Prioritizing Equity: Incorporating equity into optimal housing intervention allocations.

Trevor Gratz

Anu Zan

Abstract

The number of people experiencing homelessness in the United States is growing, and there are serious consequences for living unsheltered. At the same time, federal, state, and local governments devote a substantial amount of funding to addressing this persistent problem. Given this growth, governments may need to be more efficient with existing funding. To that end, recent advances leveraging machine learning and optimization by Kube et al., (2023) promise to more efficiently allocate scarce housing interventions to households. However, governments bound by legal and ethical constraints may not be able to implement these solutions if efficiency gains come at the cost of violating the Fair Housing Act. In this paper, we advance the work by Kube et al. (2023) in the following ways:

1. We derive the relationship between the optimal allocation of housing interventions to households and the conditional average treatment effect.
2. We document where the Department of Housing and Urban Development's policy may not align with the goal of achieving the greatest number of exits from homelessness.
3. We develop a simple exploratory data analysis technique for determining when allocating interventions to maximize exits from homelessness may violate the Fair Housing Act.
4. We develop a method for dynamically addressing potential disproportionate assignments of housing interventions by Protected Classes, thereby ensuring the Fair Housing Act is not violated.

Problem Statement

On a single night in January of 2023, more than half a million people in the United States were homeless, and the number of people experiencing homelessness is growing (De Sousa et al., 2023). Homeless housing situations matter: people experiencing homelessness have a 3.5 times all-cause mortality rate higher than otherwise observably similar housed individuals (Meyer et al, 2023). In recognition of these stark differential outcomes, governments across the United States have allocated substantial funds to assist households in finding stable housing. For instance, in 2023 the federal grant program to fund local homeless programs, Continuums of Care, allocated $3.16 billion dollars (HUD, 2024), which is on top of the local dollars raised by governments and philanthropy (Lee 2021). Despite the significant amount of funds, funding levels remain far below the need (Kim, 2023), and there is widespread national and local recognition that homelessness remains a persistent problem. Until these funds materialize, governments will need to do more with less, that is, assist more people in finding stable housing with the same amount of resources.

To that end, recent applications of machine learning have attempted to aid service providers by A) predicting who might be more likely to exit homelessness absent services and B) given a fixed amount of services help prioritize which households experiencing homelessness are most likely to benefit from those services (Kube et al., 2023). Kube et al. accomplish this by leveraging a Bayesian Additive Regression Tree to predict the likelihood that a household experiencing homelessness would exit homelessness if they received services and if they did not receive services. These probabilities are used as weights in an "assignment problem" optimization algorithm with fixed capacity (i.e., the number of available housing interventions). The authors find that were homeless services allocated based on predicted outcomes and optimized in accordance with which services are available, then 5.5% fewer households would request additional homelessness services within the next two years.

An additional consideration the authors note is that reallocation of services is not pareto-improving; in other words, while the system overall exits more households from homelessness, not all households are better off after reallocation. For example, the allocation of resources differed under historical assignment and the author's optimal assignment by gender[1]: one of seven federally Protected Classes under the Fair Housing Act.[2] For governments bound by ethics, public opinion, and the legal system, an inequitable allocation of resources may limit the adoption of algorithmic solutions that improve the system overall but have deleterious impacts on specific Protected Classes. Indeed, this is not just a theoretical consideration: evidence suggests that the Vulnerability Index used to construct many Continuum of Cares' prioritization algorithms may exhibit gender and racial bias (Kithulgoda et al., 2022). Thus, the primary question/problem this project seeks to address is:

---

[1] Both males and females are predicted to be better off under optimal assignment, however, relative to females males housing outcomes are expected to improve more.

[2] For more information see: https://www.hud.gov/program_offices/fair_housing_equal_opp/fair_housing_act_overview

1. Is there a reformulation of the algorithm described by Kube et al. that can both maximize the number of households exiting homelessness and introduce flexibility for governments to balance optimizing the overall system's performance with equity concerns for sub-populations?

Background and Data

Data on homelessness in this paper come from a large Continuum of Care. Data on homelessness is maintained in the Homelessness Management Information System (HMIS). HMIS maintains records on individuals when they enter and exit services for homelessness, the type of service, where they exit services to, demographic information, household information, and employment information.

When a client contacts a provider of homeless services about needing assistance, the client enters Coordinated Entry. The Coordinated Entry process works as follows: 1) the client is assessed to determine if they are literally homeless or fleeing domestic violence, if so, 2) a diversion conversation follows, in which an alternative housing arrangement absent assistance from a provider is sought, if none can be found 3) a prioritization interview is conducted, 4) the results of the interview generate a prioritization score for housing interventions, and if 5) another household leaves services and a spot opens up for housing interventions, the household with the highest prioritization score is selected for that intervention. That is, periodically throughout the year, the Coordinated Entry system decides which households will get scarce housing interventions based on a household's prioritization score. Consistent with HUD's notice CPD-17-01, prioritization scores in this Continuum of Care are intended to reflect "barriers to accessing housing" and "vulnerability factors". However, as we show in our methods section, the criteria outlined in CPD-17-01 may be inconsistent with the goal of maximizing the number of households exiting homelessness.

The three primary housing interventions available for households on the prioritization list are rapid rehousing (RRH), transitional housing (TSH), and permanent supportive housing (PSH). Rapid rehousing provides short and medium-term rental assistance, TSH offers supportive services and housing for up to 24 months, and PSH provides long-term housing with no set end date.[3] Due to the scarcity of slots in these programs, most households who enter the Coordinated Entry System and are literally homeless do not receive any of these three housing interventions.

Per policy, providers must contact households that do not receive an intervention within 90 days of their prioritization interview to conduct an exit interview or to reaffirm their continued request for a housing intervention. However, a significant proportion of households in the training data cannot be re-contacted and are recorded as having an unknown exit destination in HMIS. Missing data in the homelessness information management systems is a well-documented and challenging problem (Chelmis et al., 2021). Nevertheless, we follow HUD's performance metric definitions and define a successful exit as a confirmed household acquisition

[3] For more information see: https://www.hudexchange.info/homelessness-assistance/coc-esg-virtual-binders/coc-program-components/coc-program-components-overview/

of stable housing. We further fold in a requirement that no member of a household reenter homelessness within six months of exiting the system.

Information about the household is collected during the prioritization interview. Demographic, economic, health, and living situation information is collected during these interviews. All household members are interviewed, but the head of the household is identified during the interview. Thus, depending on the variable, information can be used from the head of the household or, in some cases, aggregated from the household. For example, the head of the household is asked about their veteran status, but so are each of the household members. A variable is constructed for the head of the household, but also one that considers whether any household members were veterans. We construct three feature sets: one with all head-of-household and household variables, including potentially colinear variables such as the veteran's status, one where variables are primarily taken from the head of the household,[4] and one where variables are primarily taken from aggregating household information.[5] The majority of these variables are categorical or binary. We consider two different encodings: weight of evidence encodings (De La Bourdonnye & Daniel, 2021) and one hot encoding. For continuous variables, we standardize the data and, in the case of skewed distributions, take log transforms.

Methodology

Learning an optimal and equitable prioritization algorithm for homelessness interventions has three primary steps:

1. Learn an algorithm to predict the likelihood of exiting homelessness, which can be used to estimate the likelihood of exiting homelessness whether a household received any of the three interventions or none (Kube et al., 2023)
2. Describe an algorithm for assigning housing interventions based on the results from step 1 (Kube et al., 2023)
3. Learn weights to be used in the algorithm described in step 2 to balance housing intervention assignment across a specific sub-population

<u>Objective Function Derivations</u>

To motivate the methodology for determining an optimal and equitable treatment assignment, we first consider a simpler objective: maximizing the number of households exiting homelessness.

Let y_i be the event that household i exits homelessness, $x_{i,j}$ be the 0/1 indicator for whether household i was assigned housing intervention j, where when j = 1 the household is assigned no interventions. Let S be the number of households overall exiting homelessness. Then:

[4] Not all variables can be taken from the head of the household alone. For example, we include a variable for the number of people in the household and the number of children.

[5] Not all variables can be reasonably aggregated from the household. For example, there isn't a reasonable way to aggregate race and ethnicity. For these we use the head-of-household information.

$$y_i = \begin{cases} 0 \; with \; probability \; 1 - p_i \\ 1 \; with \; probability \; p_i \end{cases}$$

$$S = \sum_{i=1}^{N} y_i$$

As stated above our goal is to maximize the expected number of households exiting homelessness by manipulating the treatment assignments, $x_{i,j}$.

$$\max_{x_{i,j}} E[S] = \max_{x_{i,j}} E\left[\sum_{i=1}^{N} y_i\right] = \max_{x_{i,j}} \sum_{i=1}^{N} E[y_i] = \max_{x_{i,j}} \sum_{i=1}^{N} p_i * 1 + (1 - p_i) * 0$$

$$= \max_{x_{i,j}} \sum_{i=1}^{N} p_i$$

By the law of total probability and that we require housing intervention assignments to be mutually exclusive:

$$p_i = \sum_j P(y_i = 1 | x_{i,j}) \, P(x_{i,j}) = \sum_j p_{i \,|\, x_{i,j}} \, x_{i,j}$$

However, since we are optimizing over $x_{i,j}$, i.e. determining these assignments, then $P(x_{i,j})$ is either 1 or 0 depending on whether or not we assign individual *i* to treatment *j*, thus $P(x_{i,j})$ can be replaced by $x_{i,j}$. Our objective function is the following:

$$\text{Eq.1)} \quad \max_{x_{i,j}} \textstyle\sum_{i=1}^{N} p_i = \max_{x_{i,j}} \sum_{i=1}^{N} \sum_j p_{i \,|\, x_{i,j}} \, x_{i,j}$$

Estimating Exit Probabilities

For the first step in solving this optimization problem, we need to learn the values of $p_{i \,|\, x_{i,j}}$. That is, we need to estimate household *i*'s likelihood of exiting homelessness when assigned to treatment *j* (note again that when j=1, we consider this the case most households realize, namely, no housing intervention). Specifically, we need a flexible model that considers non-linear heterogeneous treatment effects so that we can find a mapping from a household's attributes and treatment assignments to their likelihood of exit (Hill & Murray, 2020).

If Z_i is a vector of household information for household *i,* and each of the $x_{i,j}$ then our goal is to estimate:

$$\text{Eq. 2)} \quad y_i = f(Z_i)$$

Such that we can recover the $p_{i \,|\, x_{i,j}}$. For example, if we were to use logistic regression this would become:

$$y_i = \begin{cases} 0 \; if \; p(z_i) < 0.5 \\ 1 \; Otherwise \end{cases}$$

Where

$$p(z_i) = \frac{1}{1 + e^{-(z_i^T \beta)}}$$

To learn this mapping, we tested different classification algorithms, feature sets, feature encodings, features transformation, and used the combination that performed best at predicting out-of-sample observation outcomes. Specifically, we separated our dataset using an 80/20 train-test panel split[6] and performed 5-fold cross-validation on the training data over our search space. Our search space included:

- Algorithms – Elastic Net and Random Forest
- Hyperparameter tuning
- Three different feature spaces
  - Features primarily from the head of household
  - Features primarily aggregated from the households
  - Both head of household and aggregate household features
- Two different feature encodings
  - One Hot Encoding
  - Weight of Evidence
- Whether to perform feature selection
- Whether to perform principal components analysis

For principal component analysis, we kept as many components as necessary to retain 95% of the original variation in the data.

<u>Optimization Set Up and It's Relationship To Conditional Average Treatment Effects</u>

Once the mapping that corresponds to Equation 2 was built, we recovered the probability that household *i* exits homelessness if they were assigned each of the *j* housing interventions. Using these estimated probabilities, we developed an optimization model based on Equation 1.

In practice, slots for housing interventions become available at different points in time, and because we want to develop and optimize a model that incorporates equity considerations for sub-populations, we needed to modify equation 1. Following Kube et al. (2023), we optimized across weeks, though different providers may have preferences for how often to assign slots when they become available. That is the optimization problem can be defined as:

- Define x_{ijt} as an indicator variable that represents whether household *i* in week *t* is assigned intervention *j*.
- Use p_{ijt} to represent the probability that household *i* exits homelessness in week *t* if assigned intervention *j*.
- Define C_{jt} as the number of available slots for intervention *j* in week *t*.

---

[6] Households may appear multiple times in the data. Therefore, for the 80/20 split we sampled household IDs, rather than individual household enrollments.

- Define j=1 to be the "housing intervention" of no intervention, of which there are no capacity constraints.

Without incorporating equity considerations, the optimization model becomes:

$$\text{Eq. 3)} \max_{x_{ijt}} \sum_{i=1}^{N} \sum_{j} p_{ijt}\ x_{ijt}$$

$$\text{Subject to: } \sum_{i} x_{ijt} = C_{jt}\ \forall \text{j}\ \neq 1$$

$$\sum_{j} x_{ijt} = 1$$

$$x_{ijt} \in \{0, 1\}$$

This optimization problem is formulated as an Integer Program. Kube et al. (2023), show that it can be reformulated as a weighted bipartite b-matching problem that can be solved in polynomial time. However, as indicated above, this optimization formulation does not account for considerations of equity. Before moving to a weighted reformulation of Equation 3 that accounts for equity based on minimizing sub-group differential treatment assignment, we consider a simpler version of Equation 3. This simpler version sheds light on a useful exploratory data analysis technique for assessing the extent to which an unadjusted Equation 3 may lead to differential treatment assignments i.e., inequitable assignment to housing interventions.

Consider the objective function from Equation 3 without its time element and when there is only one housing intervention type available. Further suppose that there is only one housing intervention available among two households (x_1, x_2). For ease of interpretability, we will subscript the receipt of a housing intervention as *tx* and not receiving an intervention as *ntx*. The objective function then becomes:

$$\max_{x_{1,tx},\, x_{2,tx},\, x_{1,ntx},\, x_{2,ntx}} p_{1,tx}x_{1,tx} + p_{1,ntx}x_{1,ntx} + p_{2,tx}x_{2,tx} + p_{2,ntx}x_{2,ntx}$$

$$\text{Subject to: } x_{1,tx} + x_{1,ntx} = 1$$

$$x_{2,tx} + x_{2,ntx} = 1$$

$$x_{1,tx} + x_{2,tx} = 1$$

We can substitute the first two constraints into the objective function to yield:

$$\max_{x_{1,tx},\, x_{2,tx},\, x_{1,ntx},\, x_{2,ntx}} p_{1,tx}x_{1,tx} + p_{1,ntx}(1 - x_{1,tx}) + p_{2,tx}x_{2,tx} + p_{2,ntx}(1 - x_{2,tx})$$

$$\text{Subject to: } x_{1,tx} + x_{2,tx} = 1$$

Rearranging terms yields:

$$\max\ x_{1,tx}\left(p_{1,tx} - p_{1,ntx}\right) + x_{2,tx}(p_{2,tx} - p_{2,ntx}) + p_{1,ntx} + p_{2,ntx}$$

$$\text{Subject to: } x_{1,tx} + x_{2,tx} = 1$$

Since $p_{1,ntx}$ and $p_{2,ntx}$ are additive scalars, they can be dropped from the optimization without changing the solution. Thus, we are left with:

$$\max \quad x_{1,tx}\left(p_{1,tx} - p_{1,ntx}\right) + x_{2,tx}(p_{2,tx} - p_{2,ntx})$$

$$\text{Subject to: } x_{1,tx} + x_{2,tx} = 1$$

Thus, to maximize this function, we need to select the household with the largest change in their probability of exiting homelessness when they receive an intervention versus when they do not. Without loss of generality, this derivation can be carried out to show that when there are K interventions available and N households, the optimal solution is to select the K households for interventions that maximize the change in probability of successful exits from homelessness. Note that because we have a binary outcome, i.e. $y_i \sim Bernoulli(p_i)$ and a binary treatment the $p_{i,tx} - p_{i,ntx}$ represent the conditional average treatment effect. In other words, the solution to this simplified optimization problem is to select households for housing interventions with the largest conditional average treatment effects.

Hill and Murray (2020) argue that the power of machine learning techniques for estimating causal effects is due to their ability to estimate non-linear response surfaces that involve interactions between variables. However, these approaches still suffer from the traditional statistical problem of selection bias. In other words, unobserved characteristics may determine a household's likelihood of being selected for a housing intervention and their likelihood of exiting homelessness. While selection bias can never be ruled out completely, as by definition it is an unobservable process, the programmatic design of housing intervention provision makes this concern less salient. Specifically, HUD requires all Continuums of Cares to conduct intake assessments and, from those assessments, prioritize and score households on their need for housing interventions.[7] Given the much greater demand for housing interventions than capacity, conditional on observables, i.e., prioritization assessment scores, housing interventions are likely to be assigned when capacity exists. In other words, conditional on observables, the provision of housing interventions to households is pseudo-random depending upon the availability of services rather than some unobserved characteristic of the household.

However, as HUD makes clear in CPD-17-01, households should be prioritized based on the presence of significant functional impairment, high utilization of crisis services, being unsheltered, vulnerability, risk of continued homelessness, vulnerability to victimization or "other factors … that are based on the severity of need". While these need not be mutually exclusive, prioritizing households based on their severity of need is not the same as prioritizing households based on their conditional average treatment effects and is thus not entirely aligned with the goal of maximizing the number of households exiting homelessness.

Optimization with equity weights

Another challenge with this approach for allocating scarce housing interventions is that to the extent that conditional average treatment effects vary by subgroups, e.g., Protected Classes,

[7] For more details see: https://www.hud.gov/sites/documents/17-01CPDN.PDF

treatment assignment will be inequitable. The first step to understanding the potential of unequal treatment assignment is to examine the distribution of the conditional average treatment effects by subgroups. This approach can aid in understanding which potential subgroups may receive differential assignments to housing interventions so that we can incorporate these concerns into our optimization algorithm. We do just this in Figure 2 of the Evaluation section. However, we next turn to a more systematic approach to addressing potential inequitable assignment to housing interventions.

To account for equity considerations between arbitrary sub-populations *g* and keep the problem within polynomial time, we modify Equation 3 as follows.

- Define α_g as the proportion of group *g* in the total homeless population based on historical data.
- Calculate $\gamma_{gt}*$, the proportion of group *g* assigned to some treatment up to time t^*, and compare it with α_g to assess disproportionality.
- $E_{g,t^*} = \gamma_{gt^*}/\alpha_g$ is the proportion of assignments to treatment for group *g* at time t*.
- $r_{gt^*} = E_{g,t^*}/E^B_{g,t^*}$ is the risk ratio of group *g* relative to the base group B at time t*.

When the non-reference group has been assigned a housing intervention more often than the base group relative to their population size, then r_{gt^*} is greater than one. When assignment is equal then r_{gt^*} is one, and when assignment is lower than the base group's r_{gt^*} is less than one. Thus, we modify Equation 4 by adding the following term : $+(\sum_{j\neq 1} C * (1 - r_{gt^*})x_{it^*j}))$.

$$\text{Eq 4)} \max_{x_{ijt}} \sum_{i=1}^{N} \sum_j p_{ijt}\ x_{ijt} + (\sum_{j\neq 1} C * (1 - r_{gt^*})x_{it^*j}))$$

$$\text{Subject to: } \sum_i x_{ijt} = C_{jt}\ \forall \text{j} \neq 1$$

$$\sum_j x_{ijt} = 1$$

$$x_{ijt} \in \{0, 1\}$$

The problem introduces a new hyperparameter C, which is a weight designed to increase or decrease the likelihood that a household is assigned a housing intervention based on whether the group they belong to is over or underrepresented in past assignments to housing interventions. (Note the exclusion of j = 1, the state of no housing interventions). For example, if the group is overrepresented, then $1 - r_{gt^*}$ is negative and has the effect of penalizing an assignment to x_{it^*j}. With this setup, the problem becomes finding a value of C that will center the long-term risk ratios around 1 for all groups. C can be estimated via grid search. In other words, we perform the optimization problem for years 2017 – 2022 for different values of C. We then pick the C value that minimizes the risk ratio's absolute distance from 1. In particular, we pick the minimum value of C so the risk ratios are all +/- 0.05 of 1.

Evaluation

The likelihood of successfully exiting homelessness is unfortunately low, with fewer than 20% of households exiting homelessness successfully and not reentering services within six months. To evaluate the performance of different learning algorithms along with the parameters that define the search space indicated in the methods section, we used the area under the Receiver Operating Curve (AUROC) as our performance metric (Hossin, & Sulaiman 2015). Using 5-fold cross-validation in the training data, the out-of-sample mean AUROC for a random forest classifier and an L1 and L2 penalized logistic regression model (i.e., elastic net model) were 0.840 and 0.839, respectively. Focusing on the random forest classifier, it had an accuracy of 0.879, 0.627 precision, and 0.271 recall. The 5-fold cross-validated confusion matrix for this classifier is in Table 1.

Table 1: Random Forest Classifier Confusion Matrix for Validation Data

		Predicted	
		No Exit	Exit
Actual	No Exit	14105	382
	Exit	1634	606

What is clear from the recall value and the confusion matrix is that the model has a high false negative rate. Extracting the feature importances from the model, we observe that the single most predictive feature is whether a household was enrolled in rapid rehousing. Indeed, 90.3% of true positives (i.e., predicted to exit and actual exit) were enrolled in rapid rehousing. The centrality of the three housing interventions to the model's predictive performance can be observed by examining the confusion matrix from the same random forest model but with the housing interventions removed from the feature set.

Table 2: Random Forest Classifier Confusion Matrix for Validation Data, No Intervention Features

		Predicted	
		No Exit	Exit
Actual	No Exit	14422	65
	Exit	2203	37

Without these variables the model is unable to identify nearly any households that successfully exit. The differences in these two models speaks to three key important findings, 1) the importance of housing interventions for successfully exiting homelessness, 2) the lack of virtually any other features that predict exits, and 3) the degree to which measurement errors are likely impacting the results.

On this last point, further investigation has shown that of households that do not receive a housing intervention, 69% have unknown exit destinations.[8] On the other hand, of those enrolled in a housing intervention, only 21% had unknown exit destinations. Enrollment in a housing intervention helps ensure that program managers can contact households, producing less measurement error. However, since the majority of households are not enrolled in a housing intervention and do not have an exit interview completed, this inflates the importance of housing interventions for exiting homelessness. Measurement errors of this type and magnitude are unfortunately common in HMIS systems (Chelmis et al., 2021). The office of the state auditor of California in April of 2024 released a report stating that it was unable to evaluate the cost-effectiveness of California's homelessness response system in part because of the large number of unknown exit destinations observed in the state's HMIS data (Auditor of the State of California, 2024).

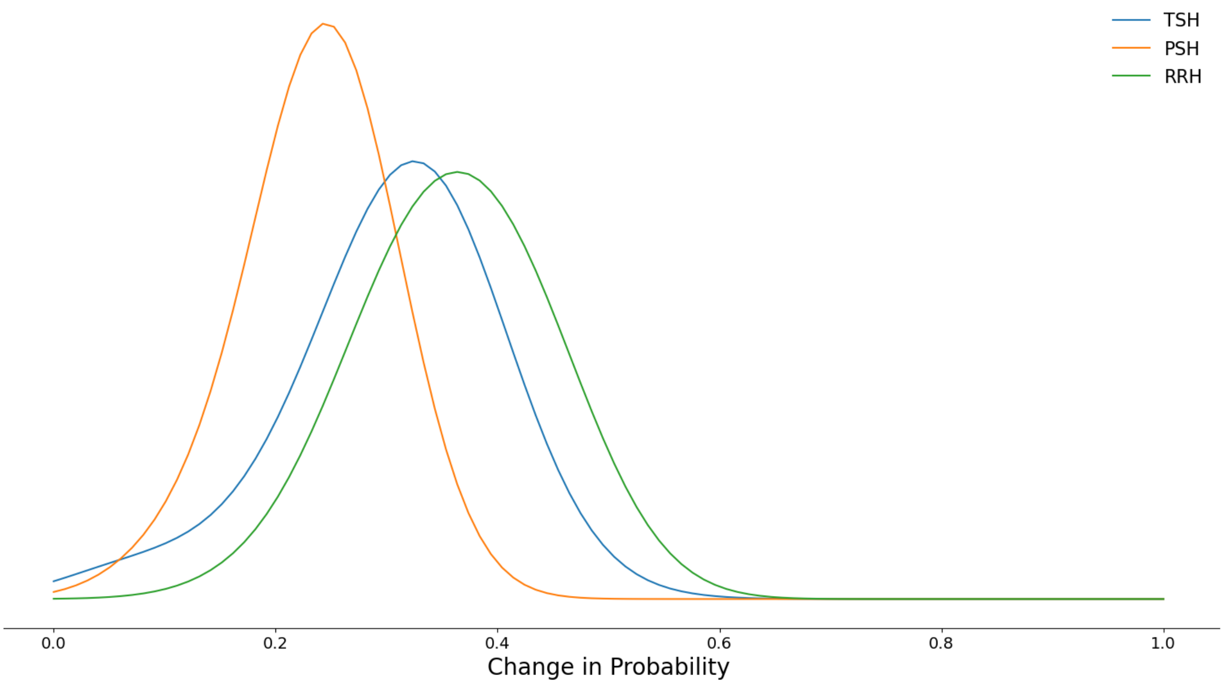
Moreover, the other component of a successful exit, lack of reentry into the system, may also suffer from measurement error. Reentry statistics require the ability of program managers to consistently track households across time. How often households are successfully tracked to the same IDs over time remains unknown. For these reasons, we do not recommend program managers pursue models to prioritize interventions until exit interviews are nearly universally completed and an audit has been performed which verifies that households are consistently linked to the same ID overtime by different program providers.

Nevertheless, we use our data to illustrate different ways to address potential inequitable assignment of housing interventions were these models to be used in practice. To be clear, these models were not used in the assignment of housing interventions, but rather, we used the model estimates to simulate the behavior of our optimization algorithm. Our random forest classifier achieved a 0.835 AUROC on the testing data, which indicates that the model does not overfit the data.

We can estimate the conditional average treatment effects with our trained model by varying the treatment assignment indicators and generating the predicted probabilities of exiting homelessness under each condition. For instance, in Figure 1 below, we create kernel density plots of the conditional average treatment effects for the three housing interventions relative to no interventions.

[8] Exit destinations can be unknown when no exit interview was completed (66.7%), the data wasn't collected (2.4%), or the client prefers not to answer or doesn't know (0.1%)

Figure 1: Conditional Average Treatment Effects of Three Housing Interventions



Interestingly, permanent supportive housing, the intervention often considered the most intensive and costs the most, exhibits the lowest conditional average treatment effect. It is possible that this intervention is less effective on average than rapid re-housing and transitional housing, or this could be an artifact of unaccounted-for confounding.

To highlight how examination of conditional average treatment effects could yield insight into potential inequitable assignments of housing interventions, we focus on one protected class: gender. In Figure 2 below, we create a kernel density plot of the conditional average treatment effects of enrollment in rapid re-housing projects by gender identity.

Figure 2: Rapid Re-Housing Conditional Average Treatment Effects by Gender

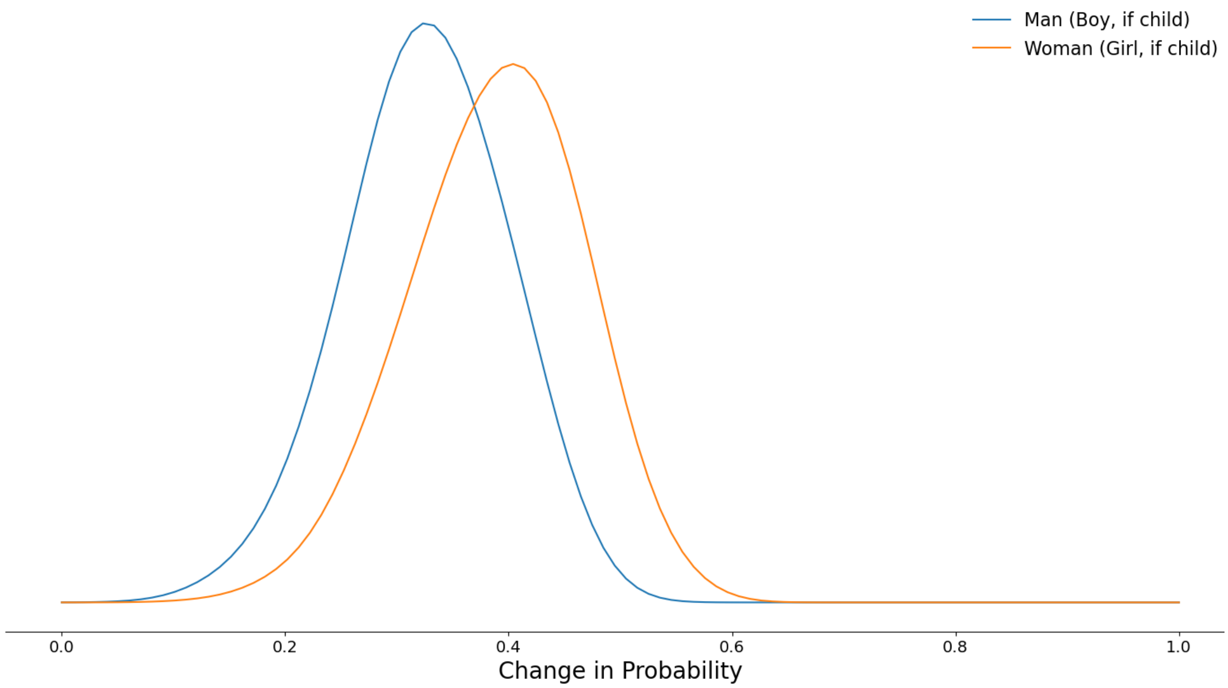
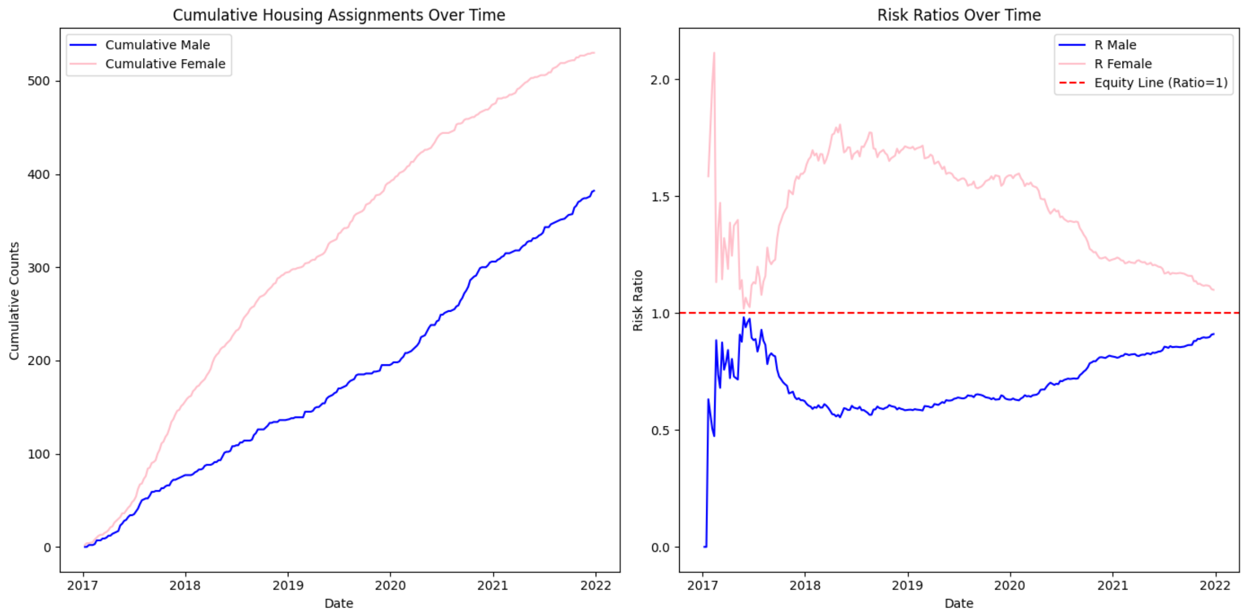


Figure 2 shows that the conditional average treatment effects for female-identifying heads of households are larger than for males. The two primary indications of this finding are A) optimizing using conditional average treatment effects could maximize the number of households exiting homelessness by preferentially selecting female-identifying heads of households and B) this would be a clear violation of the Fair Housing Act. Potential violations from the Fair Housing Act necessitate an optimization algorithm that considers differential assignment to housing interventions.

Consistent with Equation 3, our initial optimization strategy involves assigning housing interventions based on predicted probabilities of exiting homelessness, and ensuring each household is eligible for only one intervention to prevent overlap. This optimization is carried out on a weekly basis, and we track the cumulative assignment of housing interventions to male and female heads of households. We note that during our period of study, 55.8% of households that requested services were headed by female-identifying clients, compared to 44.2% of heads of households identifying as male. Figure 3 below depicts the cumulative assignment of housing interventions between 2017-2021 selected by Equation 3 to male and female-identifying heads of households and the risk ratios by gender.

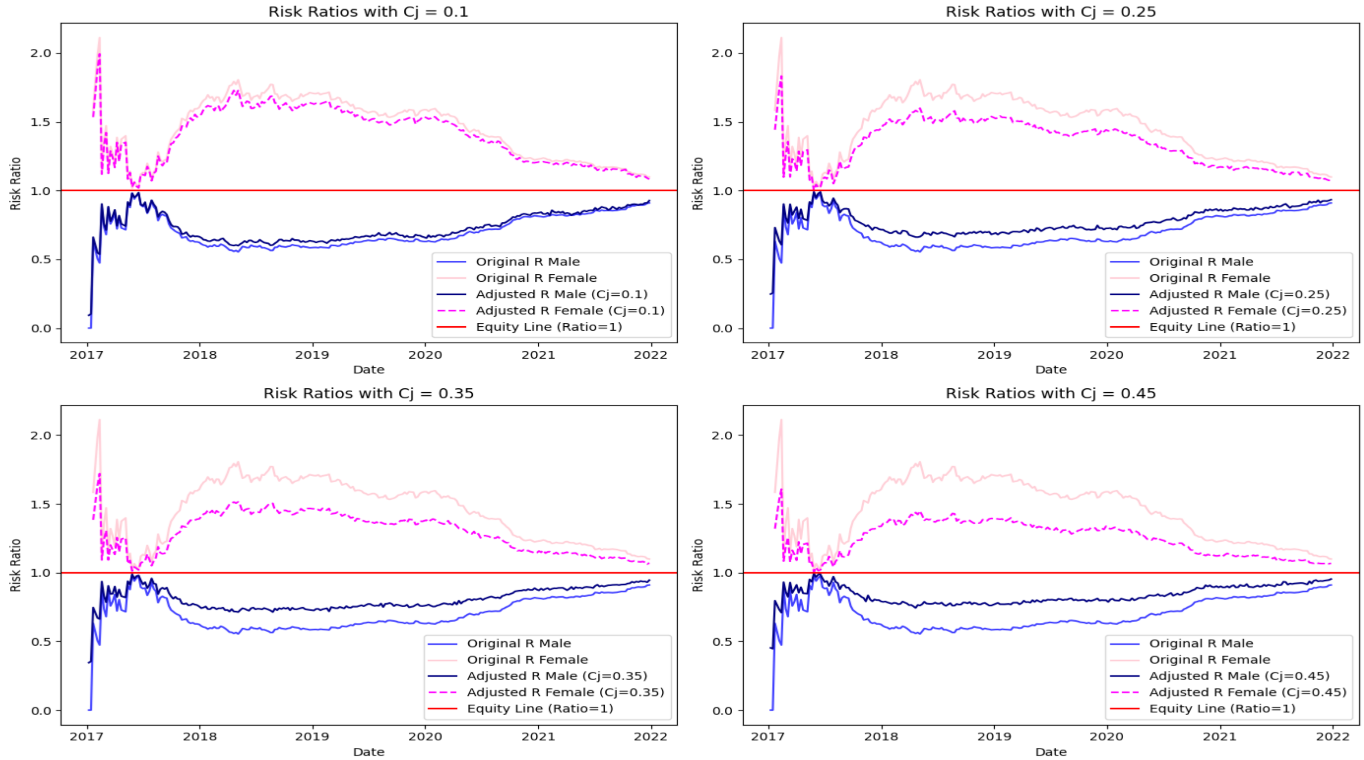
Figure 3: Unweighted Optimization of Housing Intervention Assignment By Gender



Initially, housing assignments are starkly unequal with most resources allocated to females. This is depicted by the large value of the risk ratio for female-identifying heads of households. The risk ratio for males begins at zero, indicating their exclusion from the initial housing assignments. Over time, the difference in risk ratios attenuates with the optimization algorithm selecting more households identifying as male for housing interventions. By the end of 2021, females, which make up 55.8% of households requesting services, still received a higher proportion of housing interventions, 58.11%, than males. Females' final risk ratio relative to males was 1.097, indicating that they are more likely to be assigned a housing intervention than males relative to their representation among all households requesting services. Thus, to avoid violating the Fair Housing Act, this algorithm needs to incorporate weights to account for the differential selection of housing assignments by the optimization process.

Following Equation 4, we introduce the additive weighting factor, *C*, and adjust the allocation process dynamically to balance the distribution of interventions between male and female subpopulations. The optimization process is dynamic in that we calculate the risk ratios using the cumulative assignments to housing interventions up to the current week. C is multiplied by one minus the risk ratio of the group that household *i* belongs to, $1 - r_{gt^*}$, which means that positive values of C will up weight groups with risk ratios below 1 and down weight values above 1. The optimal value of *C* is unknown, and we treat it as a hyperparameter and search over potential values of C in the range from 0.1 to 0.7. Figure 4 below depicts the cumulative assignment of housing interventions to male and female-identifying heads of households using four different weighting factors (0.1, 0.25, 0.35, and 0.45.) and the risk ratios by gender.

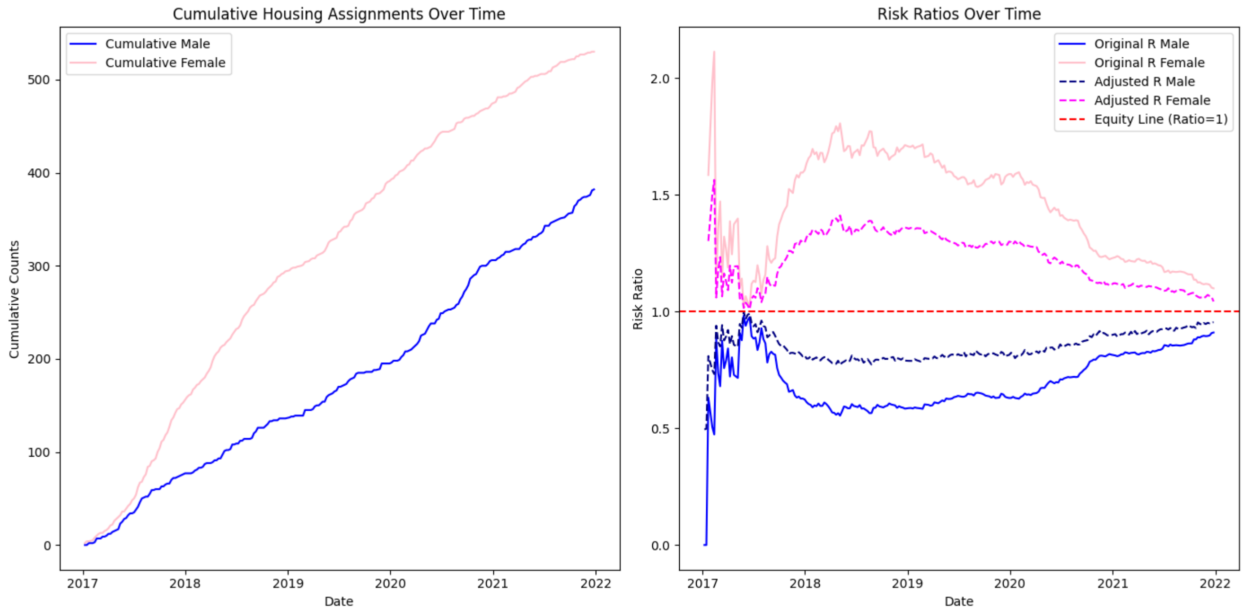
Figure 4: Weighted Optimization of Housing Intervention Assignment by Gender



In Figure 4 above, with C set to 0.35, 427 male heads of households and 484 female heads of households were assigned to housing interventions, achieving a risk ratio of 0.94 for males. Relative to males, female heads of households had a risk ratio of 1.07. This value of *C* demonstrates a balanced distribution moving towards parity, i.e., reducing the disparity in risk ratios. Increasing the value of *C* continues to move the distribution towards parity. When *C* is set to 0.45, 434 male heads of households and 477 females were assigned housing interventions with risk ratios of 0.96 and 1.06, respectively.

In theory, we could ensure balanced risk ratios by setting the weight, C, so large that housing interventions are assigned in a two-week cycle where one week, all interventions go to males, and the next week all interventions go to the female heads of households. This is an undesirable behavior in the algorithm, as a household's ability to obtain a housing intervention could be solely determined by which week they happened to request services. Thus, we have set a heuristic of selecting the minimum *C* such that the long run risk ratios are within +/- 0.05 of 1. This is achieved when *C* is set to 0.49, with the risk ratio for male heads of households at 0.96 and for female heads of households at 1.04. Figure 5 displays the cumulative housing assignments over time for males and females, along with their accompanying risk ratios.

Figure 5: Weighted Optimization (C=0.49) of Housing Intervention Assignment by Gender



Aside from the specific end values of the risk ratios, the trajectory of the risk ratios may also be important. In particular, the substantial movement of assignments and risk ratios in the first few weeks indicates that the algorithm will need a burn-in period, a period where no weighting is used. Since the noisy behavior is due to low assignment sample sizes used in the construction of the risk ratios, the historical assignments could also be used instead of a burn-in period. Moreover, these trajectories highlight that risk ratios may move around significantly based on the order in which households arrive requesting services and when. Administers of housing interventions will need to continually monitor assignments and potentially adjust the weighting factor to respond to changes in the populations requesting interventions, intervention effectiveness, and to ensure an equitable assignment of housing interventions.

Conclusion

As homelessness increases throughout the United States, local governments may be asked to do more with less: ration scarce housing interventions efficiently, so as to exit the greatest number of households experiencing homelessness. Machine learning algorithms may be uniquely suited to model the complex process households have for exiting homelessness (Kube et al., 2023). However, before they can be applied there are significant technical and legal limitations that must be addressed. Specifically, data limitations, especially data quality for outcome measures, may preclude the use of data and algorithmic-driven decisions. Moreover, we show that the benefits of many machine learning models, i.e., the ability to learn non-linear response surfaces, may unintentionally violate the Fair Housing Act. In a two-stage model, such violations can be dynamically adjusted to ensure consistent assignment to housing interventions across Protected Classes. Lastly, we point out that even if the issue of equitable assignment to housing interventions is solved, HUD policy, CPD-17-01, may be inconsistent with the goal of maximizing the number of households exiting homelessness. Any effort to alleviate the detrimental consequences of living unhoused via efficient allocation of housing interventions will need to address these data and legal limitations. If these challenges are overcome, administrators of housing interventions will need to continually monitor assignments and potentially make changes in the allocation process to ensure equitable access to scarce housing interventions.

References

Auditor of the State of California (2024, April). Homelessness in California. Report 2023-102.1

Chelmis, C., Qi, W., & Lee, W. (2021, April). Challenges and opportunities in using data science for homelessness service provision. In Companion proceedings of the web conference 2021 (pp. 128-135).

De La Bourdonnaye, F., & Daniel, F. (2021). Evaluating categorical encoding methods on a real credit card fraud detection database. arXiv preprint arXiv:2112.12024.

De Sousa, T., Andrichik, A., Prestera, E., & Rush, K., Tano, C., Wheeler, M., & Abt Associates (2023). The 2023 annual homelessness assessment report (AHAR) to Congress. *The US Department of Housing and Urban Development*

Hill, J., Linero, A., & Murray, J. (2020). Bayesian additive regression trees: A review and look forward. Annual Review of Statistics and Its Application, 7, 251-278.

Hossin, M., & Sulaiman, M. N. (2015). A review on evaluation metrics for data classification evaluations. International journal of data mining & knowledge management process, 5(2), 1.

Kim, G. (2023, January 26). Ending homelessness in king county will cost billions, regional authority says. *The Seattle Times*.

Kithulgoda, C. I., Vaithianathan, R., & Parsell, C. (2022). Racial and gender bias in self-reported needs when using a homelessness triaging tool. Housing Studies, 1-13.

Kube, A. R., Das, S., & Fowler, P. J. (2023). Fair and efficient allocation of scarce resources based on predicted outcomes: implications for homeless service delivery. Journal of Artificial Intelligence Research, 76, 1219-1245.

Lee, D. (2021). Is need enough? The determinants of intergovernmental grants to local homeless programs. *Journal of Urban Affairs*, *43*(7), 995-1009.

Meyer, B. D., Wyse, A., & Logani, I. (2023). *Life and death at the margins of society: the mortality of the US homeless population* (No. w31843). National Bureau of Economic Research.

U.S. Department of Housing and Urban Development (HUD) (2024). Biden-Harris Administration Awards $3.16 Billion in Homelessness Assistance Funding to Communities Nationwide. HUD No. 24-018