Prioritizing Homeless Assistance Using Predictive Algorithms: An Evidence-Based Approach

Halil Toros
Daniel Flaming
Economic Roundtable

Abstract

In this article, we present a predictive model for identifying homeless persons likely to have high future costs for public services. We developed the model by linking administrative records from 2007 through 2012 for 7 Santa Clara County, California agencies and identifying 38 demographic, clinical, and service utilization variables with the greatest predictive value. We modeled records for 57,259 individuals from 2007 to 2009, and the algorithm was validated using 2010 and 2011 records to predict high-cost status in 2012. A business case scenario shows that two-thirds of the top 1,000 high-cost users predicted by the model are true positives, with estimated posthousing cost reductions of more than $19,000 per person in 2011. The model performed very well in giving low scores to homeless persons with one-time cost spikes, achieving the desired result of excluding cases with single-year rather than ongoing high costs.

Overview

Homelessness is a major social problem in the United States, with large public health impacts affecting millions of individuals and families and costing billions of dollars. The most recent annual numbers available from the U.S. Department of Housing and Urban Development (HUD) are for 2016 and show 1,421,196 people used an emergency shelter or a transitional housing program at some point during the year (HUD, 2016). The most recent point-in-time numbers are for 2017 and show, on a night in January, 553,742 people were homeless. Among individuals, 24 percent were chronically homeless (HUD, 2017).

The health, personal, and economic challenges that chronically homeless individuals experience and the lack of effective, coordinated services to address these problems often lead to a vicious cycle of diminished well-being with serious implications for their service utilization patterns

(Economic Roundtable, 2015a). The impairments of some of these individuals might impede access to needed health services and other support systems, such as employment services. Consequently, they cycle through costly emergency-driven public systems without getting the ongoing care they need to address severe mental illness, substance use disorders, or chronic health conditions (Caton, Wilkins, and Anderson, 2007; Folsom et al., 2005).

The number of individuals experiencing homelessness substantially exceeds the number of affordable housing units available for them. This shortfall includes permanently affordable housing with supportive services that is needed for many chronically homeless individuals. Managing the gap between housing supply and demand is a challenge for city and county housing agencies. The predictive algorithm we describe provides a fair, objective tool for triage—prioritizing which individuals may receive Housing First.

Growth in homelessness over the last three decades has been exacerbated by economic downturns, loss of affordable housing and foreclosures, stagnating wages, an inadequate safety net, and the closing of state psychiatric institutions. In response to this growing need, the federal response to homelessness shifted in 2009 from uncoordinated short-term responses to avert homelessness—primarily using shelters—to long-term housing solutions. Permanent housing subsidies have since been shown to significantly increase housing stability, food security, and child well-being (Gubits et al., 2016).

The first component of the federal strategy shift was providing permanent supportive housing (PSH), that is, housing that is permanently affordable combined with ongoing supportive services for people experiencing chronic homelessness, and prioritizing those individuals with the most severe disabilities for assistance. The second component was connecting PSH to street outreach, shelter, and institutional "in-reach" to identify and engage people experiencing chronic homelessness. The third component was communitywide adoption of Housing First to provide permanent housing as quickly as possible in order to end chronic homelessness and prevent its recurrence (USICH, 2015).

The Housing First model was introduced by Pathways to Housing, a New York City nonprofit, to provide homeless intervention services to adults with psychiatric diagnoses and substance abuse problems. The nonprofit provided immediate housing and services to homeless adults with co-occurring diagnosis as a matter of right, with no preconditions. It also incorporated a harm reduction approach to psychiatric and substance abuse treatment and empowered the consumers of services to make choices about housing and services (Greenwood, Stefancic, and Tsemberis, 2013).

The Housing First approach makes housing stabilization the centerpiece of homeless assistance and recognizes that some people need more than housing assistance to stabilize. A small but highly visible segment of the chronically homeless population has substantial service needs. PSH with a Housing First approach enables chronically homeless individuals with disabilities that interfere with maintaining housing on their own to become stable renters.

PSH with a Housing First approach is an effective intervention for enabling chronically homeless individuals to permanently exit homelessness. However, because housing resources are limited, one of the key challenges is identifying and targeting the "highest priority" individuals so as to allocate scarce housing in a way that produces the greatest benefit. It is well documented that costly interventions, such as PSH, are not likely to generate cost offsets equal or higher than the cost of the interventions, except for the most costly users (Culhane, 2008; Poulin et al., 2010).

A 2015 study in Santa Clara County, California, confirmed that chronic homelessness is very costly. The 10 percent with the highest costs, the 10th decile, accounted for 61 percent of all public costs for homelessness, and the top 5 percent accounted for 47 percent of all costs (Economic Roundtable, 2015b). Studies in Los Angeles County found that PSH provided to chronically homeless individuals in the 10th decile generated large enough cost offsets to cover the costs of housing and services (Economic Roundtable, 2015b, 2009). However, the scarce supply of PSH is often rented out to the eligible population based on crude screening processes that rely on self-reported data. Given that PSH is proven to have a large impact on reducing chronic homelessness and associated public costs, a strong argument can be made for using more accurate screening tools to identify individuals who should have first priority for access to permanently affordable housing.

This article extends previous research applying predictive models to homelessness and high-cost service users. The model presented in this article predicts who will or will not become a high-cost public service user in the next year, given various person-level characteristics in the current year and previous year, providing a predictive score (probability) for each individual in order to determine housing priorities across large numbers of individuals.

Prioritizing high-cost homeless persons for whom the solution of housing costs less than the problem of homelessness improves the efficiency of PSH. Cost offsets from reduced service use after high-cost people are stably housed can be stretched across a larger pool of homeless people whose housing can be subsidized with those offsets.

This is a triage tool for connecting homeless persons who are high-cost users of public services with permanently affordable housing, community-based healthcare, and support services. The tool applies a statistical predictive model to administrative data in order to prioritize homeless adults with the highest needs and public costs. It provides highly accurate predictions comparable to those developed through studies of high-cost health system users. Because no other models predict high-cost service users within the homeless population, health sector models provide the closest comparison. These models identify patients at high risk of readmission to a hospital based on demographics, prior hospital admissions, and clinical conditions (Ash et al., 2001; Billings et al., 2013, 2006; Chechulin et al., 2014; Fleishman and Cohen, 2010; Moturu, Johnson, and Liu, 2010; Tamang et al., 2016).

This tool improves on earlier predictive models for identifying homeless individuals in the 10th cost decile (Economic Roundtable, 2012, 2011). Several other studies have also used predictive models to assess homeless risks. Byrne et al. (2016) estimated predictors of homelessness and developed methods for more efficiently targeting homeless prevention services. A recent study of the Home Base prevention program for families in New York City[1] showed that adoption of

[1] Home Base was a homeless prevention program operated by the New York City Department of Homeless Services from 2004 to 2008. Workers interviewed applicants about potential risk factors for homelessness, including human capital, housing conditions, disability, interpersonal discord, childhood experiences, and shelter history. The study compared the accuracy of judgments made by workers in determining eligibility for services to the results produced by a screening model in predicting whether families would enter a homeless shelter in the following 3 years.

an empirical model for deciding which families to serve can make homeless prevention more efficient (Shinn et al., 2013). Also, the U.S. Department of Veterans Affairs (VA) has explored using predictive models in screening homeless veterans (Montgomery et al., 2013). However, to date, no studies have examined the relationship of past service utilization to future high-cost homelessness using predictive algorithms to prioritize which homeless people get housing.

In this article, we describe the predictive modeling methodology used to develop a triage tool to prioritize housing access for an efficient and cost effective PSH program. After presenting the results and validation of the model, we develop a business scenario to estimate the cost savings after implementation of the triage tool. The article ends with a discussion of ways to use the tool, limitations of it, and recommendations.

Chronic Homelessness

The majority of people who become homeless remain so for less than a year. A smaller number of people, however, remain homeless much longer, experiencing continuous and chronic homelessness. According to federal guidelines, an individual is chronically homeless if he or she has a diagnosed disability—such as serious mental illness, substance use disorder, posttraumatic stress disorder, cognitive impairments or chronic physical illness or disability—and has been homeless and lives in a place not meant for human habitation, a safe haven, or in an emergency shelter for at least 1 continuous year or has experienced at least four episodes of homelessness in the past 3 years where the cumulative total of the four occasions is at least 1 year.[2]

Needs of chronically homeless individuals that are essential for their well-being go unmet, including connections to housing, income, family, and health. This leads to stress, anxiety, depression, deprivation, and chaos, thus destabilizing their lives. Over time, chronically homeless individuals have increasingly complex and costly needs, including serious health and mental health conditions and disabilities that result in cycling in and out of hospitals, jails, prisons, psychiatric hospitals, and homeless shelters.

Several studies describe the clinical and social characteristics and patterns of service utilization among people who are chronically homeless. The majority of individuals have a serious mental illness such as schizophrenia, bipolar disorder, or major depression. They also experience high rates of substance abuse disorders, physical disability, or chronic disease. Many experience co-occurring mental illness and substance use problems (Burt, 2002; Caton et al., 2005; Caton, Wilkins, and Anderson, 2007; Folsom et al., 2005; Rosenheck, 2000). In addition to serious disability, the lives of chronically homeless people are compromised by persistent unemployment and lack of earned income forcing dependence on public assistance for sustenance, healthcare, and, if fortunate, an eventual exit from homelessness (Caton et al., 2005; Caton, Wilkins, and Anderson, 2007). Moreover, chronically homeless individuals often have a long arrest history, cycling through jail and prison (Caton et al., 2005; Kushel et al., 2005; Metraux and Culhane, 2004; Zugazaga, 2004).

Chronically homeless individuals spend a disproportionate number of days in the shelter system (Kuhn and Culhane, 1998; Metraux et al., 2001). In addition, because of their complex and

[2] "Homeless Emergency Assistance and Rapid Transition to Housing: Defining 'Chronically Homeless.'" *Federal Register* 80 (75791–75806). 2015. https://www.hudexchange.info/resource/4847/hearth-defining-chronically-homeless-final-rule/.

co-occurring disabling conditions, poor health status, and elevated rates of unintentional injuries and traumatic injuries from assault, chronically homeless persons have high rates of hospital emergency rooms use and hospitalization, and longer hospital stays for mental health and substance abuse problems (Culhane, Metraux, and Hadley, 2002; Folsom et al., 2005; Kuno et al., 2000; Kushel et al., 2002). As the chronically homeless population ages, its utilization of emergency rooms and hospital rooms increases (Caton, Wilkins, and Anderson, 2007). High incarceration rates, coupled with heavy use of mental health and medical facilities in jails and prisons are also well documented (Kushel et al., 2005; McNiel, Binder, and Robinson, 2005; Metraux and Culhane, 2004).

Heavy use of acute and behavioral healthcare, criminal justice involvement, and use of social services may cost tens of thousands of dollars per individual annually (Ly and Latimer, 2015; Culhane, Metraux, and Hadley, 2002; Gilmer et al., 2009; Larimer et al., 2009; Martinez and Burt, 2006; McLaughlin, 2011). Although chronically homeless people represent only 20 percent of shelter users, they consume the largest share of health, social, and justice services with enormous costs (Ly and Latimer, 2015). In Los Angeles County, among homeless recipients of General Relief cash aid,[3] the highest cost decile accounted for 56 percent of all public costs for homeless single adults (Economic Roundtable, 2011, 2009). A recent study using Santa Clara County data also showed that public costs for homelessness are heavily skewed toward a comparatively small number of frequent users of public and medical services. Among residents experiencing homelessness in 2012, the 10th decile, accounted for almost two-thirds of costs, and the top 5 percent accounted for almost half of costs (Economic Roundtable, 2015b).

Federal funding for homeless programs increased from $3.7 billion in 2010 to nearly $5.5 billion in 2016 (USICH, 2016). In addition, federal expenditures for homeless individuals are also distributed through Medicaid, Medicare, and the VA, as well as large expenditures by state and county governments and institutions such as hospitals, jails, and social service agencies.

Although public outlays to address chronic homelessness have been growing since 2010, the prevalence and costs of homelessness remain high. With finite resources for homeless assistance, prevention services and cost-effective interventions, such as PSH, have attracted growing interest from policymakers and academic research over the past decade (Apicello, 2010; Burt and Pearson, 2005; Byrne et al., 2014; Culhane, Metraux and Byrne, 2011).

Preventive Services and Permanent Supportive Housing

The logic of prevention requires definition of what is to be prevented (such as chronic homelessness) and specification of the association (preferably causal) between the intervention and prevention of the undesirable condition. Several frameworks have been suggested for developing prevention strategies for homelessness (Burt and Pearson, 2005). The high-risk framework is the most appropriate framework for conceptualizing how to design homeless prevention policies because it draws attention to the need for direct intervention among those individuals at greatest risk. This framework focuses on alleviating the causes of homelessness for the most vulnerable subpopulations (Apicello, 2010).

[3] General Relief is a cash aid program that provides a maximum of $221 a month for destitute adults. Roughly two-thirds of the caseload is estimated to be homeless. This program is called General Assistance in other California counties.

To be successful, prevention strategies for high-risk individuals need to be both effective and efficient (Burt and Pearson, 2005; Culhane, Metraux, and Byrne, 2011; Shinn, Baumohl, and Hopper, 2001). In this context, effectiveness refers to how capable a program is of facilitating the desired goal—prevention of homelessness with reasonable costs. Effectiveness should be evaluated with robust designs by comparing a treatment group of persons who received services to a control group of individuals not subject to the intervention. Otherwise, the effect of the services in preventing homelessness cannot be assessed accurately, because it is unrealistic to assume that all the people who received services would have become or stayed homeless in the absence of those services. It is also possible that the effect of services might have not been significant; homelessness might have been merely postponed; or the ranks of high-risk individuals might simply have been reshuffled, allowing some to "jump the queue" and push others back in the line (Shinn, Baumohl, and Hopper, 2001).

As noted previously, recent research has shown that PSH, using a Housing First approach, is a very effective homeless prevention service and has led to widespread and successful efforts to reduce chronic homelessness (Byrne et al., 2014; Culhane, Metraux, and Hadley, 2002; Greenwood, Stefancic, and Tsemberis, 2013; Larimer et al., 2009; Rog et al., 2014; Tsemberis and Eisenberg, 2000; USICH, 2015, 2010). Based on increasing evidence, the U.S. federal government has endorsed PSH using a Housing First approach as the "clear solution" to chronic homelessness, and PSH has become an important priority for HUD. The number of beds in PSH projects increased nearly 60 percent between 2007 and 2014, when an estimated 285,400 people lived in PSH (HUD, 2014; USICH, 2010).

Research has also demonstrated the effectiveness of PSH in generating cost offsets. Many studies have shown that PSH and Housing First interventions for chronically homeless individuals lead to cost savings through reduced shelter costs, decreases in both psychiatric and medical inpatient hospitalization costs, lower emergency room visit costs, reduced substance abuse treatment costs, and reduced criminal justice costs due to fewer arrests, detentions, and court appearances (Culhane and Byrne, 2010; Henwood et al., 2015; Ly and Latimer, 2015; Martinez and Burt, 2006; Shinn, Baumohl, and Hopper, 2001; Shinn et al., 2013; Toros and Stevens, 2012). Cost savings from providing PSH to homeless people with mental disorders was shown to be substantial (Culhane, Metraux, and Hadley, 2002; Gilmer et al., 2009; Larimer et al., 2009; McLaughlin, 2011; Sadowski et al., 2009).

Despite such successes, the high cost of PSH would limit its availability to chronically homeless individuals with the greatest service needs if cost offsets are the benchmark for determining eligibility. Culhane (2008) reviewed several studies and concluded PSH is not likely to generate cost offsets equal to the cost of the interventions, except for the most costly users. Other studies also support the view that only frequent users of higher-cost services are likely to have sufficiently high costs to fully or mostly offset the costs of a PSH placement. Some research indicates that group may be limited to the most costly 10 percent of the chronically homeless (Poulin et al., 2010; Rosenheck, 2000). Moreover, since homeless people are typically placed in PSH programs at times when they are in crisis and have had relatively high service use, regression to the mean results in decreasing costs for many of these people, even if they are not placed in PSH (Ly and Latimer, 2015).

Hence, the research demonstrates that, although PSH is effective in reducing chronic homelessness and yields significant cost offsets, to be efficient, it should target high-cost homeless persons so that offsets will cover program and housing costs. In the context of homeless prevention, efficiency refers to targeting high-risk individuals. Efficient targeting is critical in the design and success of prevention services (Apicello, 2010; Burt and Pearson, 2005; Culhane, Metraux, and Byrne, 2011; Shinn, Baumohl, and Hopper, 2001). An efficient program should use empirically and/or theoretically derived risk factors to identify high-risk individuals who are likely to stay homeless and use costly public services unless they receive the prevention services.

However, the efficiency criterion introduces a serious challenge. Predictive models and screening tools are subject to the well-known tradeoff between sensitivity (the probability of correctly identifying true positives, or those individuals who will remain or become low-cost persons in Santa Clara County in the absence of the prevention program) and specificity (the probability of correctly identifying true negatives, or those individuals who would stay as low-cost homeless persons). If a low cutoff is selected, while the sensitivity increases and the model capturing more true positives, the specificity decreases leading to higher numbers of false positives. On the other hand, fewer false positives occur if the targeting cutoff is increased but many true positives are missed. This difficult tradeoff is at the core of the efficiency issue, as savings realized through placing a high-cost homeless person in PSH will be washed out if many low-cost homeless persons are also placed (Culhane, Metraux, and Byrne, 2011).

In the literature, it is argued that the common failing of many prevention efforts is their targeting inefficiency, which leads to ineffective programs (Burt and Pearson, 2005). It is also argued in the literature that available screening models are not sensitive or accurate enough to yield high hit rates without missing a large number of high-risk persons who would benefit from the program while producing cost savings (Apicello, 2010; Shinn, Baumohl, and Hopper, 2001). However, recent technological advances in the fields of predictive analytics and data mining together, with the availability of digital integrated administrative datasets with rich service utilization fields, allow significant improvement in prediction ability over earlier approaches and models (Larson, 2013).

This article presents the Silicon Valley Triage Tool. The County of Santa Clara supported the development of this tool so that it could identify homeless individuals in jails, hospitals, and clinics who have continuing crises in their lives that create very high public costs, and also give them first priority for access to PSH. This effort took roughly 2 years and included linking records of homeless clients across county agencies, analyzing attributes and costs for these individuals, and developing the triage tool. The model is very robust and accurate, taking advantage of advanced prediction methodologies and a unique and exceptionally valuable database created by Santa Clara County, home to Silicon Valley, linking service and cost records across county departments for the entire population of residents who experienced homelessness over a 6-year period—a total of 104,206 individuals. The tool accurately identifies individuals experiencing homelessness whose acute needs create the greatest public costs and is expected to serve as a screening tool for efficient and effective PSH programs.

Methods

Four steps were involved in developing the Silicon Valley Triage Tool: first, linking agency records to create an integrated dataset; second, analyzing the data and developing the triage tool; third, testing and validating the tool; and fourth, developing a business model to project cost savings from using the tool. Each step is described in the following sections.

Data

By collaborating in linking their client records, seven agencies in Santa Clara County[4] provided information on medical care (inpatient and outpatient), Emergency Medical Services (EMS), drug and alcohol treatment services, mental health treatment services (inpatient and outpatient), incarceration (arrest, court, and medical and mental health services in custody), and HUD-funded social and homeless services (Economic Roundtable, 2015b).

The Silicon Valley Triage Tool was developed using records for a subgroup of the total population that experienced homelessness. This subgroup included 57,259 individuals who used a homeless service and also had a linked record in another agency during the 6-year study window from 2007 through 2012. This subgroup of records was used to develop the triage tool so as to avoid using records that may have had incomplete data because of uncompleted linkages across some agencies.

We benchmarked the tool against the total population that was homeless during the 6-year time window rather than just against individuals who were documented as being homeless at a specific time. We considered this time period because the problems that result in chronic homelessness are usually structural conditions in people's lives—mental illness, trauma, debilitating health conditions, addiction, absence of qualifications or opportunities for employment, extreme poverty, and absence of sustaining personal connections. These problems do not go away just because someone is not documented as being homelessness in a given month; rather, they are drivers for the person's life trajectory.

To accurately identify high-cost homeless individuals, the triage tool must use multi-year information about individuals, assessing service encounters over a larger rather than narrower interval in their lives. It is likely that individuals with the highest 5 percent of costs move in and out of institutional care settings without being consistently documented as homeless. In addition, homeless individuals who are admitted to private hospitals, state psychiatric facilities, or incarcerated by the state correctional system would not be documented as homeless in county data systems.

Because of these data gaps, the homeless and persistently homeless status of individuals in the top 5 percent often is not evident, so we made the assumption that individuals documented as having been homeless who have ongoing public costs in the top 5 percent are likely to be persistently homeless.

Linked datasets provided information about factors that affect the outcome of interest: being a high-cost user next year. These included demographic variables such as age, gender, and ethnicity; clinical variables such as ICD-9 (International Classification of Diseases, Ninth Revision) medical

[4] The seven agencies participating in the record linkage were: the HUD Continuum of Care Board, Criminal Justice Information Control system of the Sheriff Department, Department of Alcohol and Drug Services, Emergency Management System, Mental Health Department, Social Services Agency, and Valley Medical Center.

diagnoses, and utilization variables for all service types from the current and previous year, including number of clinic or emergency room visits, number of hospitalizations and number of arrests, as well as the cost of services. The variables used in the model are listed in exhibit 1.

Exhibit 1

Averages of Model Variables for High-Cost and Other Homeless Persons (Validation Sample) (1 of 2)

Variable	High Cost (N = 5,726)	Other (N = 51,533)
Demographics (%)		
Age less than 18	5	10
Age 18–45	56	55
Age 46–65	36	31
Age 65+	3	4
Female	42	54
Criminal justice		
100+ days of probation in the last 2 years (%)	18	5
Arrested in last 2 years (%)	46	16
Jail booking in last 2 years (%)	23	9
Jail security classification of 3 or 4 (that is, high risk) this year (%)	10	1
Arrested for inebriation and released within 48 hours—this year (%)	8	1
Mean number of arrests this year	0.78	0.16
Mean number of days in jail this year	32.9	5.2
Health diagnoses		
Diagnosed with chronic medical condition; Chronic Condition Indicator for ICD-9-CM diagnosis codes by HCUP (%)	68	35
Medical encounter with diagnosis of adjustment reaction ICD-9 309 in last 2 years (%)	11	3
Medical encounter with diagnosis of heart disease ICD-9 401-429 in last 2 years (%)	6	2
Mean number of medical encounters with diagnosis of organ failure ICD-9 569-573, 576-578, 585-594, or 596 in last 2 years	0.6	0.1
Medical encounter with diagnosis of schizophrenia ICD-9 295 in last 2 years (%)	14	2
Mean number of medical encounters with diagnosis of neoplasm (ICD-9 140 to 239) in last 2 years	0.4	0.1
Medical encounter with diagnosis of “other ill-defined and unknown causes of morbidity and mortality” (ICD-9 799) in last 2 years (%)	17	4
Medical encounter with diagnosis of high-cost ICD-9 in last 2 years (%)	52	20
Health and emergency services		
EMS encounters this year (%)	30	7
EMS encounters last year (%)	29	7
Two or more EMS encounters in last 2 years (%)	12	1
Admitted as hospital inpatient via emergency unit admission or transfer from psychiatric facility in last 2 years (%)	20	4
Outpatient psychiatric emergency services or ambulatory surgery this year (%)	41	15
Mean number of hospital inpatient admissions this year	0.30	0.06
Mean number of hospital inpatient days in last 2 years	3.7	0.6
Non-inpatient (ER or clinic visit) health system encounter this year (%)	68	43
Mean number of non-inpatient (ER or clinic visits) encounters this year	6.2	2.3
11 or more non-inpatient (ER or clinic visits) health system encounters this year (%)	20	6

Exhibit 1

Averages of Model Variables for High-Cost and Other Homeless Persons (Validation Sample) (2 of 2)

Variable	High Cost (N = 5,726)	Other (N = 51,533)
Behavioral health		
Mean number of mental health outpatient days in the last 2 years	11.1	2.1
Two or more mental health outpatient visits in the last 2 years (%)	27	9
Mean number of mental health inpatient admissions this year	17.6	1.2
Two or more mental health inpatient admission in the last 2 years (%)	20	6
Substance abuse indicated by any recorded medical diagnosis or justice system charge (%)	61	31
Mean number of drug abuse and alcohol service encounters in the last 2 years	14.3	3.9
HUD-funded homeless services and county public assistance		
Chronic homeless flag in any HUD-funded homeless service provider record (%)	27	11
Public assistance benefits received this year (%)	46	40
Two or more months of food stamp payments received in the past 2 years (%)	47	44

EMS = Emergency Medical Services. ER = emergency room. HCUP = Healthcare Cost and Utilization Project. HUD = U.S. Department of Housing and Urban Development. ICD-9 = International Classification of Diseases, Ninth Revision.

The binary target variable indicated whether or not homeless persons were in the top 10 percent of high-cost users in 2009 (training cohort) and 2012 (validation cohort). In order to identify high-cost status, costs were summed across all service types and then ranked separately for the training and validation cohorts.

Analysis

The model predicts who is in the 10 percent of the homeless persons with highest public services costs in 2009, using data from 2007 and 2008. The model was validated by using 2010 and 2011 records to predict high-cost status in 2012. The sample size for the training and validation cohorts was 57,259 records. The target group was 5,726 homeless individuals who made up the 10 percent with the highest costs. It was important to test the model using data for 2010 to 2012 in order to assess its out-of-sample predictive power. Strong predictive power is often observed based on in-sample performance if the model over-fits the data. When that is the case, the model only effective for explaining the training data, and out-of-sample performance is very poor. Because a predictive model is intended to be applied to new data with unknown outcomes, validation is needed to assess a model's performance.

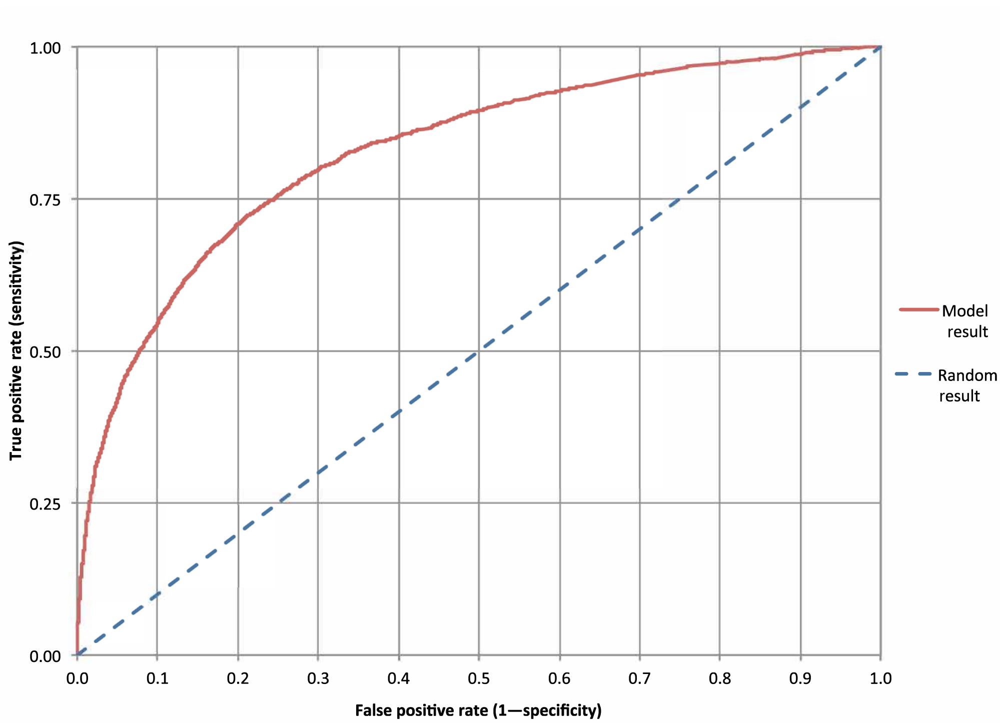
Model development was conducted in two stages. In the preprocessing stage, potential variables that might have an effect on becoming a high-cost user were identified based on earlier research and a series of F-tests (for categorical variables) and t-tests (for continuous variables). This step generated the first iteration of variable selection after eliminating redundant and irrelevant factors with p-values greater than 0.25. The initial set of selected variables was transformed and prepared for model development using several techniques such as binning continuous variables, clustering categorical variables, and generating binary and count variables. All variables were generated for the current and previous years, and a total of 256 input variables were selected to be included in the model development.

Several models for predicting high-cost users were developed and their performance was assessed using the SAS Enterprise Miner platform (Sarma, 2013; SAS, 2013). Several regression techniques were implemented to build models predicting the status of each person in the dataset as a high-cost user in the next year. We tested three techniques—logistic regression, least-angle regression, and decision tree models that are capable of explaining the classification or decision process, rather than using machine-learning algorithms that do not explain how given types of information are used to make predictions.

A comparison of the models' performance based on the receiver operating characteristic (ROC) curve led to selecting a logistic regression model as the champion model. The ROC curve shown in exhibit 2 plots the tradeoff between sensitivity or true positive rate (probability of true prediction) and specificity or false positive rate (probability of false prediction). The ROC curve is used to quantify how accurately a model can discriminate between two states—typically referred to as "event" and "nonevent"—or to compare two alternative models predicting the same event. In the final phase, this model was fine-tuned, introducing interactions between variables, testing the nonlinearity of variables and applying a sensitivity analysis to decrease the number of variables—particularly testing if current and previous year variables could be aggregated into a single variable without sacrificing the model's performance.

Exhibit 2

ROC Curve



ROC = receiver operating characteristic.

The final model was validated using the 2010–2012 cohort to assess the out-of-sample predictive power of the model. Sensitivity, specificity, positive predictive value (PPV), and accuracy measures, as well as the area ROC curve, were used to assess the out-of-sample model performance (Gonen, 2007).

The sensitivity statistic measures the proportion of high-cost homeless persons correctly identified by the model with high scores. It is also known as the true positive rate and reflects how well the model performs in capturing those homeless persons with high future costs. If the level is too low, a large number of high-cost homeless persons would not be provided with PSH.

The specificity statistic measures the proportion of not-high-cost homeless persons correctly identified by the model with low scores. If the level is too low, this is translated into to a high false positive rate (1-specificity), meaning a large number of homeless persons with low public costs would be provided with PSH.

The PPV statistic estimates the accuracy of the model by measuring the proportion of true positives (correctly classified high-cost homeless persons) within the population of all persons identified as high-cost persons. In other words, it is the probability that persons with a high score (above a defined cost threshold) truly are high-cost persons. Finally, the accuracy statistic measures the proportion of true positives and true negatives out of all persons.

The validated model was later utilized to estimate the potential costs and benefits of applying the model under several cutoff thresholds, using experience-based assumptions about costs of PSH and likely reduction in service use attributable to PSH placement.

Results

The final model had 38 variables with main effects out of 256 input variables tested and 11 variables with interactions. The descriptive values of model variables are shown in exhibit 1. The significance of the parameter estimates (*p*-values) and odds ratios are presented in exhibit 3. As shown in exhibit 1, high-cost homeless persons in Santa Clara County represent a higher proportion of males than the overall population that experienced homelessness, and are slightly older. Their rate of engagement in the criminal justice system is very high relative to the rest of the population. Almost half of them were arrested during the previous 2 years compared to only 16 percent for the rest of the population. Their average number of days in jail is more than six times greater than the rest of the population—32.9 days versus 5.2 days.

After testing 970 3-digit ICD-9 medical diagnoses, 43 diagnostic groups, and 18 body system diagnostic categories, the model retained six effective diagnosis codes or groups—adjustment reaction, organ failures, heart diseases, schizophrenia, neoplasm, and other ill-defined and unknown causes of morbidity and mortality. In addition, two other factors were included, which are the aggregations of chronic medical conditions and high-cost ICD-9. The high-cost homeless group shows much higher rates of encounters with these diagnoses whereas overall averages vary between 6 percent (heart diseases) and 68 percent (chronic medical condition). More than one-half of the high-cost group had been diagnosed with 1 or more of the 59 high-cost ICD-9s, while only a fifth of the lower-cost population had any of these diagnoses.

Exhibit 3

Logistic Regression Adjusted Odds Ratios and 95-Percent Confidence Limits for Predictor Variables (Validation Sample)

Variable	Odds Ratio	95-Percent Confidence Limits
Demographics (%)		
Age 18–45 versus less than 18*	1.21	1.06–1.38
Age 46–65 versus less than 18	0.98	0.85–1.13
Age 65+ versus less than 18***	0.88	0.69–1.14
Female versus male***	1.07	1.00–1.14
Female	42	54
Criminal justice		
100 or more days of probation in the last 2 years*	1.15	1.03–1.28
Arrested in last 2 years*	1.74	1.58–1.92
Jail booking in last 2 years*	1.14	1.04–1.26
Jail security classification of 3 or 4 (that is, high risk) this year*	1.63	1.41–1.89
Arrested for inebriation and released within 48 hours this year*	1.48	1.26–1.73
Number of arrests this year**	1.06	1.01–1.11
Number of days in jail this year*	1.007	1.005–1.009
Health diagnoses		
Diagnosed with chronic medical condition*	1.21	1.10–1.33
Diagnosed with adjustment reaction in last 2 years*	1.26	1.06–1.49
Diagnosed with heart disease in last 2 years*	1.41	1.15–1.72
Number of medical encounters with diagnosis of organ failure in last 2 years*	1.08	1.06–1.11
Diagnosed with schizophrenia in last 2 years**	1.23	1.03–1.46
Number of medical encounters with diagnosis of neoplasm in last 2 years*	1.05	1.03–1.07
Diagnosed with “other ill-defined and unknown causes of morbidity and mortality” in last 2 years **	1.28	1.05–1.58
Diagnosed with high-cost ICD-9 in last 2 years**	1.120	1.009–1.240
Health and emergency services		
EMS encounter this year*	1.27	1.14–1.41
EMS encounter last year*	1.26	1.14–1.40
Two or more EMS encounters in last 2 years*	1.34	1.12–1.60
Admitted as hospital inpatient via emergency unit admission in last 2 years*	1.35	1.19–1.54
Outpatient psychiatric emergency services or ambulatory surgery this year*	1.21	1.11–1.33
Number of hospital inpatient admissions this year*	1.16	1.09–1.25
Number of hospital inpatient days in last 2 years*	1.011	1.006–1.016
Non-inpatient (ER or clinic) health system encounter this year*	1.20	1.10–1.32
Mean number of non-inpatient (ER or clinic visits) encounters this year	6.2	2.3
11 or more non-inpatient (ER or clinic visits) health system encounters this year (%)	20	6
Number of non-inpatient (ER or clinic visits) encounters this year*	1.024	1.015–1.033
11+ non-inpatient (ER or clinic) health system encounters this year*	1.27	1.07–1.51
Behavioral health		
Number of mental health outpatient days in the last 2 years*	1.013	1.010–1.015
Two or more mental health outpatient visits in the last 2 years*	1.40	1.23–1.59
Number of mental health inpatient admissions this year*	1.002	1.002–1.003
Two or more mental health inpatient admission in the last 2 years*	1.28	1.08–1.51
Substance abuse indicated by any recorded medical diagnosis or justice system charge*	1.63	1.51–1.76
Number of drug abuse and alcohol service encounters in the last 2 years*	1.002	1.002–1.002
HUD-funded homeless services and county public assistance		
Chronic homeless flag in any HUD-funded homeless service provider record*	1.28	1.17–1.39
Public assistance benefits received in the current year*	1.36	1.18–1.57
Two or more months of food stamp payments received in the past 2 years*	0.68	0.59–0.79

* $p < .01$. ** $p < .05$. *** $p < .10$.

EMS = Emergency Medical Services. ER = emergency room. HUD = U.S. Department of Housing and Urban Development.
ICD-9 = International Classification of Diseases, Ninth Revision.

The high-cost group also shows higher rates of engagement with health and emergency services. Group differences were large for EMS encounters (30 percent versus 7 percent), hospital inpatient admissions via emergency room admission or transfer from a psychiatric facility (20 percent versus 4 percent), and outpatient psychiatric emergency services or ambulatory surgery (41 percent versus 15 percent). The number of admissions and days of inpatient hospitalization and the number of outpatient encounters are also significantly higher for high-cost homeless persons.

Finally, behavioral health data show more frequent encounters for the high-cost group. Both mental health (inpatient and outpatient) and substance abuse service rates are higher. The prevalence of documented substance abuse, as indicated by any drug-related medical diagnosis or justice system charge, is twice as high for the high-cost group—61 percent versus 31 percent for the balance of the population. In contrast, the public assistance and homeless service participation rates differ only slightly.

Adjusted odds ratios presented exhibit 3 reflect the differences we observe from descriptive comparisons. Odds ratio for continuous variables are adjusted by controlling for all other variables. As a result, odds ratios for binary variables (for example, arrested or not) are generally higher than the odds ratios for continuous variables (for example, days in jail) and are interpreted differently. For example, the odds ratios show that persons who have been arrested in the past 2 years are 1.74 times more likely to be in the high-cost group than those who have not been arrested. On the other hand, the odds ratio for each additional arrest is only 1.06, increasing the likelihood (or odds) of being in the high-cost group by 6 percent.

Odds ratios analysis reveals that being arrested in the last 2 years, higher jail security and substance abuse are among the strongest binary predictors of becoming a high-cost homeless resident, followed by being arrested for inebriation and released within 48 hours, heart disease, two or more EMS encounters, being admitted as a hospital inpatient via the emergency room, two or more mental health outpatient visits, and receiving public assistance benefits. All factors included in the model increase the likelihood of becoming a high-cost homeless person with adjusted ratios in the range of 1.05 and 1.28, with the exception of receiving 2 or more months of food stamp payments, which has an odds ratio of 0.68, indicating that receiving food stamps benefits makes it less likely to be in the high-cost group. The adjusted odds ratios for continuous variables all have values ranging from 1.002 (number drug abuse and alcohol services encounters) to 1.16 (number of hospital admissions), and all increase the likelihood of becoming a high-cost homeless person.

General performance of the model was evaluated using *C*-statistic to assess the predictive ability of the model. The *C*-statistic (sometimes called the "concordance" statistic or *C*-index) is a measure of goodness of fit for binary outcomes in a logistic regression model. It gives the probability that a randomly selected subject who experienced an event (for example, became high-cost user) had a higher risk score than a subject who had not experienced the event. It is equal to the area under the ROC curve and has values ranging from 0.5 to 1.0.

The model achieved a very strong *C*-statistic: 0.83. *C*-statistic is the probability that predicting the outcome is better than chance. Models are typically considered reasonable when the *C*-statistic is higher than 0.7 and strong when C-statistic exceeds 0.8 (Hosmer and Lemeshow, 2000). Overall, the model predicts high-cost homeless persons with a very good fit.

Exhibit 4 shows the predictive performance of the model for different scenarios: the top 1, 5, and 10 percent and the top 1,000 homeless persons with the highest risk of becoming a high-cost service user. The predictive performance measures were defined previously in the methods section.

If the 2,864 persons in the top 5 percent at greatest risk of becoming high-cost homeless service users are followed, the achieved sensitivity and specificity are 32.6 and 97.3 percent, respectively. These values suggest very reasonable predictive power, indicating that the model picks up 33 percent of all high-cost service users and correctly identifies 97 percent of those who are not high-cost users. The PPV value of 51 percent and accuracy value of 92.3 percent for the top 5 percent are also very high. If we follow a subset within the top 5 percent, the 1,000 cases with the highest probability scores for being in the high-cost group (1.75 percent of all cases), we see even more accurate prediction outcomes. The model achieves a PPV result of 67 percent, meaning that out of 1,000 persons that the model identified as high-cost persons, two-thirds are true positives and the remaining one-third are false positives. PPV is an important measure for assessing the cost-effectiveness of the model.

Another measure of the effectiveness of a predictive model is the "lift," which is calculated as the ratio between the results obtained with and without the predictive model for all thresholds. Exhibit 5 illustrates the lift of the model, which is quite high for cases with a high probability of being in the high-cost group. For example, for the top 5 percent, the model generates a lift of 6.5. This means that model generates 6.5 times more correctly identified high-cost homeless persons (true positives) than random selection, which is presented as the baseline: a lift of 1 or 0. At slightly lower thresholds, such as the top 10 percent, lift drops to 4.7 because in order to capture more true positives, the model concurrently includes more false positives. Conversely, the number of false positives decreases as the probability of being in the high-cost group increases.

The most common way of assessing the predictive power of a model in the data mining literature is the area under the ROC curve. ROC shows the tradeoff between true positives (sensitivity) and false positives (1-specificity) at all possible thresholds. The ROC curve for the model is shown in exhibit 2. The accuracy of the model depends on how well it separates high-cost individuals from lower-cost individuals. Accuracy is measured by the AUC (Area Under the Curve, the ROC curve) or *C*-statistic. The model generated a fairly high AUC of 0.83, indicating an 83-percent probability that a randomly selected homeless person with high future costs will receive a higher model score

Exhibit 4

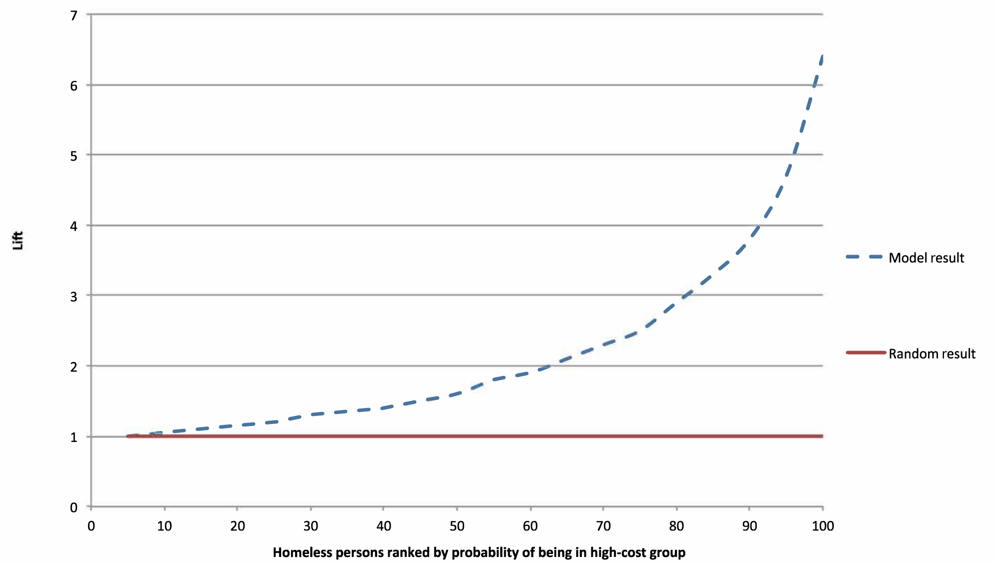
Predictive Performance of the Model

Measure	Top 1%	Top 5%	Top 10%	Top 1,000	Formula
Sensitivity (%)	9.3	32.6	47.7	14.9	True positive / (true positive + false negative)
Specificity (%)	99.7	97.3	93.2	99.4	True negative / (false positive + true negative)
PPV (%)	72.9	51.0	37.4	66.8	True positive / (true positive + false positive)
Accuracy (%)	92.6	92.3	89.5	92.7	(True positive + true negative) / number

PPV = positive predictive value.

Exhibit 5

Lift Chart



than a randomly selected homeless person without high future service costs. In the predictive analytics literature, models with AUC exceeding 0.8 are accepted as models with good predictive power, and AUC values below 0.7 indicate poor model performance.

Because the model provides a probability score ranging from 0 to 1, we have to select a cutoff score or a threshold, above which homeless persons will be offered PSH. Choice of a cutoff level introduces the tradeoff between the correct identification of high-cost service users and false alarm rates. The ROC curve illustrates this tradeoff between true positives—finding as many homeless persons as possible who would be high-cost service users next year—and false positives—decreasing potential cost savings by including homeless persons who would not be high-cost service users next year.

Business Scenario and Cost Savings

Although the performance of the triage tool presented in this article is very high in statistical terms, it is still necessary to translate this performance into a pragmatic business scenario, showing how the tool contributes to the efficiency of PSH programs by prioritizing the population to be housed. The tradeoff to be weighed in using the triage tool is between, on the one hand, using lower selection thresholds in order to find as many high-cost homeless individuals as possible but accepting a substantial number of lower-cost individuals as part of the mix, and, on the other hand, using higher selection thresholds to identify a smaller population in which a higher proportion of individuals will be high-cost service users. This tradeoff is critical to the efficiency of a PSH program as elaborated previously. The model is highly accurate in distinguishing high-cost from low-cost users, however it is still necessary to calibrate the cutoff level based on goals for saving costs by offering PSH to the targeted population. The following analysis explores the cost efficiency of providing PSH to targeted high-cost homeless persons under different cutoff levels.

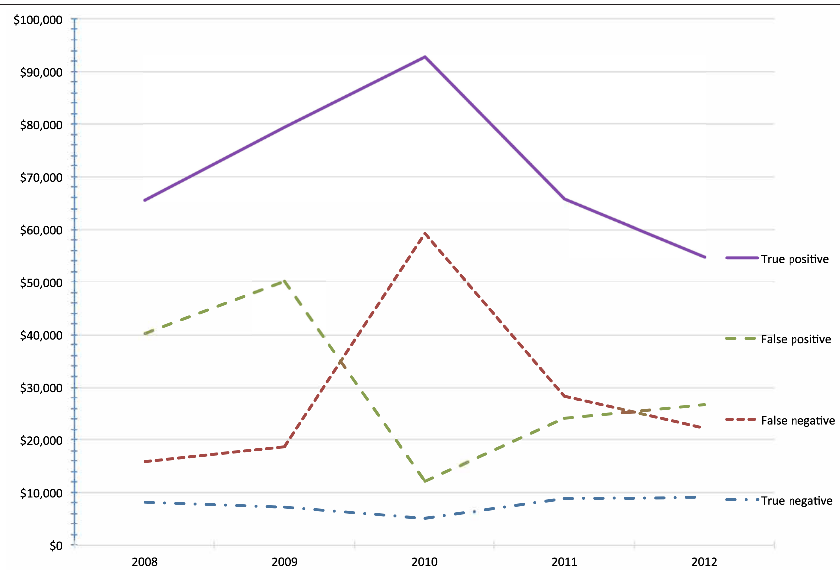
With 5 years of actual cost data, from 2008 through 2012, we used 2008 and 2009 data to produce probability scores for the likelihood of each individual being in the highest-cost group in 2010 and then track the accuracy and financial outcomes of these predictions over the next 2 years. Any placement decision has cost implications. If the homeless person predicted to be a high-cost user was correctly identified (true positive), the reduction in posthousing use of public services is likely to be roughly two-thirds. However, if the homeless person predicted to be a high-cost user was a false positive, then the expected cost savings would not be realized. The balance between the positive and negative savings generated by these two groups determines the efficiency of a PSH program.

One of the challenges the model must contend with is abrupt changes in costs in the scoring year—the year following the 2 years for which health conditions and service utilization are known. Some conditions are one-time events, resulting in costs that spike and then decline. Hence, the assessment of cost offsets should be done in the postplacement period, when the actual service utilization of true positives and false positives becomes evident. Some homeless persons who were true positives at the time of scoring year became low-cost users in subsequent years due to regression to the mean. On the other hand, some false positives that were predicted to be high-cost users but were low-cost users in the scoring year turned out to have higher costs in subsequent years.

These dynamics are shown in exhibit 6 for predictions of the top 5 percent. Looking at 2 years of post-scoring-year cost data (adjusted to 2014 U.S. dollars), the model successfully differentiates

Exhibit 6

Average Annual Costs for Triage Tool Prediction Groups



the highest cost cases from other cases, even though average costs decline because of regression to the mean. The low cost levels of true negatives verify the high specificity of the model. Another critical observation is that public costs for individuals experiencing homelessness vary significantly from one year to the next with important implications for the efficiency measure. False positives represent homeless persons with high service utilization prior to the scoring year of 2010, which led to high probability scores. However, in 2010, their service costs were low, making them false positives. On the other hand, their postprediction trend is positive, more than doubling between 2010 and 2012. If a person was predicted at the top 5 percent and had no service utilization at 2 years post scoring year, the person was labeled as false positive. If he or she was not predicted at the top 5 percent, the person was labeled as true negative, because he or she remained a low-cost user after the scoring year.

Note also that false negatives, the group with low service utilization prior to 2010 and high costs in 2010, the scoring year, typically had one-time cost spikes. Their long-term trend is negative and subsequent to the scoring year their cost levels declined substantially. Hence, omitting them as high-cost users contributes to the efficiency of the program significantly as presented below exhibit 6 suggests that cost savings should be assessed not at the year of scoring but rather in the postscoring years in order to capture the long-term service utilization of scored individuals.

The triage tool works to assign high scores to high-cost users, but different probability cutoff levels will exhibit different proportions of true positives with expected savings and false positives with no expected cost savings. Our estimation of net savings at different cutoff levels is based on the estimated cost savings for true positives after taking into account the housing and service costs for false positives. The results are sensitive to the probability score threshold, cost of housing, and the rate of anticipated reduction in service utilization and costs following placement in housing. As the probability score threshold increases, the ratio of true positives to false positives also increases, resulting in increased savings.

This analysis looks at financial outcomes based on two probability score thresholds, 0.37 and 0.53, for the predicted probability of having high costs in 2010, based on 2008 and 2009 information. The 0.37 cutoff level identifies approximately 5 percent of the test population as high-cost users. The 0.53 cutoff level identifies the top 1,000 high-probability service users in our test population. A different probability cutoff can be selected based on the requirements of specific initiatives to address homelessness. If the goal is to house a larger number of high-cost homeless persons, lower cutoff levels may be selected, resulting in lower savings per person. On the other hand, if the supply of housing is limited and a smaller number of high-cost homeless persons can be housed, than a higher cutoff level may be selected, resulting in higher savings per person.

It is assumed that the annual cost of PSH is $17,000 per person per year, based on rent subsidy and supportive service costs in Los Angeles. We used Los Angeles data for housing costs and posthousing cost savings because, at the time of our study, Santa Clara County did not have enough high-cost individuals who had been housed for a long enough interval to produce comparable data. This high-side estimate of housing costs is based on $11,000 annually for rental subsidy, including first-year costs for temporary housing and benefits advocacy, and $6,000 annually for supportive services. Actual costs may be lower based on the level of subsidies built into different affordable housing projects and the level of long-term supportive services needed by tenants after they are stabilized in housing.

The posthousing reduction in service costs is assumed to be 68 percent for homeless persons in the 10th decile based on a study from Los Angeles (Economic Roundtable, 2009). Most other studies estimate service cost reductions for homeless persons in PSH for the whole population, rather than the top decile (Culhane, 2008; Culhane and Byrne, 2010). It is also assumed that there will not be any cost reduction for false positives—individuals below the 10th decile. This is a conservative assumption because an earlier study also found posthousing cost reductions among individuals in the 5th through 9th cost deciles (Economic Roundtable, 2011). However for purposes of the cost estimates shown here, net savings are -$17,000 for cost groups below the 10th decile because no cost savings are applied to them.

Exhibit 7 presents estimated cost savings for 2011 for the two selected cutoff levels (0.37 and 0.53). Posthousing costs are calculated as 32 percent of homeless costs for individuals in the 10th cost decile, and then $17,000 is added for each person in the group to cover the cost of housing and supportive services. Net savings are calculated by subtracting estimated posthousing costs from actual homeless costs for the year. All analysis was conducted in 2014 prices. Since actual costs in 2011 and 2012 were used, regression to the mean, that is, the tendency of extreme outcomes to be closer to the average when measured a second time, has been incorporated into the estimates.

Cost differences were estimated for four probability-cost groups, which each show different cost dynamics. If a score was above the selected cutoff (0.37 or 0.53) and 2010 costs were in the top decile, the record is a true positive. However, in subsequent years, true positives in 2010 may remain high-cost or become low-cost service users. The long-term cost status of individuals was

Exhibit 7

Cost Savings for 2011 at the Cutoff Levels of 0.37 and 0.53

Status	2010 Costs (Prediction Year) ($)	2011 Costs (1 Year After Prediction) ($)	2011 Cost Savings ($)	2011 Net Savings ($)	2011 Total Savings ($)	Number of Cases
Cutoff level: 0.37						
True positives—low-cost users	90,989	10,932	0	– 17,000	– 4,335,000	255
True positives—high-cost users	93,196	83,661	56,889	39,889	30,635,068	768
False positives—low-cost users	11,444	8,511	0	– 17,000	– 8,823,000	519
False positives—high-cost users	13,029	46,551	31,655	14,655	5,085,204	347
Total / average				11,944	22,562,272	1,889
Cutoff level: 0.53						
True positives—low-cost users	111,580	11,496	0	– 17,000	– 2,074,000	122
True positives—high-cost users	96,892	86,947	59,124	42,124	22,367,823	531
False positives—low-cost users	12,427	8,829	0	– 17,000	– 3,094,000	182
False positives—high-cost users	13,579	43,560	29,621	12,621	2,082,432	165
Total / average				19,282	19,282,255	1,000

evaluated based on their actual cost rankings in 2011 or 2012. If they were in the top decile in 2011 or 2012, they were identified as long-term high-cost users. Otherwise, they were identified as low-cost users.

If a score was above the selected cutoff (0.37 or 0.53) and 2010 costs were not in the top decile, the record is a false positive. False positives may also become high- or low-cost service users in the future. We tested this possibility by observing actual costs in 2011 and 2012 and identifying cases that moved into the true positive cost category. Exhibit 7 shows that, at the 0.37 cutoff level, out of the 1,123 individuals who were true positives, 255 became low-cost users in 2011. This cost shift was more than offset by 347 false positives that turned out to be high-cost users in 2011. In sum, out of 1,889 individuals, 1,115 (60 percent) were high-cost users in 2011.

If the 5 percent (0.37 cutoff level) with the highest probability of being high-cost service users were housed permanently with supportive services, savings of more than $22 million were estimated in 2011. Even though 40 percent of individuals were low-cost users in 2011 and would not be generating any cost savings, the net savings from the remaining 60 percent shows the feasibility of the intervention. The analysis shows a cost reduction of almost $12,000 per housed homeless person for the top 5 percent of the population identified by the triage tool as having the greatest probability of high future costs.

The results are even more positive when a higher cutoff level is selected, because the accuracy of the tool in predicting high-cost users improves as the probability level increases. The 2011 cost analysis for 1,000 persons in the test population with the highest probability scores, scores at or above 0.53, shows that almost two-thirds (653 individuals) were true positives. Evaluating actual costs in 2011, it is observed that 122 of them became low-cost users, whereas more than four-fifths (531) remained high-cost users. In addition, 165 false positives turned out to be high-cost users in 2011. In sum, out of 1,000 individuals, 696 (70 percent) were high-cost users in 2011. As expected, the feasibility of the intervention is higher at the 0.53 threshold than at the 0.37 threshold, with an estimated cost reduction for this group of more than $19,000 per person in 2011.

A separate analysis estimated savings in 2012 for both cutoff levels. Because lower cost levels were observed in 2012 due to the regression to the mean, lower cost savings were estimated. At the 0.37 level, cost savings were estimated to be almost $16 million, which corresponds to more than $8,000 per housed individual. At the 0.53 level, savings per individual were estimated to be $16,000, with cumulative savings for 2011 and 2012 estimated to exceed $35 million. Over the 2 years of postprediction data that we have for Santa Clara County, we see a year-to-year decline in actual costs for individuals with a high probability of having high costs. However, this may be the first phase of a longer-term cost cycle in which costs begin to increase again. This scenario is plausible considering that most individuals in this population have serious medical and mental health disorders that are likely to become more acute as they age. Indications of a longer-term cycle in which costs decline and then increase were found in an earlier cost study in Los Angeles (Economic Roundtable, 2009).

As noted previously, our cost savings analysis assumed that the annual cost of PSH is $17,000 per person per year and that the posthousing reduction in service costs is 68 percent for homeless persons in the 10th decile. Because both of these assumptions are made based on data and recent studies from Los Angeles, a separate sensitivity analysis was carried out to see how total net cost savings estimates change if these cost assumptions change. The analysis showed that at the 0.37 cutoff level, the break-even point is reached when the annual cost of PSH is $29,000 or the posthousing reduction in service costs is 40 percent. These are the highest annual cost of PSH and the lowest percentage of service cost reduction that still yield net cost savings.

If instead of cost savings as a goal, a community seeks to break even against current or projected costs, a probability threshold of 0.20 is estimated to produce “break-even” cost outcomes, with cost savings from reduced service use equal to the cost of housing and services. An estimated 70 percent of the population captured at this probability threshold is chronically homeless, and they represent an estimated 21 percent of all individuals who are chronically homeless in a given year.

Discussion

This study is the first attempt in Santa Clara County and one of the first studies to develop and validate a predictive model for identifying homeless persons who are likely to become high-cost users of public service. This model was developed using an integrated database built by linking seven agencies’ administrative records, which provided information on risk factors such as demographics, clinical variables, and service utilization variables for the current and previous years as well as cost of service data. The cost study that was used to develop the triage tool provided key evidence supporting Measure A, a $950 million affordable housing bond measure approved by voters in 2016.

The model is particularly strong when using high probability cutoff levels, generating small numbers of false positives and high numbers of true positives. For the top 1,000 high-cost users predicted by the model, two-thirds of them are true positives. A key strength of this study is that it assessed the overall effectiveness of predictions made by the tool, looking at costs over the 3 years following the 2 years that were the source of data used to make the prediction. This assessment showed that many false positives became high-cost or close-to-high-cost users in the second year after the prediction. In addition, a majority of the false negatives were actually true negatives over the next 2 years because their high-cost level in the scoring year represented a one-time cost spike. One of the challenges the model must contend with is abrupt changes in costs from one year to the next. Some conditions are one-time events, resulting in costs that spike and then decline. The tool performed very well by giving low scores to homeless persons with one-time cost spikes.

Another key strength of the study is information it provided for identifying distinctive attributes of high-cost individuals. Individuals in this group are the most likely to be diagnosed with a mental disorder, in particular, a disorder that takes the form of a psychosis, and a psychosis that takes the form of schizophrenia. They are also the most likely to be given a maximum or high-medium security jail classification because of the safety risk they are perceived to present. They are the most likely to have been continuously homeless for 3 years. They are most likely to be diagnosed with a skin disease such as cellulitis or an endocrine disease such as diabetes. They are most likely to be tri-morbid—diagnosed with a mental disorder, a chronic medical condition and to abuse drugs or alcohol. Demographically they are most likely to be male and to be in the middle of their lives—35 to 44 years old. Also, they are most likely to frequent users of hospital emergency rooms and inpatient beds, emergency psychiatric facilities, mental health inpatient facilities, and to be incarcerated in a jail mental health cell block.

This composite profile can help hospital and jail discharge planners and homeless service providers identify high-cost individuals. However, significant diversity is in the demographic attributes and types of crisis services needed by individuals in this population. The triage tool weighs the likely cost impact of each individual's characteristics and uses this information to identify subgroups that fall outside this profile. For example, young women with acute mental illnesses and endocrine diseases who have ongoing high costs even though they are not substance abusers or involved in the justice system.

We further validated the model by developing a business analysis to assess its cost effectiveness. With 0.37 selected as the optimal cutoff level, which identifies the highest-cost 5 percent of the population that experienced homelessness over a 6-year period as the target group, the model assessed cost savings by comparing total housing and service costs ($17,000 annually) with the estimated 68 percent cost savings for true positives—those correctly identified as high-cost service users. The results confirmed that anticipated cost savings from true positives far exceed the total costs of housing, yielding net savings of $20,000 per person over the next 2 years, after the total population with a probability score of 0.37 or higher enters PSH. Using 0.53 as the minimum probability threshold for the target group, the estimated annual savings are $32,000 per person, after paying for housing and supportive services. On the other hand, using 0.20 as the probability threshold, we achieve break-even financial results, with the cost of providing housing and supportive services fully offsetting cost savings from reduced service use.

The optimal cutoff is not simply an empirical decision. In the context of PSH, it depends on the number of people who can be housed in available housing. However, in the context of a long-term strategy to address homelessness, the tradeoff between costs and savings in the population needing housing provides evidence that jurisdictions can use to validate local policy initiatives, such as affordable housing bond measures to expand the inventory of available housing.

It is often argued that the feasibility of prevention services, such as PSH, would not be attained without a strategy of balancing the costs with some degree of cost offsets. One of the most significant strengths of this study is its strong performance in identifying homeless persons with a high probability of having high ongoing public costs that will substantially exceed the cost of PSH.

The predictive performance of the Silicon Valley Triage Tool was compared to the performance of two earlier triage tools developed in Los Angeles by running all of the models on records of homeless persons from both Los Angeles and Santa Clara Counties. The tools were assessed based on the proportion of high-cost homeless persons correctly identified by each model and the proportion of persons predicted to be high-cost homeless who truly were high-cost persons. The Silicon Valley tool demonstrated comparable or higher accuracy when run on Los Angeles data and much higher accuracy when applied to the Santa Clara data. This comparison verifies that the Silicon Valley tool demonstrates strong predictive performance in multiple metropolitan regions.

Limitations

This analysis and the model developed in this study are also subject to some limitations that need to be acknowledged, and most of these limitations are inherent to analysis involving administrative datasets. Our study is limited by the usual shortcomings of research based on linked administrative records, including errors in the underlying data sources, such as missing data and data entry errors. Matching inaccuracies prevented the use of the full homeless population for the analysis. The tool was developed using data for roughly 55 percent of the population that experienced homelessness, 57,259 persons. These were individuals with at least one record linked to an agency during our 6-year study window from 2007 through 2012. Since administrative databases usually are not designed to collect data for research, information about some critical risk factors is often missing. For example, in developing this tool, we did not have access to data about income and employment. Moreover, some service costs were missing for some years and had to be estimated. For some services, when individual-level costs were not available, average costs per unit of service were used.

Another shortcoming related to the use of administrative data is incomplete and sometimes inaccurate information about the timing of homeless episodes. Because complete information about the duration of homelessness was not available, the study population was assumed to be either homeless or at risk of homelessness while predicting high-cost users, assuming that individuals would use more services when they were experiencing homelessness. In addition, the administrative datasets did not show the mobility of homeless individuals in and out of the county, which would impact their utilization of services in county facilities.

The business scenario that estimated cost savings was also subject to some limitations. First, it assumed that PSH costs $17,000 a year, which needs to be verified when the county has a larger body of postsupportive housing cost data. Second, because posthousing costs of homeless persons were not available for this study, cost offsets were based on a saving factor of 68 percent, which was derived from an earlier study conducted in Los Angeles. Actual cost savings may be different after the implementation of the program. On the other hand, service reductions measured here represent a conservative assessment of the impact of the PSH on service use and costs because it was assumed that homeless persons with costs below the 10th decile would not experience any service reductions after being housed, so that PSH costs were not adjusted with any cost offsets for this group.

Finally, the Silicon Valley Tool is a system-based tool; that is, it requires detailed healthcare and justice system information about each individual that is available only from those institutional systems. This includes medical diagnoses, accurate details of encounters with healthcare providers, and details about stints of incarceration. Cooperation of both healthcare and justice system agencies is necessary to protect the privacy of personal information while providing the data required for the tool. Santa Clara County agencies agreed to authorize a research unit in the Behavioral Health Services Department to link records across county agencies and then to de-identify the linked records so that they could be used by the Economic Roundtable to develop this triage tool.

Because of the level of effort required to obtain and integrate the necessary data, the most efficient use of the tool is for regular, ongoing system-wide screening of linked records rather than screening clients individually. By predicting how likely each person in the entire identified population of homeless resident is to have high future costs, it is possible to prioritize individuals for access to the scarce supply of PSH. For example, targeted individuals can be flagged in client databases so that housing can be offered to them the next time they seek services.

The Silicon Valley Tool can also be used to screen cases individually. A version of the tool for individual screening in Excel format as well as software code for screening entire client databases can be downloaded at https://economicrt.org/publication/silicon-valley-triage-tool/.

Because the tool does not correctly identify all high-cost individuals, the screening process for either individuals or groups should include an option to override the triage tool probability score based on the clinical judgment of healthcare professionals. For example, if a patient has recently been diagnosed with a high-cost, chronic medical condition, this would warrant overriding a negative result from the triage tool and including the patient in the high-cost group that receives access to PSH. Allowing overrides permits service providers to adapt to changing populations and conditions and to be responsive to unique circumstances.

The tool also has practical value for identifying patients served by health plans and private hospitals who have high ongoing costs, and whose health outcomes will improve and costs decrease if they are housed. Local government safety net resources can be augmented through collaborative care for frequent users who are also served by private hospitals.

Using the triage tool raises the broader ethical issue of making decisions about who gets into housing and who is left out. We see the tool as an interim means of prioritizing need in the context of an inadequate supply of affordable housing and insufficient human service interventions for reducing the flow of people into chronic homelessness. In this context, the tool prioritizes individuals based on public costs, which reflect frequency of service-intensive crises, and are closely linked to (but not identical with) level of distress. Use of the triage tool may be the approach that houses the greatest number of people because public agencies achieve the highest level of cost avoidance by housing high-cost individuals, opening the possibility of using those savings to pay for other crucial services.

Conclusion and Future Research

Needs within the homeless population vary significantly. Although the Silicon Valley Triage Tool is effective for prioritizing access to PSH for the small number of high-cost individuals who account for the majority of public costs, other tools are needed to target services for less disabled segments of the population. Less expensive interventions may be effective for individuals with less acute needs. This includes preventive care for children who have experienced homelessness, integrated outpatient healthcare, readily available and effective behavioral and mental health services, temporary affordable housing, and employment services. Without effective early intervention, the risk that individuals will become chronically homeless and that their problems will worsen to the extent that they become high-cost homeless is real.

Acknowledgments

The authors thank the County of Santa Clara and Destination: Home for carrying out the cross-agency linkage of records for homeless clients, for providing the funding that made this research possible, and also for their commitment to using data strategically for improving the lives of people experiencing homelessness.

Authors

Halil Toros is statistical analytics consultant at the Economic Roundtable.

Daniel Flaming is President of the Economic Roundtable.

References

Apicello, Jocelyn. 2010. "A Paradigm Shift in Housing and Homeless Services: Applying the Population and High-Risk Framework to Preventing Homelessness," *The Open Health Services and Policy Journal* 3: 41–52.

Ash, Arlene S., Yang Zhao, Randall P. Ellis, and Marilyn Schlein Kramer. 2001. "Finding Future High-Cost Cases: Comparing Prior Cost Versus Diagnosis-Based Methods," *Health Services Research* 36 (6 pt. 2): 194–206.

Billings, John. 2006. *Identifying High Cost Patients for Interventions to Improve Health and Social Care Services*. New York: NYU Center for Health and Public Service Research.

Billings, John, Theo Georghiou, Ian Blunt, and Martin Bardsley. 2013. "Choosing a Model To Predict Hospital Admission: An Observational Study of New Variants of Predictive Models for Case Finding," *BMJ Open*: 1–9. DOI: 10.1136/bmjopen-2013-003352.

Burt, Martha R. 2002. "Chronic Homelessness: Emergence of a Public Policy," *Fordham Urban Law Journal* 30: 1267–1279.

Burt, Martha R., Carol L. Pearson, and Ann Elizabeth Montgomery. 2005. *Strategies for Preventing Homelessness*. Report prepared for the U.S. Department of Housing and Urban Development. Washington, DC: Urban Institute; Walter R. McDonald and Associates.

Byrne, Thomas, Jamison Fargo, Ann Elizabeth Montgomery, Ellen Munley, and Dennis P. Culhane. 2014. "The Relationship Between Community Investment in Permanent Supportive Housing and Chronic Homelessness," *Social Service Review* 88: 234–263. DOI: 10.1086/676142.

Byrne, Thomas, Dan Treglia, Dennis P. Culhane, John Kuhn, and Vincent Kane. 2016. "Predictors of Homelessness Among Families and Single Adults After Exit From Homelessness Prevention and Rapid Re-Housing Programs: Evidence From the Department of Veterans Affairs Supportive Services for Veteran Families Program," *Housing Policy Debate* 26: 1, 252–275. DOI: 10.1080/10511482.2015.1060249.

Caton, Carol, Carol Wilkins, and Jacquelyn Anderson. 2007. "People Who Experience Long-Term Homelessness: Characteristics and Interventions." http://www.aspe.hhs.gov/hsp/homelessness/symposium07/caton.

Caton, Carol L.M., Boanerges Dominguez, Bella Schanzer, Deborah S. Hasin, Patrick E. Shrout, Alan Felix, Hunter McQuistion, Lewis A. Opler, and Eustace Hsu. 2005. "Risk Factors for Long-Term Homelessness: Findings From a Longitudinal Study of First-Time Homeless Single Adults," *American Journal of Public Health* 95: 1753–1759. DOI: 10.2105/AJPH.2005.063321.

Chechulin, Yuriy, Amir Nazerian, Saad Rais, and Kamil Malikov. 2014. "Predicting Patients With High Risk of Becoming High-Cost Healthcare Users in Ontario (Canada)," *Healthcare Policy* 9: 68–79. DOI: 10.12927/hcpol.2014.23710.

Culhane, Dennis P. 2008. "The Cost of Homelessness: A Perspective From the United States," *European Journal of Homelessness* 2: 97–114. http://repository.upenn.edu/spp_papers/148.

Culhane, Dennis P., and Thomas Byrne. 2010. Ending Chronic Homelessness: Cost-Effective Opportunities for Interagency Collaboration. Working paper. Philadelphia: University of Pennsylvania School of Social Policy and Practice. http://repository.upenn.edu/spp_papers/143.

Culhane, Dennis P., Stephen Metraux, and Thomas Byrne. 2011. "A Prevention-Centered Approach to Homelessness Assistance: A Paradigm Shift?" *Housing Policy Debate* 21: 295–315. DOI:10.1080/10511482.2010.536246.

Culhane, Dennis P., Stephen Metraux, and Trevor Hadley. 2002. "Public Service Reductions Associated With Placement of Homeless Persons With Severe Mental Illness in Supportive Housing," *Housing Policy Debate* 13: 107–163. DOI: 10.1080/10511482.2002.9521437.

Economic Roundtable. 2015a. *All Alone: Antecedents of Chronic Homelessness*. Los Angeles. DOI: 10.13140/RG.2.1.4067.9281.

———. 2015b. *Home Not Found: The Cost of Homelessness in Silicon Valley*. Los Angeles. DOI: 10.13140/RG.2.1.4780.6327.

———. 2012. *Hospital to Home: Triage Tool II for Identifying Homeless Hospital Patient in Crisis*. Los Angeles.

———. 2011. *Crisis Indicator: Triage Tool for Identifying Homeless Adults in Crisis*. Los Angeles. DOI: 10.13140/RG.2.1.4788.8246.

———. 2009. *Where We Sleep: The Costs of Housing and Homelessness in Los Angeles*. Los Angeles. DOI: 10.13140/RG.2.1.2624.0887.

Fleishman, John A., and Joel W. Cohen. 2010. "Using Information on Clinical Conditions To Predict High-Cost Patients," *Health Services Research* 45: 532–552. DOI: 10.1111/j.1475-6773.2009.01080.x.

Folsom, David P., William Hawthorne, Laurie Lindamer, Todd Gilmer, Anne Bailey, Shahrokh Golshan, Piedad Garcia, Jürgen Unützer, Richard Hough, and Dilip V. Jeste. 2005. "Prevalence and Risk Factors for Homelessness and Utilization of Mental Health Services Among 10,340 Patients With Serious Mental Illness in a Large Public Mental Health System," *American Journal of Psychiatry* 162: 370–376. DOI: 10.1176/appi.ajp.162.2.370.

Gilmer, Tod P., Willard G. Manning, and Susan L. Ettner. 2009. "A Cost Analysis of San Diego County's REACH Program for Homeless Persons," *Psychiatric Services* 60: 445–450.

Gonen, Mithat. 2007. *Analyzing Receiver Operating Characteristics With SAS*. SAS Press Series. Cary, NC: SAS Institute.

Greenwood, Ronnie Michelle, Ana Stefancic, and Sam J. Tsemberis. 2013. "Pathways Housing First for Homeless Persons With Psychiatric Disabilities: Program Innovation, Research, and Advocacy," *Journal of Social Issues* 69: 645–663. DOI: 10.1111/josi.12034.

Gubits, Daniel, Marybeth Shinn, Michelle Wood, Stephen Bell, Samuel Dastrup, Claudia D. Solari, Debi McInnis, Tom McCall, and Utsav Kattel. 2016. *Family Options Study: 3-Year Impacts of Housing and Services Interventions for Homeless Families*. Washington, DC: U.S. Department of Housing and Urban Development. huduser.gov/portal/publications/Family-Options-Study.html.

Henwood, Benjamin F., Howard Dichter, Robert Tynan, Christine Simiriglia, Krista Boermer, and Adam Fussaro. 2015. "Service Use Before and After the Provision of Scatter-Site Housing First for Chronically Homeless Individuals With Severe Alcohol Use Disorders," *International Journal of Drug Policy* 26: 883–886. DOI: 10.1016/j.drugpo.2015.05.022.

Hosmer, David W., and Stanley Lemeshow. 2000. *Applied Logistic Regression*, 2nd ed. New York: John Wiley and Sons.

Kuhn, Randall, and Dennis P. Culhane. 1998. "Applying Cluster Analysis To Test a Typology of Homelessness by Pattern of Shelter Utilization: Results From the Analysis of Administrative Data," *American Journal of Community Psychology* 26: 207–232. http://repository.upenn.edu/spp_papers/96.

Kuno, Eri, Aileen B. Rothbard, June Averyt, and Dennis Culhane. 2000. "Homelessness Among Persons With Serious Mental Illness in an Enhanced Community-Based Mental Health System," *Psychiatric Services* 51: 1012–1016. DOI: 10.1176/appi.ps.51.8.1012.

Kushel, Margot B., Judith A. Hahn, Jennifer L. Evans, David R. Bangsberg, and Andrew R. Moss. 2005. "Revolving Doors: Imprisonment Among the Homeless and Marginally Housed Population," *American Journal of Public Health* 95: 1747–1752. DOI: 10.2105/AJPH.2005.065094.

Kushel, Margot B., Sharon Perry, David Bangsberg, Richard Clark, and Andrew R. Moss. 2002. "Emergency Department Use Among the Homeless and Marginally Housed: Results From a Community-Based Study," *American Journal of Public Health* 92: 778–784. DOI: 10.1186/s13722-015-0038-1.

Larimer, Mary E., Daniel K. Malone, Michelle D. Garner, David C. Atkins, Bonnie Burlingham, Heather S. Lonczak, Kenneth Tanzer, Joshua Ginzler, Seema L. Clifasefi, William G. Hobson, and G. Alan Marlatt. 2009. "Health Care and Public Service Use and Costs Before and After Provision of Housing for Chronically Homeless Persons With Severe Alcohol Problems," *Journal of American Medical Association* 301: 1349–1357. DOI: 10.1001/jama.2009.414.

Larson, Eric B. 2013. "Building Trust in the Power of 'Big Data' Research To Serve the Public Good," *Journal of American Medical Association* 309: 2443–2444. DOI: 10.1001/jama.2013.5914.

Ly, Angela, and Eric Latimer. 2015. "Housing First Impact on Costs and Associated Cost Offsets: A Review of the Literature," *Canadian Journal of Psychiatry* 60: 275–287.

Martinez, Tia E., and Martha R. Burt. 2006. "Impact of Permanent Supportive Housing on the Use of Acute Care Health Services by Homeless Adults," *Psychiatric Services* 57: 1–8. DOI: 10.1176/ps.2006.57.7.992.

McLaughlin, Thomas Chalmers. 2011. "Using Common Themes: Cost-Effectiveness of Permanent Supported Housing for People With Mental Illness," *Research on Social Work Practice* 21: 404–411. DOI: 10.1177/1049731510387307.

McNiel, Dale E., Renee L. Binder, and Jo C. Robinson. 2005. "Incarceration Associated With Homelessness, Mental Disorder, and Co-Occurring Substance Abuse," *Psychiatric Services* 56: 840–846. DOI: 10.1176/appi.ps.56.7.840.

Metraux, Stephen, and Dennis P. Culhane. 2004. "Homeless Shelter Use and Reincarceration Following Prison Release: Assessing the Risk," *Criminal Public Policy* 3: 201–222. http://repository.upenn.edu/spp_papers/116.

Metraux, Stephen, Dennis P. Culhane, Stacy Raphael, Matthew White, Carol Pearson, Eric Hirsh, Patricia Ferrell, Steve Rice, Barbara Ritter, and J. Stephen Cleghorn. 2001. "Assessing Homeless Population Size Through the Use of Emergency and Transitional Shelter Services in 1998: Results From the Analysis of Administrative Data From Nine U.S. Jurisdictions," *Public Health Reports* 116: 344–352. http://repository.upenn.edu/spp_papers/85.

Montgomery, Ann Elizabeth, Jamison D. Fargo, Thomas H. Byrne, Vincent R. Kane, and Dennis P. Culhane. 2013. "Universal Screening for Homelessness and Risk for Homelessness in the Veterans Health Administration," *American Journal of Public Health* 103: S210–S211. DOI: 10.2105/AJPH.2013.301398.

Moturu, Sai T., William G. Johnson, and Huan Liu. 2010. "Predicting Future High-Cost Patients: A Real-World Risk Modeling Application," *International Journal of Biomedical Engineering and Technology* 3: 114–132. DOI: 10.1504/IJBET.2010.029654.

Poulin, Stephen R., Marcella Maguire, Stephen Metraux, and Dennis P. Culhane. 2010. "Service Use and Costs for Persons Experiencing Chronic Homelessness in Philadelphia: A Population-Based Study," *Psychiatric Services* 61: 1093–1098. DOI: 10.1176/ps.2010.61.11.1093.

Rog, Debra J., Tina Marshall, Richard H. Dougherty, Preethy George, Allen S. Daniels, Sushmita Shoma Ghose, and Miriam E. Delphin-Rittmon. 2014. "Permanent Supportive Housing: Assessing the Evidence," *Psychiatric Services* 65: 287–294. DOI: 10.1176/appi.ps.201300261.

Rosenheck, Robert. 2000. "Cost-Effectiveness of Services for Mentally Ill Homeless People: The Application of Research to Policy and Practice," *American Journal of Psychiatry* 157: 1563–1570. DOI: 10.1176/appi.ajp.157.10.1563.

Sadowski, Laura S., Romina A. Kee, Tyler J. VanderWeele, and David Buchanan. 2009. "Effect of a Housing and Case Management Program on Emergency Department Visits and Hospitalizations Among Chronically Ill Homeless Adults: A Randomized Trial," *Journal of the American Medical Association* 301: 1771–1778. DOI: 10.1001/jama.2009.561.

Sarma, Kattamuri S. 2013. *Predictive Modeling With SAS Enterprise Miner: Practical Solutions for Business Applications*. Cary, NC: SAS Institute.

SAS. 2013. *Getting Started With SAS Enterprise Miner* 13.1. Cary, NC: SAS Institute.

Shinn, Marybeth, Jim Baumohl, and Kim Hopper. 2001. "The Prevention of Homelessness Revisited," *Analyses of Social Issues and Public Policy* 1: 95–127. DOI: 10.1111/1530-2415.00006.

Shinn, Marybeth, Andrew L. Greer, Jay Bainbridge, Jonathan Kwon, and Sara Zuiderveen. 2013. "Efficient Targeting of Homelessness Prevention Services for Families," *American Journal of Public Health* 103: S324–S330. DOI: 10.2105/AJPH.2013.301468.

Tamang, Suzanne, Arnold Milstein, Henrik Toft Sørensen, Lars Pedersen, Lester Mackey, Jean-Raymond Betterton, Lucas Janson, and Nigam Shah. 2015. *Improving the Foundation of Population-Based Spending Arrangements by Predicting 'Cost Blooms' in Denmark: A Longitudinal Population-Based Study*. Palo Alto, CA: Stanford University. http://statweb.stanford.edu/~ljanson/papers/Predicting_Patient_Cost_Blooms_In_Denmark-Tamang_ea-2016.pdf.

Toros, Halil, and Max Stevens. 2012. *Project 50: The Cost Effectiveness of the Permanent Supportive Housing Model in the Skid Row Section of Los Angeles County*. Los Angeles: County of Los Angeles, CEO.

Tsemberis, Sam, and Ronda F. Eisenberg. 2000. "Pathways to Housing: Supported Housing for Street-Dwelling Homeless Individuals With Psychiatric Disabilities," *Psychiatric Services* 51: 487–493. DOI: 10.1176/appi.ps.51.4.487.

U.S. Department of Housing and Urban Development (HUD). 2016. *The 2016 Annual Homeless Assessment Report (AHAR) to Congress. Part 1: Point-in-Time Estimates of Homelessness*. Report prepared for Congress. Washington, DC: U.S. Department of Housing and Urban Development. https://www.hudexchange.info/resources/documents/2016-AHAR-Part-1.pdf.

———. 2017. *The 2017 Annual Homeless Assessment Report (AHAR) to Congress. Part 1: Point-in-Time Estimates of Homelessness*. Report prepared for Congress. Washington, DC: U.S. Department of Housing and Urban Development. https://www.hudexchange.info/resources/documents/2017-AHAR-Part-1.pdf.

———. 2016. *The 2016 Annual Homeless Assessment Report (AHAR) to Congress. Part 2: Estimates of Homelessness in the United States*. Report prepared for Congress. Washington, DC: U.S. Department of Housing and Urban Development. https://www.hudexchange.info/resources/documents/2016-AHAR-Part-2.pdf.

U.S. Interagency Council on Homelessness (USICH). 2016 *The President's 2016 Budget: Fact Sheet on Homelessness Assistance*. Washington, DC. https://www.usich.gov/resources/uploads/asset_library/2016_Budget_Fact_Sheet_on_Homelessness_Assistance.pdf.

———. 2015. *Opening Doors: Federal Strategic Plan to Prevent and End Homelessness*. Washington, DC. https://www.usich.gov/opening-doors.

———. 2010. *Opening Doors: Federal Strategic Plan To Prevent and End Homelessness*. Washington, DC.

Zugazaga, Carole. 2004. "Stressful Life Event Experiences of Homeless Adults: A Comparison of Single Men, Single Women, and Women With Children," *Journal of Community Psychology* 32: 643–654. DOI: 10.1002/jcop.20025.