational intensiveness of probabilistic matching by only considering realistic matches.

Next, we chose matching algorithms from the *TheFuzz* package (previously known *as FuzzyWuzzy*) developed by SeatGeek.[15] This package includes a number of different name-matching algorithms with various use cases (e.g., matching a shorter sequence of text within a longer sequence, matching two sequences when the order is inverted, etc.). We chose the following two algorithms:

1. A simple 0–100 score based on Levenshtein distance, which considers the number of changes (e.g., the "edit distance") needed to convert one name into another (Levenshtein 1966). (This is referred to as *ratio* in the results shown below.)
2. A composite 0–100 score that takes the highest scores from a number of algorithms, which alternatively consider things like shorter text sequences, inverted order, and so on.[16] (This is referred to as *wratio* in the results shown below.)

We made these selections for two reasons. First, they are the "best match possible" approach. Their composite score is more lenient in finding two names to be considered a match, which provides a good contrast to the simple score. Second, they represent two plausible choices for algorithms used in the real world.

Matching Thresholds

Finally, we focus on matching *threshold*. The threshold is essentially the confidence level at which two similar names are considered a match. For example, while "John Doe" and "John W. Doe" may be the same person, because the text strings are not identical, the threshold must be lowered from 100 (which indicates an exact match) to categorize them as a unique individual. Threshold levels are important because they define what level of uncertainty is acceptable. Lowering a threshold can allow for differences in a middle initial, a suffix, or a minor data-entry error—with an understanding that there is an acceptable level of risk that two records are incorrectly matched. Such false positives could lead to criminal or eviction histories being wrongly inflated or incorrectly ascribed to an individual in their tenant screening report. We use 85 as the lowest threshold that provides defensible matches, and 90, 95, and 100 to explore how threshold changes affect outcomes. Box 2 describes how matching thresholds affect results in practice.

BOX 2

Findings from Name-Matching Results

Examining name matches at each threshold level (e.g., 85, 90, 95) uncovers important differences in the quality of matches and raises important questions about what constitutes a "valid" match. Our manual review of name matches reveals some key considerations for deciding which threshold level should count as a valid match. (All names in this box have been changed to preserve anonymity, but we have included the same number of letters and formatting to reflect identical differences to those observed in the original data.)

- While higher threshold scores may lead to many fewer cases of obvious nonmatches between names (e.g., John Doe vs. James Doe), they may also exclude more ambiguous cases that could be argued are valid matches (e.g., David P Smith vs. David Paul Smith).
- The length of names can significantly affect match scores. For example, when two names differ by only one character, matches that involve longer names will have a higher score than matches that involve shorter names, because the one-character difference is a smaller percentage of the

entire name (e.g., Christopher Smith vs. Cristopher Smith will score higher than Chris Smith vs. Cris Smith).

- For threshold scores around 95, most name matches differ only by a missing middle initial (e.g., Amy Jones vs. Amy K Jones).

While the average difference between *ratio* and *wratio* scores for each potential name match was small (1.1 points), there were hundreds of cases where the *wratio* score would have met one of our threshold levels, whereas the *ratio* score did not. These mainly consisted of cases where names were identical except for an additional hyphenated last name (e.g., Jane Doe vs. Jane Doe-Smith). We use specific examples, detailed below, throughout the report to demonstrate how changes in data practice may have real-world consequences.

Following are examples of nonexact matches at each threshold level:

- 85
  - Oliver Jackson | Olivia Jackson
  - Sophia Welch | Sophia Anne Welch
  - Frankline Romero | Frank Romero
  - Diana Guerrero | Dayna Guerrero
- 90
  - Elena White | Elana White
  - Raul Garza | Raul Garza-Ortiz
  - Erik Daniel | Eric Daniel
  - Dylan Phillips | Dylan Jay Phillips
- 95
  - Elliot L Lily | Elliot Lily
  - Victoria Sanchez | Victoria Sanchez-Ortega
  - Reeve James | Reeves James
  - Brian Hill | Brian P Hill
- 100 (This threshold level only allows for exact matches.)

Following are examples of similar names where *ratio* and *wratio* scores differ by more than 15 points:

- Jason Smith | Jason Smith-Alexander
- Jorge Rivera | Jorge Rivera-Ramirez
- Pete Moore | Josef Pete Moore
- Isabella Cook | Isabella Cook-Robinson

Constructing Final Data for Analysis

Our final data for analysis aims to replicate how landlords view and use data to make decisions about which tenants to approve or deny for housing. We appended a number of variables to the harmonized criminal and landlord-tenant data to perform a sensitivity analysis. And we limited our look-back period to seven years, using the offense disposition date on the criminal data and the docket filing date on the landlord-tenant data.[17]

From here, we created several binary variables indicating whether certain types of charges and seriousness of the offenses were present for each individual within the look-back period. We also examined the time since the offense or record.

To analyze the impacts on racial equity, we relied on race and ethnicity as reported in the data from AOPC (box 3). Race and ethnicity are generally unavailable in landlord-tenant case records, whereas for criminal cases, they are much more complete. However, conversations with AOPC officials indicate that they are likely "observed" race and ethnicity, or as recorded by police officers initiating criminal complaints with the court, rather than self-reported by defendants. Quality of the data varies; sometimes the racial and ethnic designation differs even within records that our matching approach identified as being the same person. Therefore, we could only conduct an assessment of racial equity with limited precision, particularly for categories with fewer people represented. While we want to avoid overreliance on flawed data, there are still useful inferences that can be made about the racial impacts of tenant screening practices.

BOX 3

Disaggregating Results by Race and Ethnicity

Our findings in this brief are dependent on limitations and quirks in race/ethnicity information we received from AOPC. First, we do not report ethnicity because it is missing for 55–64 percent of individuals, depending on data assumptions. Second, many individuals appear multiple times in criminal records with different reported races. This is probably because different police or court officers infer different identities when initiating complaints, not because of biracial identity, which is a separate included category. We classify this group, which by definition appears across different cases, as "Ambiguous." And because these individuals all have multiple cases tied to them, they tend to have more extensive histories over the seven-year look-back period than other groups. While most race/ethnicity identifiers are missing in the landlord-tenant records, we believe the same to be true about "observed" race when it is present.

Third, defendant's race is categorized as "Unknown or Unreported" for a larger majority of landlord-tenant records (between 79–84 percent) and a smaller number of criminal records (2–3 percent). For this reason, individuals are more likely to have criminal histories and less likely to have eviction histories according to race. Finally, due to limitations in sample size, we aggregate Asian, Asian/Pacific Islander, and Native Hawaiian/Pacific Islander people into an "AANHPI" category.

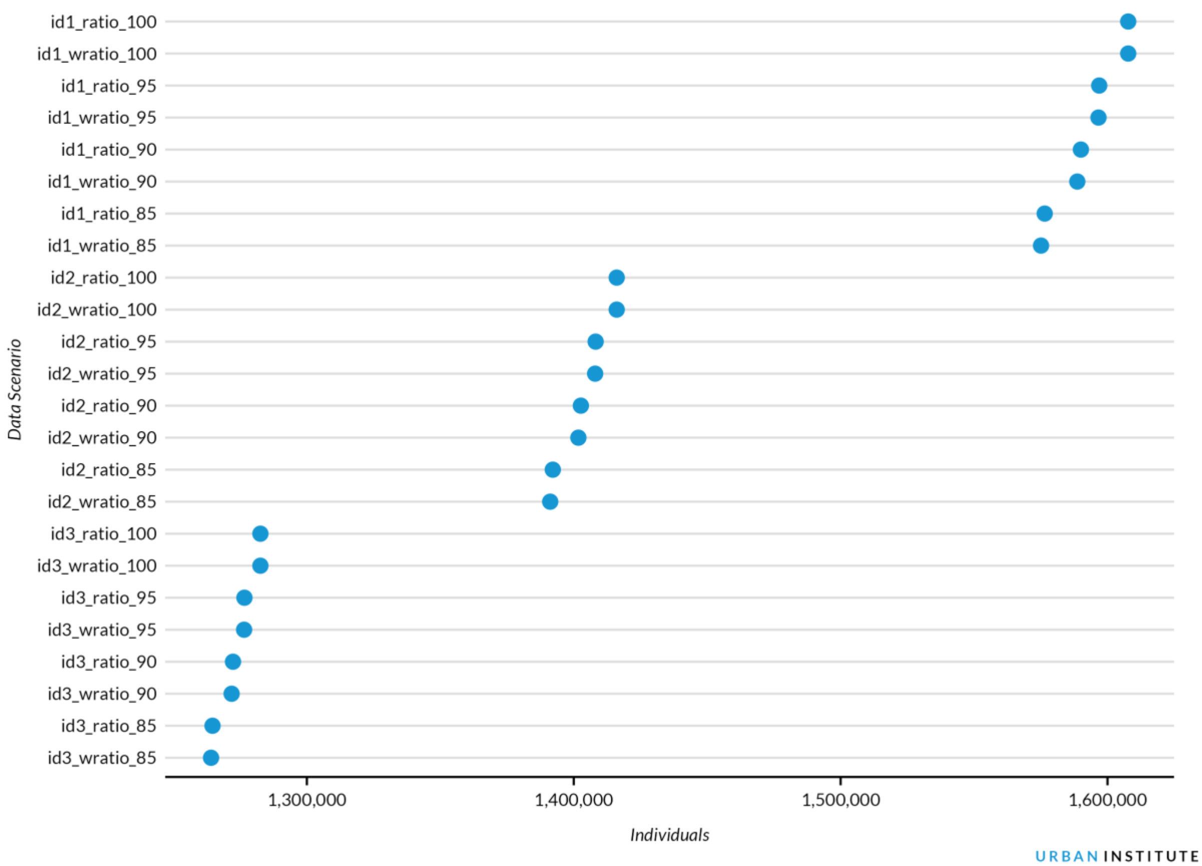
Findings

How Many People Are Captured in Court Data?

Findings from the 24 data scenarios demonstrate that decisions about name matching, thresholds, and identification create vastly different outcomes for tenant data. Figure 4 displays the total number of individuals in our dataset, which ranges from 1.26 to 1.61 million, depending on matching decisions. This is a swing of about 350,000 individuals. Moreover, between 40–52 percent have an eviction record and between 56–69 percent have a criminal record. This represents between 10–12 percent of the total Pennsylvania population.

FIGURE 4

Number of Individuals Captured in Court Data Varies Widely Based on Data Matching Assumptions



Source: Authors' analysis of data from the Administrative Office of Pennsylvania Courts (AOPC).
Notes: Within each of the 24 data scenarios, "id" refers to how records are collapsed within datasets, "ratio/wratio" refers to algorithm used for matching, and the numbers refer to the matching threshold.

The type of name de-duplication method used within the individual datasets matters a great deal: there were 27 percent more distinct people captured in our strictest matching scenario, compared with our most lenient name-only matching scenario. This matters in the real world, because if companies practice name-only matching, they may be incorrectly associating a large number of cases with a single individual. And multiple case records may flag the individual as medium or high risk, which could result in landlords rejecting their housing application. As detailed above, these data management decisions are subjective, and rental applicants and landlords currently have no visibility into the methods used.

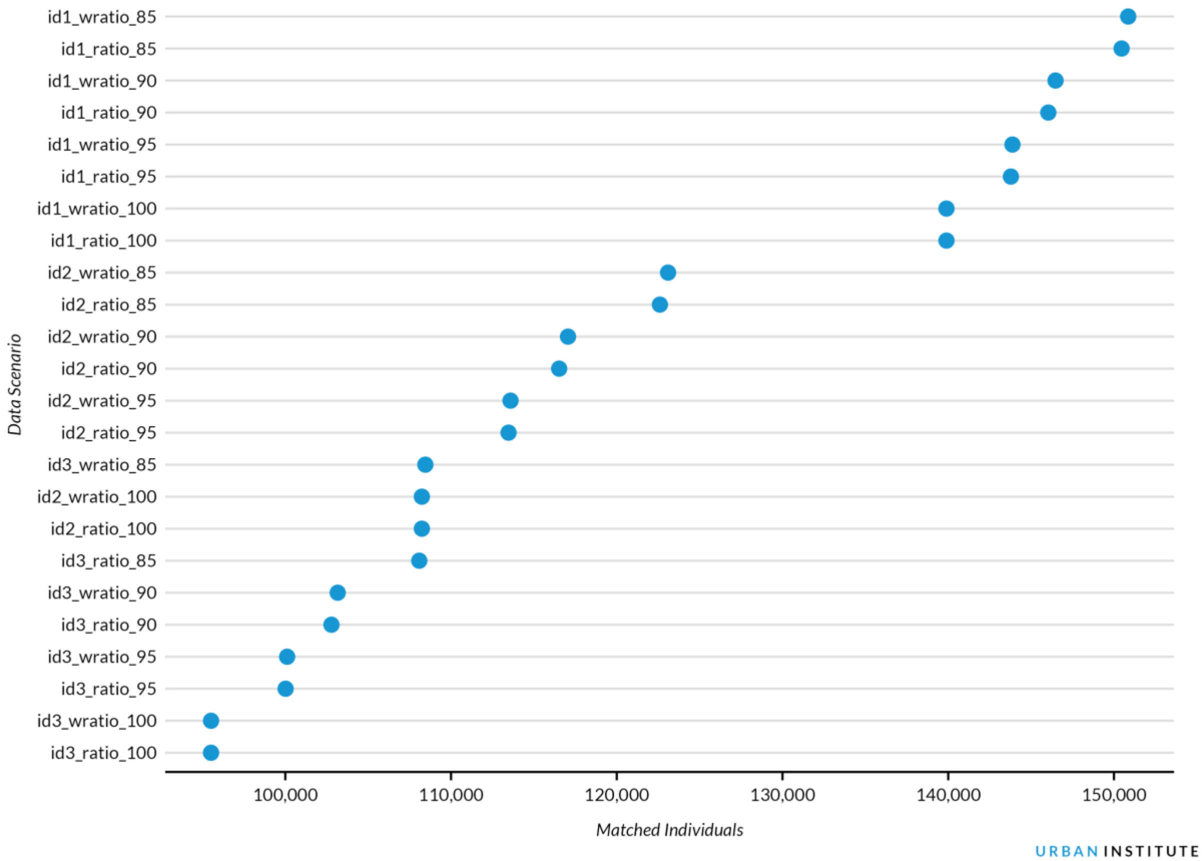
Importantly, our data includes only individuals with some sort of criminal and/or eviction record, whereas tenant screening companies, in practice, collect data on all rental applicants. Therefore, the proportion of the total Pennsylvania population captured in tenant screening data is likely far higher, which also means higher risk of attributing court records to the wrong individual.

After limiting the view to individuals who appear in both datasets, between 96,000 and 151,000 people appear as defendants, which is a swing of 55,000 people in terms of the number of individuals who have *both* eviction and criminal histories (figure 5). This highlights the large variation in how different, potentially disqualifying records can be linked to the same individual. Lenient name de-duplication methods and match thresholds lead to significant jumps in sample size—and sometimes that jump is sudden and nonlinear based on quirks in the data.

Another key pattern emerges in figures 4 and 5: out of the three dimensions in our 24 data scenarios, unique IDs tend to have the highest impact on the total number of individuals. It is also noteworthy that in terms of the composition of the data, matching decisions *within* the criminal and landlord-tenant records play a more substantial role than matching decisions *between* the two datasets.

FIGURE 5

Number of Individuals Found in Both Criminal and Landlord-Tenant Records Are Highly Dependent on Data Assumptions



Source: Authors' analysis of data from the Administrative Office of Pennsylvania Courts (AOPC).
Notes: Within each of the 24 data scenarios, "id" refers to how records are collapsed within datasets, "ratio/wratio" refers to algorithm used for matching, and the numbers refer to the matching threshold.

How Matching Thresholds May Affect Associated Criminal and Eviction Histories

Decisions around which entries count as a different or "unique" individual and which score thresholds constitute a valid match across datasets can significantly affect the associated criminal and eviction history for certain people. The examples below illustrate how variations across both ID levels and score thresholds can impact the way individuals may be linked when matching data. (All names have been changed to preserve anonymity, but we have included the same naming patterns to reflect identical differences to those observed in the original data.)

VARIATIONS IN DE-DUPLICATION ID LEVELS

Considerations around which entries are attributed to unique individuals can greatly affect the number of criminal or eviction records associated with a person. Under our strictest data de-duplication (ID1) and matching threshold (score of 100), a person in our dataset named "Jose Gomez" had a record that included one guilty misdemeanor disposition from 2020, one guilty felony disposition from 2021, and one eviction filing that was eventually settled from 2018. However, when we kept the scoring threshold at 100 but moved the de-duplication scenario from our strictest (ID1) to our most lenient (ID3), the number of records associated with one individual significantly increased. Furthermore, when we used exact name-only matching under our most lenient scenario, the name "Jose Gomez" was associated with ten misdemeanors and eight felonies from 2013 to 2024 and fourteen evictions from 2015 to 2023, although it is unlikely that all these criminal records belong to the same person, given the common name.

Data de-duplication decisions can have significant implications for people, so it is important to separate records for more common names to distinguish individuals. But limited demographic variables or other information may make it difficult to know when records should be assigned to a single individual. This can be particularly problematic for groups like Latinx people who have lower surname diversity than other groups.

VARIATIONS IN NAME-MATCHING SCORE THRESHOLDS

Increases in associated records can occur when the name-matching score threshold is relaxed even by a small amount. Under our strictest data de-duplication and matching scenario, a person with the name "Cameron Nash" was linked to two guilty misdemeanor dispositions from 2015 and 2022 and two eviction filings from 2022 and 2023. When we decreased the name-matching score threshold from 100 to 95, this person was linked to two additional guilty felony dispositions from 2019 and 2023 that were associated with the name "Cameron Oliver Nash," as well as an additional guilty misdemeanor disposition from 2023 associated with the name "Cameron O Nash." Therefore, our original "Cameron Nash" went from being linked to two misdemeanors and two evictions to being linked to three felonies, two misdemeanors, and two evictions. Because we did not change the ID level in this example, all three names—"Cameron Nash," "Cameron O Nash," and "Cameron Oliver Nash"—shared the same date of birth and zip code, making it much more likely that all records associated with these names belong to the same person.

While anecdotal, these examples highlight the need for careful consideration around both which records should count as unique individuals and which score threshold should be used when asserting matched records across datasets. A high name-matching score threshold may work to exclude obvious nonmatches, but it may also prevent more ambiguous cases, such as those involving middle names or initials, that could likely be the same person from being linked. Given the impact that tenant screening reports have on housing access, the risks and trade-offs should be balanced when making decisions on thresholds.

In much of the analysis below, we narrow our exploration to just variations across unique IDs (ID1, ID2, and ID3) to focus on the clearest data-driven explanations for the changes in matches. Any calculations and results displayed for a particular ID are averages across all match algorithms and thresholds *that use that ID*; in other words, we aggregate our 24 scenarios into just three clusters of scenarios visible in figure 4 to tell a clearer story.

Matching Technique Affects Racial and Ethnic Composition of Data

Data management practices can skew the representation of people within the AOPC dataset. As a result, the people captured in the data are disproportionately more likely to be Black. For example, while Black people make up only 11 percent of the state population, about 30 percent of criminal records and 37 percent of landlord-tenant records (with nonmissing race) are attributed to Black defendants. In contrast, white people make up 75 percent of the state population but only 68 percent of criminal records and 62 percent of landlord-tenant records for which race is available.

Representation also varies, albeit slightly, by matching processes. Table 1 shows how the number of people by race changes between the most lenient and most strict data scenarios. The strictest scenario captures a larger number of individuals across all racial groups, except for the "Ambiguous" category in which multiple races are reported for the same individual. The number of AANHPI, biracial, Black, Native American/Alaskan Native, and white defendants all see a 4 to 7 percent decrease from most strict to most lenient scenarios. The major shifts are for individuals whose race is unknown or ambiguous. The number of individuals in the "Unknown" category shrinks substantially with lenient matching assumptions, as more records that are missing race are collapsed and attributed to the same individual. And the number of individuals in the "Ambiguous" category, while a small number overall, nearly triple. This is likely because as more records are aggregated, there is a greater chance that reported race might differ among the records and therefore moved into this category. The thousands of ambiguous records speak to the challenge of maintaining reliable racial data in court files, even when race is not missing.

TABLE 1

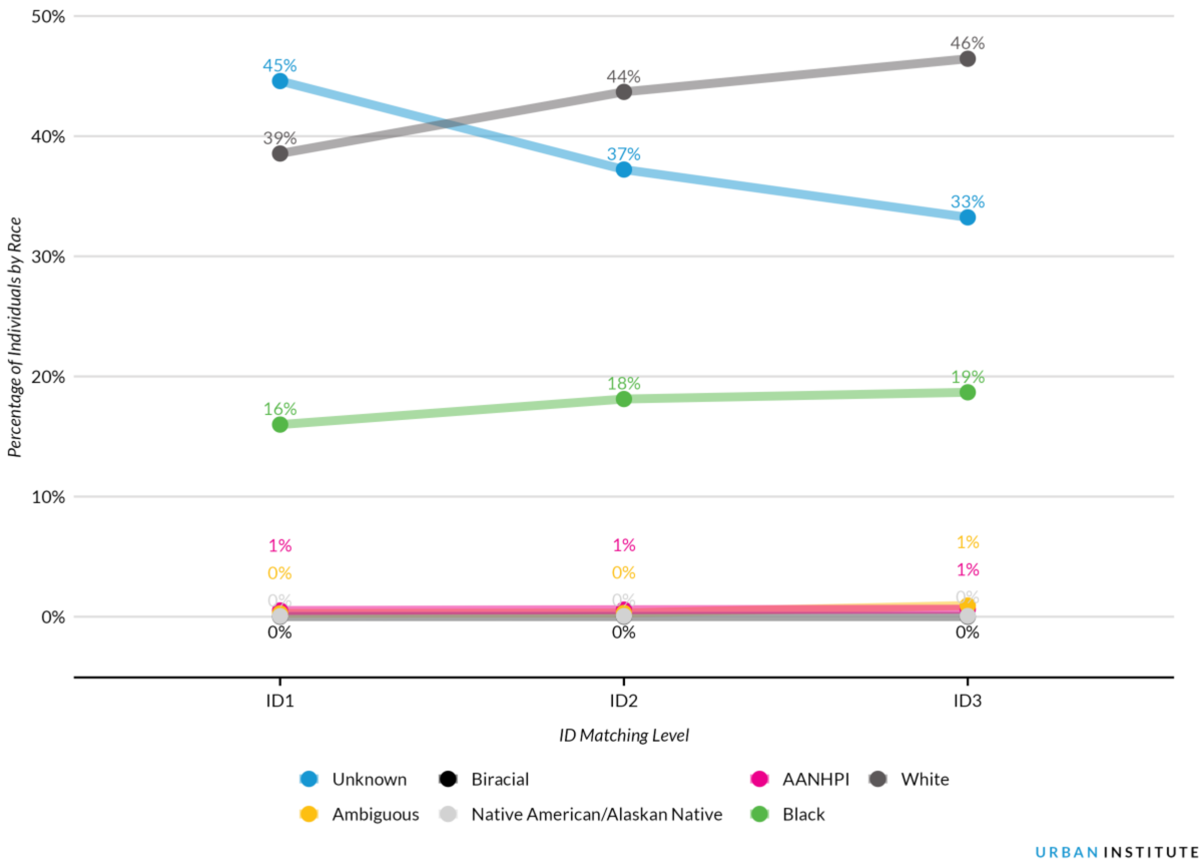
Number of Individuals and Percentages by Race in Lenient and Strict Scenarios

Race	Most Strict	Most Lenient	Difference (%)
AANHPI	8,264	7,945	–319 (–4%)
Ambiguous	4,205	11,999	7,794 (185%)
Biracial	259	248	–11 (–4%)
Black	255,227	237,532	–17,695 (–7%)
Native American/Alaskan Native	786	749	–37 (–5%)
Unknown	724, 223	414,695	–309,528 (–43%)
White	614,791	590,924	–23,867 (–4%)

Figure 6 shows the racial demographics of individuals across the three ID types—again, aggregating from our original 24 data scenarios. We can see major jumps for Black and white racial groups as name de-duplication method changes from ID1 to ID2 to ID3. While we see from table 1 that the ID assumption does not disproportionately affect any particular racial category, figure 6 illustrates that the representation of white and Black people (the two largest nonmissing categories) increases moderately as data assumptions become more lenient. The key takeaway is that as assumptions are relaxed, more records previously identified as "Unknown" are collapsed together, lowering the overall share in favor of other racial categories. Rather than indicating misrepresentation for any one group, these data reveal fundamental limitations in how race is measured across records, individuals, and case types, which are only amplified under our different matching scenarios. Researchers have applied analytical techniques, such as race imputation and data linkage, to the court and credit data used by tenant screening companies to look more deeply at outcomes by race and ethnicity (Farrell et al. 2020; Stern et al. 2025). Future work should similarly explore how to bolster the quality of demographic information in administrative data.

FIGURE 6

Representation of Black and White Defendants Increases as Data Matching Assumptions Become More Lenient



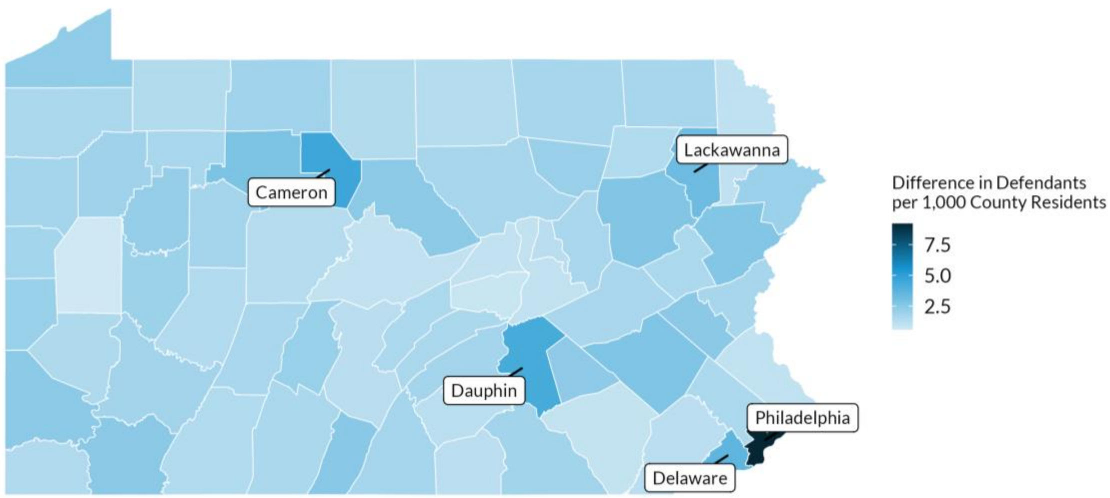
Source: Authors' analysis of data from the Administrative Office of Pennsylvania Courts (AOPC).
Notes: "ID" refers to how records are collapsed within datasets, and each point averages across all data scenarios for that ID. "Ambiguous" is used when the same individual is classified as different races across records. "Unknown" is used when race is not reported in the data.

Figures 7 and 8 provide a geographic look at how the number of individuals with criminal and eviction histories vary by county between our most strict and most lenient matching assumptions in all our original 24 data scenarios. Figure 7 shows that the approach to matching has a minimal impact on the number of people with a criminal history, with single-digit changes per 1,000 county residents. Philadelphia County has the largest change, suggesting that residents in that county may be more at risk of being linked to criminal records under more lenient standards for data matching.

FIGURE 7

Number of Criminal Defendants by County Is Similar Regardless of Data Matching Scenario

Difference between most strict scenario and most lenient scenario



URBAN INSTITUTE

Source: Authors' analysis of data from the Administrative Office of Pennsylvania Courts (AOPC).
Notes: Individuals who appear across multiple geographies are only counted for the first geography in which they appear. Court records are across the entire 2014–24 sample. Landlord-tenant case records are not available for Philadelphia County. Calculations are based on the number of individuals per 1,000 residents in the county according to the 2020 Decennial Census. Zip code–level calculations are allocated to counties based on the US Department of Housing and Urban Development's United States Postal Service residential data.

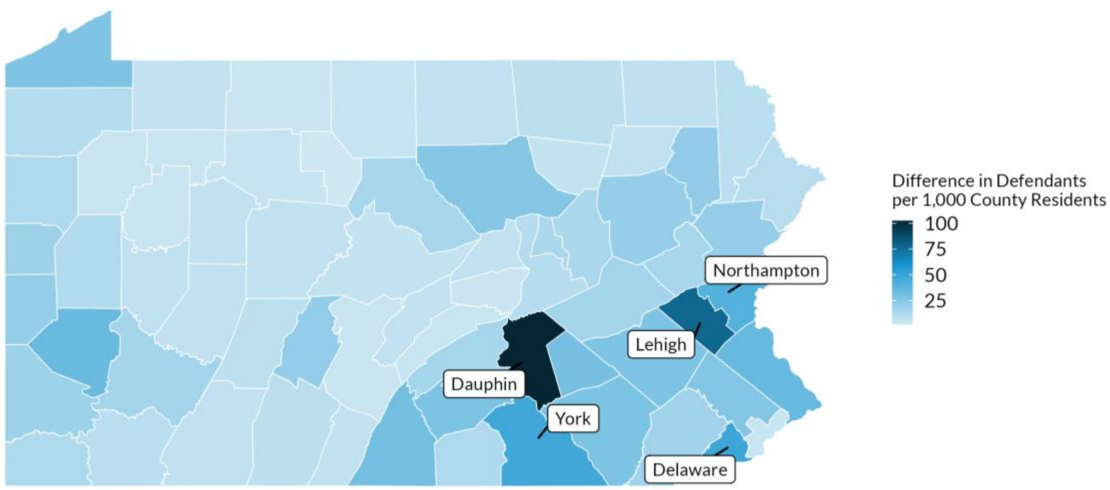
Figure 8 shows a much larger geographic variation in the concentration of individuals with eviction filings between the most strict and most lenient matching scenarios. The difference between the two scenarios is most pronounced in Dauphin, Lehigh, York, Delaware, and Northampton counties in the eastern part of the state.[18]

The larger change in eviction records is likely because of lower-quality records in eviction filings. The difference between eviction and criminal records highlights the larger role data matching approaches play when, for instance, more relevant variables are missing or are not useful as individual-level identifiers.

FIGURE 8

Number of Landlord-Tenant Defendants by County Is More Dependent on Data Matching Assumptions

Difference between most strict scenario and most lenient scenario



URBAN INSTITUTE

Source: Authors' analysis of data from the Administrative Office of Pennsylvania Courts (AOPC).

Notes: Individuals who appear across multiple geographies are only counted for the first geography in which they appear. Court records are across the entire 2014–24 sample. Landlord-tenant case records are not available for Philadelphia County. Calculations are based on the number of individuals per 1,000 residents in the county according to the 2020 Decennial Census. Zip code–level calculations are allocated to counties based on the US Department of Housing and Urban Development's United States Postal Service residential data.

Captured Criminal and Eviction Histories: Variation in Number and Recency by Matching Approach

To examine how different matching approaches affect individuals and their tenant screening reports, we analyzed how data matching assumptions influence the number of people flagged for histories that may be particularly relevant to landlords. Specifically, for criminal histories, we look at individuals who are linked to a felony or misdemeanor charge, a guilty disposition, or a subset of charges that appear to be more salient for landlords—for example, arson, sexual crimes, or production of methamphetamines—which are also often barred for federally assisted housing.[19] On the landlord-tenant record side, we look at individuals who had an eviction filed against them and had an eviction actually carried out. We limit the look-back period to seven years, which is the period that background reports usually disclose civil judgements, lawsuits, housing court cases, and arrest records, though not all reports are limited to seven years.[20]

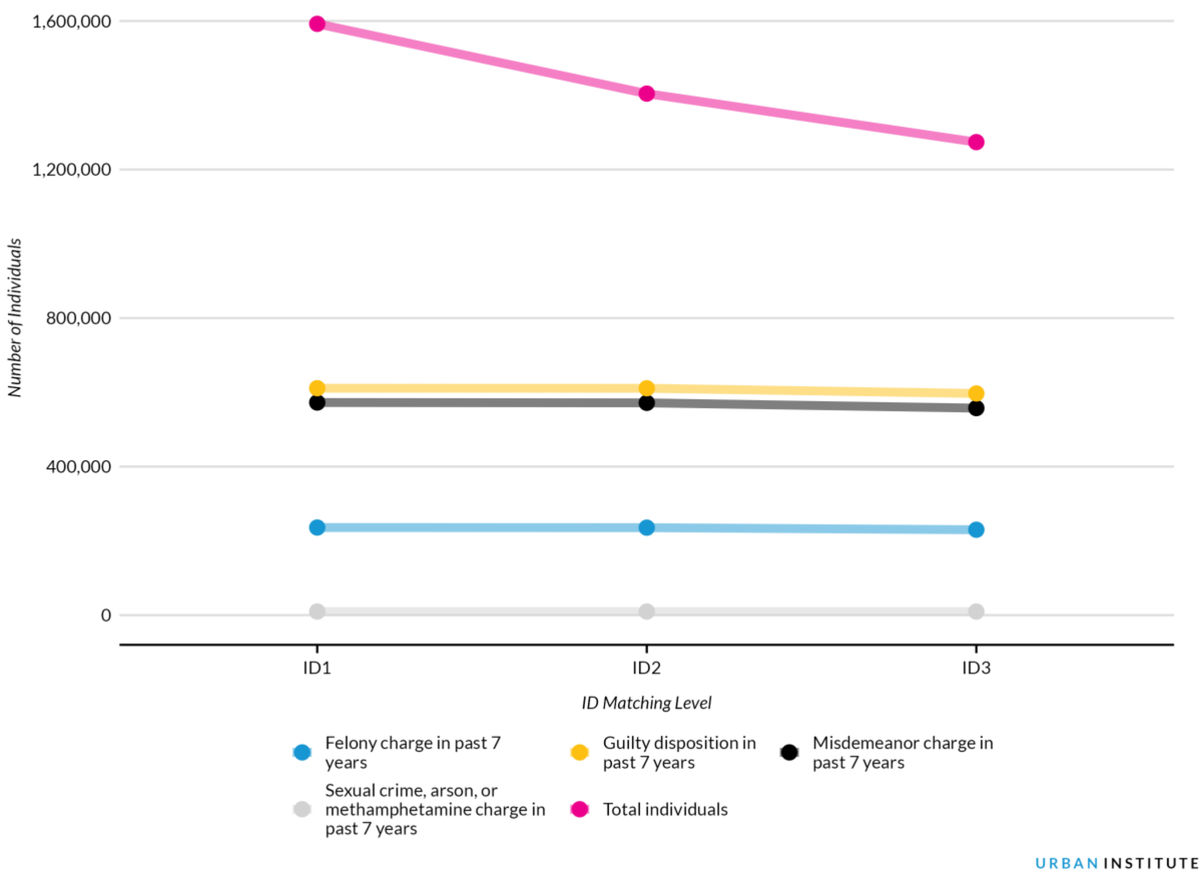
Figures 9 and 10 illustrate patterns in the total number of individuals who have the incidents mentioned above on their records, as well as the total number of individuals in our dataset, across the three ID types. Figure 9 shows that the total number of individuals with each relevant criminal history remains largely stable, even as the overall population declines because of record consolidation. In contrast, figure 10 shows the number of individuals with an eviction filing or eviction disposition drops significantly. These declines are even steeper than the overall population decrease in our dataset: a 39 percent drop in filings and a 37 percent drop in dispositions from ID1 to ID3, compared with just a 20 percent decline in total population.

The most probable reason for this is a difference in data quality: eviction records simply have fewer fields besides a name that can generate reliable matches, whereas criminal records have a near-complete date of birth, which generally offers a strong and reliable signal of a match (i.e., the odds of two records sharing a name and a date of birth and not belonging to the same individual are low). Landlord-tenant records were missing the date of birth for more than three-quarters of cases, which left us with just a name and a zip code (a much weaker indicator) or a name alone to analyze. Ultimately, relaxed data assumptions make more of a difference on the landlord-tenant side, because these fields are the *only* way to find matches in many cases where data are lacking. The sheer variation in our aggregate data on eviction histories indicates how unreliable administrative data can be for putting together tenant screening reports. This highlights the risk that certain tenants, especially those with common names and living in large zip codes, may be wrongly tied to histories that are not theirs.

FIGURE 9

Number of Specific Criminal Charges and Outcomes Hold Steady Even as Data Assumptions Are Relaxed

Criminal charges



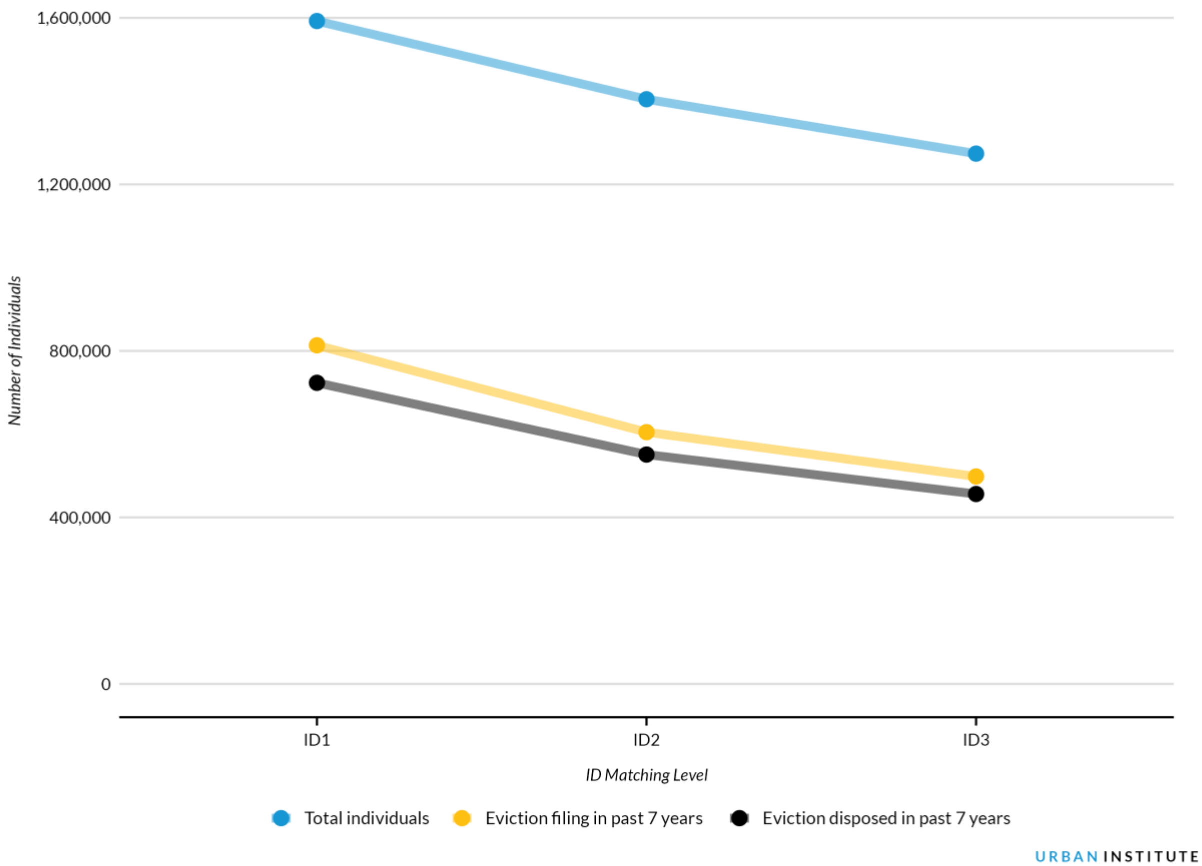
Source: Authors' analysis of data from the Administrative Office of Pennsylvania Courts (AOPC).
Notes: "ID" refers to how records are collapsed within datasets, and each point averages across all data scenarios for that ID. Offense disposition date is used to determine the seven-year look-back period. Guilty dispositions include "Guilty" and "No Contest" records.

In figure 10, we focus on landlord-tenant records, specifically the presence of an eviction filing and whether an eviction was carried out. The figure shows that the more lenient identification significantly *decreases* the proportion of individuals with eviction histories during the look-back period; in other words, more lenient matching reduces the percentage of individuals with an eviction filed or disposed against them. This is likely because the same individual may have many case records with eviction filings or dispositions that are collapsed together under more lenient assumptions. For example, because some eviction records have higher missingness for key matching variables like date of birth, eviction data may be more sensitive to variations in outcomes when assumptions are relaxed.

FIGURE 10

Number of Specific Eviction Histories Decrease Significantly as Data Assumptions Are Relaxed

Eviction charges



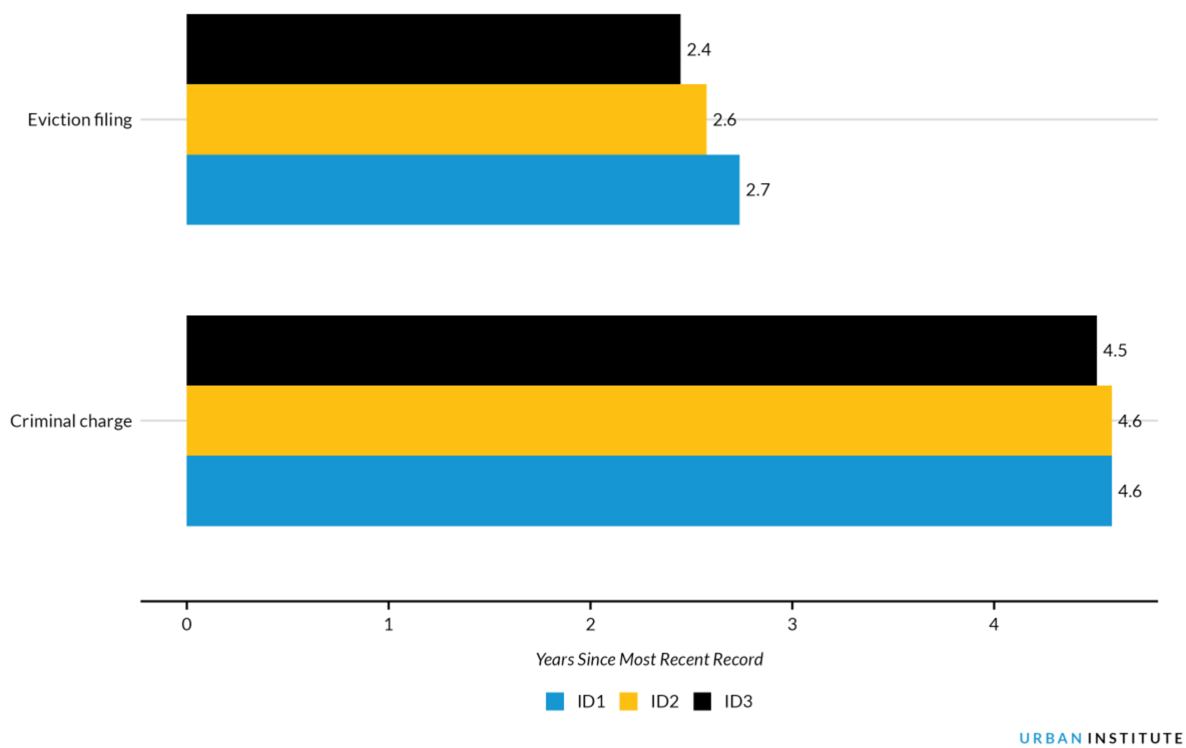
Source: Authors' analysis of data from the Administrative Office of Pennsylvania Courts (AOPC).
Notes: "ID" refers to how records are collapsed within datasets, and each point averages across all data scenarios for that ID. Docket filing date is used to determine the seven-year look-back period for the landlord-tenant data.

Notably, while both criminal and landlord-tenant records experience shifts in overall *percentage levels*, depending on the identification method, the *raw numbers* of individuals with each type of criminal history hold steady, with most of the change coming from the number of unique individuals with eviction histories. These diverging patterns suggest that the overall risk profiles within a tenant screening dataset are highly sensitive—especially on the landlord-tenant side—to subjective decisions about which cases belong to which individual. Each of the 24 data scenarios considers the same set of cases, but figure 9 clearly outlines how those cases can be attributed to individuals in ways that would likely alter any rental applicant's risk score.

Figure 11 shows the average number of years since last offense for both eviction filings and criminal data. The average is not highly sensitive to data matching: they are approximately 4.5–4.6 years for individuals with a criminal history and just 2.4–2.5 years for individuals with an eviction record. For clean slate or fair chance laws that prescribe criminal look-back periods of no longer than five years, it is noteworthy that the average criminal defendant comes fairly close to that threshold. This can indicate to landlords that the average profile of a criminal defendant is not that of a recent offender, but of someone with almost five years since their last offense. Future work should examine this cutoff more granularly to determine the effect specific policies might have on landlord-tenant records and how they impact potential renters.

FIGURE 11

Average Number of Years since Last Criminal Charge Is at Least 4.5 Years, Regardless of Matching Assumptions



Source: Authors' analysis of data from the Administrative Office of Pennsylvania Courts (AOPC).
Notes: "ID" refers to how records are collapsed within datasets, and each point averages across all data scenarios for that ID. Docket filing date and offense disposition date are used to determine the seven-year look-back period for the landlord-tenant and criminal data, respectively. Guilty dispositions include "Guilty" and "No Contest" records.

Implications for Practice and Policy

Overall, our findings show that there is a great deal of variability in court records, and this likely has real-world consequences for both landlords and potential renters. We highlight that there is a large variation in the number of records that may be either attributed to the same individual or split across multiple individuals based on decisions made around data management, data matching, and the thresholds chosen for matching. Across our 24 data scenarios, this pattern was most prominent during name de-duplication—moving from the most strict matching approach (ID1), which requires same name and date of birth, to the least strict matching approach, which requires only a same name (ID3).

The variation in records is largest when the data source itself is of lower quality and less reliable. For example, based on data cleaning and matching, there were larger differences in the landlord-tenant data, which had less concrete information, than in criminal records. However, we were still unable to draw strong conclusions about tenant racial equity questions in the criminal records, because of the low-quality race and ethnicity data and measurement limitations.

When companies use court records, they are not just objectively reporting data but making a slew of highly subjective decisions about the data that affect who is identified in the tenant screening reports, what records those identified are associated with, and how much of a risk they pose for the landlord. Even with the immense amount of data we collected and analyzed, data quality issues and the lack of transparency in tenant screening company practices made it difficult to pin down exactly how rental screening reports affect some tenants. Given the rise of tenant screening companies and the important role tenant screening reports play for potential renters, it is critical that companies create more accurate reports to protect renters, better serve landlords, and reduce harm. We recommend focusing on transparency, setting clearer and higher expectations for data quality, and limiting data access.

Increase Transparency in Reporting

There are challenges to data quality, completeness, and accuracy in court records, but landlords and tenants are currently in the dark about the quality of the data being used in tenant screening reports, posing a potential business, safety, and housing access risk to both parties. Increasing the transparency of the data assembly process used by tenant screening companies and setting limits on what types of data can legally be used in tenant screening reports can benefit landlords and tenants alike.

Clear and transparent reporting on sourcing and quality of data can help landlords make a more informed choice between tenant screening companies, remain compliant with state mandates such as clean slate laws, and update their own screening criteria to match what is reliably provided by the screening reports. For example, if a matched record is of lower quality or taken from a less reliable source, the landlord should have that information so they can maintain a fuller pool of potential renters, rather than one that is artificially reduced through poor data practices.

Tenant screening companies should be clear and transparent about where and how data are acquired and where there are gaps in high-quality data. For example, companies should be clear when data are gathered from a reliable source, such as the state court system, and when they are taken from a less reliable source.

This transparency should extend to making sure that sealed and expunged data are not included in databases or reports. Currently, it is unclear how tenant screening companies monitor and update cases that have been sealed, expunged, or otherwise affected. In a separate analysis of clean slate laws, we find that sealing and expungement have a large impact on the number of case records and should be reflected by companies as they report on people's history and background.[21]

Companies should also be transparent about the quality and strictness of the data cleaning and matching processes used. Especially since, as demonstrated in our analysis, the processes of data linking, model selection for name matching, and thresholds for accuracy significantly affect who is matched within datasets. Even if companies cannot disclose proprietary information about the algorithms they use, they can provide the standards for cleaning data, the software packages, and the matching thresholds that were chosen (especially in absence of unique identifiers). These are all imperative to minimizing the risk of incorrectly matching individuals.

Both renters and landlords, for example, should know whether a data match is based on a unique identifier or based on an estimate from name-matching or other probabilistic procedures. Again, this may affect how landlords decide to use the information and make decisions about renters.

While attention is rightfully paid to the "black box" of tenant screening reports, equal attention should be paid to the data gaps and accuracy challenges that influence what is in the reports. Federal policymakers and landlords can focus on establishing and enforcing thresholds for accuracy when matching data. (The choices presented in this brief serve as just one set of examples.) The approach used to implement the thresholds can and should be based on peer-reviewed, statistically rigorous methods for record linkage common in social science research. Probabilistic record linkage and its evaluation have been an area of study formalized over 50 years ago (Fellegi and Sunter 1969). Best practices have evolved and are now widely used across many large-scale government agency studies involving sensitive administrative data (Brown et al. 2022; CDC 2024). Tenant screening companies can look to fellow organizations that use proprietary administrative data and their published matching methodologies and validation steps, such as those used by the JPMorgan Chase Institute (Farrell et al. 2020).

More research is needed to determine how specific algorithms may be impacted by varying quality of matching in the tenant screening datasets. It would be helpful to establish a common set of quality, missingness, and accuracy standards for the data that tenant screening companies use to match individuals for their algorithmic assessments. This way both landlords and tenants can have increased confidence in the tenant screening process.

Set Stricter Standards for Data Matches without Unique Identifiers

Beyond transparency of data handling, there is a concern about individual-level matches when no unique identifier exists. In other words, in the absence of a social security number or a similar unique number in the data of interest, there is always a risk that data across sources may be incorrectly matched to one individual or assumed to belong to two individuals when they actually belong to the same individual. Our findings highlight the broad variation in how many records can be attributed to unique individuals based on data handling and what level of "fuzziness" is permitted by tenant screening companies. Data without a unique identifier require a combination of name and often-limited demographic information, and this lack of information leaves the door open to inaccurate or misleading matches (e.g., mistakes between junior and senior, punctuation, or missing fields).

Some tenant screening companies have already started proactively updating their data practices because of the data quality issues without unique identifiers. For example, the three major consumer reporting agencies—Experian, TransUnion, and Equifax—do not report eviction histories, because those records lack unique identifiers. It would be helpful for other tenant screening companies to follow this private industry trend.[22] Landlords can also choose to work with tenant screening companies that provide clearer details on how they collect, clean, and match data, and aim to only work with companies that use better practices.

State policymakers can encourage stronger data practices by limiting the amount of low-quality data that is released to the public and that can be used by companies. States like California have implemented laws that seal court records upon filing and only unseal them when a guilty verdict is reached (Fung et al. 2023). States like Illinois, which passed the Personal Information Protection Act and the Illinois Data Privacy and Protection Act, aim to establish requirements for both protecting personal information and setting clearer standards for how data is collected, processed, and transferred.

Policymakers can also shift the burden of proof for error correction in tenant screening reports from applicants to companies themselves. This would mean that in cases where there is no unique identifier, companies would be required to engage with the applicant in question and receive a confirmation that the information is accurate. Currently, it is up to the applicants to contest any inaccurate information. Transferring the burden of proof to companies would be less costly to individuals and less likely to create housing instability.

Assess Data Quality and Provide Guidance on Data Sources

Mirroring how credit scores are currently monitored and regulated in the mortgage market, federal policymakers have an opportunity to play a greater role in monitoring what types of data—and what quality of that data—can be used to generate tenant screening reports. While credit scores theoretically predict the likelihood of default on a loan and have been rigorously tested (Chatterjee et al. 2023), tenant screening reports have not been found to predict performance as a tenant or lease violation. There is little evidence around which criminal and landlord data variables predict lease adherence and/or high-quality tenancy. In fact, one study in Minnesota found that most types of prior criminal offenses do not affect housing outcomes, except for major drug crimes, fraud, assault, and property crimes, and that the relationship between these crimes and outcomes diminish quickly over time (Warren 2019). As researchers continue to generate evidence on the variables that are most predictive of on-time rent payment and good tenancy, tenant screening companies and landlords should adjust the data they use accordingly, especially if certain variables have high missingness or other known limitations.

Given the lack of evidence on the relationship between recency of criminal history and negative housing outcomes, as well as the potential privacy risks, interventions that aim to reduce the use of complete criminal filings could have an outsized impact for renters and landlords. These interventions include shortening look-back periods or limiting the criminal records to only guilty verdicts on certain offenses. States with automatic and comprehensive clean slate laws, like Pennsylvania, are likely effective in reducing the use of lower-level offenses and nonconvictions in tenant screening reports. This could be one way to reduce harmful errors in tenant screening reports.

Similarly, there is little evidence that a past eviction is predictive of a future eviction, especially given that data on eviction only indicates a filing, not an executed eviction or a guilty verdict.[23] Furthermore, an instance of an eviction is not exclusively related to a tenant's inability to pay rent, as certain landlords can engage in serial eviction-filing behavior to turn over units, which puts some individuals at higher risk of receiving an eviction notice (Garboden and Rosen 2019; Immergluck et al. 2019; Polk 2020). Without clear outcome or disposition data to explain how an eviction filing was ultimately resolved, it is challenging to justify the inclusion of such low-quality data in tenant screening algorithmic assessments. Policymakers should, therefore, look to limit the inclusion of eviction filings in tenant screening reports, including by banning the use of filings without a guilty verdict. States like California, for example, have erred on the side of privacy by focusing on ways to keep court cases that do not result in a guilty disposition sealed.

Resolve the Tension between Data Reliability and Data Privacy

An overarching tension in this work is whether landlords and tenants would benefit from more rigorous, detailed data or reduced access to low-quality data. Some organizations like Pew Charitable Trusts have discussed ways that court data could be made more reliable, more consistent, and more aggregated.[24] With increased aggregation of data to assist landlords assess a prospective tenant's likelihood of paying rent, tenant screening data should be more reliable and representative of all people, including by reducing the need for companies to match individuals to criminal and eviction histories.

Yet, these moves toward more comprehensive, reliable court data are in tension with privacy concerns for individuals, as aggregation of highly sensitive criminal, eviction, and credit data could increase risks of identity theft and data misuse.[25]

Overall, the low-quality, decentralized nature of court data is gaining attention, as is the tension over the end goal of tenant screening reports—which is to make fair, rigorous, and transparent decisions that benefit the entire rental housing market.

Next Steps

In this brief, we highlighted some of the challenges and concerns associated with using court data in tenant screening reports. But our analysis was limited by two considerations. First, there is lack of transparency into how tenant screening companies are obtaining, handling, cleaning, and matching data. To address this concern, we conducted outreach to multiple individuals doing work on tenant screening reports, including legal organizations that represent tenants who have found errors in their reports. However, we are ultimately unable to speak comprehensively on what tenant screening companies are doing, given the proprietary nature of their work.

Second, there is limited data availability. Our analysis of Pennsylvania, for example, is based on what the state collects and its laws guiding data sharing, including the clean slate laws that have been passed in recent years.

As the number of tenant screening companies increases, and the impact of tenant screening becomes clearer, it is critical to focus on ways to improve data quality, matching, and management processes within tenant screening reports. There are clear opportunities for action, which requires a combination of more transparency, higher standards for accuracy, and improving what tenant screening reports are evaluating. While we continue to build a body of evidence in this understudied domain, these reforms would benefit landlords and tenants—as well as advance housing justice, further equity in the housing market, and ensure housing stability and the quality-of-life improvements associated with it.

Appendix

TABLE A.1

Fields Captured within the Court System Data on Criminal Cases and Landlord Tenant Cases

Criminal Case Data	Percent Complete	Landlord-Tenant Case Data	Percent Complete
Docket number	100%	Docket number	100%
Defendant name	100%	Defendant name	100%
Defendant DOB	99.4%	Defendant DOB	22.4%
Defendant gender	99.2%	Defendant gender	22.5%
Defendant race	98.1%	Defendant race	16.2%
Defendant ethnicity	37.9%	Defendant ethnicity	NA
Defendant city	99.8%	Defendant city	100%
Defendant state	99.8%	Defendant state	100%
Defendant zip code	99.8%	Defendant zip code	100%
Offense sequence number	100%	Case disposition	100%
Title	99.9%	Claim amount	100%
Section	99.9%	Judgement amount	95.0%
Subsection	99.9%	Monthly rent	100%
Offense description	99.9%	Judgement component type	95.0%
Offense grade	97.8%	Judgement component amount	95.0%
Offense disposition	91.2%		
Offense disposition date	91.2%		
Case disposition	92.7%		
Case disposition date	92.7%		

Notes

[1] Jung Hyun Choi, Laurie Goodman, and Daniel Pang, “The Real Rental Housing Crisis Is on the Horizon,” *Urban Wire* (blog), Urban Institute, March 11, 2022, https://www.urban.org/urban-wire/real-rental-housing-crisis-horizon.

[2] Choi, Goodman, and Pang, “The Real Rental Housing Crisis Is on the Horizon”; and Lauren Kirchner and Matthew Goldstein, “How Automated Background Checks Freeze Out Renters,” *New York Times*, May 28, 2020, https://www.nytimes.com/2020/05/28/business/renters-background-checks.html.

[3] Kirchner and Goldstein, “How Automated Background Checks Freeze Out Renters.”

[4] Peter Hepburn, Renee Louis, and Matthew Desmond, “Racial and Gender Disparities among Evicted Americans,” *Eviction Lab Updates*, December 16, 2020, https://evictionlab.org/demographics-of-eviction/.

[5] Choi, Goodman, and Pang, “The Real Rental Housing Crisis Is on the Horizon.”

[6] Many companies use "wild card” searches, which gathers different names that start with the same few letters. This is a problem for people who have similar surnames; for example, more than 12 million Latinx nationwide share just 26 surnames. See Kirchner and Goldstein, “How Automated Background Checks Freeze Out Renters.”

[7] See Consumer Financial Protection Bureau, “Tenant Background Checks,” accessed February 25, 2025, https://www.consumerfinance.gov/rules-policy/tenant-background-checks/; and “Review Your Rental Background Check,” updated October 11, 2023, https://www.consumerfinance.gov/rules-policy/tenant-background-checks/review-your-rental-background-check/. See Federal Trade Commission, “What Tenant Background Screening Companies Need to Know about the Fair Credit Reporting Act,” October 2016, https://www.ftc.gov/business-guidance/resources/what-tenant-background-screening-companies-need-know-about-fair-credit-reporting-act; and “Tenant Background Checks and Your Rights,” March 2024, https://web.archive.org/web/20250121080539/https://consumer.ftc.gov/articles/tenant-background-checks-and-your-rights. See US Department of Housing and Urban Development, “Guidance on Application of Fair Housing Act to the Screening of Applicants for the Rental Housing,” April 29, 2024, https://web.archive.org/web/20250117202458/https://www.hud.gov/sites/dfiles/FHEO/documents/FHEO_Guidance_on_Screening_of_Applicants_for_Rental_Housing.pdf; and “Tenant Background Checks and Your Rights,” March 2024, https://web.archive.org/web/20250215175220/https://www.hud.gov/sites/dfiles/FHEO/documents/HUD_Tenant_Background_Checks_and_Your_Rights.pdf.

[8] See Conn. Fair Hous. Ctr v. CoreLogic Rental Prop. Sols., 3:18-cv-705-VLB (D. Conn. Jul. 20, 2023), https://casetext.com/case/conn-fair-hous-ctr-v-corelogic-rental-prop-sols-4.

[9] See Federal Trade Commission, “RealPage, Inc.,” last updated October 16, 2018, https://www.ftc.gov/legal-library/browse/cases-proceedings/152-3059-realpage-inc.

[10] See City of Philadelphia, “Renters’ Access Act: Tenant Screening Guidelines,” last updated September 11, 2023, https://www.phila.gov/documents/renters-access-act-tenant-screening-guidelines/#:~:text=The%20Renters'%20Access%20Act%20supports,810%20of%20the%20Philadelphia%20Code.

[11] See Federal Trade Commission, “AppFolio, Inc.,” last updated December 8, 2020, https://www.ftc.gov/legal-library/browse/cases-proceedings/1923016-appfolio-inc.

[12] Rebecca John, Katie Fallon, Judah Axelrod, Sonia Torres Rodríguez, and Brendan Chen, “Do Clean Slate Laws Reduce Housing Barriers?” Housing Matters, Urban Institute, December 4, 2024, https://housingmatters.urban.org/articles/do-clean-slate-laws-reduce-housing-barriers.

[13] So (2022) provides a list of data fields in residential and criminal histories that are available to 18 tenant screening models used in the industry. This list helped guide our research with additional data collection.

[14] See Consumer Financial Protection Bureau, "Advisory Opinion on Name-Only Matching," November 4, 2021, https://www.consumerfinance.gov/rules-policy/final-rules/fair-credit-reporting-name-only-matching-procedures/.

[15] Adam Cohen, "FuzzyWuzzy: Fuzzy String Matching in Python," *ChairNerd*, SeatGeek, July 8, 2011, https://chairnerd.seatgeek.com/fuzzywuzzy-fuzzy-string-matching-in-python/.

[16] Specifically, we used the *wratio* algorithm available at https://github.com/seatgeek/thefuzz/blob/83bea3d4a109a5d0c2e07334b504953cda4959c5/thefuzz/fuzz.py#L119.

[17] See Consumer Financial Protection Bureau, "How Long Can Information, Like Eviction Actions and Lawsuits, Stay on My Tenant Screening Record?" last reviewed July 1, 1021, https://www.consumerfinance.gov/ask-cfpb/how-long-can-information-like-eviction-actions-and-lawsuits-stay-on-my-tenant-screening-record-en-2104/. According to federal guidance, lawsuits and judgments, including evictions, can be reported up to seven years; bankruptcies can stay on people's records for ten years; and there is no limit for criminal convictions.

[18] Philadelphia County uses a different system than the rest of the state for tracking landlord-tenant cases, so most evictions in that county are not included in this analysis. This lower absolute number of landlord-tenant records translates to a smaller difference in figure 8 than would otherwise be expected.

[19] See US Department of Housing and Urban Development, "Are Applicants with Felonies Banned from Public Housing or Any Other Housing Funded by HUD? Do the Public Housing Agencies (PHAs), State, or Landlords Have Any Discretion in the Process that Could Bar Certain Felonies?" *HUD Exchange*, January 2022, https://www.hudexchange.info/faqs/programs/housing-choice-voucher-program/eligibility-determination-and-denial-of-assistance/background-screening/are-applicants-with-felonies-banned-from-public-housing-or-any-other/.

[20] See US Department of Housing and Urban Development, "Tenant Background Checks and Your Rights," March 2024, https://web.archive.org/web/20250215175220/https://www.hud.gov/sites/dfiles/FHEO/documents/HUD_Tenant_Background_Checks_and_Your_Rights.pdf.

[21] John, Fallon, Axelrod, Torres Rodríguez, and Chen, "Do Clean Slate Laws Reduce Housing Barriers?"

[22] See National Consumer Law Center's July 2017 announcement on policy updates. Chi Chi Wu, "Big Changes for Credit Reports, Improving Accuracy for Millions of Consumers," National Consumer Law Center, July 27, 2017, https://library.nclc.org/article/big-changes-credit-reports-improving-accuracy-millions-consumers.

[23] Lillian Leung, Peter Hepburn, and Matthew Desmond, "Serial Eviction Filings: How Landlords Use the Courts to Collect Rent," *Eviction Lab Updates*, September 15, 2020, https://evictionlab.org/serial-eviction-filings/#:~:text=We%20often%20associate%20%E2%80%9Ceviction%E2%80%9D%20with,prevalence%20across%20the%20United%20States.&text=In%20assembling%20the%20Eviction%20Lab's,landlords%20file%20serially%20for%20eviction.

[24] Pew Charitable Trusts, "How to Standardize Court Data for Greater Transparency and Ongoing Improvement: Steps for Making Civil Courts More Open," Fact Sheet, September 27, 2023, https://www.pewtrusts.org/en/research-and-analysis/fact-sheets/2023/09/how-to-standardize-court-data-for-greater-transparency-and-ongoing-improvement.

[25] Equifax, "The Risk of Technology and Identity Theft," press release, October 23, 2014, https://investor.equifax.com/news-events/press-releases/detail/535/the-risk-of-technology-and-identity-theft.

References

Barak, Gregg, Paul Leighton, and Jeanne Flavin. 2007. *Class, Race, Gender, and Crime: The Social Realities of Justice in America*. Lanham, MD: Rowman & Littlefield Publishers.

Bhatia, Shivangi. 2020. "To 'Otherwise Make Unavailable': Tenant Screening Companies' Liability under the Fair Housing Act's Disparate Impact Theory." *Fordham Law Review* 88 (6): 2551–83.

Brown, Derek, Mithuna Srinivasan, Courtney Zott, Kala Wilson, Prashila Dullabh, and Scott Smith. 2022. *Medicare Data Linkages for Conducting Patient-Centered Outcomes Research on Economic Outcomes*. Washington, DC: US Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation.

CDC (Centers for Disease Control and Prevention). 2024. *The Linkage of National Center for Health Statistics Survey Data to Centers for Medicare & Medicaid Services Transformed Medicaid Statistical Information System Claims Data (2014–2019): Matching Methodology and Analytic Considerations*. Hyattsville, MD: CDC, National Center for Health Statistics, Division of Analysis and Epidemiology.

CFPB (Consumer Financial Protection Bureau). 2022. *Tenant Background Checks Market*. Washington, DC: CFPB.

Chatterjee, Satyajit, Dean Corbae, Kyle Dempsey, and José-Víctor Ríos-Rull. 2023. "A Quantitative Theory of the Credit Score." *Econometrica* 91 (5): 1803–40. https://doi.org/10.3982/ECTA18771.

Couloute, Lucius. 2018. Nowhere to Go: Homelessness among Formerly Incarcerated People. Northampton, MA: Prison Policy Initiative.

Desmond, Matthew, and Monica Bell. 2015. "Housing, Poverty, and the Law." *Annual Review of Law and Social Science* 11:15–35. https://doi.org/10.1146/annurev-lawsocsci-120814-121623.

Desmond, Matthew, and Tracey Shollenberger. 2015. "Forced Displacement from Rental Housing: Prevalence and Neighborhood Consequences." *Demography* 52 (5): 1751–72.

Duke, Annette, and Andrea Park. 2019. *Evicted for Life: How Eviction Court Records Are Creating a New Barrier to Housing*. Boston: Massachusetts Law Reform Institute.

Eisenberg, Alexa, and Kate Brantley. 2024. *Record Costs: Collateral Consequences of Eviction Court Filings in Pennsylvania*. Ann Arbor: University of Michigan, Housing Solutions for Health Equity.

Farrell, Diana, Fiona Greig, Chris Wheat, Max Liebeskind, Peter Ganong, Damon Jones, and Pascal Noel. 2020. *Racial Gaps in Financial Outcomes: Big Data Evidence*. New York: JPMorgan Chase Institute.

Fellegi, Ivan P., and Alan B. Sunter. 1969. "A Theory for Record Linkage." *Journal of the American Statistical Association* 64 (328): 1183–210. https://doi.org/10.1080/01621459.1969.10501049.

Fung, Lauren, Isabella Remor, Katie Fallon, and Nyla Holland. 2023. "Masking the Scarlet 'E': A Study on California's Attempt to Mask Eviction Records through AB 2819." Washington, DC: Urban Institute.

Garboden, Philip M. E., and Eva Rosen. 2019. "Serial Filing: How Landlords Use the Threat of Eviction." *City & Community* 18 (2): 638–61. https://doi.org/10.1111/cico.12387.

Housing Justice Center. 2021. *Opening the Door: Tenant Screening and Selection*. St. Paul, MN: Housing Justice Center; Minneapolis, MN: Family Housing Fund.

Immergluck, Dan, Jeff Ernsthausen, Stephanie Earl, and Allison Powell. 2019. "Evictions, Large Owners, and Serial Filings: Findings from Atlanta." *Housing Studies* 35 (5): 903–24. https://doi.org/10.1080/02673037.2019.1639635.

Kleysteuber, Rudy. 2007. "Tenant Screening Thirty Years Later: A Statutory Proposal to Protect Public Records." *The Yale Law Journal* 116:1344–88.

Lake, Jaboa, and Leni Tupper. 2021. *Eviction Record Expungement Can Remove Barriers to Stable Housing*. Washington, DC: Center for American Progress.

Levenshtein, V. I. 1966. "Binary Codes Capable of Correcting Deletions, Insertions, and Reversals." *Soviet Physics—Doklady* 10 (8): 707–10.

Lewis, Brittany, Molly Calhoun, Cynthia Matthias, Kya Conception, Thalya Reyes, Carolyn Szczepanski, Gabriela Norton, et al. 2019. *The Illusion of Choice: Evictions and Profit in North Minneapolis*. Minneapolis: University of Minnesota, Center for Urban and Regional Affairs.

Nellis, Ashley. 2021. *The Color of Justice: Racial and Ethnic Disparity in State Prisons*. Washington, DC: The Sentencing Project.

Pasley, Tex, Henry Oostrom-Shah, and Eric Sirota. 2021. *Screened Out: How Tenant Screening Reports Undermine Fair Housing Laws and Deprive Tenants of Equal Access to Housing in Illinois*. Chicago: Shriver Center on Poverty Law.

Pinard, Michael. 2013. "Criminal Records, Race and Redemption." *Journal of Legislation and Public Policy* 16 (4): 963–97.

Polk, Katelyn. 2020. "Screened Out of Housing: The Impact of Misleading Tenant Screening Reports and the Potential for Criminal Expungement as a Model for Effectively Sealing Evictions." *Northwestern Journal of Law & Social Policy* 15 (3): 338–56.

Porton, Adam, Ashley Gromis, and Matthew Desmond. 2020. "Inaccuracies in Eviction Records: Implications for Renter and Researchers." *Housing Policy Debate* 31 (3–5): 377–94. https://doi.org/10.1080/10511482.2020.1748084.

Reosti, Anna. 2020. "'We Go Totally Subjective': Discretion, Discrimination, and Tenant Screening in a Landlord's Market." *Law & Social Inquiry* 45 (3): 618–57. http://doi.org/10.1017/lsi.2019.81.

———. 2021. "The Costs of Seeking Shelter for Renters with Discrediting Background Records." *City & Community* 20 (3): 235–59. https://doi.org/10.1177/15356841211012483.

Roman, Caterina Gouvis, and Jeremy Travis. 2004. *Taking Stock: Housing, Homelessness, and Prisoner Reentry*. Washington, DC: Urban Institute.

Shiffer-Sebba, Doron. 2021. "Understanding the Divergent Logics of Landlords: Circumstantial versus Deliberate Pathways." *City & Community* 19 (4): 1011–37. https://doi.org/10.1111/cico.12490.

So, Wonyoung. 2022. "Which Information Matters? Measuring Landlord Assessment of Tenant Screening Reports." *Housing Policy Debate* 33 (6): 1484–510. https://doi.org/10.1080/10511482.2022.2113815.

Stanley-Becker, Tom. 2022. "Breaking the Cycle of Homelessness and Incarceration: Prisoner Reentry, Racial Justice, and Fair Chance Housing Policy." *Journal of Law and Public Affairs* (University of Pennsylvania) 7 (2): 257–314.

Stern, Alena, Manuel Alcalá Kovalski, Leah Hendey, Elizabeth Burton, Sandra Ambrozy, and Carlos Manjarrez. 2025. *Unveiling Disparities in Eviction: A Novel Approach to Demographic Imputation in Cleveland*. Washington, DC: Urban Institute.

TechEquity. 2024. *Screened Out of Housing: How AI-Powered Tenant Screening Hurts Renters*. Oakland, CA: TechEquity.

Warren, Cael. 2019. *Success in Housing: How Much Does Criminal Background Matter?* St. Paul, MN: Amherst H. Wilder Foundation, Wilder Research.

Weiss, Elayne. 2018. "Housing Access for People with Criminal Records." In *2019 Advocates' Guide*, 6-27–6-33. Washington, DC: National Low Income Housing Coalition.

About the Authors

Judah Axelrod is an associate director of data science at the Urban Institute. He works in collaboration with Urban's Technology and Data Science team and Racial Equity Analytics Lab with a particular focus on the equitable use of AI and automated systems. Axelrod holds a BA in economics from Rutgers University and an MSc in data science from the London School of Economics.

Brendan Chen is a research analyst in the Housing and Communities Division at the Urban Institute. His research interests include housing affordability and homelessness. Chen holds a BA in economics and international development studies from the University of California, Los Angeles.

Katie Fallon is a principal policy associate in the Research to Action Lab at the Urban Institute. She applies a mixed-methods approach to understanding issues related to housing access, stability, and finance. Fallon holds a BA in public and international policy from Princeton University and a PhD in sociology from the University of Wisconsin—Madison.

Sonia Torres Rodríguez is a research associate in the Race and Equity Division at the Urban Institute. She is affiliated with Urban's Racial Equity Analytics Lab and the Community Engagement Resource Center. She applies econometrics and data science methods to study policies and programs that advance equity and inclusion, particularly related to housing, neighborhood change, racial wealth equity, and automated decisionmaking systems. Torres Rodríguez holds a BA in mathematical economic analysis from Rice University and an MS in applied economics from Johns Hopkins University.

Acknowledgments

This brief was funded by Chan Zuckerberg Initiative and Conrad N. Hilton Foundation. 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.

URBAN INSTITUTE

500 L'Enfant Plaza SW
Washington, DC 20024

www.urban.org

ABOUT THE URBAN INSTITUTE

The nonprofit Urban Institute is a leading research organization dedicated to developing evidence-based insights that improve people's lives and strengthen communities. For 50 years, Urban has been the trusted source for rigorous analysis of complex social and economic issues; strategic advice to policymakers, philanthropists, and practitioners; and new, promising ideas that expand opportunities for all. Our work inspires effective decisions that advance fairness and enhance the well-being of people and places.