Title
Predicting Exits from Permanent Supportive Housing in Los Angeles

Author
Scherling, Adam

Publication Date
2018

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Predicting Exits from Permanent Supportive Housing in Los Angeles

A thesis submitted in partial satisfaction

of the requirements for the degree

Master of Science in Statistics

by

Adam Edward Scherling

2018

ABSTRACT OF THE THESIS

Predicting Exits from Permanent Supportive Housing in Los Angeles

by

Adam Edward Scherling

Master of Science in Statistics

University of California, Los Angeles, 2018

Professor Mark Stephen Handcock, Chair

Permanent supportive housing programs, which provide high-need homeless individuals with long-term housing and supportive services, are thought to be crucial for addressing chronic homelessness. However, many individuals who enroll into permanent supportive housing programs exit within a short period of time, often to unsuitable destinations. This paper utilizes a random survival forest model to predict the outcomes of permanent supportive housing programs in Los Angeles County.

The model demonstrates moderate success out-of-sample, with a concordance of 75% between expected risk of exit and observed length of stay. The identification of negative outcomes is similarly successful, with an AUC of 0.7. Organization-level covariates are found to be the most important predictors. Other important factors include age, previous homeless experience, and variables related to client income and benefits. On the other hand, most demographic variables, client health, and client disabilities are found to play relatively small roles in predicting outcomes.

The thesis of Adam Edward Scherling is approved.

Erin K. Hartman

Till von Wachter

Mark Stephen Handcock, Committee Chair

University of California, Los Angeles

2018

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

Many thanks to my advisor, Mark Handcock, for the opportunity to work with this rich data set and for much useful advice along the way. I am grateful to Erin Hartman for keeping me honest about the interpretation of my results, and to Till von Wachter for his assistance with a variety of topics, including navigating the complexities of HMIS.

A few brief conversations with Elior Cohen were vital in helping me pick a topic. Thanks are due as well to Ana Ramirez, Harsh Gupta, Irfan Malik, Janey Rountree, and Nathan Hess at the California Policy Lab/California Center for Population Research, and to Steven Rocha at LAHSA, for their help and support. I am also grateful to Glenda Jones in the Department of Statistics for her assistance throughout my time at UCLA. Lastly, thanks are due to my parents, John and Sandy, and my brother, Alex, for setting me on the road that brought me here and supporting me every single step of the way.

This work was supported by the UCLA Transdisciplinary Seed Grant program of the Office of the Vice Chancellor for Research. Specific funding was under National Center for Advancing Translational Sciences UCLA CTSI Grant UL1TR001881.

1 Introduction

1.1 Homelessness and Permanent Supportive Housing in Los Angeles

Los Angeles is in the midst of a homelessness crisis. A point-in-time homeless count in the County of Los Angeles in 2017 recorded 55,000 homeless individuals, an increase of 17% from the previous year (Smith, 2018). In April 2018, Los Angeles Mayor Eric Garcetti proclaimed in his State of the City address that "homelessness isn't *an* issue... it is *the* issue" (Garcetti 2018, emphasis in original).

Voters in the City of Los Angeles responded to the growing crisis in November 2016 by approving Measure HHH, providing $1.2 billion in funding for the construction of 10,000 units of permanent supportive housing (PSH) for the homeless (Chiland, 2016). PSH provides long-term project- or tenant-based housing along with supportive services for homeless individuals and families, focusing in particular on those with disabilities (HUD, 2017).

Proponents of PSH argue that providing high levels of support to the individuals with highest need is an efficient allocation of resources that can be cost-neutral or even cost-saving (Cultrane and Metraux, 2008). Several studies in Los Angeles County have identified large cost offsets in public expenditures that may be unlocked by enrolling high-cost individuals in PSH (Flaming et al., 2009; Flaming et al., 2013; Hunter et al., 2017). Others argue for the merit of PSH even if it is not cost-saving, based on the higher quality of life it affords to those in need (Kertesz et al., 2016). For these reasons, PSH is described by the City of Los Angeles Comprehensive Homeless Strategy as being essential for keeping the chronically homeless off the streets (City of Los Angeles, 2016).

1.2 Exits from Permanent Supportive Housing

Unfortunately, the experiences of individuals in PSH do not always match these high expectations. Wong et al. (2006) find that clients with mental illness who enrolled in PSH in Philadelphia exited in substantial numbers after even a few months. Although some such clients moved from PSH into suitable permanent living situations, many returned to temporary shelters or to places not meant for human habitation. Further, Wong and coauthors found that it was not possible to predict upon entry how long an enrollment would last, or what the end destination would be.

Many other studies have examined supportive housing programs to identify factors which increase the risk of exit (e.g., Lipton et al., 2000; Martinez and Burt, 2006). Some such studies focus in particular on PSH (e.g., Bernet et al., 2015; Collins et al., 2013; Pearson et al., 2009). However, little if any attempt is made in the literature to determine whether any risk factors can effectively predict program exits out-of-sample.

Predictive analytics has shown promise in related applications, such as identifying chronically homeless, high-cost users of public services for whom enrollment in PSH would be cost-saving (Toros and Flaming, 2016). This paper creates a predictive model for exits from PSH. A random survival forest is used to identify individuals at high risk, and to predict the exit type of those who drop out. The model demonstrates that outcomes of individuals enrolling in PSH can be predicted with a moderate level of accuracy using information available at time of entry.

2 Data

2.1 HUD, LAHSA, and HMIS

The U.S. Department of Housing and Urban Development (HUD) is a major source of funding for homeless housing and services programs. Through its Continuum of Care Grants and Emergency Solutions Grants, HUD funds a number of different program types, including PSH (U.S. Dept. of Housing and Urban Development, cited May 27, 2018). As a condition of receiving these funds, homeless service providers must provide enrollment and service records to the Housing Management Information System (HMIS) maintained by the local Continuum of Care (U.S. Dept. of Housing and Urban Development, cited May 28, 2018). Organizations that do not receive these funds are also encouraged to report to HMIS (U.S. Dept. of Housing and Urban Development, 2016a).

A Continuum of Care is a regional planning body that coordinates homeless housing and services. The Los Angeles Homeless Services Authority (LAHSA) is the lead agency for the Los Angeles Continuum of Care, which represents all of Los Angeles County with the exceptions of Glendale, Long Beach, and Pasadena (LAHSA, 2018). The data used in this paper are from the Los Angeles HMIS, which is maintained by LAHSA.

To allow for consistent reporting across regions, HUD requires each HMIS to comply with certain data standards. These standards are available online, and serve as a key resource for understanding the data (U.S. Dept. of Housing and Urban Development 2016a, 2016b, 2016c).

2.2 Population of Interest and Available Sample

The population of interest for this analysis is the set of all enrollments of single adults into PSH programs in the Los Angeles Continuum of Care from 2010 through 2016. Reporting rates of PSH enrollments into HMIS are less than 100%, but are fairly comprehensive. More than 12,000 enrollments of single adults into PSH are recorded in this time frame.

However, some filtering is necessary to ensure that every enrollment labeled as an enrollment into PSH stands up to scrutiny. Several programs which are labeled in HMIS as providing PSH stretch the limits of the imagination; one such project had an average length of enrollment of two days, and provided many clients nothing but transportation services. In other cases, a client might enroll into PSH, but their entry into housing may be delayed for a short or moderate period of time. For these reasons, entry into PSH is assumed to occur only once the client has a record in the system of receiving housing. Due to inconsistent reporting practices, this unfortunately excludes many true enrollments into PSH for which receipt of housing was not recorded. The remaining sample size consists of a little more than 6,400 enrollments.

Since exit from PSH does not always represent a bad outcome for the client, system performance measures from HUD (Office of Community Planning and Development, 2018) are used to classify exits as positive or negative. Outcomes such as a private rental by the client or a permanent living situation with family or friends are regarded as positive. Outcomes such as entry into an emergency shelter, any living situation with temporary status, or an unknown destination are regarded as negative. Lastly, if the client is deceased or residing in a hospital or long-term care facility, this is regarded as extraneous. These cases are excluded from the analysis, bringing the sample size to just under 6,200 observations.

2.3 Properties of the Sample

Table 1 compares the remaining sample to those enrollments which were excluded from the sample due to the lack of a record of receiving housing. (The "Excluded" sample, like the "Included" sample, is limited to single adults who did not have an extraneous destination type.) The demographic properties of the two groups do not differ too dramatically, although it can be seen that the sample used for analysis contains a substantially larger percentage of individuals with reported mental illness, and a slightly higher rate of reported substance abuse. Additionally, it contains a higher percentage of males and blacks (and fewer females and whites). It seems plausible that these groups are more likely to have received housing due to the fact that they were regarded as relatively high-need.

The characteristics of the projects that clients are enrolled into reveal greater discrepancies across the two samples. Clients in the included sample were generally enrolled into projects run by larger organizations. The clients in the included sample were also enrolled into projects with larger numbers of housing units, and larger numbers of HMIS participating beds. Perhaps the most dramatic finding, however, is that clients in the included sample were far more likely to be enrolled into a project that was being funded by a Continuum of Care grant at the time of enrollment. Many of these factors will prove to be key predictors of the outcomes of PSH.

2.4 Response of Interest

There are two response variables of interest: length of stay in PSH and the type of exit (qualitatively positive or negative). For length of stay, the start date used is the first date

Table 1: Sample Characteristics

		Included (n = 6,162)	Excluded (n = 5,717)
Race	Asian	2%	2%
	Black	62%	54%
	Native American	2%	3%
	Pacific Islander	1%	1%
	White	32%	39%
Ethnicity	Hispanic	17%	20%
Gender	Female	34%	37%
	Male	64%	59%
	Transgender	1%	1%
	Not reported	1%	3%
Age	25th percentile	41	39
	75th percentile	56	56
Veteran status	Yes	6%	8%
Disabling condition	Chronic health	28%	31%
	Mental health	69%	47%
	Physical disability	25%	28%
	Substance abuse	27%	22%
Project characteristics	Organization size (mean)	1387	582
	Project size (mean)	167	154
	Unit inventory (mean)	9	5
	HMIS beds (mean)	10	4
	CoC grant	70%	26%
	VASH grant	3%	0%
	Other grant	54%	65%

at which there is a record of the client receiving housing. While this usually coincides with the program enrollment date, in some cases there is a moderate discrepancy.

Determining the end date is more complex. Although HMIS does record program exits, the set of recorded exits is known to be incomplete. Therefore I determine that an exit has occurred if at least one of the following criteria are met:

1. An exit is recorded in HMIS.
2. The client has no record of receiving housing during the last month of the observation period, December 2016.

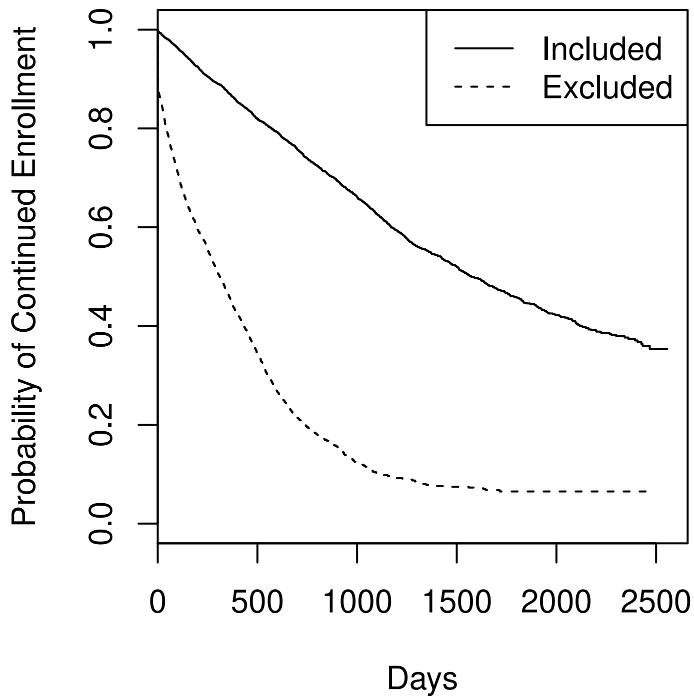
Although the second criteria is not perfect, it is nonetheless quite reliable. Receipt of housing is recorded in HMIS on a daily basis. Amongst those clients with a record of receiving housing from a PSH program, it is quite abnormal for there to be any gaps. Furthermore, this method identifies only 48 exits without a record; in comparison, there are 2,349 recorded exits.

Length of stay in PSH is then defined as the number of days from the first date at which a client has a record of receiving housing to the last. For individuals who are deemed not to have exited the program, the response is considered to be right-censored: the true length of stay is unknown.

In cases where an exit is inferred but was not recorded, the destination is considered to be missing. I follow the criteria provided by HUD, under which unknown destinations are classified as a negative outcome. Otherwise, I follow the criteria as mentioned above and classify exits to permanent housing situations as positive and exits to temporary or otherwise unsuitable housing situations as negative.

Figure 1 plots the estimated survival functions (further described in the Design section) for the included and excluded samples. For the included group, exits and length of stay are as defined above. For the excluded group, an exit is inferred if a service was not received in the second half of 2016; length of stay is calculated as the time from the first recorded service to the last. The curve for the excluded sample starts below a probability of 1 because

Figure 1: Estimated Sample Survival Functions



some individuals enrolled into PSH had no record of receiving any services.

The dramatic difference between the two curves is likely attributable to the fact that a substantial portion of the excluded group did not receive housing. However, such a hypothesis cannot be confirmed with the available data. It may be the case that many of these individuals did receive housing, but that other factors were responsible for their exit. The remaining analysis will focus on those who had a record of receiving housing. Although the exit rate for this group is much smaller, a sizable portion still exit from PSH within a short time frame; fully 12% exit within one year of enrollment.

2.5 Covariates of Interest

HMIS contains a wide variety of information about the clients which are enrolled into PSH, including demographics as well as information on health, disabilities (including mental health and substance abuse issues), domestic violence, income and benefits, and homeless/residence history. Most of these variables are self-reported by the client. Additionally, HMIS contains information about the projects that provide PSH and the organizations that run them.

Several variables were not directly available in HMIS, but were generated using the available information. This includes client age; day of the week and month of the year in which the client was enrolled; organization and project size; project inventories at the time of enrollment; grants funding the project at the time of enrollment; and previous enrollments by the client into emergency shelters and other programs in HMIS.

A full list of the covariates used in this analysis is given in the Appendix, along with brief descriptions and summaries of how many observations exhibit nonresponse for each covariate. Other variables with extremely large portions of missing data are excluded entirely, and are not listed in the Appendix.

2.6 Missing Data

The question of how to deal with missing data is an important one. There are many ways of treating missing data, all with advantages and disadvantages. The easiest, omitting any observations that have missing covariates, will generally lead to bias. This would also make it impossible to form a prediction for any observation with a missing covariate.

Multiple imputation (Rubin, 2004) is a relatively common approach with far better statistical properties. It is also more computationally intensive, because it requires the model to be fit on each of many imputed datasets. For a random forest, predictions of the response and estimates of variable importance would need to be combined across each of the many forests as well. Multiple imputation also relies upon the assumption that missingness occurs at random (after conditioning on any observed covariates).

While multiple imputation is promising, I take the alternate approach of treating missingness as its own type of response. This is justifiable for a number of reasons, in addition to mere convenience. Firstly, my focus is on prediction, not parameter estimation. Multiple imputation is unlikely to improve the predictive power of my model since it does not add any information that is not already available in the dataset. Further, this approach is also justifiable if nonresponse happens for a reason that might be of scientific interest.

As it is, HMIS contains several codes for nonresponse. Most variables include categories for "Client refused," "Client doesn't know," and "Data not collected." Where a response is missing, I add another category for "No record." This has the added benefit of making interpretation more transparent. A variable for substance abuse status, for example, should be interpreted as the client's response to the question of whether they suffer from substance abuse, not as the actual truth of whether they suffer from it.

Treating nonresponse as its own category does present some difficulty in the case of continuous variables with missing data. In the current analysis, this occurs for the project inventory variables and the reported income variables. I convert these from numerical to categorical; bins are created by taking the log and rounding to the nearest integer. A log transform is reasonable because these variables are very highly skewed to the right. However, it may be the case that multiple imputation would be a preferred method for these particular variables.

3 Design

3.1 Survival Analysis: An Overview

The study of censored time-to-event variables has long been an important subfield of statistics. Due to the censoring of the variable of interest, most traditional statistical methods cannot be used. In the field of survival analysis, focus shifts from the distribution of the time-to-event variable to the survival and hazard functions. A basic overview is given below; for a more detailed treatment see an introductory text, e.g. Klein and Moeschberger (2005).

For enrollment i into PSH, $i = 1, ..., n$, let the length of stay be $T_i = \min(T_i^o, C_i^o)$, where T_i^o is the length of time until an exit occurs and C_i^o is the length of time until the end of the observation period. Let the exit type be $\delta_i^o \in \{1, 2\}$, where 1 indicates an exit to an unknown or unsuitable destination, and 2 indicates an exit to a known destination with permanent tenure. The observed exit type is then $\delta_i = \delta_i^o I(T_i^o \leq C_i^o)$. An exit type is only observed if the exit occurs prior to the end of the observation period; otherwise, $\delta_i = 0$ indicates that the client is still enrolled at the end of observation period. Lastly, let x_i be a vector of covariates for observation i.

The survival function (0.1) gives the probability that length of stay is at least t:

$$S(t) = P(T > t) = \int_t^{\infty} f(t)dt \tag{0.1}$$

Given an observed set of survival times, the survival function is commonly estimated using the Kaplan-Meier estimator, given in (0.2), where Y_i is the number of observations that have not exited before time t, d_i is the number of exits that occurred at time t, and $t_1, \ldots, t_m$ are the observed lengths of stay.

$$\hat{S}(t) = \begin{cases} 1 & \text{if } t < t_1 \\ \prod_{t_i \leq t}[1 - \frac{d_i}{Y_i}] & \text{if } t_1 \leq t \end{cases} \tag{0.2}$$

The hazard function (0.3) is the exit rate at time t conditional upon not having exited before time t:

$$\alpha(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} = \frac{f(t)}{S(t)} \tag{0.3}$$

The cumulative hazard function is then the cumulative hazard up to time t (0.4). The cumulative hazard function is commonly estimated by the Nelson-Aalen estimator (0.5).

$$H(t) = \int_0^t \alpha(t) \tag{0.4}$$

$$\hat{H}(t) = \begin{cases} 0 & \text{if } t < t_1 \\ \sum_{t_i \leq t} \frac{d_i}{Y_i} & \text{if } t_1 \leq t \end{cases} \tag{0.5}$$

The hypothesis of equal hazards for some populations 1 and 2 can be tested using the log-rank statistic (0.6). In particular, (0.6) tests the difference between the hazard for population 1 and the hazard for the pooled population of groups 1 and 2. The statistic consists of the weighted sum (across all time periods t_i, $i = 1, ..., m$) of the difference between the estimated hazard at time t_i for subpopulation 1 and the estimated hazard at time t_i for the entire population. The weights $W(t_i)$ are set equal to the number of observations in group 1 which survive until at least time t_i, i.e. $W(t_i) = Y_{i1}$, resulting in (0.7). An estimator for the variance of this statistic is given by (0.8). This test will come in handy for fitting a random survival forest, as described below.

$$L = \sum_{i=1}^{m} W(t_i)\Big[\frac{d_{i1}}{Y_{i1}} - \frac{d_i}{Y_i}\Big] \tag{0.6}$$

$$L = \sum_{i=1}^{m} \Big[d_{i1} - Y_{i1}\Big(\frac{d_i}{Y_i}\Big)\Big] \tag{0.7}$$

$$\hat{\sigma}^2 = \sum_{i=1}^{m} d_i \frac{Y_{i1}}{Y_i}\Big(1 - \frac{Y_{i1}}{Y_i}\Big)\Big(\frac{Y_i - d_i}{Y_i - 1}\Big) \tag{0.8}$$

3.2 Random Survival Forests for Competing Risks

The random forest (Breiman 2001) is a machine learning tool that is notable for its predictive power, flexibility, and ease of use. A good introduction can be found in James et al. (2017). Random forests can be used for regression, classification, and (with some modification) survival analysis (Ishwaran et al., 2008).

Briefly, a random forest consists of a collection of decision trees. For each tree a new dataset is created by sampling observations from the original data, with replacement, until the sample drawn is the same size as the original dataset. The tree is then built by performing recursive binary splits of this dataset. For each split, a random subset of the covariates are chosen as split candidates. Amongst these, the split is chosen which optimizes some outcome of interest. This recursive splitting continues until some stopping criterion is met.

Fitted values or predictions for a given observation can be found by dropping the observation down every fitted tree. By applying the tree's decision rules to an observation's covariates, the observation is placed into a single terminal node in the tree. The prediction of the response variable is then calculated using the observed responses of the observations in this terminal node in the fitted tree. The predictions of each tree are averaged together to create the prediction of the forest as a whole.

In this paper, a random survival forest for competing risks is grown using the randomForestSRC package (Ishwaran and Kogalur, 2018) for the R language (R Core Team, 2018). Ishwaran et al. (2014) defines the procedure used by the package. Additional information is provided online by Kogalur and Ishwaran (cited May 26, 2018).

Splits in the random forest are chosen by maximizing the composite log-rank test statistic (0.9):

$$L(x,c) = \frac{\sum_{j=1}^{J} \hat{\sigma}_j^2(x,c) L_j(x,c)}{\sqrt{\sum_{j=1}^{J} \hat{\sigma}_j^2(x,c)}} \tag{0.9}$$

Each component $L_j(x,c)$ is a modified version of the log-rank statistic (0.7) that is calculated using only exit type j (e.g., positive or negative), conditional on performing a split on variable x at value c. The variance $\hat{\sigma}_j^2(x,c)$ is similarly the variance of the log-rank statistic (0.8) for exit type j after conditioning on such a split. Maximizing this composite log-rank statistic thus picks the split which maximizes the combined difference between exit type-specific hazard rates.

To build the random forest, decision trees are grown on bootstrapped versions of the dataset using this splitting criteria. Predicted survival functions and hazard functions for a particular observation with covariates x_i can then be computed from the forest, using the Kaplan-Meier and Nelson-Aalen estimators on the populations in the terminal node corresponding to x_i in each tree.

3.3 Study Design and Evaluation

This paper seeks to determine whether exits from PSH can be effectively predicted and classified. The data are therefore split into training and test sets. 85% of the data ($n = 5{,}238$) will be used to grow the random forest, and the remaining 15% ($n = 924$) will be used to test its accuracy out-of-sample. Model parameters, including the number of trees, the number of splitting points to consider for each variable, and the terminal node size, are optimized using the out-of-bag observations in the training data. I use 2000 trees, 6 randomly selected split points, and a terminal node size of 6; however, varying these parameters was found to have only a marginal impact on the results.

Once the random forest is constructed, a statistic is calculated to estimate the total risk of each exit type for each individual enrollment. The particular quantity used is the integrated cause-specific cumulate incidence function. For each exit type j, the cause-specific cumulative incidence function gives the probability that an exit of type j will occur by time t. It is estimated by the Aalen-Johansen estimator (0.10), where $\hat{S}$ is the Kaplan-Meier estimator of the survival function and $d_j(t_k)$ is the number of exits of type j at time t_k.

$$\hat{F}_j(t) = \sum_{t_1 < t_k \leq t} \hat{S}(t_{k-1}) \frac{d_j(t_k)}{Y(t_k)} \tag{0.10}$$

The quantity of interest is then $\mathcal{M}_j = \int_0^{t_m} F_j(s)ds$, which is estimated by (0.11). This statistic is used as a measure of the risk of exit type j.

$$\hat{\mathcal{M}}_j = \sum_{k=1}^{m} \hat{F}_j(t_k)(t_{k+1} - t_k) \tag{0.11}$$

These predicted risk values are then used to predict exits and exit types. The overall risk of an exit is estimated as $\hat{\mathcal{M}} = \hat{\mathcal{M}}_1 + \hat{\mathcal{M}}_2$. The accuracy of $\hat{\mathcal{M}}$ can be evaluated using Harrell's C-Index or "concordance" (Schmid et al., 2016), defined in (0.12). This quantity represents the probability that, of two randomly selected enrollments, the enrollment with a higher predicted risk $\hat{\mathcal{M}}$ exits first.

$$C = \frac{\sum_{i,i'} I(T_i > T_{i'}) \cdot I(\mathcal{M}_i < \mathcal{M}_{i'}) \cdot I(T^o_{i'} < C^o_{i'})}{\sum_{i,i'} I(T_i > T_{i'}) \cdot I(T^o_{i'} < C^o_{i'})} \tag{0.12}$$

Similarly, the event-specific concordance can also be calculated for each exit type. For example, for negative exits, the concordance can be calculated between $\hat{\mathcal{M}}_1$ and the lengths of enrollment, where exits to a positive destination are considered to be censored observations.

Predictions for the type of exit can be made by selecting observations with the highest quantile of risk for a particular exit type. The predictions that result from different choices of quantile are plotted in a Receiver Operating Characteristic (ROC) curve, which displays the trade-off that can be made between sensitivity (e.g., the portion of negative exits which are classified as such) and specificity (e.g., the portion of censored observations and positive exits which are not classified as negative).

Lastly, the identification of important covariates is performed using a "variable importance" measure. The importance of a given variable x is determined by calculating the forest's prediction error using two methods. First, the prediction error is calculated using all trees in the random forest, as usual. Then, the prediction error is calculated using a modified forest in which each split that occurs on variable x is replaced by random assignment from the parent node into two daughter nodes. The increase in prediction error that arises from using the second method gives the importance of variable x. In order to calculate variable importance separately for each exit type, prediction error is calculated using a weighted event-specific concordance measure, described in Ishwaran et al. (2014).

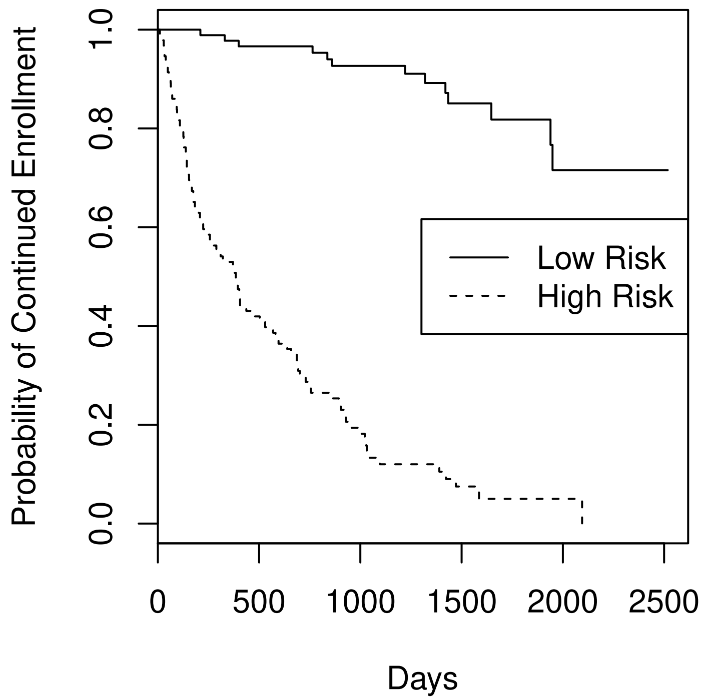
4 Results

4.1 Predictive Performance

The random forest demonstrates a moderate capability of predicting survival times out-of-sample: the concordance of the estimated overall risk $\hat{\mathcal{M}}$ and the actual survival times in the test data is 75%. Conditional on an exit, the average length of stay of enrollments predicted to be in the highest decile of risk is 460 days; for those in the lowest decile of risk, the figure is 1,100 days. This distinction is visualized in Figure 2, which plots the estimated survival functions for the enrollments in the test set which were predicted to be in the highest and lowest deciles of risk.

The model also demonstrates a modest capability of identifying particular exit types. The event-specific concordance statistics for negative and positive exits are 73% and 78%,

Figure 2: Estimated Survival Functions by Predicted Risk

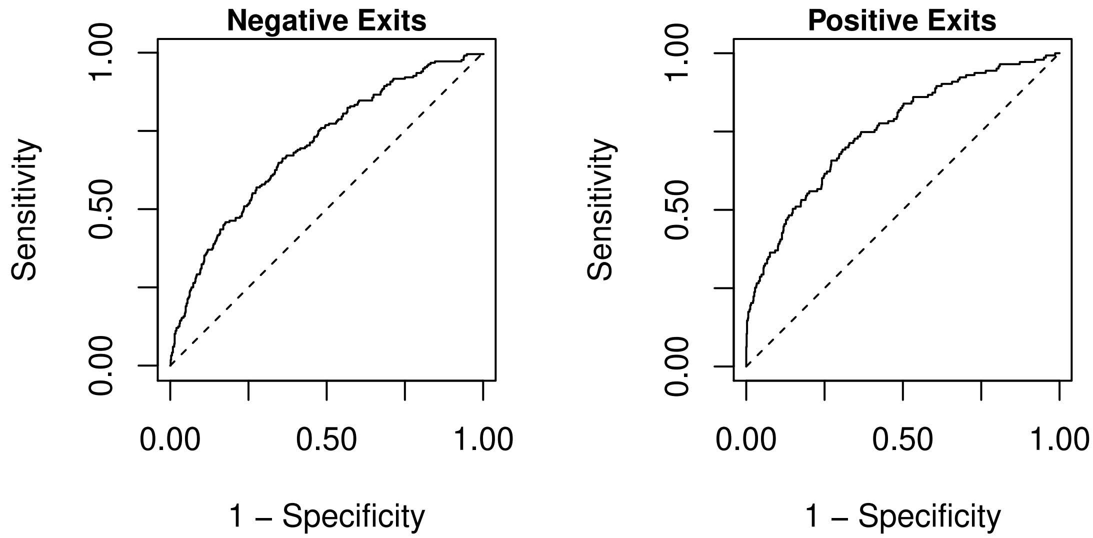


respectively. Figure 3 displays ROC curves for negative and positive exit types. In the left-hand chart, the y-axis represents the proportion of true negative exits which are classified as such; the x-axis represents the proportion of positive exits or censored observations which are misclassified as a negative exit. The right chart is analogous, focusing on positive exits. The area under the curve (AUC) values for negative and positive exits are 0.70 and 0.75, respectively.

The marginal distributions of $\hat{\mathcal{M}}_1$ and $\hat{\mathcal{M}}_2$, the predicted risk of negative exits and positive exits, are shown in Figure 4. Both distributions are skewed to the right, with the risk of a positive exit more strongly so. Figure 5 shows a scatterplot of the same variables.

The correlation of the predicted risk of a negative exit and the predicted risk of a positive exit is 0.28. Thus, for most enrollments, the model sees positive exits and negative exits as more alike than different. At the right tails, however, it is clear from Figure 5 that some enrollments have particularly high risk of a negative exit and some have particularly high

Figure 3: ROC Curves for Type-Specific Predictions



risk of a positive exit. The larger right tail for positive exits explains the higher AUC value. There is a small cohort of enrollments with a very high chance of a positive exit; otherwise, negative exits are more likely.

4.2 Variable Importance

The variable importance estimates are plotted in Figure 6. The 20 most important covariates for each exit type are given in Table 2. It can be seen that the most important variables, for any type of exit, are factors related to the project that the client is enrolled into, or the organization that runs the project. A surprisingly important variable, for negative exits in particular, is the client's response to whether they have insurance from any source. Age is found to be the most significant demographic variable, by a wide margin. Previous enrollments into HMIS also appear to have significant predictive power. Other classes of variables are found to play comparatively limited roles.

For context, Tables 3 through 8 show the outcomes of enrollments in the full sample (n = 6,162) by certain variables of interest. In Table 3, organization size is defined as the number of enrollments into PSH projects run by each organization. The table shows that the largest organizations have particularly high negative exit rates. Medium- to large-sized

Figure 4: Distributions of Predicted Risk by Exit Type

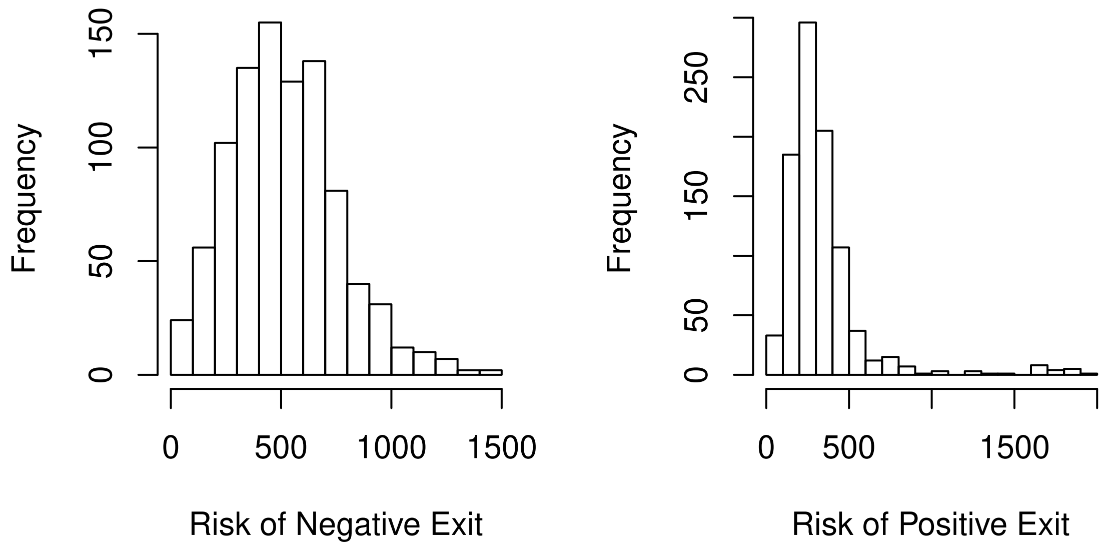
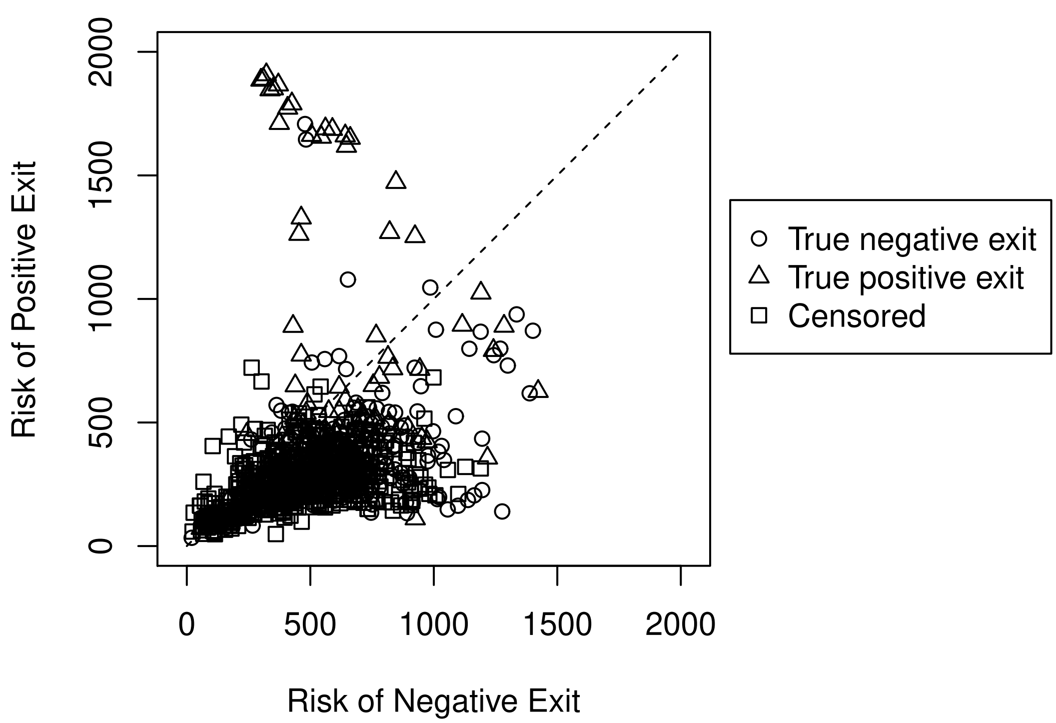


Figure 5: Scatterplot of Predicted Risk by Exit Type



organizations (those with 100 to 1000 enrollments into PSH in the sample period) perform the best.

Table 4 shows the outcomes by response to the question of whether the client has any health insurance. Those who report having insurance have lower exit rates than those who do not. However, a larger gap is found between clients for whom data was collected and those for whom it was not collected. A similar pattern can be found for the response to employment status, shown in Table 6.

The pattern of outcomes by unit inventory level is shown Table 5. Projects with only 1 unit ($1 = \exp(0)$) have the lowest overall exit rates. Those with several dozen units ($\exp(4) \approx 55$) have remarkably high positive exit rates. Those projects that fall in between these two groups seem to perform relatively poorly, as do those projects with no inventory record.

Table 7 shows the outcomes by age group. Transition-age youth (ages 18-25) have the highest exit rates, with the rates decreasing rather steadily as age increases. Lastly, Table 8 shows the outcomes by the number of previous emergency shelter stays. A large number of previous shelter stays appears to be quite strongly correlated with greater exit rates, and negative exits in particular.

5 Discussion

5.1 Proper Interpretation and Implications for Future Work

As demonstrated above, the random survival forest model has moderate predictive capability. Using information available at the time of enrollment, the model can determine with reasonable accuracy whether the client is at high risk or at low risk of experiencing a negative outcome. The model also identified which factors play the largest roles in creating this prediction.

The model does not, however, provide insight as to why these factors are important, or what should be done about it. In addition, no attempt was made to examine the experiences

Table 2: Variable Importance: Top 20 Factors, by Exit Type

	Hazard of Negative Exit	Hazard of Positive Exit
1	OrganizationID	OrganizationID
2	InsuranceFromAnySource	ProjectID
3	OrganizationSize	OrganizationSize
4	UnitInventory	HMISParticipatingBeds
5	HMISParticipatingBeds	UnitInventory
6	ProjectID	ProjectSize
7	Age	InsuranceFromAnySource
8	ProjectSize	Employed
9	PreviousEnrollmentN	BedInventory
10	Employed	RentalAssistanceOngoing
11	PreviousShelterEnrollmentN	ContinuumProject
12	ContinuumProject	NotEmployedReason
13	BedInventory	DisabilityResponse_MentalHealth
14	HousingStatus	Age
15	DisabilityResponse_SubstanceAbuse	PreviousShelterEnrollmentN
16	OtherIncomeAmount	HousingStatus
17	NotEmployedReason	PreviousEnrollmentN
18	MonthsHomelessPastThreeYears	MonthsHomelessPastThreeYears
19	SocSecRetirementAmount	SSDIAmount
20	OutOfState	CoCGrant

Figure 6: Variable Importance

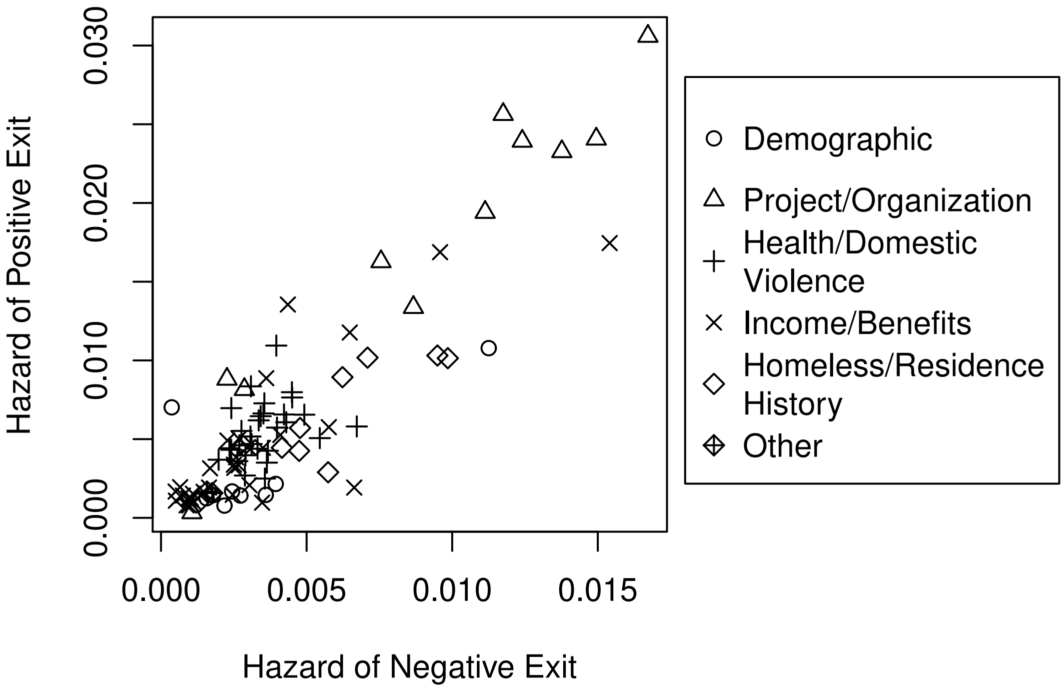


Table 3: Outcomes by Organization Size

Organization Size	No Exit	Negative Exit	Positive Exit	Number of Enrollments	Number of Organizations
≥ 1000	54%	28%	18%	3565	2
500-999	76%	13%	10%	1186	2
100-499	73%	15%	12%	435	3
30-100	66%	19%	15%	847	13
< 30	58%	19%	23%	129	14

Table 4: Outcomes by Insurance Response

Has Insurance	No Exit	Negative Exit	Positive Exit	Number of Enrollments
No	65%	23%	12%	590
Yes	70%	19%	11%	3544
Client Doesn't Know	62%	29%	8%	24
Client Refused	43%	43%	14%	7
Data Not Collected	44%	30%	25%	1992
No Record	80%	0%	20%	5

Table 5: Outcomes by Unit Inventory

Unit Inventory (log)	No Exit	Negative Exit	Positive Exit	Number of Enrollments	Number of Projects
0	73%	18%	9%	1954	93
1	56%	27%	16%	758	18
2	48%	36%	16%	416	21
3	51%	29%	20%	333	23
4	38%	18%	44%	351	11
5	71%	25%	4%	28	1
No Record	59%	23%	17%	2322	113

Table 6: Outcomes by Employment Response

Employed	No Exit	Negative Exit	Positive Exit	# Enrollments
No	64%	22%	14%	4562
Yes	67%	16%	17%	281
Client Refused	60%	40%	0%	5
Data Not Collected	28%	37%	35%	797
No Record	82%	10%	8%	517

Table 7: Outcomes by Age

Age	No Exit	Negative Exit	Positive Exit	# Enrollments
18-25	42%	34%	24%	317
26-35	56%	26%	18%	739
36-45	56%	28%	16%	1169
46-55	60%	23%	17%	2349
56-65	71%	16%	13%	1405
66-75	79%	13%	7%	156
76-85	71%	14%	14%	21
86-95	83%	17%	0%	6

Table 8: Outcomes by Number of Previous Shelter Stays

# Shelter Stays	No Exit	Negative Exit	Positive Exit	# Enrollments
0	64%	19%	17%	3047
1	58%	24%	17%	1304
2-5	57%	28%	15%	1552
6-10	61%	31%	9%	232
> 10	67%	30%	4%	27

of clients while they were enrolled in PSH (for example, by considering the services received). There are also many sample selection issues which preclude a causal interpretation of the results. The assignment of clients to programs was not performed randomly; neither was the decision of programs to provide housing to some individuals and not to others. Many enrollments into PSH were also excluded from the sample because there was no record in HMIS of the client receiving housing; this group may be characteristically different from the group that was analyzed.

Nonetheless, the results from this analysis may be useful in a number of ways. From an academic perspective, the identification of high risk enrollments provides a focus for future research. The importance of organization- and project-level covariates in determining risk bears further investigation; more work is required to determine why these differences exist. Organizations may perform poorly for many reasons, including the characteristics of the clients they enroll, the location they are in, the amount of funding they receive, the quality of the services they offer, and many other factors. These would need to be studied in a causal framework in order to provide a true understanding of the relationship between these factors and the risk of negative exits.

From a policy perspective, high-risk enrollments are identified as a group that may be deserving of greater attention and greater resources. The appropriate means for addressing these high-risk enrollments is not made clear from this analysis. However, it is worth knowing that there are predictable differences in the risk of exit, and that the risk is associated with factors such as the organization and project into which clients are enrolled, as well as age, previous enrollments in HMIS, and client response to having insurance. Policymakers and researchers should collaborate further to better understand these associations.

5.2 Statistical Contributions

The results demonstrate that random survival forests are an effective tool for predicting the outcomes of PSH. For such a purpose, the random survival forest has considerable advantages compared to a traditional survival analysis model. However, it also has some significant disadvantages.

A standard competing risks model would have difficulty fitting such a large number of covariates. Variable selection would need to be performed, which can be a time-consuming and imperfect process. In addition, any collinearity in the data could have potentially large impacts on the parameter estimates. Further, in a traditional model, any interaction effects or nonlinearities would have to be known and explicitly accounted for. Lastly, such a model would also be working with the assumption of proportional hazards, which may or may not be justified.

Random forests, on the other hand, are well-adapted to building predictive models that utilize large numbers of covariates. Collinearity is not a very serious issue: even the inclusion of a duplicated variable will affect the forest only by doubling the probability that the variable is a candidate for selection. Furthermore, the non-parametric nature of random forests allows them to account for nonlinearities and interactions in the data without having to explicitly identify them up front. In addition, this is accomplished without any significant overfitting. In this analysis, the model was found to have essentially equivalent prediction success in the out-of-bag sample and in the test set.

On the other hand, random forests do have serious limitations compared to more traditional statistical methods. Perhaps the most evident limitation is the inscrutability of the results. Although variable importance gives a sense of how large a role each covariate plays, it does not reveal much about the nature of the association between the covariate and the response. The direction of any effect, as well as any nonlinearities or interactions, are not readily seen. Lastly, any model fit on observational data will have significant challenges if the goal is causal interpretation of effects. Controlled and randomized experiments remain the gold standard for answering particular questions of scientific interest.

6 Conclusion

Homelessness is a complex issue, and one that the Los Angeles City and County governments are actively struggling to confront. Permanent supportive housing programs provide one promising avenue for keeping chronically homeless individuals off the street. However, the success of these programs is mixed, with many clients exiting PSH to return to emergency shelters, the streets, or other undesirable locations.

The random survival forest model utilized in this paper provides a unique and meaningful perspective for understanding these outcomes. Most notably, the model found that project- and organization-level factors, including the organization ID and organization size, proved to be the most powerful predictors of exit risk. Age and previous homeless experience were also relevant, as were the responses to particular questions regarding insurance and employment. Demographics, health, and disabilities were found to play only a small role in predicting outcomes.

Although it was not perfect, the model demonstrated significant capability of predicting exits, and identifying negative exits in particular. The model thus provides a tool which may be of interest to policymakers as they implement the expansion of permanent supportive housing programs. It also effectively demonstrates the applicability of random forests to studies of homelessness, providing a new avenue for future research.

7 Appendix: Covariates Included in the Model

Unless otherwise specified, the variable was available for every observation in the included sample.

Demographic Variables

Age Each client's date of birth is recorded in HMIS. Age in years at time of enrollment is calculated using the enrollment date and date of birth.

Ethnicity A field is available in HMIS for Hispanic ethnicity. This is not mutually exclusive with any race. A small number of observations ($< 2\%$) exhibit some form of nonresponse.

Gender A field is available in HMIS for gender. In addition to male and female there are options for transgender or other gender. A small number of observations ($< 2\%$) exhibit some form of nonresponse.

Race A separate field is available in HMIS for each of the following races: American Indian, Asian, Black, Native Hawaiian or Pacific Islander, and White. The categories are not mutually exclusive.

VeteranStatus Yes/No. A small number of observations ($< 2\%$) exhibit some form of nonresponse.

Project/Organization Variables

CoCGrant Yes/No. Is there a record of the project receiving funding through a Continuum of Care grant at the time of enrollment?

ContinuumProject Yes/No. Is the project considered to form a part of the Continuum of Care? A small number of observations ($< 10\%$) are listed in HMIS as having an unknown status.

Grants Yes/No. Is there a record of the project receiving funding through a CoC grant, a VASH grant, or another type of grant, respectively, at the time of enrollment?

Inventories The UnitInventory, BedInventory, and HMISParticipatingBeds fields in HMIS are used to estimate the project's inventory of each type at the time of enrollment. These variables are missing for over a third of the included sample. They are treated as categorical variables, including a category for missingness as well as categories corresponding to the log value of the variable, rounded to the nearest integer.

OrganizationID ID for the organization overseeing the project that the client was enrolled into.

OrganizationSize Total number of enrollments in the included sample into projects overseen by the organization. The size is split into broad bins: greater than 1000, less than 1000, less than 500, less than 100, and less than 30.

ProjectID ID for the project the client was enrolled into.

ProjectSize Total number of enrollments into the project in the included sample. The size is split into broad bins: greater than 300, less than 300, less than 100, less than 50, and less than 10.

Health/Domestic Violence Variables

DisablingCondition Yes/No. Does the client report a disabling condition?

DisabilityResponse_x For each disability type (chronic health, developmental, HIV/AIDS, mental health, physical, substance abuse), does the client report having such a disability? For substance abuse, separate responses are allowed for alcohol abuse, drug abuse, or both. All other disability types are Yes/No. These variables exhibit some form of nonresponse for less than 10% of the included sample.

IndefiniteAndImpairs_x Yes/No. For each disability type, is the client's disability indefinite and does it impair the client? For individuals with a recorded disability of the particular type, this variable is always recorded in HMIS. However, the record may be a form of nonresponse, with nonresponse sometimes exceeding 50% of the subpopulation which reports having such a disability.

ReceivingServices_x Yes/No. For each disability type, is the client receiving services related to the disability? For individuals with a recorded disability of the particular type, this variable is always recorded in HMIS. However, the record may be a form of nonresponse, with nonresponse sometimes exceeding 25% of the subpopulation which reports having such a disability.

DocumentationOnFile_x Yes/No. For each disability type, is there documentation on file regarding this disability? For individuals with a recorded disability of the particular type, this variable is always recorded in HMIS. However, the record may be a form of nonresponse, with nonresponse sometimes exceeding 60% of the subpopulation which reports having such a disability.

DomesticViolenceVictim Yes/No. This variable exhibits some form of nonresponse for less than 10% of the included sample.

WhenOccurred When did domestic violence occur? For individuals recorded as domestic violence victims, this variable is always recorded in HMIS. It is split into bins for within 3 months, more than 3 months, more than 6 months, and more than a year. This variable exhibits some form of nonresponse for approximately 6% of those listed as domestic violence victims.

CurrentlyFleeing Yes/No. Is the client currently fleeing domestic violence? For individuals recorded as domestic violence victims, this variable is always recorded in HMIS. However, it exhibits nonresponse for over 70% of those listed as domestic violence victims.

GeneralHealthStatus Is the client's general health status poor, fair, good, very good, or excellent? This variable exhibits some form of nonresponse for over 60% of the included sample.

Income/Benefits Variables

Employed Yes/No. Is the client employed? This variable exhibits some form of nonresponse for over 20% of the included sample.

EmploymentType Is the employment full-time, part-time, or seasonal? For individuals recorded as employed, this variable is always recorded in HMIS. However, it exhibits nonresponse for over 60% of those listed as employed.

NotEmployedReason If the client is not employed, are they looking for work, unable to work, or not looking for work? For individuals recorded as not employed, this variable is always recorded in HMIS. However, it exhibits nonresponse for over 50% of those listed as not employed.

Income Yes/No. Does the client report receiving income from the given source? HMIS contains fields for: IncomeFromAnySource, Earned, Unemployment, SSI, SSDI, VADisabilityService, VADisabilityNonService, TANF, GA, SocSecRetirement, OtherIncomeSource. These variables have response rates of essentially 100% in the included sample.

Income Amounts How much income does the client receive per month from the given source? This is treated as a categorical variable, with categories for NA (less than 1% of all observations), and categories corresponding to the log value of the variable, rounded to the nearest integer.

BenefitsFromAnySource Yes/No. A small number of observations ($< 2\%$) exhibit some form of nonresponse.

SNAP Yes/No. Is the client enrolled in SNAP? This variable has a response rate of essentially 100% in the included sample.

RentalAssistanceOngoing Yes/No. This variable has a response rate of essentially 100% in the included sample.

OtherBenefitsSource Yes/No. Does the client receive benefits from any other source? This variable has a response rate of essentially 100% in the included sample.

InsuranceFromAnySource Yes/No. Does the client have health insurance from any source? About a third of the included sample exhibits some form of nonresponse.

Medicaid Yes/No. Is the client covered by Medicaid? This variable has a response rate of essentially 100% in the included sample.

Medicare Yes/No. Is the client covered by Medicare? This variable has a response rate of essentially 100% in the included sample.

VAMedicalServices Yes/No. Is the client covered by VA Medical Services? This variable has a response rate of essentially 100% in the included sample.

Homeless/Residence History Variables

ResidencePrior What was the client's living situation prior to enrolling in PSH? This variable has the same categories as the destination variable, including private housing, other homeless programs, or places not meant for human habitation. A small number of observations ($< 3\%$) exhibit some form of nonresponse.

ResidencePriorLengthOfStay Was the client at their previous residence for more than 1 week, more than 1 month, more than 90 days, or more than 1 year? A small number of observations ($< 5\%$) exhibit some form of nonresponse.

TimesHomelessPastThreeYears In the past three years, was the client homeless once, twice, three times, or more than three times? About two-thirds of the included sample exhibits some form of nonresponse.

MonthsHomelessPastThreeYears For how many months was the client homeless in the last three years? Responses of greater than one year are included as a single group. About three-quarters of the included sample exhibits some form of nonresponse.

HousingStatus Was the client homeless, at imminent risk of losing housing, at-risk of homelessness, stably housed, homeless only under other federal statutes, or fleeing domestic violence? A small number of observations ($< 4\%$) exhibit some form of nonresponse.

OutOfState Yes/No. Was the client's last residence outside of California? This variable is missing for approximately 20% of the included sample.

OutOfCounty Yes/No. Was the client's last residence outside of Los Angeles County? The answer is inferred from the ZIP code of the last reported residence. This variable is missing for approximately 20% of the included sample.

ReferredToPH Yes/No. Is there a record in HMIS of the client being referred to permanent housing from another homelessness project?

PreviousEnrollmentN The number of previous project enrollments by this client that are recorded in HMIS.

PreviousShelterEnrollmentN The number of previous emergency shelter enrollments by this client that are recorded in HMIS.

<u>Other Variables</u>

EntryMonth Month of the year in which the client was enrolled.

EntryDayOfWeek Day of the week in which the client was enrolled.

8 References

Bernet, A., Warren, C., and Adams, S. (2015), "Using a Community-Based Participatory Research Approach to Evaluate Resident Predictors of Involuntary Exit from Permanent Supportive Housing," *Evaluation and Program Planning*, Vol. 49, Apr. 2015, pp 63-69.

Breiman, L. (2001), "Random Forests," *Machine Learning*, 45, pp. 5-32.

Chiland, E. (2016), "Measure HHH: Angelenos OK $1.2 Billion Bond to Tackle Homelessness," *Curbed Los Angeles*, Nov. 9, 2016. Available at https://la.curbed.com/2016/11/9/13574446/homelessness-ballot-measure-hhh-housing-bond-pass.

City of Los Angeles (2016), *Comprehensive Homeless Strategy*, Los Angeles, CA: Author. Available at https://cityclerk.lacity.org/lacityclerkconnect as Council File 15-1138-S1.

Collins, S. E., Malone, D. K., and Clifasefi, S. L. (2013), "Housing Retention in Single-Site Housing First for Chronically Homeless Individuals with Severe Alcohol Problems," *American Journal of Public Health*, 103, S2, pp. S269-S274.

Culhane, D. P., and Metraux, S. (2008), "Rearranging the Deck Chairs or Reallocating the Lifeboats?: Homeless Assistance and Its Alternatives," *Journal of the American Planning Association*, Vol. 74, Iss. 1. Available at http://works.bepress.com/dennis_culhane/51/.

Flaming, D., Burns, P., and Matsunaga, M. (2009), "Where We Sleep: Costs when Homeless and Housed in Los Angeles," Los Angeles, CA: Economic Roundtable.

Flaming, D., Lee, S., Burns, P., and Sumner, G. (2013), "Getting Home: Outcomes from Housing High Cost Homeless Hospital Patients," Los Angeles, CA: Economic Roundtable.

Garcetti, E. (2018), "State of the City 2018," City of Los Angeles. Available at https://www.lamayor.org/sites/g/files/wph446/f/landing_pages/files/2018_SOTC_AS_PREPARED_FOR_DELIVERY.pdf.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2017), *An Introduction to Statistical Learning With Applications in R* (7th printing), New York: Springer Texts in Statistics. Available at http://www-bcf.usc.edu/~gareth/ISL/.

Hunter, S. B., Harvey, M., Briscombe, B., and Cefalu, M. (2017), "Evaluation of Housing for Health Permanent Supportive Housing Program," Santa Monica, CA: RAND Corporation. Available at https://www.rand.org/pubs/research_reports/RR1694.html.

Ishwaran, H., Gerds, T. A., Kogalur, U. B., Moore, R. D., Gange, S. J., and Lau, B. M., (2014), "Random Survival Forests for Competing Risks," *Biostatistics*, Vol. 15, No. 4, pp. 757-773.

Ishwaran, H., and Kogalur, U. B. (2018), "Random Forests for Survival, Regression, and Classification (RF-SRC)," R package version 2.6.1.

Ishwaran, H., Kogalur, U. B., Blackstone, E. H., and Lauer, M. S. (2008), "Random Survival Forests," *The Annals of Applied Statistics*, Vol. 2, No. 3, pp. 841-860.

Klein, J. P., and Moeschberger, M. L. (2005), *Survival Analysis: Techniques for Censored and Truncated Data*, Springer Science & Business Media.

Kogalur, U. B. and Ishwaran, H. [cited May 26, 2018], "Theory and Specifications," *Random Forests for Survival, Regression, and Classification: A Parallel Package for a General Implementation of Breiman's Random Forests*, Authors. Available at https://kogalur.github.io/randomForestSRC/theory.html.

Los Angeles Homeless Services Authority [cited May 25, 2018], "Los Angeles Continuum of Care." Available at https://www.lahsa.org/coc/.

Lipton, F. R., Siegel, C., Hannigan, A., Samuels, J., and Baker, S. (2000), "Tenure in Supportive Housing for Homeless Persons with Severe Mental Illness," *Psychiatric Services*, 51, 4, pp. 479-486.

Martinez, T. E., and Burt, M. R. (2006), "Impact of Permanent Supportive Housing on the Use of Acute Care Health Services by Homeless Adults," *Psychiatric Services*, 57, 7, pp. 992-999.

Office of Community Planning and Development, U.S. Department of Housing and Urban Development (2018), "System Performance Measures Programming Specifications," Version 2.2. Available at https://www.hudexchange.info/resources/documents/System-

Performance-Measures-HMIS-Programming-Specifications.pdf

Pearson, C., Montgomery, A. E., Locke, G. (2009), "Housing Stability Among Homeless Individuals With Serious Mental Illness Participating in Housing First Programs," *Journal of Community Psychology*, 37, 3, pp. 404-417.

R Core Team (2018), "R: A language and environment for statistical computing," Vienna, Austria: R Foundation for Statistical Computing. Available at https://www.R-project.org/.

Rubin, D. B. (2004), *Multiple Imputation for Nonresponse in Surveys* (Vol. 81), John Wiley & Sons.

Schmid, M., Wright, M. N., and Ziegler, A. (2016), "On the Use of Harrell's C for Clinical Risk Prediction via Random Survival Forests," *Expert Systems with Applications*, 63, pp. 450-459.

Smith, D. (2018), "Statistical Error Inflated Last Year's Homeless Count by More Than 2,700," *Los Angeles Times*, May 18, 2018. Available at http://www.latimes.com/local/lanow/la-me-ln-homeless-count-correction-20180518-story.html.

Toros, H. and Flaming, D. (2016), "Silicon Valley Triage Tool: Identifying and Housing High-Cost Homeless Residents," Los Angeles, CA: Economic Roundtable.

U.S. Department of Housing and Urban Development (2016a), "2014 HMIS Data Standards: Data Manual," Version 5.1. Available at https://www.hudexchange.info/resource/3824/hmis-data-dictionary/.

U.S. Department of Housing and Urban Development (2016b), "HMIS Data Standards: Data Dictionary," Version 5.1. Available at https://www.hudexchange.info/resource/3824/hmis-data-dictionary/.

U.S. Department of Housing and Urban Development (2016c), "HMIS CSV Format Specifications," Version 5.1. Available at https://hudhdx.info/VendorResources.aspx.

U.S. Department of Housing and Urban Development (2017), "People in Permanent Supportive Housing in the United States," 2016 Annual Homeless Assessment Report to

Congress: Part 2. Available at https://www.hudexchange.info/programs/hdx/guides/ahar.

U.S. Department of Housing and Urban Development [cited May 27, 2018], "Continuum of Care (CoC) Program Eligibility Requirements." Available at https://www.hudexchange.info/programs/coc/coc-program-eligibility-requirements/.

U.S. Department of Housing and Urban Development [cited May 28, 2018], "HMIS Requirements." Available at https://www.hudexchange.info/programs/hmis/hmis-requirements/.