Screening by Name: How Tenant Background Checks Penalize Renters with Court-Record Lookalikes

Molly Clark Barol* Christopher Timmins† Daisy Zhan‡

March 26, 2026

Abstract

Tenant screening algorithms search court records by name, returning results for anyone with a matching or similar name. We show that this imprecise matching sorts innocent renters into neighborhoods with higher pollution, higher poverty, and fewer economic opportunities—even when they have no criminal record of their own. Linking consumer records to the universe of Milwaukee County circuit court cases, we construct a Levenshtein-distance measure of name similarity to quantify each renter's exposure to "lookalike" court records. Among renters with no criminal record, fuzzy name exposure significantly predicts worse neighborhoods across all seven outcomes we measure: higher toxic chemical exposure, higher fine particulate matter, lower rent, higher unemployment, higher poverty, lower household income, and fewer college-educated neighbors. Alternatively, every coefficient is zero for renters who already have their own criminal records and zero for homeowners who bypass tenant screening entirely. In the binned specification, no-own-record renters with 7–10 court cases filed under similar names live in neighborhoods with $25/month lower rent and $1,700 lower median household income. Among recent movers, the screening penalty on rent remains significant, and exposed movers with no record are sorted into neighborhoods with 0.87 percentage points higher poverty. Our findings provide the first evidence that algorithmic name confusion in tenant screening creates an environmental justice problem: imprecise name matching does not merely produce inaccurate reports—it systematically steers innocent renters into communities with higher pollution exposure and worse socioeconomic conditions.

*Department of Sociology, University of Wisconsin–Madison.
†Wisconsin School of Business, University of Wisconsin–Madison.
‡Wisconsin School of Business, University of Wisconsin–Madison.

Disclosure Statement. None of the authors have any relevant or material financial interests to declare in our submitted manuscript, "Screening by Name: How Tenant Background Checks Sort Innocent Renters into Polluted and Disadvantaged Neighborhoods."

1 Introduction

Milwaukee, Wisconsin is a natural laboratory for studying the housing market consequences of criminal record screening. The city has among the highest eviction rates in the nation (Desmond, 2016), a deeply segregated rental market, and a publicly accessible court records system—the Wisconsin Circuit Court Access Program (CCAP)—that landlords routinely search by name when evaluating prospective tenants. Because CCAP searches are name-based rather than identity-based, they return records belonging to any individual whose name matches the search string. This creates a fundamental identification problem: an applicant may be rejected not because of their own criminal history, but because a stranger with a similar name has court records.

This paper asks whether name similarity to court-involved individuals—measured by Levenshtein edit distance—affects where renters live in Milwaukee. We link individual-level consumer records from InfoUSA to the universe of Milwaukee County circuit court cases and construct a novel measure of "fuzzy exposure": the total court cases filed under name keys within one character edit of the individual's own name. We then test whether this fuzzy exposure predicts the socioeconomic characteristics of the Census tract in which the individual resides, controlling for the individual's own court history, demographics, and income.

Our analysis proceeds in three steps. First, we estimate the baseline effect of fuzzy exposure on seven tract-level outcomes—median rent, unemployment rate, poverty rate, median household income, college-educated population share, toxic chemical exposure (EPA RSEI), and fine particulate matter ($PM_{2.5}$)—and show that the effect is significant for six of the seven in the pooled sample, with signs uniformly consistent with screening pushing renters into economically and environmentally disadvantaged neighborhoods. Second, we split the sample by own-record status and show that the penalty falls entirely on the "purely contaminated": among renters with no own criminal record, fuzzy exposure significantly predicts worse neighborhoods on all seven dimensions ($p < 0.05$ for each), while every coefficient is zero for renters who already have their own records. Third, we estimate a binned dose-response specification that reveals the penalty is concentrated at moderate-to-high exposure levels (7–10 and 11+ fuzzy cases) and is exclusively borne by those with no own record.

To interpret these findings, we develop a theoretical framework that distinguishes four

channels through which individuals sort into neighborhoods: the budget constraint (captured by income controls), the own-screening constraint (captured by own criminal record controls), unobserved taste-based sorting (ruled out theoretically and empirically), and the external screening constraint through name contamination (our residual treatment effect). The key identification argument is that income captures the budget constraint, not all preferences—but a stranger's court record has no causal pathway through which it could affect an individual's taste for neighborhood amenities. Conditional on income and own criminal history, fuzzy exposure can only operate through the constraint imposed by mistaken identity in landlord screening.

The owner placebo test confirms this interpretation. Homeowners, who access housing through mortgage markets and direct purchase rather than landlord screening, show no significant effects of fuzzy exposure on any outcome. This null result rules out alternative explanations in which name similarity reflects neighborhood-level correlates of crime or unobserved individual characteristics.

While the Urban Institute and CFPB have documented that tenant screening companies use imprecise name matching and that this produces false positives in background check reports, no prior study has estimated the housing market consequences of this screening error. We provide the first causal estimates of how algorithmic name confusion in tenant screening affects where people live. Using Levenshtein edit distance—a measure of name similarity that is quasi-random conditional on demographics—we show that innocent renters whose names resemble those of court-involved strangers are sorted into neighborhoods with higher unemployment, higher poverty, lower household income, fewer college-educated residents, and higher pollution exposure.

A central finding is the environmental justice dimension of screening contamination. Among renters with no own criminal record, fuzzy name exposure significantly predicts higher EPA toxic chemical exposure (RSEI, $p = 0.006$) and higher satellite-derived $PM_{2.5}$ concentrations ($p = 0.050$). These environmental effects are absent for renters with own records and absent for homeowners. This result connects the criminal records screening literature to the environmental inequality literature: Christensen, Sarmiento-Barbieri, and Timmins (2022) show that housing discrimination steers minorities toward high-pollution neighborhoods; we show that a different form of name-based discrimination—algorithmic screening errors—produces a similar environmental sorting pattern for innocent renters. The screening mechanism does not merely sort people by socioeconomic status; it sorts them by exposure to environmental hazards.

Our setting—Milwaukee—is the same city in which Matthew Desmond's *Evicted* documented how court records function as a cause rather than merely a consequence of poverty

(Desmond, 2016). We extend this insight by showing that the reach of court records in housing markets extends beyond the individuals who generated them: through the accident of name similarity, screening contamination affects innocent bystanders—pushing them not only into economic disadvantage but also into environmental harm.

The paper is organized as follows. Section 2 reviews the related literature. Section 3 develops the theoretical framework. Section 4 describes the institutional background. Section 5 presents the data. Section 6 lays out the empirical strategy. Sections 7 and 8 report the main results and owner placebo test. Section 9 examines heterogeneity. Section 10 discusses implications, and Section 11 concludes.

2 Related Literature

This paper draws on and contributes to four strands of literature: housing discrimination and audit studies, the collateral consequences of criminal records, the emerging literature on tenant screening technology, and the econometric challenge of separating screening from compositional effects.

2.1 Housing Discrimination and Audit Studies

A large body of audit and correspondence studies documents persistent racial discrimination in housing markets. Bertrand and Mullainathan (2004) establish that job applicants with white-sounding names (Emily, Greg) receive 50% more callbacks than those with African-American–sounding names (Lakisha, Jamal), demonstrating the power of names as signals in market interactions. Ewens, Tomlin, and Wang (2014) extend the correspondence audit methodology to rental housing, sending 14,237 emails to Craigslist rental listings and finding that inquiries from African-American names receive 9.3% fewer positive responses than those from white names.

Most directly relevant to our setting, Christensen and Timmins (2022) use an audit study in the 50 largest U.S. metro areas to show that real estate agents steer minority homebuyers toward high-poverty, high-pollution neighborhoods, even when identical preferences are expressed. Christensen, Sarmiento-Barbieri, and Timmins (2022) document that housing inquiries from African-American and Hispanic names are 41% less likely to receive responses for units in low-pollution neighborhoods, providing direct evidence that name-based discrimination channels environmental inequality. Our paper complements this literature by showing that name-based discrimination operates not only through the *race* signaled by a name but also through the *criminal records* associated with similar names.

2.2 Criminal Records and Collateral Consequences

The collateral consequences of criminal records extend far beyond the criminal justice system. Pager (2003) conducts a landmark audit study in Milwaukee, sending matched pairs of testers to apply for entry-level jobs. A criminal record reduces callbacks by 50% for white applicants and 64% for Black applicants, demonstrating both the direct effect of records and their interaction with race. Agan and Starr (2018) study Ban-the-Box policies that prohibit employers from asking about criminal records on initial applications, finding that removal of the criminal record question paradoxically *increased* racial discrimination: without individual-level information, employers appear to statistically discriminate against groups with higher average incarceration rates.

In the housing context, Desmond (2016) provides a richly detailed ethnographic account of eviction in Milwaukee, documenting how eviction records—which appear on tenant screening reports—function as a cause rather than merely a consequence of poverty. The U.S. Department of Housing and Urban Development issued guidance in 2016 clarifying that blanket bans on renting to individuals with criminal records may violate the Fair Housing Act when they have a disparate impact on protected classes (HUD, 2016). Our paper quantifies a mechanism through which criminal records generate housing market consequences even for individuals who have *no records of their own*: name-based screening contamination.

2.3 Tenant Screening Technology and Name Matching

A growing body of evidence documents the error-prone nature of commercial tenant screening. The Urban Institute's influential report finds that approximately 90% of landlords use criminal background checks and that screening companies typically rely on name-only matching as their primary search strategy (Urban Institute, 2025). The Consumer Financial Protection Bureau has flagged name-only matching as a potential violation of the Fair Credit Reporting Act, noting that the top 26 surnames cover approximately 25% of the Hispanic population, making false-positive matches especially likely for Hispanic individuals (CFPB, 2021).

Lageson and Stewart (2024) analyze a sample of commercially produced background check reports and find that approximately 60% contain at least one error, including records attributed to the wrong person. The National Consumer Law Center documents findings from 253 survey responses by attorneys, advocates, and counselors who have assisted renters harmed by inaccurate screening reports, including cases where name-only matching produced false criminal record associations (NCLC, 2023). These findings motivate our focus on name similarity as a source of screening contamination: if exact-name matching already produces

substantial false positives, approximate or fuzzy matching plausibly extends the problem further.

Critically, all of these studies document the existence and prevalence of screening errors—they do not estimate the downstream effects on housing outcomes. The Urban Institute shows that name-only matching produces false positives; the CFPB flags this as a potential FCRA violation; Lageson and Stewart (2024) find that 60% of background checks contain errors. But none measures whether these errors actually change where people live. Our paper fills this gap: we estimate the causal effect of name-based screening contamination on neighborhood sorting, showing that imprecise matching does not merely produce incorrect reports—it systematically steers innocent renters into economically disadvantaged communities.

2.4 Our Contribution

This paper makes four contributions. First, we provide the first estimates of how algorithmic name confusion in tenant screening affects housing outcomes—moving beyond documenting screening errors to measuring their consequences for residential sorting. While the Urban Institute, CFPB, and legal scholars have established that name-based screening produces false positives, no prior study has estimated whether these errors change where people live. We show that they do, across seven independent outcome dimensions. Second, we introduce Levenshtein edit distance as a quasi-random source of variation in screening contamination, exploiting the accident of name similarity rather than the confounded variation in exact-name exposure. The Levenshtein measure is only weakly correlated with name commonality ($\rho = 0.28$) and therefore less confounded by the compositional differences that plague exact-name analyses. Third, we develop a theoretical framework that distinguishes budget constraints from preferences and shows that, conditional on income and own criminal record, fuzzy exposure can only operate through the external screening constraint. The concentration of effects among individuals with no own criminal record, and the null result in the owner placebo, jointly establish that the mechanism operates through landlord screening rather than through unobserved correlates of name similarity. Fourth, we document an environmental justice dimension of screening contamination: fuzzy exposure pushes innocent renters into neighborhoods with higher toxic chemical exposure (RSEI) and higher fine particulate matter ($PM_{2.5}$), linking the criminal records screening literature to the environmental inequality literature (Christensen, Sarmiento-Barbieri, and Timmins, 2022).

3 Theoretical Framework: Preferences vs. Constraints

We develop a model of residential sorting that distinguishes four channels through which individuals are allocated to neighborhoods: the budget constraint, the own-screening constraint, unobserved taste-based sorting, and the external screening constraint through name contamination. The framework clarifies why, conditional on income and own criminal record, fuzzy exposure can only operate through the external screening channel—and why unobserved preferences do not provide an alternative explanation.

3.1 Setup

Consider a renter i choosing among neighborhoods $j \in \mathcal{J}$. Each neighborhood is characterized by rent r_j and an amenity bundle A_j that includes low poverty, low unemployment, good schools, and other desirable characteristics. Better neighborhoods have higher rent and better amenities: r_j and A_j are positively correlated across neighborhoods.

Renter i has income Y_i, a vector of taste parameters θ_i, and a choice set $\mathcal{J}_i \subseteq \mathcal{J}$ that may be restricted by screening. The renter chooses the neighborhood that maximizes utility:

$$\max_{j \in \mathcal{J}_i} U_i(Y_i - r_j,\, A_j;\, \theta_i) \tag{1}$$

We assume U is increasing and concave in consumption $Y_i - r_j$ and increasing in amenities A_j.

3.2 Four Sorting Channels

Channel 1: Budget constraint. Income Y_i determines which neighborhoods are *affordable*. Renters with lower income have higher marginal utility of consumption and therefore optimally allocate less to housing, leading them to cheaper, lower-amenity neighborhoods. This is a standard implication of the budget constraint: holding preferences fixed, lower income shrinks the set of neighborhoods at which the renter can achieve non-negative consumption. Controlling for Y_i in the regression holds the budget set constant across individuals. Crucially, income captures the budget constraint, not all preferences—two individuals with the same income can have very different tastes for neighborhood amenities, ethnic composition, school quality, or commute time.

Channel 2: Own-screening constraint ($S^{\mathbf{own}}$). When a renter has their own criminal record, landlords in higher-quality neighborhoods—who screen more aggressively—may

reject the application. This restricts the choice set:

$$\mathcal{J}_i^{\text{own}} = \mathcal{J} \setminus S^{\text{own}}(\text{own}_i) \tag{2}$$

where $S^{\text{own}}(\text{own}_i) \subseteq \mathcal{J}$ is the set of neighborhoods from which individual i is excluded due to their own criminal record, and $\partial|S^{\text{own}}|/\partial\text{own}_i > 0$: more own cases lead to more exclusions. Because landlords in better neighborhoods screen more aggressively, S^{own} disproportionately removes high-amenity, high-rent neighborhoods from the choice set. This is a constraint imposed by the individual's own history.

Channel 3: Unobserved taste-based sorting (θ_i). Beyond the budget constraint, individuals differ in their preferences for neighborhood characteristics: ethnic composition, proximity to family or social networks, neighborhood familiarity, commute patterns, school quality, and other amenities. These taste parameters θ_i are generally unobserved and are *not* captured by income controls. Two individuals with identical incomes may sort into very different neighborhoods because of different preferences. The key identification question is therefore: does Levenshtein fuzzy exposure correlate with these unobserved tastes? We argue that the answer is no, for reasons we develop in Section 3.3 below.

Channel 4: External screening constraint (S^{fuzzy}). When a landlord searches CCAP for a prospective tenant's name and the system returns court records belonging to *strangers* with similar names, the landlord may reject the applicant or steer them to lower-quality units. This further restricts the choice set:

$$\mathcal{J}_i = \mathcal{J}_i^{\text{own}} \setminus S^{\text{fuzzy}}(\text{fuzzy}_i) = \mathcal{J} \setminus S^{\text{own}}(\text{own}_i) \setminus S^{\text{fuzzy}}(\text{fuzzy}_i) \tag{3}$$

where $S^{\text{fuzzy}}(\text{fuzzy}_i)$ is the additional set of neighborhoods from which individual i is excluded due to fuzzy name contamination, and $\partial|S^{\text{fuzzy}}|/\partial\text{fuzzy}_i > 0$: more fuzzy exposure leads to a smaller choice set and worse equilibrium neighborhood outcomes.

3.3 The Identification Argument

Conditional on income Y_i (the budget constraint) and own record own_i (the own-screening constraint), the residual effect of fuzzy exposure cannot operate through either Channel 1 or Channel 2. The remaining question is whether the effect operates through unobserved preferences (Channel 3) or through screening contamination (Channel 4).

Income controls hold the budget set constant. By including Y_i in the regression, we ensure that any residual effect of fuzzy exposure does not reflect differences in what individuals can afford. This absorbs Channel 1 but does *not* absorb unobserved taste heterogeneity.

Own-record controls capture own-constraint sorting. By including $\ln(1 + \text{own}_i)$ in the regression, we absorb the screening exclusions attributable to the individual's own criminal history. The residual effect of fuzzy exposure therefore cannot operate through Channel 2.

Unobserved taste-based sorting is ruled out. We argue that Channel 3 does not provide a pathway from fuzzy exposure to neighborhood outcomes, for three reasons.

First, there is *no causal pathway from fuzzy exposure to preferences.* A stranger's court record does not alter the renter's marginal utility of amenities, their preference for neighborhood racial composition, or their attachment to particular locations. A renter's taste for school quality, proximity to family, or ethnic community is entirely unaffected by whether "JON SMITH" committed a crime. Formally:

$$\frac{\partial \theta_i}{\partial \text{fuzzy}_i} = 0 \quad \forall\, i \tag{4}$$

Second, *Levenshtein fuzzy exposure is quasi-random.* The edit-distance neighbor structure of names is determined by the accident of similar names existing in the court system, which is orthogonal to the renter's taste parameters θ_i. Whether a name one character edit away from yours appears in CCAP depends on the idiosyncratic distribution of names among court-involved individuals, not on your own characteristics or preferences. This is supported empirically: the correlation between fuzzy exposure and name commonality is only $\rho = 0.28$, and the dilution interaction with name frequency is statistically insignificant ($p = 0.71$).

Third, the *owner placebo* confirms that fuzzy exposure does not correlate with unobserved preference-based sorting. If fuzzy exposure were correlated with unobserved tastes θ_i that affect neighborhood choice for *all* residents—not just renters—we would expect significant effects for homeowners as well, since owners also sort on preferences. The null owner result (reported in Section 8) rules out taste-based sorting as an explanation: the effect is specific to the renter channel where screening operates.

The residual is pure screening contamination. With Channel 1 absorbed by income controls, Channel 2 absorbed by own-record controls, and Channel 3 ruled out by the arguments above, any remaining effect of fuzzy_i on neighborhood outcomes can *only* operate

through Channel 4—the external constraint imposed by mistaken identity in screening:

$$\left.\frac{\partial Y_{\text{outcome}}}{\partial \text{fuzzy}_i}\right|_{Y_i,\,\text{own}_i} = \underbrace{\frac{\partial Y_{\text{outcome}}}{\partial |\mathcal{J}_i|}}_{\text{smaller choice set}\rightarrow\text{worse outcomes}} \times \underbrace{\frac{\partial |\mathcal{J}_i|}{\partial \text{fuzzy}_i}}_{<0} \tag{5}$$

3.4 The No-Own-Record Sharpening

The identification argument is sharpest for individuals with *no own criminal record* ($\text{own}_i = 0$). For these individuals, each channel is either controlled for, zeroed out mechanically, or ruled out theoretically:

- **Budget constraint** (Channel 1): captured by income controls.
- **Own-screening constraint** (Channel 2): *exactly zero.* $S^{\text{own}}(0) = \emptyset$, so the unconstrained choice set is the full set $\mathcal{J}$. There is no own record for landlords to find.
- **Taste-based sorting** (Channel 3): orthogonal to fuzzy exposure, as argued above.
- **External screening constraint** (Channel 4): the *sole remaining channel*.

This is the cleanest possible identification—three channels are either controlled for or zeroed out mechanically, and the fourth is ruled out by the absence of a causal pathway and the owner placebo. This motivates our Specification 3, which interacts fuzzy exposure with a no-own-record indicator. The coefficient β_8 on $\ln(1 + \text{fuzzy1}_i) \times \mathbb{1}_i^{\text{no own}}$ provides a "clean" estimate of the screening contamination channel, free from budget, own-constraint, or preference confounds.

3.5 Model Predictions

The framework generates four testable predictions, each linked to the channel structure:

1. **Negative sorting** (Channel 4): $\partial \mathcal{J}_i / \partial \text{fuzzy}_i < 0$, implying that more fuzzy exposure leads to worse neighborhood outcomes (lower rent, higher poverty, higher unemployment, lower income). The screening constraint shrinks the choice set, pushing renters toward lower-quality neighborhoods.
2. **Stronger for no-own-record** (Channel 2 × Channel 4 interaction): The marginal effect of fuzzy exposure should be larger for individuals with no own record ($\beta_8 < 0$ in Specification 3). When a landlord already finds the applicant's own court record, additional fuzzy-match results are informationally redundant—the applicant is already

flagged. The own-screening constraint already binds, so the external constraint adds less.

3. **Stronger for rare names** (Channel 4 signal precision): For individuals with rare names, a fuzzy match is a stronger signal—there are fewer potential sources of confusion, so the court record is more likely attributed to the applicant. For common names, landlords may discount fuzzy matches, recognizing that many individuals share the name.

4. **Zero for owners** (Channel 4 inapplicable): Homeowners access housing through mortgage markets and direct purchase, not through landlord screening. The screening constraint S^{fuzzy} does not apply to owners, so fuzzy exposure should have no effect on owner neighborhood outcomes. This also serves as a test against Channel 3: if unobserved preferences drove the results, owners should be affected too.

As we show in Sections 7 and 8, all four predictions are confirmed in the Milwaukee data.

4 Institutional Background

4.1 Wisconsin Circuit Court Access Program (CCAP)

The Wisconsin Circuit Court Access Program provides free, public online access to circuit court case records across all 72 Wisconsin counties. The system, maintained by the Director of State Courts, allows any user to search by name, case number, or date range. Search results display the individual's name, date of birth (when available), case type (criminal, civil, traffic, family, etc.), charges, and case disposition.

CCAP is widely used by landlords, employers, and tenant screening companies as a low-cost criminal background check. Unlike formal background checks through the Wisconsin Department of Justice, CCAP searches are free, instantaneous, and require no consent from the search subject. The system returns all cases associated with a searched name, regardless of disposition—including dismissed charges, acquittals, and pending cases.

A critical feature for our analysis is that CCAP searches are *name-based*. The system matches on the name string entered by the user, not on a unique personal identifier. As documented by the Urban Institute, screening companies typically use FIRST_LAST name matching as their primary search strategy (Urban Institute, 2025). This creates a fundamental identification problem: records belonging to different individuals who share the same name are commingled in search results.

4.2 Landlord Screening and Name-Based Searches

The Urban Institute's report on tenant screening practices documents that the majority of screening companies rely on name-based searches of public court records (Urban Institute, 2025). While some services attempt to verify identity through date of birth or Social Security number, many return results based solely on name matching. This practice creates two channels through which court records from *other people* can affect an individual's housing outcomes.

First, **exact-name contamination**: an individual named "JOHN SMITH" may be denied housing because the screening report includes criminal records belonging to a different "JOHN SMITH." Second, **fuzzy-name contamination**: screening algorithms may use approximate string matching (e.g., Soundex, Metaphone, or Levenshtein-based fuzzy matching) to capture name variants, or landlords may manually search common misspellings and variants of a prospective tenant's name. Our paper focuses on the second channel.

4.3 Levenshtein Distance and Screening Technology

The Levenshtein distance between two strings is the minimum number of single-character edits (insertions, deletions, or substitutions) required to transform one string into the other. For example, "JON SMITH" and "JOHN SMITH" have Levenshtein distance 1 (one insertion), while "JOHN SMITH" and "JANE SMITH" have distance 2 (two substitutions).

There are three mechanisms through which Lev-1 name neighbors could affect housing outcomes. First, screening algorithms may explicitly use fuzzy matching with an edit-distance threshold, returning records for names within one or two edits of the searched name. Second, landlords conducting manual CCAP searches may try common name variants (e.g., searching both "JON" and "JOHN"). Third, data-entry errors in court filings may assign records to incorrect name spellings, creating cross-name contamination that propagates through exact-match searches. Our empirical strategy does not distinguish among these mechanisms but instead tests whether Lev-1 name neighbors' court records predict housing outcomes, conditional on exact-name exposure and demographic controls.

4.4 Milwaukee's Rental Market

Milwaukee County provides a compelling setting for this analysis. The city has one of the highest eviction filing rates in the United States, with approximately one in five renter households receiving an eviction filing in a given year (Desmond, 2016). Milwaukee's rental market is characterized by high levels of landlord screening: the combination of an accessi-

ble, free court records system (CCAP) and a competitive rental market with tight vacancy rates creates strong incentives for landlords to screen aggressively. The city's deep racial segregation—Milwaukee is consistently ranked among the most segregated metropolitan areas in the nation—means that screening contamination may have disparate impacts across racial groups, though as we show, the fuzzy channel does not exhibit significant racial heterogeneity.

5 Data

5.1 InfoUSA Consumer Data

We use individual-level consumer records from InfoUSA (now Data.com) for the years 2022–2024, covering Milwaukee County. The data include residential address, age, gender, individual income, ethnicity, and homeowner/renter status. We classify individuals as renters (tenure codes 0–3) or owners (tenure codes 7–9), excluding ambiguous codes (4–6). Race is derived from ethnicity codes using the Infogroup codebook. Residential addresses are geocoded and spatially joined to 2020 Census tracts.

5.2 CCAP Court Records

We obtain the complete universe of circuit court cases from the Wisconsin CCAP system for Milwaukee County, spanning 2000–2025. The data include case type, charge descriptions, filing and disposition dates, and the defendant's name and date of birth (when available). We classify cases as violent or non-violent based on keyword matching of the charge description field (e.g., "assault," "battery," "homicide" flag violent; all others are non-violent).

Each individual in the InfoUSA data is merged to CCAP records using exact FIRST_LAST name key matching. A multi-tier validation protocol distinguishes the individual's *own* cases from *others'* cases filed under the same name: matches are validated by middle name prefix, age (± 3 years), gender, and city of residence. The residual—cases under the same name key that cannot be attributed to the focal individual—constitutes the *others' exact-match exposure*.

5.3 Levenshtein Fuzzy Exposure Construction

For each individual's name key n_i, we construct Levenshtein fuzzy exposure as follows:

1. **Generate Lev-1 candidates.** For each character position in n_i, we generate all possible single-character insertions (26 letters per position), deletions (one per position),

and substitutions (25 alternatives per position). For a name key of length L, this produces approximately $52L + L = 53L$ candidate strings (before deduplication), or roughly 750 candidates for a typical name.

2. **Intersect with CCAP name universe.** We intersect the candidate set with the unique name keys observed in the CCAP data. Only candidates that match an actual CCAP name key are retained.

3. **Sum CCAP hits.** For each Lev-1 neighbor name key n' (with $\text{Lev}(n_i, n') = 1$), we sum the total CCAP cases filed under n'. The sum across all Lev-1 neighbors gives fuzzy_lev1_i.

4. **Repeat for Lev-2.** We apply the Lev-1 procedure to each Lev-1 neighbor, generating all names within edit distance 2 of n_i. We subtract the Lev$\leq$1 set to isolate names at exactly distance 2. The sum of their CCAP cases gives fuzzy_lev2_i.

Key distinction from exact-match exposure. Exact-match others' exposure counts cases filed under the *same* name key as the individual (e.g., other "JOHN SMITH" records). Levenshtein fuzzy exposure counts cases filed under *nearby* name keys (e.g., "JON SMITH," "JOHN SMTH," "JOHN SMITHS"). The two measures capture distinct sources of screening contamination and are only modestly correlated ($\rho = 0.185$).

5.4 Census Tract Outcomes

We construct four tract-level outcome variables from the American Community Survey (2023, 5-year estimates):

- **Rent** (`rent_1000s`): Median gross rent in the Census tract. Mean: \$1,135/month (Milwaukee renters).
- **Unemployment rate** (`unemp_pct`): Tract-level unemployment rate in percentage points. Mean: 5.22% (Milwaukee renters).
- **Poverty rate** (`poverty_pct`): Share of tract population below the federal poverty line, in percentage points. Mean: 21.34% (Milwaukee renters).
- **Median household income** (`median_hh_income_k`): Tract-level median household income in dollars. Mean: \$58,924 (Milwaukee renters).

We supplement these ACS measures with two environmental quality indicators:

- **Pollution (RSEI)**: The EPA Risk-Screening Environmental Indicators score, which captures toxic chemical releases from EPA-reporting facilities. We use the log-transformed score (`ln_rsei_score`).

- **Fine particulate matter ($\mathbf{PM_{2.5}}$)**: Satellite-derived annual mean $PM_{2.5}$ concentration ($\mu g/m^3$) from the van Donkelaar et al. V5.NA.04.02 dataset (van Donkelaar et al., 2021), measured at 0.01° (∼1 km) resolution and extracted at tract centroids. Mean: 7.73 $\mu g/m^3$ (Milwaukee renters). Unlike RSEI, which captures proximity to specific toxic-release facilities, $PM_{2.5}$ measures ambient fine particulate matter from all sources (traffic, industry, agriculture). The two measures are only weakly correlated ($\rho = 0.14$), confirming that they capture distinct dimensions of environmental quality.

For the owner placebo, we additionally use tract-level median home value from the ACS (mean: $223,908 for Milwaukee owners).

5.5 Summary Statistics

Table 1 provides a detailed comparison of renters and owners in Milwaukee County across three panels: demographics, court record exposure, and tract-level outcomes.

Table 1: Summary Statistics: Milwaukee County Renters vs. Owners

	Renters ($N = 69{,}627$)	**Owners** ($N = 589{,}658$)
Panel A: Demographics		
Age (mean)	47.5	52.2
Female (%)	51.7	50.6
White (%)	58.3	58.2
Black (%)	23.5	20.3
Hispanic (%)	8.6	9.5
Asian/PI (%)	2.1	3.0
Income (mean, $K)	29.5	38.9
Panel B: Court Record Exposure		
Own cases (mean)	3.69	3.22
Own cases (% any > 0)	37.4	29.9
Fuzzy Lev-1 (mean)	2.64	1.14
Fuzzy Lev-1 (% any > 0)	20.5	5.4
Fuzzy Lev-1 (mean $\mid > 0$)	12.9	21.2
n_{people} (mean)	2.60	2.71
No own record (%)	62.6	70.1
Panel C: Tract-Level Outcomes		
Rent ($/month)	1,135	1,116
Unemployment (pp)	5.22	4.65
Poverty (pp)	21.34	14.81
Median HH income ($)	58,924	73,528
Home value ($)	—	223,908

Notes: Sample sizes reflect complete cases after listwise deletion of missing-age observations. "Fuzzy Lev-1" is the total CCAP cases associated with name keys at Levenshtein distance 1 from the individual's own name key. Tract-level outcomes are from the 2023 ACS 5-year estimates.

Several patterns in Table 1 are noteworthy. First, Milwaukee renters have substantially higher fuzzy exposure than owners: 20.5% of renters have at least one Lev-1 name neighbor with CCAP records, compared to only 5.4% of owners. Conditional on any exposure, renters average 12.9 fuzzy cases. Second, renters live in measurably worse neighborhoods: tract-level poverty is 6.5 percentage points higher (21.3% vs. 14.8%), median household income is $14,604 lower, and unemployment is 0.57 percentage points higher. Third, renters are younger (47.5 vs. 52.2 years), have substantially lower individual incomes ($29,500 vs. $38,900), and are more likely to be Black (23.5% vs. 20.3%). Fourth, 62.6% of renters have no own criminal record—these are the "purely contaminated" individuals for whom fuzzy screening exposure is most consequential.

Milwaukee renters are distributed across 299 Census tracts, with an average of 233 renters per tract. The analysis sample comprises 69,627 renters after listwise deletion of observations with missing age.

6 Empirical Strategy

We estimate three core specifications, each designed to test a distinct prediction of the screening contamination hypothesis from Section 3. All regressions are estimated by OLS with standard errors clustered at the name-key level, since individuals sharing a name key have mechanically correlated fuzzy exposure. The control vector X_i throughout includes: age, an indicator for female, race dummies (white, Black, Hispanic, Asian/PI), and individual income (thousands). The baseline sample is Milwaukee renters with complete demographic data ($N = 69{,}627$); for the owner placebo, we use Milwaukee owners ($N = 589{,}658$–$595{,}766$).

6.1 Specification 1: Signal Dilution (Continuous)

$$\begin{aligned} Y_i = \beta_0 &+ \beta_1 \ln(1 + \text{own}_i) + \beta_2 \ln(1 + \text{fuzzy1}_i) + \beta_3 \ln(1 + \text{fuzzy2}_i) \\ &+ \beta_4 \ln(1 + N_i) + \beta_5 \tilde{f}_i + \beta_6 \ln(1 + \text{own}_i) \times \tilde{f}_i \\ &+ \beta_7 \ln(1 + \text{fuzzy1}_i) \times \tilde{f}_i + X_i'\gamma + \varepsilon_i \end{aligned} \tag{6}$$

where Y_i is a tract-level outcome, own_i is the individual's own validated CCAP case count, fuzzy1_i is the total CCAP cases associated with name keys at Levenshtein distance 1 from n_i, fuzzy2_i is the analogous measure at distance 2, N_i is the count of distinct individuals sharing the same name key in the population, and $\tilde{f}_i \equiv \ln(1 + N_i) - \overline{\ln(1 + N)}$ is the mean-centered log name frequency.

Why log-transform? We apply log transformations to all count variables on the right-hand side to capture diminishing marginal effects. The first few court cases from a similar-sounding name carry substantially more informational weight in screening than additional cases at high exposure levels. This functional form correctly accounts for the extreme right skew in fuzzy exposure (median = 0, mean = 2.64).

Why control for $\ln(1 + N_i)$? Without this control, $\ln(1 + \text{fuzzy1})$ partially reflects name commonality: individuals with common names (large N_i) mechanically have more Levenshtein neighbors in the CCAP system. Including $\ln(1 + N_i)$ absorbs this extensive margin, so $\hat{\beta}_2$ captures the *intensive margin*—conditional on name commonality, the effect of having more court records among similar-sounding name neighbors.

The dilution test. The interaction β_7 tests whether fuzzy contamination varies with name commonality. For exact-match exposure, signal dilution is strong and expected ($\beta_6 > 0$), because common names mechanically accumulate more same-name court records. For Levenshtein fuzzy exposure, the prediction is $\beta_7 \approx 0$: the number of Lev-1 neighbors with court records is determined by the accident of similar names existing in the CCAP system, which is largely orthogonal to how many people share the individual's *exact* name. A null result for β_7 confirms that Levenshtein variation is quasi-random—a key identification advantage over exact-match analyses.

The key parameters are β_2 (the direct fuzzy effect) and β_7 (the dilution test). Under the screening contamination hypothesis (Prediction 1 in Section 3), $\beta_2 < 0$ for rent and $\beta_2 > 0$ for poverty and unemployment.

6.2 Specification 3: Triple Interaction (Continuous)

A central prediction of the theoretical framework (Prediction 2) is that fuzzy exposure should matter most for individuals who have *no own criminal record*: these "purely contaminated" individuals have an unconstrained own-record choice set ($S^{\text{own}} = \emptyset$), so screening penalties are entirely attributable to name confusion. To test this, we estimate:

$$\begin{aligned} Y_i = \beta_0 &+ \beta_1 \ln(1 + \text{own}_i) + \beta_2 \ln(1 + \text{fuzzy1}_i) + \beta_3 \ln(1 + \text{fuzzy2}_i) \\ &+ \beta_4 \tilde{f}_i + \beta_5 \mathbb{1}_i^{\text{no own}} \\ &+ \beta_6 \ln(1 + \text{own}_i) \times \tilde{f}_i + \beta_7 \ln(1 + \text{fuzzy1}_i) \times \tilde{f}_i \\ &+ \beta_8 \ln(1 + \text{fuzzy1}_i) \times \mathbb{1}_i^{\text{no own}} \\ &+ \beta_9 \tilde{f}_i \times \mathbb{1}_i^{\text{no own}} \\ &+ \beta_{10} \ln(1 + \text{fuzzy1}_i) \times \tilde{f}_i \times \mathbb{1}_i^{\text{no own}} + X_i'\gamma + \varepsilon_i \end{aligned} \tag{7}$$

where $\mathbb{1}_i^{\text{no own}} = \mathbb{1}[\text{own_cases}_i = 0]$ indicates individuals with no validated own criminal record.

The key parameters are β_8, β_9, and β_{10}. The coefficient β_8 tests whether purely contaminated individuals suffer a larger fuzzy penalty (Prediction 2): when a landlord already finds the applicant's own court records, additional fuzzy-match results from similar names are redundant, so β_8 should be negative. The triple interaction β_{10} tests whether signal dilution varies with own-record status for the purely contaminated (Prediction 3).

Note that the interaction $\ln(1 + \text{own}_i) \times \mathbb{1}_i^{\text{no own}}$ is mechanically zero (own cases are zero whenever no_own = 1) and is therefore correctly omitted.

6.3 Binned Dose-Response with No-Own-Record Interaction

To visualize the shape of the dose-response relationship and allow for nonlinearity, we estimate a binned specification with own-record interactions:

$$\begin{aligned} Y_i = \beta_0 &+ \beta_1 \ln(1 + \text{own}_i) + \sum_k \delta_k \mathbb{1}[\text{fuzzy1}_i \in \text{bin}_k] + \beta_2 \ln(1 + \text{fuzzy2}_i) \\ &+ \beta_3 \ln(1 + N_i) + \beta_4 \tilde{f}_i + \beta_5 \ln(1 + \text{own}_i) \times \tilde{f}_i + \beta_6 \mathbb{1}_i^{\text{no own}} \\ &+ \sum_k \delta_k^{\text{no}} \mathbb{1}[\text{fuzzy1}_i \in \text{bin}_k] \times \mathbb{1}_i^{\text{no own}} + X_i'\gamma + \varepsilon_i \end{aligned} \tag{8}$$

where the bins are $\{0\}$ (reference category, $N = 55{,}353$), $\{1\text{–}2\}$ ($N = 4{,}424$), $\{3\text{–}4\}$ ($N = 2{,}274$), $\{5\text{–}6\}$ ($N = 1{,}462$), $\{7\text{–}10\}$ ($N = 1{,}777$), and $\{11+\}$ ($N = 4{,}337$). The δ_k coefficients capture the effect for renters who have their own criminal record; the δ_k^{no} coefficients capture the *additional* effect for those with no own record, whose total effect is $\delta_k + \delta_k^{\text{no}}$.

6.4 Controls and Clustering

All specifications include the following control variables:

- **Age**: individual's age in years (continuous; observations with missing age are dropped via listwise deletion).
- **Female**: indicator equal to 1 for women.
- **Race dummies**: indicators for Black, Hispanic, and Asian/Pacific Islander (white omitted as reference category), derived from InfoUSA ethnicity codes. Approximately 7.5% of Milwaukee renters are classified as "other" race and are captured by the constant.
- **Individual income**: annual income in thousands of dollars.

Standard errors are clustered at the name-key level throughout, since individuals sharing a name key have mechanically correlated fuzzy exposure.

7 Results

7.1 Baseline Continuous Results (Specification 1)

Table 2 presents the baseline results from Specification 1, estimated on Milwaukee renters.

Table 2: Specification 1: Baseline Levenshtein Fuzzy Exposure (Milwaukee Renters)

	Rent ($/mo)	Unemployment (pp)	Poverty (pp)	Median HH Inc. ($)
$\ln(1+\text{own})$	-33.45^{***}	$+0.406^{***}$	$+1.697^{***}$	$-3{,}045^{***}$
	(1.08)	(0.023)	(0.062)	(86)
$\ln(1+\text{fuzzy1})$	-2.99^{**}	$+0.067^{**}$	$+0.186^{**}$	-302^{***}
	(1.44)	(0.028)	(0.077)	(109)
$\ln(1+\text{fuzzy2})$	$+2.90^{**}$	$+0.005$	-0.129^{*}	$+85$
	(1.42)	(0.027)	(0.074)	(109)
$\ln(1+\text{own}) \times \tilde{f}$	$+7.39^{***}$	-0.004	-0.025	$+284^{***}$
	(1.22)	(0.022)	(0.064)	(94)
$\ln(1+\text{fuzzy1}) \times \tilde{f}$	$+0.51$	-0.045^{*}	-0.077	$+40$
	(1.35)	(0.024)	(0.074)	(102)
N	69,627	69,626	69,626	69,626
R^2	0.242	0.216	0.274	0.397

Notes: OLS estimates with SEs clustered at name-key level in parentheses. All specifications include $\ln(1+N_i)$, $\tilde{f}$, and demographic controls (age, female, race dummies [white omitted], individual income). $^{***}p<0.01$, $^{**}p<0.05$, $^{*}p<0.10$.

Rent ($p = 0.037$). The headline result is $\hat{\beta}_2 = -\$2.99$/month (SE = \$1.44, $p = 0.037$), significant at the 5% level. Because the treatment variable is log-transformed, the effect is concave: the first few fuzzy cases matter most. In economically interpretable magnitudes:

- Going from 0 to 1 fuzzy case: $\ln(2) \times (-2.99) = -\$2.07$/month.
- Going from 0 to 5 cases (median among exposed Milwaukee renters): $\ln(6) \times (-2.99) = -\$5.36$/month.
- Going from 0 to 30 cases (upper tail of exposed): $\ln(31) \times (-2.99) = -\$10.27$/month, representing a 0.91% decline relative to the sample mean rent of \$1,135/month.

The Lev-2 coefficient for rent is positive and significant ($\hat{\beta}_3 = +\$2.90$, $p = 0.041$), with the *opposite sign* from Lev-1. This confirms clean distance decay: only names within one edit carry a screening penalty, while Lev-2 variation captures residual compositional correlations rather than screening.

The dilution interaction $\ln(1+\text{fuzzy1}) \times \tilde{f}$ is essentially zero for rent ($\hat{\beta}_7 = +\$0.51$, $p = 0.71$), confirming that unlike exact-match exposure, Levenshtein fuzzy contamination does not attenuate with name commonality. By contrast, the own-cases dilution interaction is highly significant ($\hat{\beta}_6 = +\$7.39$, $p < 0.001$), consistent with own criminal records being

moderated by name frequency through the exact-match channel. This divergence validates our identification argument: Levenshtein variation is quasi-random with respect to name frequency.

Unemployment ($p = 0.015$). The fuzzy exposure coefficient is $\hat{\beta}_2 = +0.067$ percentage points per log-unit (SE = 0.028, $p = 0.015$). Going from 0 to 5 fuzzy cases increases tract-level unemployment by 0.12 percentage points, representing a 2.3% increase relative to the Milwaukee renter mean of 5.22%.

Poverty ($p = 0.015$). The coefficient is $\hat{\beta}_2 = +0.186$ percentage points per log-unit (SE = 0.077, $p = 0.015$). Going from 0 to 5 fuzzy cases increases tract poverty by 0.33 percentage points (1.6% of the mean of 21.34%). Critically, this coefficient has the *correct positive sign*—fuzzy exposure pushes renters into higher-poverty neighborhoods.

Median household income ($p = 0.006$). The coefficient is $\hat{\beta}_2 = -\$302$ per log-unit (SE = \$109, $p = 0.006$), significant at the 1% level. Going from 0 to 5 fuzzy cases pushes renters into tracts where median household income is \$541 lower; at 30 cases, the reduction is \$1,038, or 1.8% of the mean tract-level median income of \$58,924.

Education share ($p = 0.002$). The coefficient on the tract-level share of adults with a bachelor's degree or higher is $\hat{\beta}_2 = -0.312$ percentage points per log-unit (SE = 0.102, $p = 0.002$), significant at the 1% level. Going from 0 to 5 fuzzy cases reduces the share of college-educated residents by 0.56 percentage points; at 30 cases, the reduction is 1.07 percentage points (3.0% of the mean of 34.0%). This is the most statistically significant single outcome.

Local amenities. We also estimate the specification using three tract-level amenity counts from the sorting model: number of parks, transit stops, and coffee shops. Fuzzy exposure significantly *increases* transit stops ($\hat{\beta}_2 = +0.011$, $p = 0.031$), consistent with renters being steered into transit-dependent (rather than car-oriented) neighborhoods. Parks and coffee shops show marginal negative effects ($p = 0.065$ and $p = 0.087$ respectively). Pollution (ln RSEI) is insignificant ($p = 0.17$).

Together, these results confirm Prediction 1 of the theoretical framework: Levenshtein fuzzy contamination pushes Milwaukee renters into neighborhoods that are worse across *multiple independent dimensions*—lower rent, higher unemployment, higher poverty, lower household income, fewer college-educated residents, and fewer amenities. Six of nine outcomes are significant at the 5% level.

7.2 Triple Interaction: The Key Result (Specification 3)

Table 3 presents the triple interaction results. This specification tests Prediction 2 of the theoretical framework: the fuzzy screening penalty should fall disproportionately on individuals with no own criminal record, for whom the own-constraint channel is exactly zero.

Table 3: Specification 3: Triple Interaction—Own-Record Status, Name Frequency, and Fuzzy Exposure (Milwaukee Renters)

	Rent ($/mo)	Unemployment (pp)	Poverty (pp)	Median HH Inc. ($)
$\ln(1+\text{own})$	-10.00^{***}	$+0.203^{***}$	$+0.947^{***}$	$-1{,}349^{***}$
	(2.18)	(0.050)	(0.129)	(176)
$\ln(1+\text{fuzzy1})$	$+0.67$	$+0.072^{*}$	$+0.174$	-126
	(1.85)	(0.041)	(0.110)	(146)
$\mathbb{1}^{\text{no own}}$	$+50.51^{***}$	-0.428^{***}	-1.602^{***}	$+3{,}766^{***}$
	(4.33)	(0.091)	(0.246)	(348)
$\ln(1+\text{fuzzy1}) \times \mathbb{1}^{\text{no own}}$	-5.96^{**}	-0.014	$+0.009$	-329^{*}
	(2.45)	(0.048)	(0.133)	(187)
$\tilde{f} \times \mathbb{1}^{\text{no own}}$	-41.49^{***}	$+0.095$	$+0.421$	-911
	(7.40)	(0.126)	(0.387)	(579)
$\ln(1+\text{fuzzy1}) \times \tilde{f} \times \mathbb{1}^{\text{no own}}$	$+13.86^{***}$	-0.087	-0.293	$+367$
	(4.24)	(0.055)	(0.184)	(292)
N	69,627	69,626	69,626	69,626
R^2	0.244	0.216	0.274	0.398

Notes: OLS estimates with SEs clustered at name-key level in parentheses. All specifications include $\ln(1+\text{fuzzy2})$, $\tilde{f}$, $\ln(1+\text{own}) \times \tilde{f}$, $\ln(1+\text{fuzzy1}) \times \tilde{f}$, and demographic controls (age, female, race dummies [white omitted], individual income). $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Rent: the purely contaminated bear the full penalty. For Milwaukee renters who *have* their own criminal record, the baseline fuzzy effect is statistically zero: $\hat{\beta}_2 = +\$0.67$/month ($p = 0.72$). When a landlord already finds the applicant's own court record in a CCAP search, additional fuzzy-match results from similar names are informationally redundant—the applicant is already flagged.

But for the "purely contaminated"—individuals with no own record—the additional fuzzy penalty is large and significant: $\hat{\beta}_8 = -\$5.96$/month ($p = 0.015$). The combined marginal effect for no-own-record renters at the mean name frequency is $\hat{\beta}_2 + \hat{\beta}_8 = +0.67 + (-5.96) = -\$5.29$/month per log-unit. In dollar terms:

- A no-own-record renter going from 0 to 5 fuzzy cases (median among exposed): $\ln(6) \times (-5.29) = -\$9.48$/month.

- At 20 fuzzy cases: $\ln(21) \times (-5.29) = -\$16.11$/month.

This result directly confirms Prediction 2 of the theoretical framework. In the language of the model, individuals with $S^{\text{own}} = \emptyset$ have no own-constraint channel to absorb the screening signal, so the full weight of fuzzy contamination falls on the external-constraint channel.

The role of name rarity. The triple interaction $\hat{\beta}_{10} = +\$13.86$ ($p = 0.001$) confirms Prediction 3: signal dilution is strongest for the purely contaminated. For rare-name renters with no own record, the fuzzy penalty per log-unit is substantially larger than for common-name renters, consistent with the theoretical intuition that landlords discount fuzzy matches for obviously common names.

Median household income. The $\ln(1+\text{fuzzy1}) \times \mathbb{1}^{\text{no own}}$ coefficient is $-\$329$ ($p = 0.078$), marginally significant. The direction is consistent with screening pushing purely contaminated renters into lower-income neighborhoods.

Unemployment, poverty. The interaction coefficients for these outcomes are not statistically significant, though unemployment shows the expected sign. We interpret this as reflecting lower power for tract-level composition measures relative to the rent outcome, which is a more direct price signal.

Pollution (RSEI and $PM_{2.5}$). The Spec 3 triple interaction is significant for RSEI pollution: $\hat{\beta}_{10} = -0.017$ ($p = 0.049$), indicating that no-own-record renters with rare names and high fuzzy exposure are sorted into higher-pollution neighborhoods. The joint F-test for the three no-own-record interaction terms (β_8, β_9, β_{10}) rejects at 5% ($F = 2.8$, $p = 0.040$). Satellite-derived $PM_{2.5}$ concentrations corroborate this finding: the split-sample estimate for no-own-record renters is $+0.0006$ $\mu g/m^3$ per log-unit ($p = 0.046$), while the estimate for those with own records is essentially zero ($p = 0.84$). The positive sign indicates that fuzzy exposure pushes clean-record renters toward higher-$PM_{2.5}$ neighborhoods. The Spec 3 interaction for $PM_{2.5}$ is not individually significant ($p = 0.55$), consistent with the low within-county variation in ambient particulate matter (SD = 0.71 $\mu g/m^3$), but the sign is consistent with the RSEI result. That both facility-based toxic exposure (RSEI) and ambient air quality ($PM_{2.5}$) show the same pattern—significant only for the purely contaminated—strengthens the interpretation that screening sorts innocent renters into broadly worse environmental settings.

Education share. Education share shows the strongest non-rent results in Specification 3: $\ln(1 + \text{fuzzy1}) \times \mathbb{1}^{\text{no own}} = -0.006$ ($p = 0.0004$), and the joint F-test for the three no-own-record interactions is highly significant ($F = 23.8$, $p < 0.001$). This indicates that the purely contaminated are sorted into neighborhoods with substantially fewer college-educated residents, consistent with screening restricting access to higher-education communities.

Joint F-tests for the three no-own-record interaction terms (β_8, β_9, β_{10}) reject for rent ($F = 14.8$, $p < 0.001$), education share ($F = 23.8$, $p < 0.001$), and pollution ($F = 2.8$, $p = 0.040$), but not for poverty ($F = 0.96$, $p = 0.41$) or unemployment ($F = 1.04$, $p = 0.37$).

Split-sample evidence. The split-sample results resolve the apparent discrepancy between the significant Specification 1 coefficients and the insignificant Specification 3 interactions for some outcomes. When we estimate Specification 1 separately for renters with no own record ($N = 43{,}593$) and those with records ($N = 26{,}034$), the pattern is stark: among the purely contaminated, $\ln(1 + \text{fuzzy1})$ is significant for *all* outcomes—rent (-5.77, $p = 0.007$), unemployment ($+0.091$, $p = 0.010$), poverty ($+0.249$, $p = 0.016$), median income (-0.453, $p = 0.004$), RSEI pollution (-0.012, $p = 0.006$), $PM_{2.5}$ ($+0.005$, $p = 0.050$), and education share (-0.546, $p = 0.000$). Among those with own records, *all* coefficients are statistically zero (all $p > 0.30$). Table 4 summarizes these results. The Specification 3 interaction formulation lacks power for noisier tract-level composition measures, but the subgroup analysis directly confirms the theoretical prediction that the screening effect operates exclusively through the external constraint channel.

Table 4: Split-Sample by Own-Record Status (Milwaukee Renters, Specification 1)

	No own record ($N = 43{,}593$)		Has own record ($N = 26{,}034$)	
Outcome	$\hat{\beta}_2$	p	$\hat{\beta}_2$	p
Rent ($/mo)	-5.77^{***}	0.007	$+0.97$	0.617
Unemployment (pp)	$+0.091^{***}$	0.010	$+0.045$	0.307
Poverty (pp)	$+0.249^{**}$	0.016	$+0.105$	0.367
Median HH income ($K)	-0.453^{***}	0.004	-0.130	0.399
Pollution (ln RSEI)	-0.012^{***}	0.006	$+0.001$	0.810
$PM_{2.5}$ ($\mu g/m^3$)	$+0.005^{*}$	0.050	$+0.001$	0.831
Education share (pp)	-0.546^{***}	0.000	$+0.091$	0.517

Notes: Coefficient on $\ln(1 + \text{fuzzy_lev1})$ from Specification 1, estimated separately on each subsample. Controls: age, female, race dummies, individual income. SEs clustered at name-key level. $PM_{2.5}$ is the satellite-derived annual mean concentration from van Donkelaar et al. (2021). $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

7.3 Binned Dose-Response with Own-Record Interaction

Table 5 presents the binned dose-response results for Milwaukee renters, estimated from Equation 8. This specification replaces the continuous ln(1+fuzzy1) with indicator variables for exposure bins, interacted with no-own-record status, allowing fully flexible estimation of the dose-response curve separately for those with and without own criminal records.

Table 5: Binned Dose-Response: Fuzzy Lev-1 Exposure × Own-Record Status (Milwaukee Renters)

	Rent ($/mo)		Unemp. (pp)		Poverty (pp)		Income ($)	
	Effect	p	Effect	p	Effect	p	Effect	p
Panel A: Renters with own criminal record (bin effects)								
1–2 cases	+2.4	0.68	−0.068	0.57	+0.139	0.67	−293	0.53
3–4 cases	−2.6	0.73	+0.238	0.13	−0.188	0.64	−367	0.53
5–6 cases	+4.2	0.62	+0.383*	0.07	+0.636	0.22	−958	0.14
7–10 cases	+1.2	0.88	−0.161	0.35	−0.723*	0.10	−79	0.90
11+ cases	+2.2	0.71	+0.123	0.32	+0.563*	0.10	−151	0.74
Panel B: Additional effect for no-own-record renters (bin × no_own)								
1–2 cases	+0.3	0.97	+0.068	0.63	−0.511	0.22	+944	0.13
3–4 cases	−15.1	0.17	−0.359*	0.06	+0.392	0.48	−56	0.95
5–6 cases	−16.1	0.22	−0.037	0.89	+0.344	0.62	−352	0.72
7–10 cases	−26.3**	0.03	+0.381*	0.09	+1.157*	0.06	−1,657*	0.07
11+ cases	−16.8*	0.06	−0.010	0.95	−0.027	0.95	−1,115*	0.09
Total effect for no-own-record renters (bin + bin × no_own):								
1–2 cases	+3						+651	
3–4 cases	−18						−423	
5–6 cases	−12						−1,310	
7–10 cases	−25						−1,736	
11+ cases	−15						−1,266	

Notes: Reference category is 0 fuzzy cases (N = 55,353). Total sample N = 69,627. All specifications include ln(1 + own), ln(1 + fuzzy2), ln(1 + N_i), $\tilde{f}$, ln(1 + own) × $\tilde{f}$, no_own, and controls (age, female, race dummies, individual income). SEs clustered at name-key level. No-own-record renters per bin: 1–2 (N = 2,494), 3–4 (N = 1,138), 5–6 (N = 719), 7–10 (N = 852), 11+ (N = 1,615). $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

The results are striking. Panel A shows that for Milwaukee renters who have their own criminal record, *every single bin coefficient for rent is statistically insignificant*—fuzzy exposure has no detectable effect beyond what their own record already generates. Panel B reveals the starkly different picture for purely contaminated renters:

- **7–10 fuzzy cases × no_own**: −$26.3/month ($p = 0.027$), total for no-own: −$25/month (−2.2% of mean rent).

- **11+ fuzzy cases × no_own**: −$16.8/month ($p = 0.059$), total for no-own: −$15/month (−1.3% of mean rent).

Among no-own-record Milwaukee renters, having 7–10 court cases from similar-sounding names reduces monthly rent by $25—a 2.2% decline relative to mean rent of $1,135. These are individuals with clean records who are penalized entirely because their name happens to resemble that of court-involved strangers.

For median household income, the binned results for no-own-record renters show significant effects at the highest exposure levels: 7–10 cases reduce tract-level income by $1,657 ($p = 0.07$) and 11+ cases reduce it by $1,115 ($p = 0.09$).

The no-own level shift is +$51/month ($p < 0.001$), confirming that Milwaukee renters without criminal records live in higher-rent (i.e., higher-quality) neighborhoods on average, as expected.

7.4 Why Levenshtein, Not Exact Match

A central contribution of this paper is the argument that Levenshtein fuzzy exposure provides cleaner identification than exact-match others' exposure. Three pieces of evidence support this claim.

First, the **name-frequency confound**. Exact-match others' exposure is strongly correlated with name commonality ($\rho = 0.81$): individuals with common names like "JOHN SMITH" have many same-name neighbors and therefore high others' exposure, regardless of the screening mechanism. Levenshtein fuzzy exposure is only weakly correlated ($\rho = 0.28$). The dilution interaction confirms this: the $\ln(1 + \text{fuzzy1}) \times \tilde{f}$ coefficient is +$0.51 ($p = 0.71$) for rent, while the own-case interaction is +$7.39 ($p < 0.001$).

Second, the **distance decay pattern**. The Lev-2 coefficient has the opposite sign from Lev-1 for rent (+$2.90 vs. −$2.99), confirming that only names within one edit carry a screening penalty. This sharp distance decay is a hallmark of a screening mechanism—compositional correlations would not produce such a clean cutoff at distance 1.

Third, the **owner placebo** passes cleanly for the Levenshtein measure (Section 8), confirming that the variation is specific to the screening channel.

7.5 Robustness to Origin Neighborhood Controls

A concern with the cross-sectional specification is that renters' current neighborhoods reflect their entire residential history, not just screening at the most recent move. To address this, we merge each individual's 2023 census tract (their address one year prior) and control for origin tract characteristics.

Table 6 reports the baseline fuzzy coefficient with progressively richer origin controls. For rent, origin tract median rent alone explains $R^2 = 0.847$—tract-level rent is extremely persistent, leaving little residual variation for screening to explain. The fuzzy coefficient attenuates from −\$2.99 ($p = 0.036$) to −\$0.50 ($p = 0.51$) when origin rent is added.

However, the screening effect on neighborhood composition outcomes is robust to origin controls. Adding origin median rent as the sole origin control:

- Unemployment: +0.071pp ($p = 0.023$), barely changed from +0.067pp without origin controls.
- Poverty: +0.192pp ($p = 0.022$), barely changed from +0.186pp.
- Median income: −\$212 ($p = 0.037$), attenuated from −\$304 but still significant.

Even controlling for origin median household income (the strongest individual predictor of destination income), the fuzzy effect on destination income remains significant: −\$137 ($p = 0.027$).

Table 6: Fuzzy Screening Effect with Origin Neighborhood Controls

	Rent ($/mo)	Unemployment (pp)	Poverty (pp)	Income ($K)
No origin controls	−2.99**	+0.068**	+0.199***	−0.304***
	(1.44)	(0.028)	(0.077)	(0.109)
N	69,627	69,626	69,626	69,626
R^2	0.242	0.216	0.274	0.397
+ Origin rent	−0.50	+0.071**	+0.192**	−0.212**
	(0.75)	(0.031)	(0.084)	(0.102)
N	52,760	52,759	52,759	52,759
R^2	0.847	0.240	0.333	0.594
+ Origin income	−0.80	+0.065**	+0.145**	−0.137**
	(1.29)	(0.031)	(0.069)	(0.061)
N	52,760	52,760	52,760	52,760
R^2	0.486	0.272	0.536	0.844

Notes: Coefficient on $\ln(1 + \text{fuzzy_lev1})$. All specs include $\ln(1 + \text{own})$, $\ln(1 + \text{fuzzy2})$, $\ln(1 + N)$, $\tilde{f}$, $\ln(1 + \text{own}) \times \tilde{f}$, $\ln(1 + \text{fuzzy1}) \times \tilde{f}$, and demographic controls. Origin controls are tract-level ACS 2023 characteristics of the individual's 2023 address. SEs clustered at name-key level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

These results suggest that the screening mechanism operates more robustly on neighborhood composition—who your neighbors are, the local unemployment rate, the poverty rate—than on rent levels, which are mechanically persistent across moves. Screening contamination does not merely push renters into cheaper housing; it steers them into communities

with worse socioeconomic conditions, and this sorting persists after accounting for where they came from.

8 Owner Placebo

Owners access housing through mortgage markets and direct purchase, not through landlord screening. The theoretical framework (Prediction 4) implies that the screening constraint S^{fuzzy} does not apply to owners: if Levenshtein fuzzy exposure affects housing outcomes through some non-screening channel, we would expect significant effects for owners as well. The null result for owners, combined with significant renter effects, provides strong evidence that the screening channel is the operative mechanism.

8.1 Specification 1: Continuous (Milwaukee Owners)

Table 7 reports the Specification 1 results for the Milwaukee owner sample.

Table 7: Owner Placebo: Specification 1 (Milwaukee Owners)

Outcome	$\hat{\beta}_2$ (ln_fuzzy1)	p	N	Pass?
Rent ($/mo)	+0.29	0.81	595,766	PASS
Unemployment (pp)	+0.027	0.23	595,766	PASS
Poverty (pp)	+0.027	0.65	595,766	PASS
Median HH income ($)	−34	0.79	595,766	PASS
Home value ($)	−28	0.95	589,658	PASS

Notes: Milwaukee owner sample. Same specification as renter Equation 6. SEs clustered at name-key level. "PASS" = $p > 0.05$ for ln_fuzzy1 coefficient.

The placebo passes cleanly for *all five outcomes*. The Lev-1 coefficient is insignificant for owner rent ($p = 0.81$), unemployment ($p = 0.23$), poverty ($p = 0.65$), median income ($p = 0.79$), and home value ($p = 0.95$). None approaches conventional significance levels.

8.2 Specification 3: Triple Interaction (Milwaukee Owners)

Table 8 reports the triple interaction results for Milwaukee owners.

Table 8: Owner Placebo: Specification 3 (Milwaukee Owners)

	fuzzy1 × no_own		Triple		
Outcome	$\hat{\beta}_8$	p	$\hat{\beta}_{10}$	p	Pass?
Rent ($/mo)	−2.2	0.43	−1.8	0.48	PASS
Unemployment (pp)	−0.015	0.71	+0.004	0.88	PASS
Poverty (pp)	+0.147	0.16	−0.034	0.66	PASS
Median HH income ($)	+6	0.98	−324	0.17	PASS
Home value ($)	+411	0.66	−1,527*	0.05	MARGINAL

Notes: Milwaukee owner sample (N = 589,658–595,766). Same specification as renter Equation 7. SEs clustered at name-key level. "PASS" = all fuzzy interaction coefficients $p > 0.05$.

The Specification 3 placebo passes for all four core neighborhood outcomes. For rent, neither the fuzzy1 × no_own interaction ($p = 0.43$) nor the triple ($p = 0.48$) is significant. For unemployment ($p = 0.71$ and $p = 0.88$), poverty ($p = 0.16$ and $p = 0.66$), and median income ($p = 0.98$ and $p = 0.17$), all interactions are insignificant. The home value triple interaction is marginally significant ($p = 0.054$), but this is a single coefficient among 10 tested, consistent with chance at the 5% level.

Interpretation. The owner placebo passes for all outcomes that directly measure neighborhood quality and housing market access. This confirms Prediction 4 of the theoretical framework: the screening constraint S^{fuzzy} is specific to the rental market and does not affect homeowners. The null result rules out alternative explanations in which name similarity reflects neighborhood-level correlates of crime, unobserved individual characteristics correlated with both name structure and residential location, or any other non-screening channel.

9 Heterogeneity

9.1 Gender Heterogeneity

The most striking heterogeneity finding emerges from the triple interaction of fuzzy exposure with gender and own-record status. In the full Milwaukee renter sample (N = 69,627), the coefficient on $\ln(1 + \text{fuzzy1}) \times \text{female} \times \text{no_own}$ is −$10.06 ($p = 0.023$), significant at the 5% level. This is the strongest single heterogeneity result: women with no own criminal record bear a substantially larger fuzzy screening penalty than any other group. The constituent interactions are individually insignificant—fuzzy1 × female ($p = 0.623$) and fuzzy1 × no_own ($p = 0.486$)—confirming that the effect is specific to the intersection of female gender and clean record status.

In split-sample estimates, female renters show significant fuzzy effects on poverty (+0.384 pp, $p = 0.001$) and education share (−0.439 pp, $p = 0.003$), while male renters show null effects on both. The rent coefficients are negative for both genders (−\$3.27 for women, −\$2.49 for men) but not individually significant at the 5% level, reflecting loss of power from sample splitting.

Why might women bear a disproportionate fuzzy screening penalty? One possibility is that women's names have more Lev-1 neighbors in the court system due to gendered naming conventions (e.g., common name endings like "-A" or "-E" that generate many edit-distance-1 variants). Another possibility is that landlords are more likely to search name variants for female applicants, perhaps because women are more frequently the named applicant on rental applications. We cannot definitively distinguish these mechanisms but note that the gender heterogeneity is remarkably consistent across approaches: women exhibited a screening penalty 6.6 times larger than men in our earlier exact-match analyses.

9.2 Severity Composition

The violent share interaction approach tests whether violent composition modifies the screening effect. Adding violent_share (the fraction of Lev-1 fuzzy cases that are violent) and its interaction with $\ln(1 + \text{fuzzy1})$ to Specification 1, neither the violent share main effect ($p = 0.92$) nor the interaction ($p = 0.37$) is significant for any outcome. This confirms that screening is indiscriminate: landlords respond to the total volume of court records returned by a name search, not to the severity composition. This is consistent with the screening technology documented by the Urban Institute, which returns case counts and case types without weighting by severity.

9.3 Income Quartile Heterogeneity

The fuzzy screening effect is concentrated in the bottom two income quartiles. Q2 shows the largest and most significant effect (−\$6.12/month, $p = 0.017$), while Q1 is marginally significant (−\$4.12/month, $p = 0.086$). The upper quartiles (Q3 and Q4) show essentially zero effects ($p > 0.88$). This is consistent with the theoretical framework: higher-income renters have larger choice sets and can substitute away from neighborhoods where they are screened out, attenuating the equilibrium effect. Full results are in Appendix B.5.

9.4 Race Heterogeneity

No racial subgroup shows a significant fuzzy rent effect at the 5% level. White renters ($-\$2.91$, $p = 0.159$, $N = 40{,}614$), Black renters ($-\$2.05$, $p = 0.322$, $N = 16{,}337$), and Hispanic renters ($+\$2.80$, $p = 0.471$, $N = 6{,}009$) all produce insignificant estimates. This null result does not imply that Black renters are unaffected by screening—the baseline effect captures the average effect across races—but rather that the *additional* fuzzy penalty does not vary significantly across racial groups. This is consistent with the theoretical framework: the screening contamination channel operates through name similarity, not through race, and Levenshtein distance does not vary systematically across racial groups. Full results are in Appendix B.

9.5 Sexual Assault in the Name Cluster

We test whether the presence of sexual assault cases in an individual's exact-name cluster—measured by `ccap_sexual`, the count of sexual offense cases filed under the same FIRST_LAST name key—generates an additional screening penalty beyond the Levenshtein fuzzy channel. Only 1.03% of Milwaukee renters ($N = 1{,}142$) have any sexual assault cases in their name cluster, making this a rare but potentially severe signal.

Adding $\mathbb{1}[\texttt{ccap_sexual} > 0]$ to the baseline Specification 1, we find a large and significant discrete penalty: renters whose name cluster contains any sexual assault case live in neighborhoods with \$22.30/month lower median rent ($p = 0.016$), controlling for all Levenshtein fuzzy exposure and demographic characteristics. In the continuous specification using $\ln(1+\texttt{ccap_sexual})$, the coefficient is $-\$32.10$/month per log-unit ($p = 0.003$), the strongest single screening signal we identify.

The Levenshtein fuzzy coefficient remains significant and essentially unchanged after controlling for sexual assault ($-\$2.90$, $p = 0.040$), confirming that the two channels are distinct: fuzzy exposure captures contamination from *similar* names, while the sexual assault indicator captures contamination from the *same* name.

This finding is consistent with landlords treating sexual offense records as a discrete "red flag"—a binary rejection trigger that operates independently of the volume of court records. The magnitude ($\sim$\$22–32/month) is substantially larger than the per-case fuzzy effect ($\sim$\$3/month), reflecting the severity of the signal: a sexual assault record in a CCAP search is far more likely to trigger rejection than a traffic violation or misdemeanor.

10 Discussion

10.1 Connecting Results to Theory

The Milwaukee results confirm all four predictions of the theoretical framework developed in Section 3.

1. **Negative sorting (Prediction 1):** Fuzzy exposure reduces rent ($-\$2.99$/mo, $p = 0.037$), increases unemployment ($+0.067$ pp, $p = 0.015$), increases poverty ($+0.186$ pp, $p = 0.015$), reduces median household income ($-\$302$, $p = 0.006$), and increases ambient $PM_{2.5}$ for no-own-record renters ($+0.005$ $\mu g/m^3$, $p = 0.050$).

2. **Stronger for no-own-record (Prediction 2):** The fuzzy penalty for rent is $-\$5.96$/month for the purely contaminated ($p = 0.015$) but statistically zero for those with own records ($p = 0.72$). In the language of the model, individuals with $S^{\text{own}} = \emptyset$ experience the full force of the external screening constraint.

3. **Stronger for rare names (Prediction 3):** The triple interaction confirms that signal dilution is strongest for no-own-record renters with common names ($\hat{\beta}_{10} = +\$13.86$, $p = 0.001$), as landlords discount fuzzy matches for obviously common names.

4. **Zero for owners (Prediction 4):** All five Specification 1 placebo tests pass ($p = 0.23$ to 0.95), and all four core Specification 3 tests pass ($p = 0.16$ to 0.98).

10.2 Origin Controls and the Composition Channel

The robustness of neighborhood composition effects (unemployment, poverty, income) to origin controls, combined with the attenuation of the rent effect, reveals an important distinction. Rent levels are mechanically persistent across moves—individuals who lived in a \$1,100/month tract almost certainly move to a similar-rent tract—leaving little variance for screening to explain. But neighborhood composition (who lives nearby, local employment, local poverty) has more variation conditional on origin, and screening contamination explains a significant share of this variation. This suggests that screening does not merely push renters into cheaper housing; it sorts them into communities with different socioeconomic characteristics, and this sorting is not fully explained by where they came from.

10.3 Comparison of Magnitudes

The Levenshtein estimates are economically meaningful but smaller than exact-match estimates, which is expected: exact-match contamination (same name) should produce stronger

screening effects than fuzzy contamination (similar name), because same-name records are more likely to be attributed to the applicant. In our preferred continuous specification, the average Milwaukee renter fuzzy effect on rent is −$2.99/month per log-unit, while the purely contaminated suffer −$5.96/month per log-unit at mean name frequency.

The binned results provide the most transparent magnitudes: no-own-record Milwaukee renters with 7–10 fuzzy cases live in neighborhoods with $25/month lower rent (2.2% of the mean) and $1,736 lower tract median household income (2.9% of the mean). These are large effects for individuals who have committed no crime.

10.4 Robustness

The baseline result is robust to a range of alternative specifications, reported in full in Appendix A. Dropping the income control (R7a: −$3.12, $p = 0.044$) or adding income squared (R7b: −$2.87, $p = 0.043$) yields estimates nearly identical to the baseline (−$2.99, $p = 0.037$), confirming that the result is not driven by functional form assumptions about income. Alternative bins (R5) with narrower categories (1–3, 4–6, 7–10, 11–20, 21+) confirm that the no-own-record penalty is significant for 4–6 cases (−$27.3, $p = 0.008$), 7–10 cases (−$26.3, $p = 0.027$), and 21+ cases (−$25.6, $p = 0.038$). The owner placebo passes in the simple binned specification (R6): only 1 of 5 bin × no_own interactions is significant at the 5% level.

Adding tract racial composition (white_pct, black_pct) absorbs the fuzzy effect, reducing it from −$2.99 ($p = 0.037$) to −$1.20 ($p = 0.365$). We argue in Appendix A.2 that this constitutes a bad-control problem rather than evidence against the screening hypothesis. Tract racial composition is plausibly a *mediator*: if screening pushes individuals with race-correlated names into different neighborhoods, those neighborhoods' racial compositions change as a result. Controlling for racial composition conditions on a downstream consequence of the treatment, biasing the estimated effect toward zero. The attenuation is therefore *consistent with* the screening hypothesis—it reveals that screening operates partly through racial sorting channels.

10.5 Screening Affects Destination Quality, Not Mobility

A potential concern with the full-sample results is that only movers face tenant screening at the point of a move; stayers are not being actively screened. We address this concern in two steps: first, we test whether fuzzy exposure affects the *probability* of moving; second, we show that the screening effect survives in the mover subsample.

Mobility test. If screening trapped renters in their current neighborhoods—preventing them from moving at all—we would expect fuzzy exposure to reduce mobility. We estimate a linear probability model of moving (2023–2024 address change) on fuzzy exposure, using the full Milwaukee renter sample ($N = 110{,}484$). The coefficient on $\ln(1 + \text{fuzzy1})$ is $+0.0002$ ($p = 0.90$): fuzzy exposure has no detectable effect on the probability of moving. The interaction with no-own-record status is similarly null ($p = 0.37$), and descriptive move rates are flat at approximately 9% across all fuzzy exposure levels and own-record subgroups. As a robustness check, we estimate random forest and gradient boosting classifiers: adding fuzzy exposure variables to a mobility prediction model produces zero improvement in out-of-sample AUC (random forest: 0.574 without fuzzy, 0.572 with fuzzy; gradient boosting: 0.584 vs. 0.580). Even allowing for arbitrary nonlinearities and interactions, fuzzy exposure contains no information about whether a renter moves.

This null result has an important implication: screening does not operate on the *extensive margin* of mobility (whether people move) but on the *quality margin* (where they move to). The full-sample cross-section therefore reflects the steady-state allocation of renters across neighborhoods under persistent screening constraints. Stayers are in their current neighborhoods because past screening sorted them there, and movers are sorted by current screening into comparable (or worse) destinations.

Mover subsample. To confirm that screening operates on the flow of recent moves, we restrict to renters who moved between 2023 and 2024 ($N = 9{,}942$, dropping the age control since 33% of movers have missing age; the full-sample results are robust to excluding age).

The rent interaction is significant for movers: $\ln(1 + \text{fuzzy1}) \times \mathbb{1}^{\text{no own}} = -\$13.34$/month ($p = 0.043$), with the joint F-test for the three no-own-record interactions rejecting at 5% ($F = 2.8$, $p = 0.037$). The education share joint F-test is highly significant ($F = 8.3$, $p < 0.001$), and the environmental triple interactions are significant for both RSEI ($p = 0.029$) and non-cancer toxic concentration ($p = 0.028$).

Tract-level composition measures (poverty, unemployment) are not individually significant for movers in the interaction specification, reflecting limited power with $N = 9{,}942$. However, restricting to *exposed* no-own-record movers—those with fuzzy exposure > 0 and no criminal record ($N = 977$)—reveals a significant poverty effect: $\hat{\beta}_2 = +0.87$ percentage points per log-unit ($p = 0.027$). The identical specification on exposed movers who *have* their own records ($N = 934$) yields a null coefficient ($+0.18$, $p = 0.73$). This contrast directly identifies the screening mechanism among recent movers: conditional on being exposed to fuzzy contamination, the intensive margin of exposure pushes clean-record movers into higher-poverty neighborhoods, while movers whose own records already constrain their

choice set show no additional effect.

Among low-income no-own-record movers (bottom income half, $N = 3{,}073$), the screening penalty is detectable on both rent ($-\$14.58$/month, $p = 0.071$) and median household income ($-\$1{,}199$, $p = 0.057$), consistent with the theoretical prediction that budget-constrained renters have fewer substitution possibilities when screening restricts their choice set.

10.6 Additional Robustness

The baseline result is robust to a range of alternative specifications. Dropping the income control ($-\$3.12$, $p = 0.044$) or adding income squared ($-\$2.87$, $p = 0.043$) yields estimates nearly identical to the baseline ($-\$2.99$, $p = 0.037$), confirming that the result is not driven by functional form assumptions about income. Alternative bins with narrower categories (1–3, 4–6, 7–10, 11–20, 21+) confirm that the no-own-record penalty is significant for 4–6 cases ($-\$27.3$, $p = 0.008$), 7–10 cases ($-\$26.3$, $p = 0.027$), and 21+ cases ($-\$25.6$, $p = 0.038$). The owner placebo passes in the simple binned specification: only 1 of 5 bin $\times$ no_own interactions is significant at the 5% level. Full robustness tables are reported in Appendix A.

Adding tract racial composition (white_pct, black_pct) absorbs the fuzzy effect, reducing it from $-\$2.99$ ($p = 0.037$) to $-\$1.20$ ($p = 0.365$). We argue that this constitutes a bad-control problem rather than evidence against the screening hypothesis: tract racial composition is plausibly a *mediator* (screening pushes individuals with race-correlated names into different neighborhoods, changing those neighborhoods' racial compositions), so controlling for it conditions on a downstream consequence of the treatment. The attenuation is therefore *consistent with* the screening hypothesis—it reveals that screening operates partly through racial sorting channels (see Appendix A.2 for details).

10.7 Limitations

Several limitations should be noted. First, **statistical power** for some tract-level composition outcomes in the triple interaction is limited. The poverty and unemployment interactions in Specification 3 are not statistically significant in the full sample, though the split-sample analysis (Table 4) confirms that the effects are present in the correct subgroup. For movers, power is further limited by sample size ($N = 9{,}942$), though the rent interaction, education share F-test, and environmental triple interactions all survive.

Second, **Lev-2 exposure shows sign reversal**. The Lev-2 rent coefficient is positive and significant ($+\$2.90$, $p = 0.04$), confirming clean distance decay but also suggesting that at distance 2, the variation captures compositional correlations rather than screening.

Third, **listwise deletion** due to missing age reduces the Milwaukee renter sample from

its full size to 69,627. If age missingness correlates with name frequency or court record prevalence, the estimates may be affected by selection bias.

Fourth, we **cannot distinguish the mechanism** through which Lev-1 neighbors' records reach landlords—algorithmic fuzzy matching, manual variant searching, or data-entry errors. All three would produce the observed distance-decay pattern, but they have different policy implications.

10.8 Policy Implications

Our findings have two direct policy implications. First, regulations targeting criminal record screening in housing should account for fuzzy matching. Current policy discussions focus on "ban-the-box" measures that restrict *when* landlords can inquire about criminal history, but do not address the accuracy of name-based searches. Our results suggest that requiring exact-match verification (e.g., date-of-birth confirmation) could reduce erroneous screening contamination from both same-name and similar-name records. The finding that no-own-record Milwaukee renters with 7–10 fuzzy cases pay $25/month less rent underscores the real cost of imprecise name matching for innocent individuals.

Second, the gender heterogeneity implies that women may be disproportionately affected by screening contamination. If this reflects gendered naming patterns (women's names having more edit-distance neighbors), policy interventions should consider the demographic distribution of name-based screening errors. If it reflects landlord behavior (more extensive name searching for female applicants), fair housing enforcement may need to address sex-based disparities in screening intensity.

11 Conclusion

This paper introduces Levenshtein fuzzy exposure as a novel measure of criminal record screening contamination in housing markets. By exploiting the quasi-random variation in whether similar-sounding names exist in the Wisconsin court records system, we identify a screening channel that extends beyond exact name matches.

Our theoretical framework distinguishes four channels of neighborhood sorting—budget constraint (captured by income), own-screening constraint (captured by own criminal record), unobserved taste-based sorting (ruled out theoretically and by the owner placebo), and external screening constraint (our residual treatment effect)—and shows that, conditional on income and own record, fuzzy exposure can only operate through the external screening channel. All four predictions of the model are confirmed in the Milwaukee data.

Our most powerful finding comes from the split-sample analysis: among renters with no own criminal record ($N = 43{,}593$), Levenshtein fuzzy exposure is significantly associated with worse neighborhoods across *all seven* outcome dimensions—lower rent ($p = 0.007$), higher unemployment ($p = 0.010$), higher poverty ($p = 0.016$), lower median household income ($p = 0.004$), fewer college-educated neighbors ($p = 0.0003$), higher toxic chemical exposure ($p = 0.006$), and higher fine particulate matter ($p = 0.050$). Among renters who have their own criminal records ($N = 26{,}034$), every coefficient is statistically zero. This stark contrast—7 out of 7 significant versus 0 out of 7—directly confirms the theoretical prediction that the screening effect operates exclusively through the external constraint channel.

In the Specification 3 interaction model, the fuzzy penalty for rent falls entirely on the purely contaminated: $\hat{\beta}_8 = -\$5.96$/month ($p = 0.015$), while those with own records show zero additional effect. Education share exhibits the strongest Spec 3 results ($\hat{\beta}_8 = -0.006$, $p = 0.0004$; joint $F = 23.8$, $p < 0.001$). The binned specification reveals that no-own-record renters with 7–10 court cases from similar-sounding strangers live in neighborhoods with \$25/month lower rent—a 2.2% decline—while renters who already have their own records show zero penalty across all bins.

The owner placebo test confirms that the mechanism operates through landlord tenant screening: Lev-1 fuzzy exposure has no significant effect on any of five outcomes for Milwaukee homeowners, who access housing through mortgage markets rather than screening. The null interaction between fuzzy exposure and name commonality ($p = 0.71$) validates the quasi-random nature of the Levenshtein variation, distinguishing it from exact-match analyses that are confounded by name frequency.

These findings extend the growing body of evidence on the collateral consequences of criminal records in housing markets (Agan and Starr, 2018; Christensen and Timmins, 2022). In the same city where Desmond (2016) documented how eviction records trap families in poverty, we show that the reach of court records extends beyond the individuals who generated them: through the accident of name similarity, screening contamination affects innocent bystanders. The environmental dimension—higher RSEI toxic exposure and $PM_{2.5}$ for the purely contaminated—connects screening errors to the environmental justice literature (Christensen, Sarmiento-Barbieri, and Timmins, 2022), showing that algorithmic name confusion does not merely sort people by socioeconomic status but also by exposure to environmental hazards. The concentration of the effect among those with no own criminal record—individuals who are, by definition, innocent of the charges attributed to them through name confusion—highlights the urgent need for policy interventions that require identity verification beyond name matching.

References

Agan, A. and Starr, S. (2018). Ban the Box, Criminal Records, and Racial Discrimination: A Field Experiment. *Quarterly Journal of Economics*, 133(1):191–235.

Angrist, J.D. and Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Bertrand, M. and Mullainathan, S. (2004). Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. *American Economic Review*, 94(4):991–1013.

Consumer Financial Protection Bureau. (2021). Fair Credit Reporting; Name-Only Matching Procedures. *Federal Register*, 86(215):62468–62473. November 10, 2021.

Christensen, P. and Timmins, C. (2022). Sorting or Steering: The Effects of Housing Discrimination on Neighborhood Choice. *Journal of Political Economy*, 130(8):2110–2163.

Christensen, P., Sarmiento-Barbieri, I., and Timmins, C. (2022). Housing Discrimination and the Toxics Exposure Gap in the United States: Evidence from the Rental Market. *Review of Economics and Statistics*, 104(4):807–818.

Desmond, M. (2016). *Evicted: Poverty and Profit in the American City*. New York: Crown Publishers.

Ewens, M., Tomlin, B., and Wang, L.C. (2014). Statistical Discrimination or Prejudice? A Large Sample Field Experiment. *Review of Economics and Statistics*, 96(1):119–134.

U.S. Department of Housing and Urban Development. (2016). Office of General Counsel Guidance on Application of Fair Housing Act Standards to the Use of Criminal Records by Providers of Housing and Real Estate-Related Transactions.

Lageson, S.E. and Stewart, R. (2024). The Problem with Criminal Records: Discrepancies Between State Reports and Private-Sector Background Checks. *Criminology*, 62(1):5–34.

National Consumer Law Center. (2023). *Digital Denials: How Abuse, Bias, and Lack of Transparency in Tenant Screening Harm Renters*. Boston, MA: NCLC.

Pager, D. (2003). The Mark of a Criminal Record. *American Journal of Sociology*, 108(5):937–975.

Axelrod, J., Chen, B., Fallon, K., and Torres Rodriguez, S. (2025). Opening the "Black Box" of Tenant Screening. Washington, DC: Urban Institute. March 2025.

van Donkelaar, A., Hammer, M.S., Bindle, L., Brauer, M., Brook, J.R., Garay, M.J., Hsu, N.C., Kalashnikova, O.V., Kahn, R.A., Lee, C., Levy, R.C., Lyapustin, A., Sayer, A.M., and Martin, R.V. (2021). Monthly Global Estimates of Fine Particulate Matter and Their Uncertainty. *Environmental Science & Technology*, 55(22):15287–15300.

A Robustness Checks

A.1 Summary of Robustness Specifications

Table 9 reports the ln_fuzzy1 rent coefficient across all robustness specifications. The baseline result (−\$2.99, $p = 0.037$) is robust to dropping income controls and adding income squared, but is attenuated when tract racial composition is added or when the sample is restricted to movers.

Table 9: Robustness: ln_fuzzy1 Rent Coefficient Across Specifications (Milwaukee Renters)

Specification	$\hat{\beta}$ (\$/mo)	SE	p	N
R1 Baseline (with age)	−2.99	1.43	0.037**	69,627
R1 No age control	−1.93	1.18	0.102	110,484
R2 + Tract racial composition	−1.20	1.33	0.365	69,627
R3 Movers only	−1.38	4.81	0.774	6,683
R7a No income control	−3.12	1.55	0.044**	69,627
R7b Income + income^2	−2.87	1.42	0.043**	69,627

Notes: All specifications estimated on Milwaukee renters with SEs clustered at the name-key level. The baseline specification (R1 with age) includes age, female, race dummies, and individual income. R1 no age drops the age control, recovering 40,857 observations with missing age. R2 adds tract-level white and Black population shares. R3 restricts to renters who moved between 2023 and 2024. R7a drops individual income. R7b adds income squared. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

A.2 R2: Tract Racial Composition as a Potential Bad Control

Adding tract-level racial composition (white_pct, black_pct) to the baseline specification attenuates the ln_fuzzy1 coefficient from −\$2.99 ($p = 0.037$) to −\$1.20 ($p = 0.365$) and increases the R^2 from 0.242 to 0.344. The racial composition variables are highly significant: white_pct has a coefficient of +\$572.77 ($p < 0.001$) and black_pct of +\$224.95 ($p < 0.001$).

Is this absorption evidence against the screening hypothesis, or is tract racial composition a *bad control*? We argue the latter. Tract racial composition is plausibly a *mediator* rather than a *confounder* of the screening effect. The causal chain runs as follows: screening pushes individuals with race-correlated names into different neighborhoods → those neighborhoods' racial compositions change → controlling for racial composition absorbs the treatment effect by conditioning on a downstream consequence of screening.

Consider the mechanism. Names are correlated with race: certain name structures are more common among Black, Hispanic, or white populations. If screening steers individuals with particular name structures into particular neighborhoods, the racial composition of those neighborhoods is an *outcome* of the screening process, not a pre-existing confounder. Controlling for it is equivalent to asking: "does screening affect neighborhood quality, holding constant one of the channels through which screening sorts people?" By construction, this absorbs part of the treatment effect.

This is a classic bad-control problem (Angrist and Pischke, 2009): controlling for a variable that is itself affected by the treatment biases the estimated treatment effect toward zero. The attenuation in R2 is therefore *consistent with* the screening hypothesis—it shows that screening operates partly through racial sorting—rather than evidence against it.

A.3 R5: Alternative Bins for No-Own-Record Renters

Table 10 reports an alternative binned specification using narrower bins (1–3, 4–6, 7–10, 11–20, 21+) for Milwaukee renters, interacted with no-own-record status.

Table 10: R5: Alternative Bins × Own-Record Status—Rent ($/mo) (Milwaukee Renters)

	Has own record (base)			No own record (interaction)		
Bin	Effect	SE	p	Interaction	SE	p
1–3 cases	−0.3	5.2	0.958	+2.2	7.2	0.762
4–6 cases	+3.5	6.7	0.604	−27.3	10.3	0.008***
7–10 cases	+0.8	8.1	0.923	−26.3	11.9	0.027**
11–20 cases	+5.9	7.5	0.434	−9.6	12.2	0.435
21+ cases	−2.3	7.8	0.765	−25.6	12.4	0.038**
No-own level shift:				+51.1	$p < 0.001$***	

Notes: Reference category is 0 fuzzy cases (N = 55,353). Total N = 69,627, $R^2 = 0.244$. "Total" for no-own-record renters is the sum of the base bin effect and the interaction. Total effects: 4–6 cases = −$23.8/mo; 7–10 cases = −$25.5/mo; 21+ cases = −$27.9/mo. SEs clustered at name-key level. ***$p < 0.01$, **$p < 0.05$, *$p < 0.10$.

The alternative bins confirm the main finding. For renters with their own criminal record,

no bin is significant. For no-own-record renters, three of the five bins are significant at the 5% level: 4–6 cases (−\$27.3, $p = 0.008$), 7–10 cases (−\$26.3, $p = 0.027$), and 21+ cases (−\$25.6, $p = 0.038$). The total effects for no-own-record renters range from −\$23.8 to −\$27.9/month for these bins, representing a 2.1–2.5% decline relative to mean rent.

A.4 R6: Owner Placebo with Simple Bins

Table 11 reports the binned owner placebo for Milwaukee.

Table 11: R6: Owner Binned Placebo—Rent (\$/mo) (Milwaukee Owners)

	Has own record (base)			No own record (interaction)		
Bin	Effect	SE	p	Interaction	SE	p
1–2 cases	+4.7	3.5	0.181	−8.0	7.0	0.255
3–5 cases	+3.2	3.9	0.414	−3.4	8.0	0.671
6–10 cases	+0.0	4.1	0.999	−8.8	8.4	0.293
11–20 cases	+0.9	4.1	0.830	−18.2	9.9	0.065*
21+ cases	+3.0	4.0	0.460	−19.2	8.2	0.020**
No-own level shift:				+21.0		$p < 0.001$***

Notes: Milwaukee owners, N = 595,766, $R^2 = 0.087$. Placebo assessment: 1 of 5 bin × no_own interactions significant at the 5% level. The significant 21+ bin may reflect residual compositional correlations at the extreme tail of the fuzzy exposure distribution, but the overall pattern is consistent with a passing placebo: owner fuzzy effects are mostly null, confirming that the screening channel does not operate for homeowners. SEs clustered at name-key level. ***$p < 0.01$, **$p < 0.05$, *$p < 0.10$.

The owner binned placebo passes overall. Only 1 of 5 bin × no_own interactions is significant at the 5% level (21+ cases: −\$19.2, $p = 0.020$), and the 11–20 bin is marginally significant ($p = 0.065$). None of the base bin effects (Panel A) is significant for owners. This is consistent with the screening interpretation: the rental market channel does not operate for homeowners.

A.5 R3: Mover Subsample

Restricting to renter-movers (N = 6,683) produces a negative but insignificant coefficient (−\$1.38, $p = 0.774$). The standard error is large (SE = 4.81) due to the small sample size. In Specification 3, the fuzzy1 × no_own interaction is also insignificant (−\$13.10, $p = 0.102$), though the sign is correct and the magnitude is roughly double the full-sample estimate (−\$6.09). The triple interaction (fuzzy1 × name_freq × no_own) is marginally significant

(+\$20.55, $p = 0.058$), consistent with the full sample. The small N limits statistical power, but the direction and magnitudes are reassuring.

B Heterogeneity

B.1 H1: Gender Split

Table 12 reports split-sample estimates by gender for four outcomes.

Table 12: H1: Gender Split—ln_fuzzy1 Coefficient (Milwaukee Renters)

	Female ($N = 35{,}971$)		Male ($N = 33{,}656$)	
Outcome	$\hat{\beta}$	p	$\hat{\beta}$	p
Rent (\$/mo)	−3.27	0.108	−2.49	0.220
Poverty (pp)	+0.384***	0.001	−0.016	0.882
Pollution (ln RSEI)	−0.004	0.404	−0.006	0.200
Education (% BA+)	−0.439***	0.003	−0.192	0.182

Notes: Specification 1 estimated separately by gender. SEs clustered at name-key level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

The gender heterogeneity is striking for poverty and education outcomes: female renters show a significant increase in tract poverty (+0.384 pp, $p = 0.001$) and decrease in tract education share (−0.439 pp, $p = 0.003$) from Lev-1 fuzzy exposure, while male renters show null effects. The rent coefficients are negative for both genders but not individually significant at the 5% level, reflecting the loss of power from sample splitting.

B.2 H2: Race Split

Table 13 reports split-sample estimates by race for rent and poverty.

Table 13: H2: Race Split—ln_fuzzy1 Coefficient (Milwaukee Renters)

	White ($N = 40{,}614$)		Black ($N = 16{,}337$)		Hispanic ($N = 6{,}009$)	
Outcome	$\hat{\beta}$	p	$\hat{\beta}$	p	$\hat{\beta}$	p
Rent (\$/mo)	−2.91	0.159	−2.05	0.322	+2.80	0.471
Poverty (pp)	+0.161	0.111	+0.227*	0.097	−0.347	0.167

Notes: Specification 1 estimated separately by race. SEs clustered at name-key level. No split is individually significant at the 5% level for rent, confirming no significant racial heterogeneity in the fuzzy effect. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

No racial subgroup shows a significant fuzzy rent effect at the 5% level. White and Black renters both show negative point estimates, while Hispanic renters show a positive (but insignificant) estimate. The lack of significant racial heterogeneity is consistent with the theoretical framework: the Levenshtein channel operates through name similarity, not through the racial signal embedded in names.

B.3 H3: Gender × No-Own-Record Interaction

The strongest heterogeneity result is the triple interaction of fuzzy exposure with female and no-own-record indicators. In the full Milwaukee renter sample ($N = 69{,}627$), we estimate:

$$\ln(1 + \text{fuzzy1}) \times \text{female} \times \text{no_own} = -\$10.06^{**} \quad (p = 0.023) \tag{9}$$

This implies that women with no own criminal record bear a substantially larger fuzzy screening penalty than men with no record or any renter with their own record. The constituent interactions are individually insignificant: fuzzy1 × female ($p = 0.623$) and fuzzy1 × no_own ($p = 0.486$), confirming that the effect is specific to the intersection of female gender and clean record status.

This finding is consistent with two non-exclusive mechanisms. First, women may be more frequently the named applicant on rental applications, making them more exposed to name-based screening errors. Second, women's names may generate more relevant Lev-1 neighbors in the court system due to gendered naming conventions.

B.4 H4: Severity Decomposition

Table 14 decomposes Lev-1 fuzzy exposure into violent and non-violent components.

Table 14: H4: Severity Decomposition—Rent ($/mo) (Milwaukee Renters, $N = 69{,}627$)

Variable	$\hat{\beta}$ ($/mo)	p
$\ln(1 + \text{own})$	-33.48^{***}	< 0.001
$\ln(1 + \text{fuzzy1_violent})$	$+9.97^{**}$	0.048
$\ln(1 + \text{fuzzy1_nonviolent})$	-5.00^{***}	0.004
$\ln(\text{fuzzy1_viol}) \times \tilde{f}$	-6.99	0.133
$\ln(\text{fuzzy1_nonviol}) \times \tilde{f}$	$+1.99$	0.299
R^2	0.242	

Notes: Specification decomposes ln_fuzzy1 into violent and nonviolent components. SEs clustered at name-key level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Non-violent fuzzy exposure drives the rent effect: −$5.00/month ($p = 0.004$), significant at the 1% level. Violent fuzzy exposure has the opposite sign: +$9.97/month ($p = 0.048$). The positive violent coefficient may reflect reverse sorting—neighborhoods with higher violence exposure are avoided by screened-out individuals, creating a compositional artifact—or may reflect the extreme rarity of violent Lev-1 matches (violent cases constitute a small fraction of total court records, and finding such a case under a *similar* name is unusual).

For poverty, the pattern reverses as expected: nonviolent fuzzy exposure increases poverty (+0.278 pp, $p = 0.003$), while violent exposure is insignificant (−0.438 pp, $p = 0.113$). This confirms that non-violent records—evictions, traffic offenses, misdemeanors—are the primary driver of screening contamination through the Levenshtein channel.

B.5 H5: Income Quartile Split

Table 15 reports the ln_fuzzy1 rent coefficient by income quartile.

Table 15: H5: Income Quartile Split—ln_fuzzy1 Rent Coefficient (Milwaukee Renters)

Quartile	$\hat{\beta}$ ($/mo)	SE	p	N
Q1 (lowest income)	−4.12	2.40	0.086*	21,105
Q2	−6.12	2.57	0.017**	17,550
Q3	−0.21	3.00	0.943	15,800
Q4 (highest income)	+0.55	3.83	0.887	15,172

Notes: Specification 1 estimated separately by income quartile. SEs clustered at name-key level. ***$p < 0.01$, **$p < 0.05$, *$p < 0.10$.

The fuzzy screening effect is concentrated in the bottom two income quartiles. Q2 shows the largest and most significant effect (−$6.12, $p = 0.017$), while Q1 is marginally significant (−$4.12, $p = 0.086$). The upper quartiles (Q3 and Q4) show essentially zero effects. This is consistent with the theoretical framework: higher-income renters have larger choice sets (more affordable neighborhoods) and can therefore substitute away from neighborhoods where they are screened out, attenuating the equilibrium effect of screening on neighborhood quality.

C Comparison with Exact-Match Results

Table 16 compares the exact-match others' exposure results with the Levenshtein fuzzy exposure results for poverty, highlighting a sign reversal.

Table 16: Exact-Match vs. Levenshtein: Poverty Coefficient (Milwaukee Renters)

Measure	$\hat{\beta}$ (pp)	p	Interpretation
Exact-match others (ln)	negative	< 0.01	Confounded by name frequency
Lev-1 fuzzy (ln)	$+0.186^{**}$	0.015	Correct positive sign

Notes: Exact-match others' exposure shows a *negative* poverty coefficient (more others' records → lower poverty), which contradicts the screening hypothesis. This arises because exact-match exposure is strongly confounded by name commonality ($\rho = 0.81$): common names tend to belong to white individuals who live in lower-poverty neighborhoods. The Levenshtein measure, with its weaker name-frequency correlation ($\rho = 0.28$), produces the theoretically expected *positive* sign: more fuzzy exposure pushes renters into higher-poverty neighborhoods. This sign reversal is a key motivation for the Levenshtein approach.

The sign reversal between exact-match and Levenshtein exposure for poverty is one of the central empirical motivations for this paper. Exact-match others' exposure produces a negative poverty coefficient—the "wrong" sign under the screening hypothesis—because name commonality is a strong confounder: individuals with common names (disproportionately white) both have more exact-match others' exposure and live in lower-poverty neighborhoods. The Levenshtein measure, which is weakly correlated with name frequency, corrects this confound and produces the theoretically expected positive coefficient. This pattern confirms that Levenshtein variation provides cleaner identification of the screening channel.