Estimating the National Prevalence of Eviction Using Millions of Public Court Records

Eviction from rental housing is a cause, not just a consequence, of poverty (Desmond, 2012, 2016). Experiencing an eviction is associated with a number of negative consequences: poorer mental health (Desmond & Kimbro, 2015; Fowler, Gladden, Vagi, Barnes, & Frazier, 2014), job loss (Desmond & Gershenson, 2016), and relocation to more disadvantaged neighborhoods (Desmond & Shollenberger, 2015). In 2015, 74% of renting families living below the poverty line received no housing assistance,[1] leaving the majority of these families to rely on the private rental market for housing. Previous research on renting households in Milwaukee, Wisconsin, demonstrated that forced moves—most resulting from an eviction—were common, affecting more than 1 in 8 renters in the previous two years (Desmond & Shollenberger, 2015).

Yet, we do not know how many households face eviction each year in the United States, or how the prevalence of eviction varies across space. There are no official federal statistics that track housing evictions. National-level surveys measuring eviction often under count households that experienced an eviction due to how and to whom the question is asked (Desmond & Kimbro, 2015). Eviction also increases residential instability (Desmond, Gershenson, & Kiviat, 2015), which may also increase the difficulty of capturing these households in survey data.

Large-scale collection of public courts records presents a new opportunity to estimate the prevalence of eviction. This type of large-scale administrative data can provide insight into longitudinal trends in neighborhood characteristics and inequality (O'Brien, Sampson, & Winship, 2015; Sampson, 2017). A critical approach is needed when using data not created for research purposes, however, to ensure that the scope of the data is accurately defined, data are analyzed in context, and that spurious correlations are not interpreted as significant patterns (boyd & Crawford, 2012).

We use a novel dataset of over 71 million public court records to estimate the prevalence of housing eviction nationally. We supplement our analyses with an additional 10 million court records and 25,000 yearly, county-level reported eviction filings collected directly from state and county courts. By combining these data, we are able to provide the first set of comprehensive national estimates of the number of eviction cases filed, households threatened by eviction, and judgments for eviction annually for 2000-2016. We make this data publically available to encourage new research into the causes and consequences of eviction across the U.S.

Data and Methods

Primary Data Our primary court records data was purchased from two bulk record collection companies. These companies obtain court records either by visiting courts and manually recording information from publicly available documents or through bulk collection of electronic records. This data includes 72,040,362 records covering all 50 U.S. states and the District of Columbia from 2000-2016 and contains information on when the case was filed, the names and addresses of plaintiffs (landlords) and defendants (tenants) involved in the case, and actions that occurred on a case, including how the case was resolved. Eviction cases were considered to end in an eviction judgement if there was an order for restitution of the property or a monetary settlement for the landlord.

Supplementary Data We collected an additional 10,895,619 court records by making bulk data requests directly to state civil courts.[2] We received data from 15 states covering the 2000-2016 period (for all years available). These records included comparable case information to that received in the primary data described above. Additionally, we requested statistics on the number of eviction cases filed annually from

[1] U.S. Census Bureau, 2015 American Housing Survey
[2] Virginia and Philadelphia County, Pennsylvania were collected via web-scraping of online court records.

all 50 U.S. states and DC.[3] We received county-level data for 35 states for all years available during the same period (N=25,687 county-year filing counts). Figure 1 demonstrates the distribution of data coverage across the U.S.

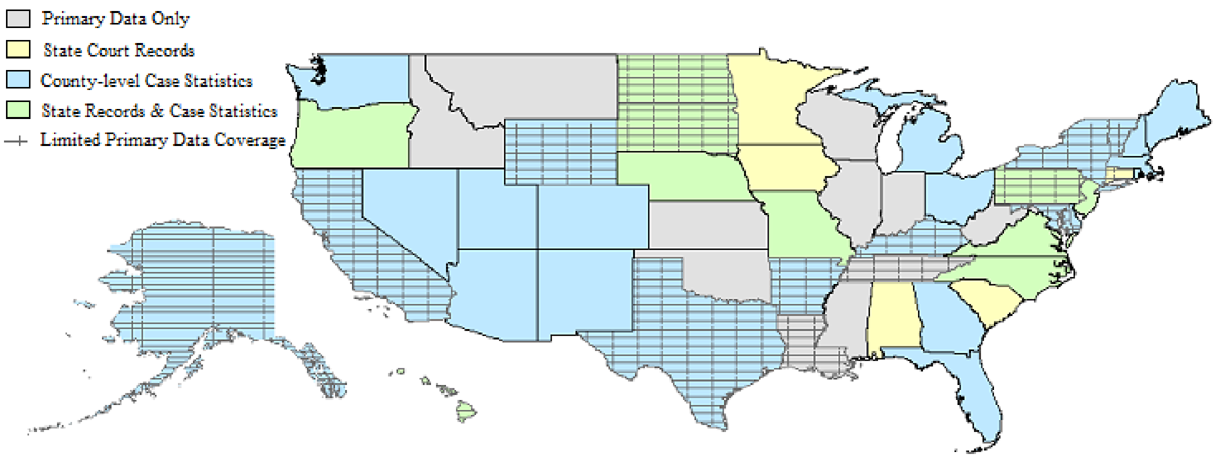

Figure 1. Primary and Supplementary Data Coverage for Eviction Court Records

Data Preparation We cleaned and standardized both the primary and supplementary court records to create aggregate counts of eviction cases filed, households threatened by eviction, and eviction judgements. To locate where an eviction case occurred, we separated tenant address components into separate fields, [4] and then geocoded the addresses. We used the geographical coordinates from the geocode to assign cases to Census block groups.[5] We assigned Census tract, county, and state identifiers by aggregating up from Census block groups. We also created a standardized representation of tenant names[6] and used the *fastLink*[7] program in R to find instances in which variations of the same name at the same address appeared across multiple cases. We linked these cases over time to create a unique household identifier. We also combined records that corresponded to the same case, as court case is our primary unit of analysis. We used case filing dates as the official case date when available. When filing dates were not available, we used the earliest date an action was recorded for a case. We excluded duplicated records and cases in which the tenant was a business, as the focus of this study is housing eviction.[8] Additional details of the data and data preparation are discussed in Desmond et al (2018).

Validation We validated our primary data using supplementary records, where available (see Figure 1). We compared both individual-level records and aggregated county-level counts to determine how data coverage varied across space and time. Triangulating data from multiple sources allows us to identify fluctuations in data coverage and adjust calculations of eviction prevalence accordingly. We collected additional information on characteristics of court systems, record collection from courts, and local demographics to better understand what factors affect variations in coverage.

[3] We did not make these requests if the number of eviction cases filed was already available online.
[4] Addresses were formatted into five separate fields: street address, apartment designation, city name, state abbreviation, and five-digit zip code. Defendant address was assumed to be the property address of the eviction case.
[5] We used block group boundaries from Census Shapefiles. Each case was assigned to the block group that contained the address coordinates.
[6] Names were formatted as Last Name, First Name, Middle Initial/Name (if present), Name Suffix (if present).
[7] https://cran.r-project.org/web/packages/fastLink/index.html
[8] The primary data included an indicator for commercial cases. For the supplementary data, we developed a list of key words commonly associated with commercial or business entities and then used regular expressions to identify tenant names that included these key words.

Imputation Collection of eviction-related court records are more difficult in some states due to dispersion of eviction cases among many small, local courts or the sealing or purging of records after resolution of a case. Furthermore, many states do not maintain comprehensive court record systems or case statistics. When no reliable data is available or we suspected available data to be undercounted, eviction prevalence is imputed using information obtained during the validation process. Cases can be missing for systematic reasons, e.g. a court does not release records of cases that were dismissed, or more stochastic reasons, e.g. the case was not collected or entered into the electronic management system.

First, we impute systematically missing cases by predicting the expected proportion of dismissed cases given outcome patterns in the data. We treat the proportion of cases to be imputed as a random normal variable with the mean set to the expected proportion of systematically missing cases in a county and the standard deviation derived from the same data. Second, we impute the randomly missing cases using a longitudinal linear random effects model with logged eviction case count[9] as the outcome and information on court systems, record collection from courts, and local demographics as covariates. It is important to distinguish between systematically and randomly missing cases as the systematically missing cases contribute to the prevalence of eviction cases filed in an area, but not the number of evictions (because the cases were dismissed), while the randomly missing cases potentially contribute to both the number of eviction cases filed and number of evictions.

After we estimate the number of eviction cases filed in an area, we predict the number of households threatened with eviction and the number of cases with eviction judgments as Poisson random variables. We borrow ratios of households threatened to eviction cases filed and eviction judgments to eviction cases filed from comparable areas with good data coverage to determine the probability with which an eviction case filed should be counted as a unique household threatened with eviction or an eviction judgment, respectively.

Renting Households We estimate the number of renting households from linear interpolation of block group level data from the 2000 and 2010 Censuses and 2016 ESRI Business Analyst. We then aggregate these data upwards to create counts of renting households at the Census tract, county, state, and national levels.

Eviction Prevalence We calculated the yearly prevalence of eviction cases filed by dividing the number of eviction cases observed in each year by the count of renting households in each state. We calculated the yearly rate of households threatened by eviction by dividing the number of unique households named in an eviction case each year by the number of renting households in each state. Finally, we calculated the eviction rate by dividing the number of eviction judgments per year by the number of renting households in each state. We did not count eviction judgements in which the same household appeared in a subsequent eviction case.[10]

Preliminary Results

We estimate that more than 3 million eviction cases were filed nationwide in 2016. Figure 2 shows that there is significant variation in the prevalence of eviction cases across U.S. states. Surprisingly, many of the states with the highest eviction filing rates are found in the Southeast, an area rarely mentioned in discussions of access to affordable housing.

[9] The case count was logged due to skew.

[10] The appearance of the same household at the same address in a subsequent eviction case implies that any previous eviction judgements to restore the property to the landlord were not executed.

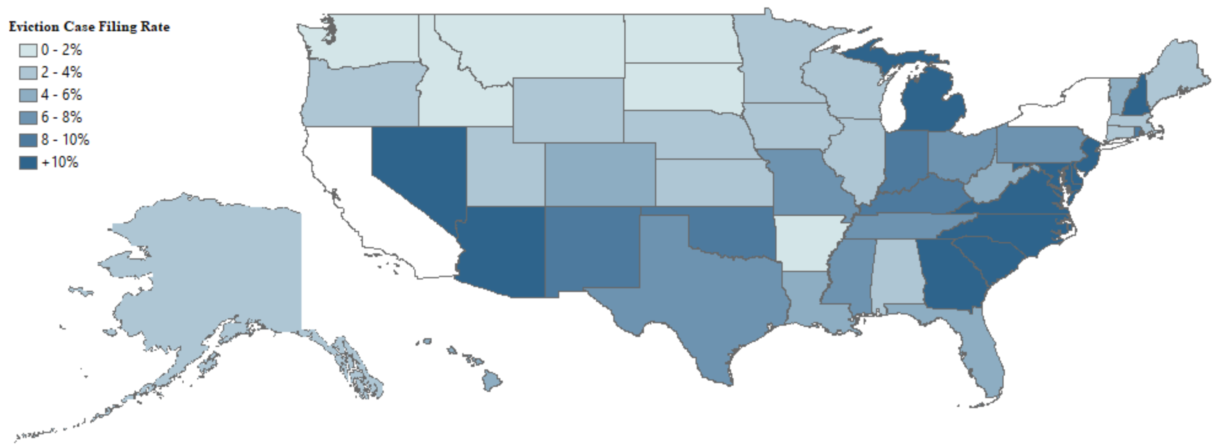

Figure 2. Estimated Eviction Court Case Filing Rates by State, 2016
Note: Results are preliminary. California and New York are still to be estimated.

Although not shown on the map, rates of households threatened by eviction and eviction judgments also show significant variation across space. Interestingly, these rates do not always move in parallel with eviction filing rates. To examine this, we calculate the absolute difference between the eviction case filing rate and the households threatened rate in each state. In many states this difference is within one to two percentage points; however, in five states the difference is greater—District of Columbia (5.7%), Delaware (4.9%), South Carolina (3.8%), Maryland (3.4%), and Virginia (2.4%). All of these states also appear among those with the highest filing rates in Figure 2. This finding appears to indicate that in some states landlords use eviction courts as a means of enforcing collection of past-due rent by repeatedly filing cases against the same tenant(s) but not formally evicting them.

This pattern is also reflected in eviction judgment rates. On average, across all states, 60% of the eviction cases filed receive an eviction judgment. Yet, there is a statistically significant, negative correlation ($r = -0.57$, $p < .01$) between the difference in rates of eviction cases filed and unique households threatened by eviction and the percentage of eviction cases filed that receive an eviction judgment. This is both an interesting finding and a cautionary tale for using administrative data: without accounting for recurring cases involving the same households over time, the eviction rate would be over-estimated in these areas.

Additional analysis will use bootstrapping to estimate the prevalence of eviction in states still missing data, as well as provide a robustness check and an estimation of confidence intervals around the eviction rates for the other states. We will also use the same strategy described here to estimate and present comparable eviction prevalence rates for the 2000-2015 period.

Implications

This study uses big data to understand the spatial distribution of a previously under-specified population—households that have been threatened with or experienced eviction across the U.S. The collection of this type of individual-level data provides an important opportunity to examine large-scale demographic and economic characteristics of a population that is hard to capture using other research methods. By releasing data aggregated at the Census block group level publically,[11] we hope that this data source will serve as an important resource for other researchers examining the causes and consequences of eviction in communities across the U.S.

[11] Data available at www.evictionlab.org.

References

boyd, danah, & Crawford, K. (2012). Critical Questions for Big Data. *Information, Communication & Society*, *15*(5), 662–679. https://doi.org/10.1080/1369118X.2012.678878

Desmond, M. (2012). Eviction and the reproduction of urban poverty. *American Journal of Sociology*, *118*(1), 88–133.

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

Desmond, M., & Gershenson, C. (2016). Housing and Employment Insecurity among the Working Poor. *Social Problems*, *63*(1), 46–67. https://doi.org/10.1093/socpro/spv025

Desmond, M., Gershenson, C., & Kiviat, B. (2015). Forced Relocation and Residential Instability among Urban Renters. *Social Service Review*, *89*(2), 227–262. https://doi.org/10.1086/681091

Desmond, M., Gromis, A., Edmonds, L., Hendrickson, J., Krywokulski, K., Leung, L., & Porton, A. (2018). *Eviction Lab Methodology Report: Version 1.1.0* (No. Version 1.1.0). Princeton, NJ: Princeton University. Retrieved from www.evictionlab.org/methods

Desmond, M., & Kimbro, R. T. (2015). Eviction's Fallout: Housing, Hardship, and Health. *Social Forces*, *94*(1), 295–324. https://doi.org/10.1093/sf/sov044

Desmond, M., & Shollenberger, T. (2015). Forced Displacement From Rental Housing: Prevalence and Neighborhood Consequences. *Demography*, *52*(5), 1751–1772. https://doi.org/10.1007/s13524-015-0419-9

Fowler, K. A., Gladden, R. M., Vagi, K. J., Barnes, J., & Frazier, L. (2014). Increase in Suicides Associated With Home Eviction and Foreclosure During the US Housing Crisis: Findings From 16 National Violent Death Reporting System States, 2005–2010. *American Journal of Public Health*, *105*(2), 311–316. https://doi.org/10.2105/AJPH.2014.301945

O'Brien, D. T., Sampson, R. J., & Winship, C. (2015). Ecometrics in the Age of Big Data: Measuring and Assessing "Broken Windows" Using Large-scale Administrative Records. *Sociological Methodology*, *45*(1), 101–147. https://doi.org/10.1177/0081175015576601

Sampson, R. J. (2017). Urban sustainability in an age of enduring inequalities: Advancing theory and ecometrics for the 21st-century city. *Proceedings of the National Academy of Sciences*, 201614433. https://doi.org/10.1073/pnas.1614433114