Research Brief

Predicting Homelessness For Better Prevention

Background

Homelessness is among the most pressing public policy challenges facing New York City. More than 125,000 individuals passed through homeless shelters in 2016, among whom more than 70 percent were in families. Much public attention has been given to the scale of the homelessness crisis in New York City and the significant challenge of addressing it. While there are some interventions that have proven effective at reducing the likelihood of shelter entry, it is difficult to reach at-risk households to deliver prevention assistance before they become homeless. Devising effective means of directing homelessness prevention services to those at greatest risk is therefore a key policy issue. To help improve existing means, the Center for Innovation through Data Intelligence (CIDI) in the Mayor's Office partnered with New York University's Furman Center to use data on human services, buildings, and neighborhoods to predict families' risk of homelessness. This brief summarizes the key insights from this work.

Method

The study uses administrative data on receipt of public benefits, including cash assistance and Medicaid, linked to information on homeless shelter applications and stays, building characteristics, and neighborhood characteristics from the years 2006 to 2015. We use machine learning methods to predict shelter application and entry in the year 2015 as a function of these characteristics in previous years. We evaluate the quality of our predictions on a withheld test sample using common machine learning performance metrics.

To better understand whether algorithm-driven predictions can enhance homeless prevention programs, we explore whether our algorithms can identify families at higher risk of homelessness than families currently seeking out and receiving prevention assistance on their own. We first estimate what share of families currently receiving prevention assistance from Homebase, the city's primary homelessness prevention program, would have become homeless had they not received assistance. We then compare this to the share that becomes homeless in an equivalently sized group of high-risk households identified by our algorithms.

Findings

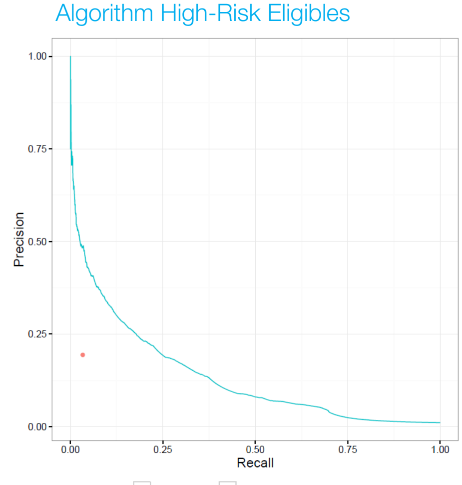
Machine learning v. Prevention Program Take-up:

- Our machine learning predicted risk scores can identify a prevention-eligible population that is roughly 1.5 times more likely to be apply for shelter within 24 months than those currently receiving prevention services through Homebase.

  Recall is the share of the actual homeless population that we predict to be homeless.

  Precision is the probability that an individual actually becomes homeless when we predict he or she will become homeless.

Current Homebase Recipients vs. Algorithm High-Risk Eligibles



We estimate that roughly 20 percent of Homebase recipients would've become homeless had they not received assistance, compared to 35 percent of high-risk Homebase-eligible families identified by our prediction models.

[1]Our data covers 2006-2014 for Medicaid and Cash assistance from HRA, 2003-2015 for data from DHS, and 2006-2015 for data from the New York City Housing Court.

Findings (Continued)

Building and Neighborhood Variables:

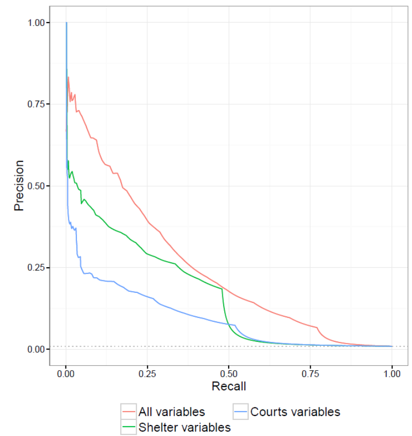
- Outreach to roughly 6,000 at-risk families identified using only individual-level variables in our predictions would correctly identify 3,300 families that would subsequently apply for shelter in the next 12 months, adding building and neighborhood variables to the prediction model allows us to correctly identify more than 4,000 families that would apply to shelter.

Building and Neighborhood Targeting:

- Our best building-level prediction model is 30% more accurate at identifying building that house families at risk of entering shelter than comparison models built just from the information currently used to direct building-level outreach. These improved predictions could be used to enhance the cost effectiveness of outreach.

- Adding building and neighborhood shelter entry rates to building-level models based on housing court variables nearly doubles the predictive accuracy of our building-level predictions.

Comparing Different Methods of Building Outreach



As in most studies involving prediction, we find that lagged measures of the variable being predicted are the best predictors. In our case, this means that measures related to prior homelessness, such as days in shelter or number of separate times staying in shelter, are the strongest predictors. We summarize the top risk factors for our individual-level analysis and building-level analysis in the following table.

Top Ten Predictors, Unordered

Individual-Level Prediction	
Variables	Risk
Ever Stayed in Homeless Shelter	+
Ever Received TANF	+
Eviction in t	+
Building had Previous Shelter Entrant	+
Apply for PA in t	+
Active on PA in t	+
Denied PA in t	+
Sanctioned from PA in t	+
Shelter Code 6: Hotel/Motel	+
Shelter Code 13: DV Program Housing	+

Building-Level Prediction	
Variables	Risk
NYCHA Building	+
Number of Units in Building	+
HPD Housing Code Violation	+
HPD Ordered Repair	+
HPD Litigation against Owner This Year	-
HPD Litigation against Owner Ever	+
Shelter Application This Year	+
Shelter Application Ever	+
Shelter Application Rate, Tract	+
Shelter Application Rate, Block	+

Policy implications

This research suggests several important points for homelessness assistance in New York City.

- The use of predictive targeting in the provision of homelessness prevention services can help ensure that programs find those most likely to benefit from assistance
- Outreach to neighborhoods or buildings from which families applied to shelter in the past can reach more at-risk families than outreach to neighborhoods with many non-payment filings or evictions
- Building and neighborhood characteristics can improve assessment of individual risk of homelessness.