Model-Based Triage of Permanent Supportive Housing in Santa Clara County

Jane Carlen, Halil Toros, Daniel Flaming

ECONOMIC ROUNDTABLE

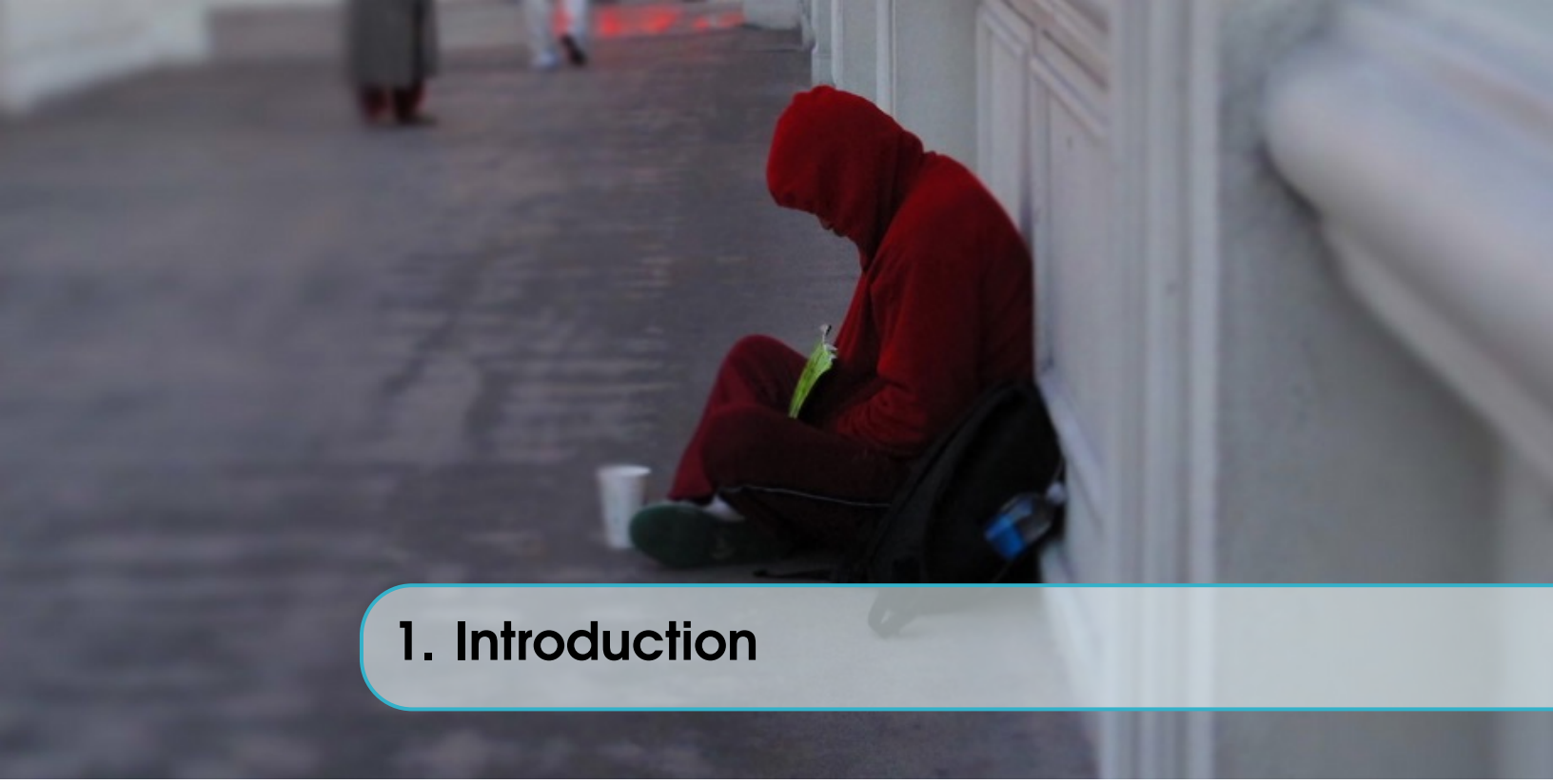
This report has been prepared by the Economic Roundtable, which assumes all responsibility for its contents. Data, interpretations and conclusions contained in this report are not necessarily those of any other organization that supported or assisted this project.

This report can be downloaded from the Economic Roundtable website: http://www.economicrt.org

Underwritten by the County of Santa Clara Office of Supportive Housing

First release, July 2018

Contents



1. Introduction

1.1 Motivation

Individuals experiencing homelessness bear an immense cost, as do public systems that must allocate limited resources to meet growing demands. The need for services is not evenly spread among individuals however, and recent evidence from Santa Clara County, California suggests that five percent of homeless people account for 47 percent of all public spending on the homeless (Flaming, Toros, and Burns, 2015). These individuals tend to have multiple obstacles to finding and retaining housing, such as physical and mental health problems, criminal justice system involvement, and social isolation.

A large body of research indicates that moving into permanent supportive housing (PSH) is a key and often primary step for stabilizing the lives of individuals who experience homelessness and are frequent users of public services. However, it is an expensive form of intervention, and demand exceeds supply. Thus, it must be allocated strategically to achieve the greatest benefit. By drawing on data from individual service records we identify those with the highest public costs so that they can be prioritized for access to PSH. This provides immense benefit to those who receive housing, and can generate significant reductions in service usage, making more resources available for the broader population.

1.2 Objective

To improve the allocation of supportive housing resources in Santa Clara County we have designed a predictive model of public spending on homeless individuals in the county. Our model predicts whether an individual will be in the top decile in terms of public cost in the coming year based on their service history and demographics. We built on our earlier work designing a model-based triage tool described in *Identifying and Housing High-Cost Homeless Residents* (2016), which used public agency service records to predict future high costs associated with homelessness. Our results are derived from de-identified linked administrative records from public agencies in Santa Clara County from 2007 to 2012.

The development of an updated triage tool was initiated by the Santa Clara County Center for Population Health Improvement (CPHI). They sought to incorporate an empirical tool for PSH triage directly into their information system, but needed it to be tailored to their available data. They now face data limitations that did not apply when we first developed a triage tool for Santa Clara County. In addition, an industry-wide switch in medical diagnostic classification systems over the last few years necessitated a conversion of all diagnostic variables previously incorporated. Making the triage tool work within the CPHI information system is critical to its adoption and evaluation.

The triage tool presented in this report reflects current restrictions on available data, up-to-date diagnostic coding, and newly incorporated predictive methods. The underlying model remains strongly predictive, despite additional data restrictions.

1.3 Background

As the homeless population in the U.S. has grown over the last three decades, the response at the local and federal level has shifted from short-term responses such as emergency shelters to long-term housing assistance. A leading school of thought encouraging this shift is Housing First, which promotes the idea that housing should be the first step in assisting chronically homeless individuals. Pathways to Housing, the New York City nonprofit that pioneered this model, also advocates empowering consumers to make choices about housing and services (Greenwood, Stefancic, and Tsemberis, 2013).

Housing provides stability that makes it easier to address other issues. For example, food security and child well-being improve as a result of permanent housing subsidies (Gubits et al., 2016). PSH adds a layer of service provision to permanent housing, including such elements as mental and physical healthcare services and substance abuse treatment. Studies have shown it decreases psychiatric and medical inpatient hospitalization, emergency room visits, substance abuse treatment costs, and criminal justice system interactions (Culhane and Byrne, 2010; Henwood et al., 2015; Ly and Latimer, 2015; Martinez and Burt, 2006; Shinn et al., 2013; Toros and Stevens, 2012). PSH as part of a Housing First approach can effectively reduce chronic homelessness (Byrne et al., 2014; Culhane, Metraux, and Hadley, 2002; Greenwood, Stefancic, and Tsemberis, 2013; Rog et al., 2014; Tsemberis and Eisenberg, 2000). In contrast, demanding that chronically homeless individuals address mental health and addiction issues before being eligible for housing puts those most in need in a frustrating catch-22.

Because of its comprehensive nature, PSH is expensive and chronically under-supplied. To achieve the greatest benefit with limited resources, it is crucial that allocation of PSH should be focused on individuals with several impediments to finding and maintaining housing on their own. Many of these individuals already generate significant costs for local service providers and agencies. Annual costs incurred by chronically homeless individuals may average in the tens of thousands of dollars due to heavy use of acute and behavioral healthcare, criminal justice involvement, shelter stays and social services. PSH is most likely to generate cost offsets when targeted to individuals in the upper tail of the distribution of public costs (Culhane, 2008; Poulin et al., 2010). Previous studies in Los Angeles have found that PSH is ultimately less expensive than not providing housing for homeless individuals in the top ten percent of costs (Flaming, Toros, and Burns, 2015).

In this report we describe the process and outcomes of building an updated triage tool for Santa Clara County. In Chapter 2 we discuss the data available for model building and how it was processed, modelling techniques, and the implementation of the resulting triage tool. In Chapter 3 we examine the predictive power of the model and consider descriptive takeaways from our final model and

competing methods. In Chapter 4 we evaluate the expected cost savings from implementing the triage tool. We conclude with a discussion of this study and future work in Chapter 5.

2. Methods

2.1 Data

Our complete data consists of de-identified linked administrative records for seven agencies in Santa Clara County from 2007 to 2012, but data from only four of these agencies is available to CPHI to implement the triage tool. The agencies providing data are the county's Continuum of Care, Criminal Justice Information Control (CJIC) system, Mental Health Department, and Valley Medical Center (VMC) Hospital and Clinics. Variables from Emergency Medical Services, the Department of Alcohol and Drug Services, and the Social Services Agency used in our original screening tool were excluded in building this tool because they will not be available during implementation. Even so, the variables included in the available data set provide extensive information in the following areas: demographics, medical and healthcare history, and interactions with the criminal justice system.

We began with a pool of 57,259 individual records of those who utilized a homeless service from at least one of the seven agencies listed above between 2007 and 2012. Santa Clara County created this rich data set through a record-linkage process.[1] The 57,259 records are a subset of over one-hundred thousand records of county residents who experienced at least a month of homelessness in Santa Clara County from 2007 to 2012. The subsetting criteria ensured the integrity of the record linkages, following some linkage problems with agency databases (Toros and Flaming, 2016).

Although many of these records do not present persistent homelessness, the chance of unrecorded stints of homelessness or return to homelessness is high, especially among those classified as high-risk by our model. In addition, individuals admitted to private hospitals, state psychiatric facilities or prisons are not documented as homeless during those stays. As we describe below, individuals with the greatest chance of becoming high-cost exhibit a combination of factors that are very likely to cause ongoing housing instability.

Our outcome variable of interest is whether an individual in our records will be in the top decile of costs in the next year. To make and test our predictions we divided our data into a training and validation set. The training set is the data that we used to assign importance (coefficients) to

[1] Obtaining linked agency records is a difficult and protracted process. Due to the time constraints of this study, we did not seek to expand our database of linked records.

individual predictors. The validation set is the data we used to check the out-of-sample predictive power of our model. Evaluating our model on the validation data set is one way we ensured that we were not overfitting our data. Overfitting amounts to reducing error on the training data at the cost of increasing error when we apply the model to new data. The training and validation sets must be separate if the validation set is to provide an external check on the strength of our model.

As with our original model, we used a two year past window to predict cost in the following year. This provides a balance between including enough service history to build predictive power, and providing accurate input information for individuals who may not have extensive service histories. As the length of the training window grows, so do the administrative costs of collecting data and the potential for inconsistencies due to changes in record-keeping methodology or local policy.

For our training data set, we used records from 2007-2008 to predict costs in 2009. Our validation set used records from 2010-2011 to predict costs in 2012. Unlike our approach for the original model, we incorporated a second training set, using records from 2008-2009 to predict cost in 2010. Including this second training set influences variable selection and improves predictive power. Our model will be more consistent in future applications if our variables display consistent coefficients from year to year. Thus, we only included in our final model predictors that were detected in both training models. We found that models constructed in this way made more sense in context than those incorporating all predictors from both training models. We also tailored our transformations of covariates toward consistency in effects over the two training sets. Due to correlation between records in each of the training sets, we combined our training models by averaging the outputs of the two models.

The variables in our data provide a wide range of information about individuals. This includes demographic information such as age, gender and ethnicity; health information such as ICD-9 (International Classification of Diseases, Ninth Revision) medical diagnoses (converted to ICD-10 for implementation), inpatient and outpatient medical care history, including mental health treatment; and criminal justice history including arrests, time incarcerated and special services while in custody.

At the outset our records included over 250 unique variables describing individuals. This reduced variable set was rigorously constructed after eliminating variables with no relationship to the outcome variable, and aggregating others (see Toros and Flaming, 2016; Toros and Flaming, 2018). For example, from roughly one thousand diagnostic codes we extracted six meaningful categories: adjustment reaction, organ failure, heart disease, schizophrenia, neoplasm, and other causes of morbidity and mortality. Other diagnostic groupings were of chronic medical conditions and high-cost conditions.

Several variables that were statistically significant in our original model were excluded *a priori* from this modelling process because they will not be available to Santa Clara County when implementing this new triage tool. They include the number of Emergency Medical Service (EMS) encounters, number of drug abuse and alcohol service encounters, number of mental health outpatient days, and whether an individual received public assistance benefits. Whether an individual received food stamps for at least two months of the previous two years was also excluded. Interestingly, this was the only downward predictor other than age in the original model as it is correlated with stability in the recipient's living situation.

One of our objectives in designing this tool was to recover as much predictive power as possible from available variables to account for these restrictions. In some cases, overlapping information from other agencies can compensate for these losses. For example, medical diagnoses or justice system charges indicative of substance abuse can compensate for the loss of information about substance abuse treatment.

Another consideration in model development was that during implementation certain health information will be fully available from Santa Clara's public Valley Medical Center, but only partially available for other hospitals using claims data. Data originating from claims will only be available for clients covered by Valley Health Plan (VHP) insurance, estimated to be about 60 percent of the target population. As a result, we sought to use indicator (binary) variables instead of counts as much as possible to capture health variables affected by these restrictions. We did this to avoid introducing bias based on an individual's type of insurance.

More generally we considered various transformations of the available variables, including binning continuous variables, clustering categorical variables, and generating indicator and count variables. We also eliminated or combined highly correlated variables. The transformations included in our final model differ somewhat from those our original model due to the additional data restrictions, and desire for consistency between our two training models.

2.2 Model

The core of the triage tool is a predictive model of public spending on homeless individuals, predicting whether they will be in the most costly decile ("high-cost") based on person-level characteristics previously observed. We considered several classes of interpretable models to predict high-cost status, including logistic regression, penalized logistic regression and decision trees. Ultimately, we elected to use the penalized regression method LASSO (least absolute shrinkage and selection operator) to generate our final model. LASSO jointly performs variable selection and coefficient estimation, but the interpretation of the coefficients is akin to a standard logistic regression model. We found that models fit with LASSO performed on par with un-penalized logistic regression models but with fewer input variables. Both of these model types outperformed decision trees.

It is important that our model not only have predictive power, but also explanatory power, indicating the relative importance of the predictors. Parsimony in the model is also valuable for clarity of interpretation and to limit the administrative burden of implementing it. For these reasons, we did not use machine learning techniques to build the tool. Though they generally deliver higher predictive power, it is at the expense of interpretability and simplicity. In Section 3.3 we compare our model output with one such technique, random forests. We also discuss insights gained from other models that were considered but not ultimately employed. They may be called upon in future work.

Modelling was conducted in `R` version 3.5.0, and LASSO was implemented using the `glmnet` package (Friedman, Hastie, and Tibshirani, 2010). As an immediate protection against overfitting we used 10-fold cross-validation when fitting our models. Although our validation data set helps to check for overfitting, it is limited in this capacity by the fact that there are overlapping individuals in the training and validation sets. Cross-validation during model fitting *prevents* overfitting by creating many subsets of mutually exclusive training and testing data. These subsets are sampled randomly, which leads to small fluctuations in the resulting estimates. To reduce these to a negligible level we repeated the cross-validated estimation process 100 times for each training set, and averaged the resulting estimates.

Cross-validated LASSO outputs a model that minimizes error on the testing sets given a certain value of a penalty term in the model. The penalty term is selected by the user and controls the number of predictors. In general, we elected to use a penalty term that was one standard deviation higher than the value that minimized error overall. We found that this made the resulting model simpler and the interpretation more reasonable in context, without sacrificing significant predictive power.

Our primary tool for comparing model fits was the receiver operating characteristic (ROC) curve, and the area under this curve (AUC). The ROC curve visualizes the ability of the model to capture true positives without including false positives, and the AUC quantifies this ability. In our context, the ROC curve plots the percentage of individuals predicted to be high-cost who truly are (true positive rate) against the percentage of individuals predicted to be high-cost who are not (false positive rate).

Each individual in our study is assigned a predictive probability of being in the top decile of costs in the next year. Implementation of the triage tool therefore demands picking a cutoff value above which individuals would be prioritized for PSH. The higher this cutoff is, the more likely it is that we only provide PSH to individuals who will exhibit high costs (true positive rate increases). The lower it is, the more likely we are to service more individuals who will exhibit high costs (false negative rate decreases), but at the cost of also servicing individuals for whom it is not cost-effective (false positive rate increases). The ROC curve visualizes how quickly the true positive and false positive rates increase as the cutoff decreases. The choice of cutoff is examined more closely in Chapters 3 and 4, and related to estimated costs of PSH.

2.3 Implementation

The triage tool developed by the Economic Roundtable will be implemented by the County of Santa Clara Center for Population Health Improvement (CPHI). Santa Clara County identifies individuals as homeless based on indicators of homelessness in the Homeless Management Information System (HMIS) or medical records. CPHI is developing a process for ensuring client consent to the linkage of client records needed to implement the tool, in accordance with medical privacy regulations. At this time of writing, CPHI was also verifying the availability of all criminal justice data employed in the triage tool model.

The triage tool will be used in combination with the Vulnerability Index Service Prioritization Decision Assistance Tool (VI-SPDAT). The VI-SPDAT assigns a score to an individual based on an in-depth interview. In the case that an individual does not have any service records or does not consent to sharing them, VI-SPDAT may be the only tool available to assess the acuity of an individual's housing needs with some level of objectivity. However, it is still quite subjective. Variation arises from qualities of the interviewer, and whether the interviewee is willing and able to provide complete information, especially in heavily stigmatized areas such as mental health and drug addiction. There is also the risk that individuals may provide inaccurate or exaggerated responses that improve their prospects for obtaining housing. The Economic Roundtable's triage tool provides a more objective predictive metric because it is based on service records. During implementation, the relative efficiency of the Economic Roundtable's triage tool and the VI-SPDAT will be evaluated.

3. Results

3.1 Model output

The final predictive model is summarized in Table 3.1. The coefficients are presented on the exponential scale, such that each value represents the multiplicative change in the odds[1] of an individual being high-cost for each additional unit of the predictor value, holding all other predictor values constant. In most cases, the predictors are indicator variables that take either the value of zero or one. For example, individuals who have had at least one arrest in the year leading up to prediction (referred to as "this year" in the table) have odds of being high cost that are 1.98 times the odds for an individual who was not arrested, all else being equal.

Whether an individual was arrested this year (1.98) and whether they served more than three months in jail (2.12) are the most powerful predictors in the model. Among the main effects, i.e., not including interaction effects between two variables, other top predictors are whether an individual had at least ten mental health inpatient admissions in both of the last two years (1.72) and whether they had more than ten non-inpatient health system encounters this year (1.69).

Other strong predictors are whether substance abuse was indicated by a medical diagnosis or justice system charge (1.68), whether an individual was housed in a jail medical facility this year (1.67), and whether they were jailed with a security classification of three or four - the top two security classifications (1.52). Most medical diagnostic factors have mild to moderate effects. The only negative predictor, i.e. a factor which indicates decreased chance of being in the highest cost decile, is for an individual to be over 45 years of age.

The interaction effects in the final predictive model are summarized in Table 3.2. By far the strongest interaction effect pertains to the inpatient hospital admissions an individual has in the year leading up to prediction (1.86). The interaction effects generally combine two factors of the same type, such as two health factors, allowing the model to capture the severity of an individual's issues in that area. The final factor in the table combines chronic homelessness and substance abuse and is more descriptive of the overlapping issues that lead to high service utilization.

The overall performance of the model is quantified using the AUC, also known as the concordance

[1] The odds of an outcome x is the ratio of the probability of the outcome to the probability of it not occurring, or $\frac{p(x)}{1-p(x)}$.

Table 3.1: Predictive variables in the triage tool model. Coefficients are presented with their exponentiated value, indicating multiplicative impact on the odds of being in the top decile of costs. If no time frame is specified a window of the previous two years is assumed.

Variable	**Odds Ratio**
Demographic and HMIS records	
Older than 45	0.86
Chronic homelessness flag in HMIS record	1.27
Criminal justice	
Arrested this year	1.98
One to three months of jail this year	1.49
More than three months of jail this year	2.12
Jail booking	1.23
Jailed with security level three or four	1.52
Jailed in medical facility type this year	1.67
Jailed in mental health facility type this year	1.24
Jailed in other facility type this year	1.27
Arrested for inebriation and released with 48 hours	1.27
At least 500 days probation indicated this year	1.36
At least 500 days probation indicated last year	1.18
Health diagnoses	
Diagnosis of schizophrenia	1.33
Diagnosis of heart disease	1.32
Diagnosis of other cause of morbidity/mortality	1.26
Diagnosis of organ failure	1.22
Diagnosis of chronic medical condition	1.18
Diagnosis of disease of the genitourinary system	1.12
Diagnosis of alcohol or drug dependence	1.03
High-cost diagnosis this year	1.20
High-cost diagnosis this and last year	1.14
Health services	
More than ten non-inpatient health system encounters this year	1.69
VMC non-inpatient health system encounter this year	1.20
Inpatient visit via ER or a psychiatric facility this year	1.40
Surgery or acute inpatient psychiatric hospitalization	1.14
Outpatient ER, psychiatric service or ambulatory surgery this year	1.07
Behavioral health	
Mental health inpatient admissions this year (at least 10)	1.38
Mental health inpatient admissions this and last year (at least 10 each year)	1.72
Mental health outpatient treatment	1.33
Substance abuse indicated by medical diagnosis or justice system charge	1.68

Table 3.2: Predictive interaction terms in the triage tool model.

Variable	Odds Ratio
Interaction effects	
Inpatient visit via ER this year and at least two inpatient visits this year	1.86
Jailed with sec. level 3-4 and more than three months of jail this year	1.33
Diagnosis of chronic condition and inpatient visit via ER this year	1.07
Diagnosis of chronic condition and inpatient visit via ER last year	1.34
Diagnosis of chronic condition and outpatient via ER this year	1.21
Diagnosis of chronic condition and schizophrenia	1.10
Outpatient ER/psych/amb. this year and diagnosis of factor influencing health e.g., disease carrier, disease contact, vaccination	1.15
Outpatient ER/psych/amb. this year and 10+ non-inp. health encounters this year	1.05
Chronic homelessness (HMIS) and substance abuse	1.14

statistic, C-statistic or C-index. It is equivalent to the probability that a random individual who became a high-cost individual had a higher probability assigned to them under the model than a non-high cost individual. The ROC curve for the model is shown in Figure 3.4, compared with a random forest curve and a baseline representing random prediction. The baseline C-statistic is 0.5. The C-statistic of our model is 0.8, which is slightly lower than the C-statistic of 0.82 obtained by the original model when applied to data excluding minors. (Without excluding minors the C-statistic of the original model is 0.83.) This is expected given the data restrictions that did not apply previously. A model with a C-statistic of at least 0.8 is considered strong.

To reduce out-of-sample error, we averaged models of two training data sets and sought consistency between the two models, as described above. Figure 3.1 depicts the similarity of the two training models including only main effects with coefficients ordered by average magnitude. Only a few major deviations appear. In terms of odds ratio, the importance of being jailed with security level three or four is about 0.5 higher in the second model, and being arrested for inebriation is about 0.25 higher in the first model. Otherwise, the deviations are generally in the range of 0 to 0.2. This is to be expected given natural stochasticity in the service needs of individuals over time. The overall agreement shows that our final model is capturing meaningful relationships in our data.

3.2 Predictive power

To verify the performance of our model we applied it to our validation data set using data from 2010 and 2011 to predict costs in 2012. The results are summarized in Table 3.3.

Table 3.3 displays the predictive power of the model and its relationship to a cutoff value for the model-assigned probability of becoming high-cost. We present cutoffs of the top one, five, and ten percent of probability scores. We also consider the top 1000 individuals, which is roughly equivalent to the top two percent. The metrics we use to quantify predictive power are sensitivity, specificity, positive predictive value (PPV) and accuracy. They are defined as follows:

- $Sensitivity = \frac{True\ Positive}{True\ Positive + False\ Negative}$; percent of correctly labeled high-cost individuals.
- $Specificity = \frac{True\ Negative}{False\ Positive + True\ Negative}$; percent of non-high cost individuals labelled as such.
- $PPV = \frac{True\ Positive}{True\ Positive + False\ Positive}$; percent of individuals labelled high-cost that actually are.

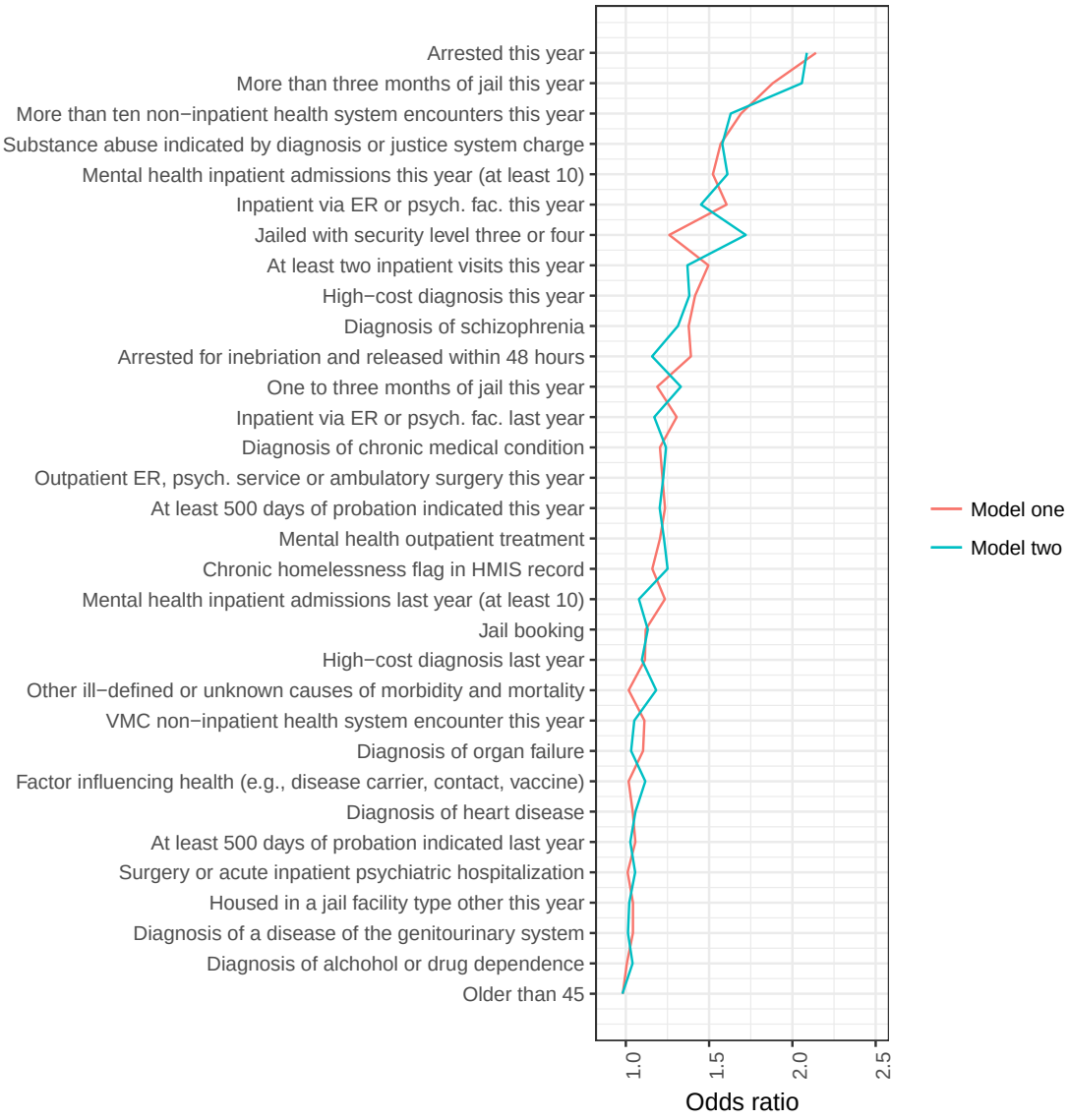


Figure 3.1: Comparison of coefficient importance in models of two training data sets. Model one uses data in 2007-2008 to predict high-cost status in 2009, and model two uses data in 2008-2009 to predict high-cost status in 2010.

- $Accuracy = \frac{True\ Positive + True\ Negative}{Total\ Cases}$; percent of correctly labeled individuals.

Because our ultimate goal is PSH triage, we are most interested in the PPV and sensitivity. High PPV leads to large costs savings. The PPV is very good for the top one percent - over two-thirds of that group is in fact part of the highest-cost decile in 2012. PPV tails off for lower cutoffs, but not so much that cost-savings are eliminated, as we find in Chapter 4.

High sensitivity indicates that a large portion of people in need of PSH will receive it, though that is limited by the available housing stock. For the top ten percent cutoff, nearly half of those truly in the top decile of costs would be assigned PSH by our model. Because the positive and negative

Table 3.3: Predictive Performance of the Model (Interactions)

	Top 1%	Top 5%	Top 10%	Top 1000
Sensitivity	8.4	28.3	43.0	14.2
Specificity	99.7	97.1	93.0	99.2
PPV	69.7	46.8	35.5	60.9
Accuracy	92.1	91.4	88.8	92.2

prediction groups are imbalanced in size, the specificity and accuracy are guaranteed to be high. Still, the specificity values well over ninety percent indicate that our model also does well at the easier task of classifying true negatives.

3.3 Descriptive takeaways

We considered and compared many predictive techniques in building the triage tool. Although we ultimately used a single penalized regression model as the predictive engine of the tool, other models offered insights into the factors that lead to high costs. In this section we consider some key takeaways from the modelling process.

We applied the same technique of penalized regression as above to separately analyze records by gender and age group (18-24, 25-45, and over 45). We did not find that this improved overall predictive power, but the models for each age group did provide insights into how factors change and build up over time. For comparison, we created subsets of the wider age groups (18-24, 35-40, and 55+) to gain more separation between the age groups and balance the number of records in each age-specific model. The top ten coefficients for each group based on models without interaction effects are shown in Figure 3.2.

Interactions with the criminal justice system are universally important. Whether an individual was arrested in the previous year is the second-strongest predictor for the 18-24 age group, and the top predictor for the other two groups. For the 18-24 age group criminal justice system interactions are especially important, making up three of the top four predictors. Whether an individual served at least three months in prison in the previous year is by far the strongest predictor for young adults, and the third-strongest predictor for adults age 35-40, but not a predictor for adults in the oldest age group.

Mental health issues appear in all models, though in slightly different forms. For young adults, mental health inpatient and outpatient treatment are both among the top ten predictors. The models for adults age 35-40 and 55+ include the more specific diagnosis of schizophrenia, a condition that may emerge after young-adulthood. (For the oldest age group it is just outside the top ten.) Substance abuse as indicated by a diagnosis or justice system charge is also consistently predictive across all age groups.

While mental health treatment is more prominent in the model of younger adults, usage of inpatient health services is more important for older adults. Whether an individual had an inpatient admission via an ER or psychiatric facility in the previous year is a strong predictor for the two older age groups. Three of the top ten variables in the oldest age group model are related to inpatient care, but no such predictors are included in the model of the youngest age group. The age-specific models show how health issues accumulate over time. Older individuals are less able to resist the damage of accumulated health problems, which therefore require more intense care.

Another predictive strategy that we considered is the decision tree. Although the performance of

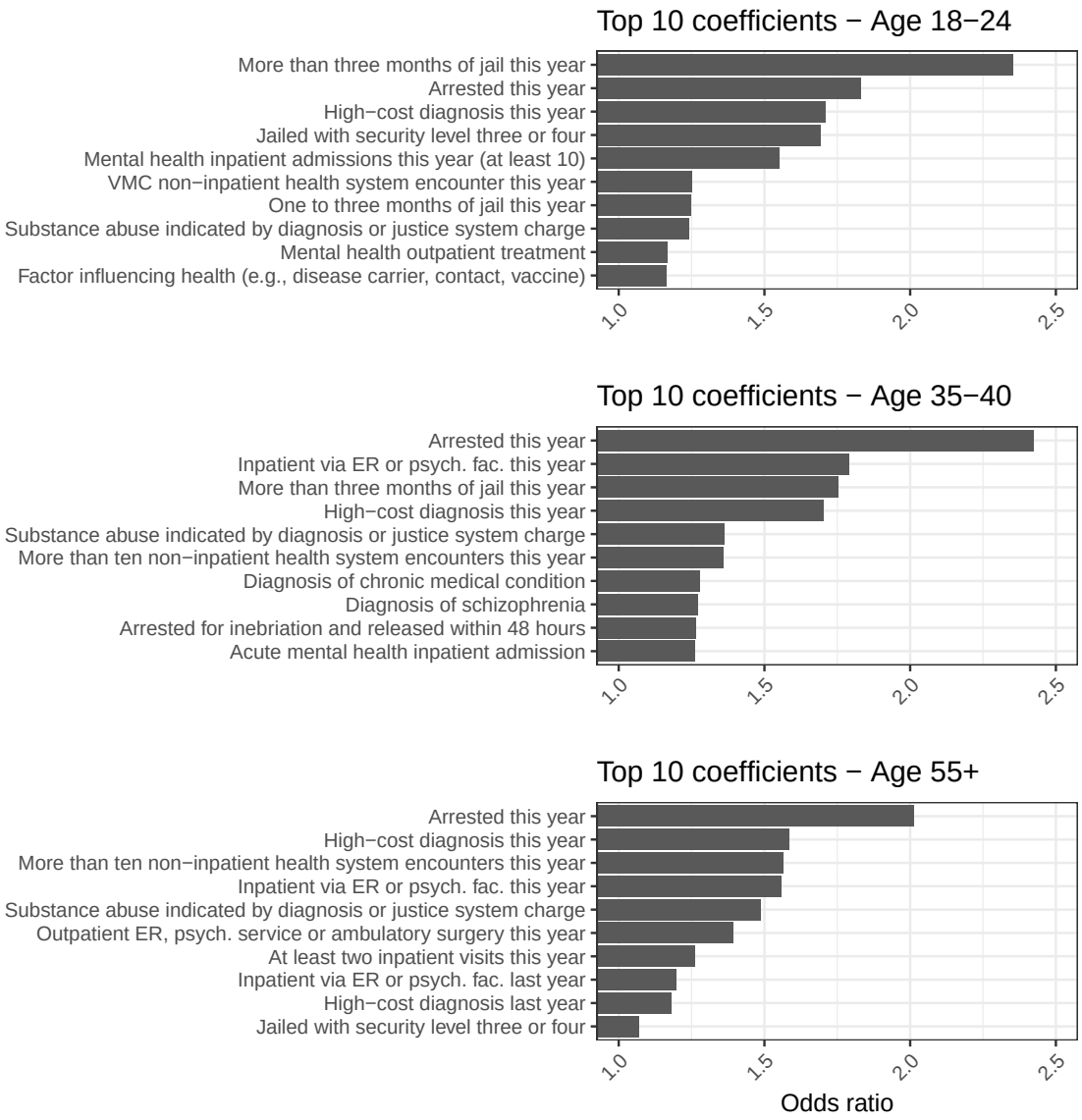


Figure 3.2: The top 10 coefficients for models of three age groups.

the decision tree is worse than regression models in terms of prediction, its structure is well-suited to visualize relationships between variables. Decision tree models iteratively divide records into groups depending on their risk of being in the high-cost group.

Figure 3.3, presents a fairly simple tree model of our first training data set. We created this in `R` using the packages `rpart` and `rpart.plot` (Milborrow, 2017; Therneau and Atkinson, 2018). It shows that there are two main types of individuals who are likely to be high-cost. The further-left paths are largely dependent on health conditions. They consists of individuals who were not arrested in the previous year. Within those paths individuals with a high-cost diagnosis and an inpatient admission via an ER or psychiatric facility are most likely to be high-cost.

The further-right paths consist of individuals who were arrested in the previous year. Within those paths, individuals who also had a high-cost diagnosis are at highest risk of being high-cost. For those without a high-cost diagnosis, higher risk is associated with having spent at least three months in jail in the previous year, and may be compounded by substance abuse, mental health issues, or chronic homelessness. These left and right paths to high-risk status are somewhat correlated with older (left) and younger (right) segments of the population. We were able to detect this as a result of our age-specific models.

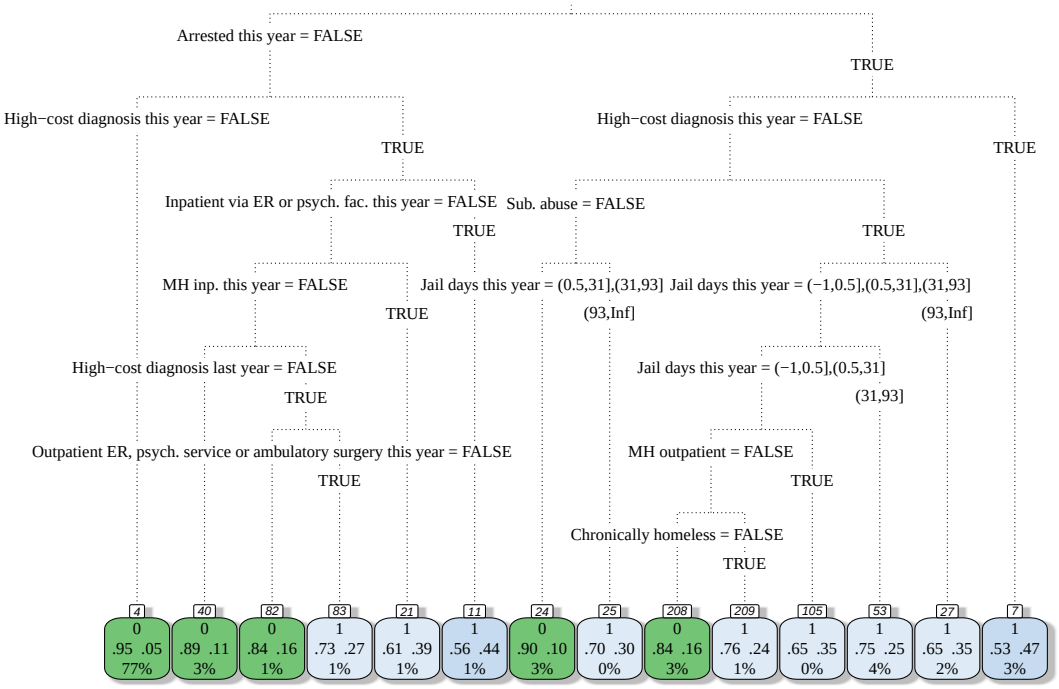


Figure 3.3: Decision tree model of the first training data set. Each node in the bottom row indicates if individuals described by the corresponding path are predicted to be low-cost (0, green) or high-cost (1, blue). The nodes also show the true split of low-cost and high-cost individuals in the node, and the overall percentage that reaches the node. MH indicates mental health.

Random forests are a type of machine learning algorithm that generalizes the notion of decision trees. Rather than using one complex tree to predict a certain outcome, random forests fit a large

number of simple trees. None of these trees does a good job of prediction on its own, but each is better than a random prediction, and they are aggregated using weights that depend on their predictive power. In this way, random forests are often compared to a decision-by-committee, in which certain votes carry more weight. We applied random forests using the package `randomForest` in `R` (Liaw and Wiener, 2002).

Figure 3.4 compares the ROC curves of the LASSO output described above with the averaged predictions from random forests on the two training sets. The random forest offers a very slight improvement over the LASSO at the low end of the curve, when false positive rates are in the practical range of less than 15 percent. Overall the ROC curve of the LASSO slightly outperforms the random forest model.

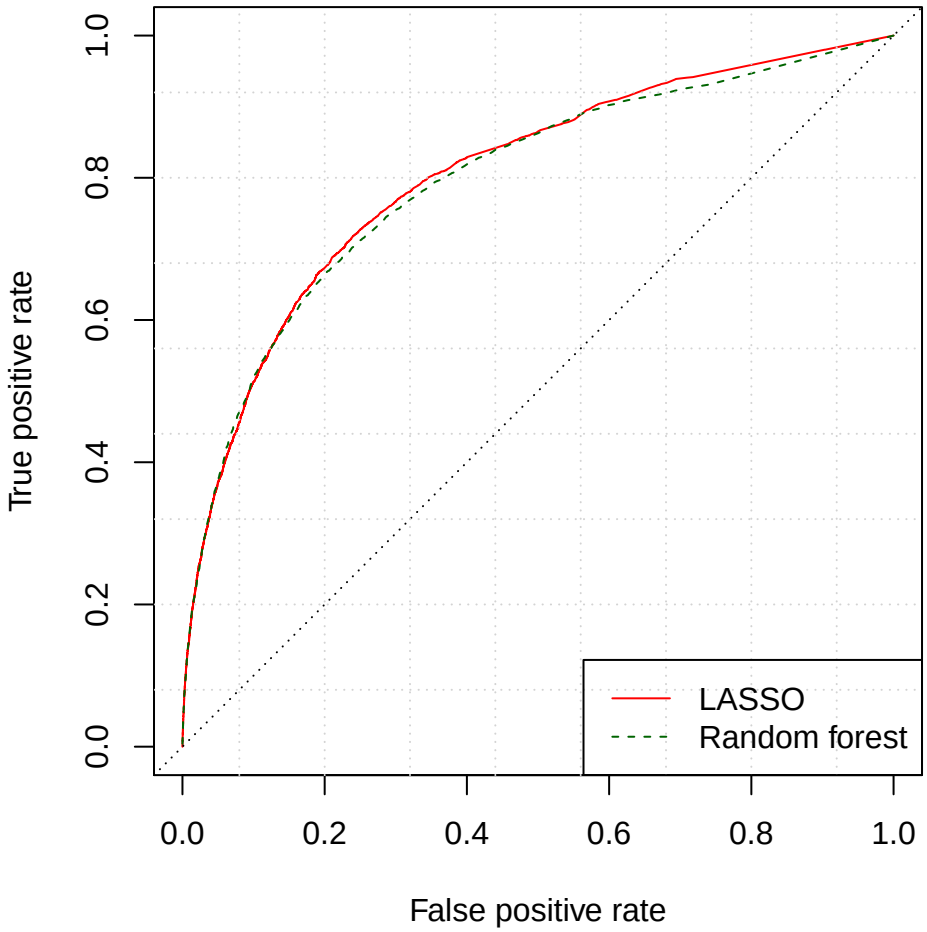


Figure 3.4: Comparison of ROC curves for the LASSO and random forest. The straight line represents a baseline model randomly assigning one in ten individuals to the highest cost decile.

Table 3.4 shows that the predictive performance of the random forest is slightly better than the LASSO overall. However, such small improvements do not justify the costs to interpretability of

using a random forest. Although techniques exist to estimate the relative importance of coefficients in a random forest, they lack the straightforward interpretation of regression coefficients. The random forest also incorporates far more predictors than the LASSO, which multiplies administrative cost.

Table 3.4: Predictive performance of the random forest

	Top 1%	Top 5%	Top 10%	Top 1000
Sensitivity	8.4	28.5	44.2	14.2
Specificity	99.7	97.1	93.1	99.2
PPV	69.3	47.0	36.5	61.5
Accuracy	92.1	91.4	89.0	92.2

4. Business Scenario and Cost Savings

4.1 Projected savings

While the benefit of PSH for high-need individuals has been documented many times over, as discussed in Chapter 1, we address here the cost-benefit analysis specific to the triage tool applied in Santa Clara County. We find that the overall savings from providing PSH to target populations identified by the triage tool are conservatively estimated to be in the range of 20 million dollars. (All estimates below are presented in 2017 dollars, the most recent year for which an annual inflation factor is available.)

To project costs, we must account for the expense of PSH, and the reduction in service use associated with a move into PSH. We assume that PSH has a flat cost of $18,500 per year, based on our original analysis. This combines a high-end estimate of $12,000 in annual rental subsidy (excluding an individual's contribution), and $6,500 for supportive services. These estimates draw on data from Los Angeles, since relevant longitudinal studies of costs associated with PSH are not yet available for Santa Clara County.

We follow our related study *Where We Sleep: The Costs of Housing and Homelessness in Los Angeles* (2009) in applying a 68 percent reduction in service costs for individuals in the top decile of costs placed in PSH. As in our earlier work, we conservatively assume no reduction in service costs for individuals outside the top decile. In fact, studies have shown cost reductions for individuals in the top half of the empirical cost distribution (*Crisis Indicator: Triage Tool for Identifying Homeless Adults in Crisis* 2011).

Table 4.1 presents a detailed projection of potential cost savings from applying the triage tool in the year after the year predicted under the model. (The values in this table use 2010 as the year predicted.) Two scenarios are presented. The first considers the top five percent of users in terms of estimated probability of being high-cost in the year predicted by the model. The second applies the more restrictive cutoff of the top 1000 individuals by probability scoring. Individuals who are predicted to be in the top decile of costs are the positives (assigned PSH), whereas all others are negatives. Those groups are further split into true and false groups depending on whether their predicted outcome matches their true outcome in the year predicted, and low- and high-cost groups

Table 4.1: Average costs by group in the year predicted, and expanded costs and savings for the following year. The top half of the table is for individuals in the top five percent of high-cost risk, and the bottom half is for the top 1000 individuals.

	Prediction Year: Cost	Following Year: Cost	Cost Savings	Net Savings	Total Savings	Cases
Cutoff level: Top 5 percent						
True positive-low-cost	74,980	11,138	0	-18,500	-5,494,500	297
True positive-high-cost	95,370	86,759	58,996	40,496	32,801,604	810
False positive-low-cost	10,893	8,065	0	-18,500	-11,137,000	602
False positive-high-cost	13,727	46,835	31,848	13,348	5,352,363	401
Total					21,522,467	2,110
Cutoff level: Top 1000						
True positive-low-cost	78,573	11,085	0	-18,500	-2,701,000	146
True positive-high-cost	99,576	91,395	62,149	43,649	21,867,939	501
False positive-low-cost	12,039	8,761	0	-18,500	-3,718,500	201
False positive-high-cost	14,666	46,101	31,349	12,849	1,953,026	152
Total					17,401,465	1,000

based on whether their service costs are in the top decile in either of the following two years.

The average service costs by group in the predicted year and following year are shown in columns one and two of Table 4.1. The corresponding average cost savings, which are $0 for low-cost users and 68 percent of service costs for high-costs users, are shown in column three. To calculate the net savings shown in column four, we subtract the annual cost of PSH, $18,500, from each value in column three. The total savings in column five equal the net savings multiplied by the number of cases shown in column six.

We use these scenarios to examine the importance of one-time costs spikes. In both scenarios, the average costs for the positive groups are highest in the year predicted. This is expected, since the model is designed to project risk in the year after a two-year service window. However, PSH is a long-term intervention, so we must verify that it is cost effective in subsequent years.

Following the predicted year, those with high costs break off into low and high-cost users in the subsequent years. In both scenarios, the number of positives who remain in the high-cost group is much larger than the number that revert to the mean and drop into the low-cost group. In addition, among the high-cost positives, the drop-off in average costs from the predicted year to the subsequent year is less than ten percent - even smaller than in our original model. This may be the result of

using cross-validation and an ensemble model to minimize out-of-sample error.

Furthermore, among the false-positives, about forty percent of the individuals are in the high-cost subgroup. They show a significant average cost increase in the year following prediction. This suggests that the factors leading those individuals to be false positives in the predicted year are meaningful and likely to result in high service use down the line. This also implies long-term cost savings from giving PSH to almost half of the "false-positive" individuals.

For both cutoffs depicted in Table 4.1, the total costs savings are sizable. We found that the break-even point for cost savings is about 6,000 users, or the top 14 percent. In other words, as long as we target the top 14 percent or fewer for PSH, we can expect to generate cost savings in the year after prediction. If the cost of PSH increases, we can expect to generate cost savings when targeting the top five percent of users as long as the average annual cost of PSH stays below twenty-eight thousand dollars per year. When targeting the top 1000 users, costs savings are generated as long as PSH costs less than about thirty-six thousand dollars per year. This is valuable information to direct the scope of future programs.

4.2 Costs over time

We have investigated the predictive power of our model and found it to be strong. A related question is how the cost patterns of the prediction groups vary over several years. To what extent do high-cost individuals continue to be high-cost, and to what extent do costs revert to the mean? Although we have seen some evidence of one-time spikes in user costs, our objective is to identify positive prediction groups that maintain higher overall costs than the negative prediction groups.

Figure 4.1 illustrates these patterns for the years 2007 to 2012. The year predicted by the model is 2010 and positives are separated from negatives by a ten percent cutoff. The figure shows that the true positive group - individuals predicted and verified to be in the top ten percent in the year predicted - does in fact exhibit higher costs than all other groups over time. The false positive group exhibits higher costs than the negative groups in the years when data is collected for the model, but costs fall in the prediction year. The false negative group - those who exhibit costs in the top ten percent in the year predicted but were not predicted to do so - exhibit one-time cost spikes in the prediction year. Following the prediction year, the false-positives and false-negatives show increasingly similar costs. This indicates that these groups are fairly similar when they are not experiencing one-time cost spikes, which makes it difficult to distinguish their long-term behavior.

From Figure 4.1 three general classes of individuals emerge: those with consistently low costs, those with usually moderate costs but occasional cost spikes, and those with consistently high costs. Only the low-cost group experiences no cost spikes. The model effectively distinguishes these three groups into true negatives, false positives and negatives, and true positives, respectively. The separation of these groups even two years after the prediction year provides further evidence that the triage tool will produce long-term reductions in costs.

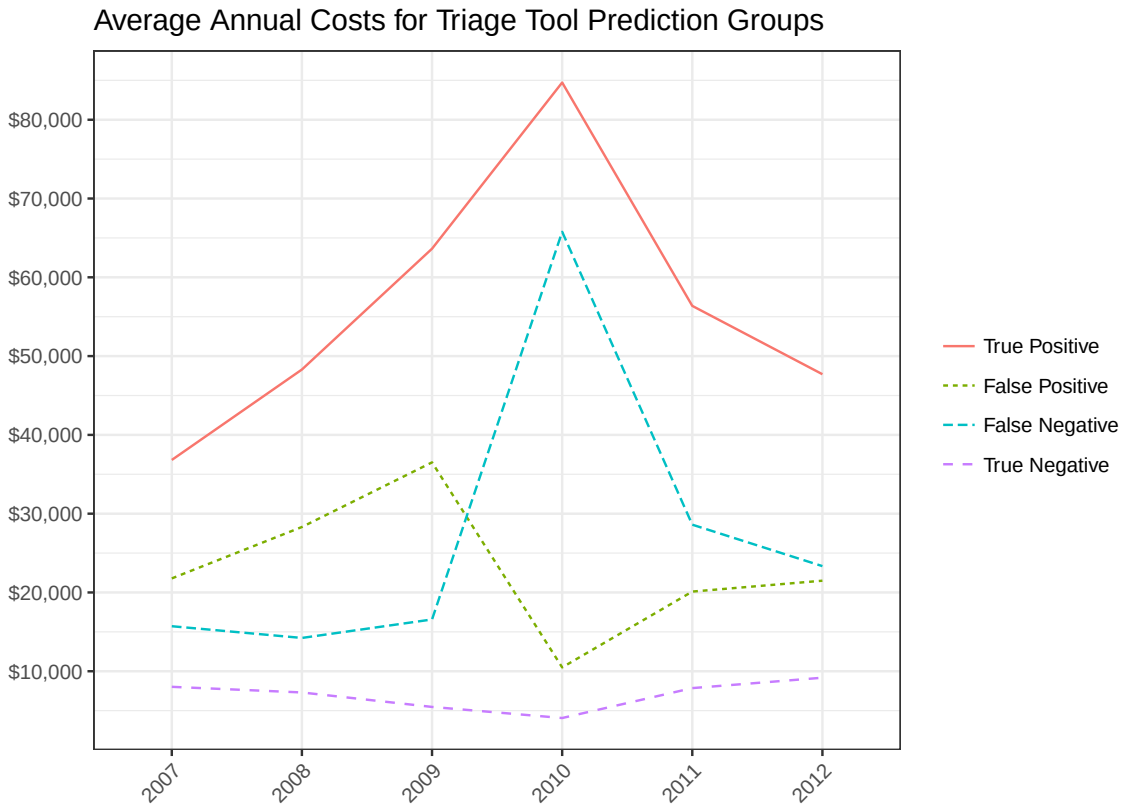


Figure 4.1: Average costs for prediction groups over time. Groups were created using a cutoff of the top ten percent in 2010.

5. Conclusion

5.1 Discussion and future work

In this report we have presented a triage tool that builds on our earlier work (Toros and Flaming, 2016; Toros and Flaming, 2018), while incorporating new techniques to reduce out-of-sample error, and responding to project-specific restrictions on available data. Despite the exclusion of several significant data fields used in our original model, we were able to maintain nearly all of the predictive power of that model, as verified on our validation data set. When fine-tuning our model we emphasized model simplicity, consistency between models of each training set, and the elimination or transformation of highly correlated variables

Our final model is an easily-interpretable logistic regression model which incorporates 31 predictive variables and 9 interaction effects. From this model we learn that the strongest predictors of being in the highest cost decile are whether an individual was arrested in the previous year, and whether they were in jail for at least three months. Also key were health variables indicating frequent receipt of inpatient, outpatient or mental health care.

We drew on other methods to help us further understand these predictive variables. By modeling our data in specific age groups we found that criminal justice and mental health variables are the most important predictors for young adults, while the importance of inpatient healthcare increases for older age groups. As individuals age, insufficiently treated health problems accumulate and are more likely to results in serious care in a hospital or emergency setting. Not only is PSH a cost-effective strategy to apply to individuals predicted to be in the highest-cost decile, our investigation reaffirms that early interventions are critical to preventing harmful outcomes for individuals. The onset of health problems for older age groups can potentially be prevented if housing and supportive services or other appropriate interventions are provided earlier in their course of homelessness.

The decision tree depicted in Figure 3.3 points to two general paths for entering the high-cost group. One depends only on high-cost medical diagnoses and medical treatment. It accounts for about a third of those predicted to be in the highest-cost decile and is more associated with older individuals. The other path depends heavily on criminal justice system involvement and is associated with younger individuals.

Our study is based on a carefully constructed database of linked administrative records. As a result, it is limited by factors that compromise agency data in general. Data may be incomplete due to clients moving in and out of the county, incomplete linking, or errors in data collection. Another limitation comes from the imputation of costs for specific services, which is necessarily an approximation. Increased precision in estimating costs will further improve the model. For example, a current estimate of the cost of PSH that is specific to Santa Clara County will refine our analysis of how widely PSH can be targeted and still generate cost savings.

Making this tool operational will present more opportunities for improvement. One area that may be ripe for improvement is in accounting for missing client data. Certain fields may be more frequently missing in client records than others. If that is the case, correlations between input variables could be used to impute missing fields. For example, a regression model of commonly missing variables on a subset of fields with low rates of missing-ness could be applied. The Economic Roundtable could work collaboratively with Santa Clara County CPHI to develop imputation models.

The evaluation of the performance of the tool both independently and in comparison to the VI-SPDAT will also provide valuable information for future refinement. It may expose areas where our data was inadequate, and offer insight into long-term outcomes.

We did not elect to use a machine learning algorithm for prediction in the triage tool because the gains in predictive power did not justify the loss in interpretability or the added administrative burden. However, once the tool is embedded in the Santa Clara County CPHI data processing systems the administrative burden may shrink to the point where machine learning algorithms are advantageous. They could be built alongside simpler regression models to harness the predictive power of the former and the interpretability of the latter. Comparison of random forests with other machine learning algorithms in this context is a subject of future work.

Ultimately, PSH provides benefits to all of its residents, but not always cost benefits to the system responsible for delivering services. The triage tool addresses this problem by ensuring that PSH is prioritized for those whose needs are most acute, and for whom receiving it leads to the greatest reductions in costs. This has the potential to make more resources available to serve a larger population. Our model indicates that it would be cost-effective to target the entire highest-risk decile for PSH. In that light, the triage tool is not only of immediate practical use, but can be used to advocate for strategically expanding the supply of PSH in Santa Clara County.

Bibliography

Byrne, Thomas et al. (2014). "The Relationship Between Community Investment in Permanent Supportive Housing and Chronic Homelessness". In: *Social Service Review* 88, pages 234–263. DOI: 10.1086/676142 (cited on page 6).

Crisis Indicator: Triage Tool for Identifying Homeless Adults in Crisis (2011). Technical report. Economic Roundtable. DOI: 10.13140/RG.2.1.4788.8246 (cited on page 23).

Culhane, Dennis P. (2008). "The Cost of Homelessness: A Perspective From the United States". In: *European Journal of Homelessness* 2, pages 97–114. URL: http://www.repository.upenn.edu/spp_papers/148 (cited on page 6).

Culhane, Dennis P. and Thomas Byrne (2010). *Ending Chronic Homelessness: Cost-Effective Opportunities for Interagency Collaboration*. Technical report. University of Pennsylvania School of Social Policy and Practice. URL: http://repository.upenn.edu/spp_papers/143 (cited on page 6).

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". In: *Housing Policy Debate* 13, pages 107–163. DOI: 10.1080/10511482.2002.9521437 (cited on page 6).

Flaming, Daniel, Halil Toros, and Patrick Burns (2015). *Home Not Found: The Cost of Homelessness in Silicon Valley*. Technical report. Los Angeles: Economic Roundtable. DOI: 10.13140/RG.2.1.4780.6327 (cited on pages 5, 6).

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani (2010). "Regularization Paths for Generalized Linear Models via Coordinate Descent". In: *Journal of Statistical Software* 33.1, pages 1–22 (cited on page 11).

Greenwood, Ronnie Michelle, Ana Stefancic, and Sam J. Tsemberis (2013). "Pathways Housing First for Homeless Persons with Psychiatric Disabilities: Program Innovation, Research, and Advocacy". In: *Journal of Social Issues* 69, pages 645–663 (cited on page 6).

Gubits, Daniel et al. (2016). *Family Options Study: 3-Year Impacts of Housing and Services Interventions for Homeless Families*. Technical report. Washington, DC: Government Publishing

Office; U.S. Department of Housing and Urban Development. URL: huduser.gov/portal/publications/Family-Options-Study.html (cited on page 6).

Henwood, Benjamin F. et al. (2015). "Service Use Before and After the Provision of Scatter-Site Housing First for Chronically Homeless Individuals with Severe Alcohol Use Disorders". In: *International Journal of Drug Policy* 26, pages 883–886 (cited on page 6).

Liaw, Andy and Matthew Wiener (2002). "Classification and Regression by randomForest". In: *R News* 2.3, pages 18–22. URL: https://CRAN.R-project.org/doc/Rnews/ (cited on page 20).

Ly, Angela and Eric Latimer (2015). "Housing First Impact on Costs and Associated Cost Offsets: A Review of the Literature". In: *Canadian Journal of Psychiatry* 60, pages 275–287 (cited on page 6).

Martinez, Tia E. and Martha R. Burt (2006). "Impact of Permanent Supportive Housing on the Use of Acute Care Health Services by Homeless Adults". In: *Psychiatric Services* 57, pages 1–8 (cited on page 6).

Milborrow, Stephen (2017). *rpart.plot: Plot 'rpart' Models: An Enhanced Version of 'plot.rpart'*. R package version 2.1.2. URL: https://CRAN.R-project.org/package=rpart.plot (cited on page 19).

Poulin, Stephen R. et al. (2010). "Service Use and Costs for Persons Experiencing Chronic Homelessness in Philadelphia: A Population-Based Study". In: *Psychiatric Services* 61.1093-1098. DOI: 10.1176/ps.2010.61.11.1093 (cited on page 6).

Rog, Debra J. et al. (2014). "Permanent Supportive Housing: Assessing the Evidence". In: *Psychiatric Services* 65, pages 287–294. DOI: 10.1176/ (cited on page 6).

Shinn, Marybeth et al. (2013). "Efficient Targeting of Homelessness Prevention Services for Families". In: *American Journal of Public Health* 103.S324-S330 (cited on page 6).

Therneau, Terry and Beth Atkinson (2018). *rpart: Recursive Partitioning and Regression Trees*. R package version 4.1-13 (cited on page 19).

Toros, Halil and Daniel Flaming (2016). *Identifying and Housing High-Cost Homeless Residents*. Technical report. Economic Roundtable. URL: https://economicrt.org/publication/silicon-valley-triage-tool/ (cited on pages 5, 9, 10, 27).

— (2018). "Prioritizing Homeless Assistance Using Predictive Algorithms: An Evidence-Based Approach". In: *Cityscape: A Journal of Policy Development and Research* 20.1, pages 117–146 (cited on pages 10, 27).

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*. Technical report. County of Los Angeles, CEO (cited on page 6).

Tsemberis, Sam and Ronda F. Eisenberg (2000). "Pathways to Housing: Supported Housing for Street-Dwelling Homeless Individuals with Psychiatric Disabilities". In: *Psychiatric Services* 51, pages 487–493. DOI: 10.1176/appi.ps.51.4.487 (cited on page 6).

Where We Sleep: The Costs of Housing and Homelessness in Los Angeles (2009). Technical report. Economic Roundtable. DOI: 10.13140/RG.2.1.2624.0887 (cited on page 23).