Socio-Economic Planning Sciences 99 (2025) 102212

Contents lists available at ScienceDirect

Socio-Economic Planning Sciences

journal homepage: www.elsevier.com/locate/seps

Tradeoffs between equity and efficiency in Coordinated Entry of homeless housing systems

Dashi I. Singham *, Mary McDonald, Robert Elliot

Operations Research Department, Naval Postgraduate School, United States of America

ARTICLE INFO

Keywords:
Queueing models
Discrete-event simulation
Homelessness
Threshold policies

ABSTRACT

Coordinated Entry is designed to provide a single access point for people experiencing homelessness to enter a Continuum of Care. Some regions, including the County of San Francisco, offer a scoring assessment to potential program participants to determine their relative need for housing. This score is then used to route participants to the appropriate housing resource. The goal is to achieve equity in assigning the most intensive resources to the most vulnerable people, while balancing efficiency in quickly housing as many people as possible. We create a queueing simulation model to directly compare policies that attempt to balance tradeoffs between equity and efficiency. In particular, we model scoring threshold policies for routing participants, as well as jockeying policies for reallocating participants as additional housing inventory becomes available. Finally, we apply an extensive experimental design to rigorously compare the policies while incorporating wide-ranging input uncertainty. This produces recommendations on how effective routing policies can be designed under changing conditions, with applications to healthcare and other tiered service systems.

1. Introduction

Homelessness remains a pervasive and complex issue across many regions in the United States, presenting significant challenges for local governments and service providers. Addressing this issue requires a comprehensive approach due to the different needs of the homeless population and the varying availability of housing resources. Each county may possess different types of housing solutions, ranging from emergency shelters to long-term supportive housing, aimed at serving populations with distinct needs. In this context, the Coordinated Entry (CE) system, which serves as a centralized process designed to streamline the management of homeless care systems, has emerged as a pivotal mechanism to manage the intake, assessment, and allocation of housing resources to the appropriate persons in the system [1].

Homelessness in San Francisco has escalated into a critical issue, driven by high housing costs, economic disparities, and systemic social challenges. Despite being a region of substantial economic power, it hosts a significant portion of the nation's unsheltered homeless population, with recent estimates indicating over 38,000 homeless individuals on any given night. Moreover, approximately 35% of the 38,000 homeless people in the city are chronically homeless, ten percentage points higher than the rest of the nation that stands at 25% [2]. The COVID-19 pandemic further exacerbated this crisis, straining already limited resources and highlighting the urgent need for innovative solutions and substantial investment in housing and supportive services.

This paper will build a queueing simulation model based on the San Francisco CE system to test the effects of routing policies. In this section, we describe CE systems, present our research contributions, and provide a brief literature review of related research.

1.1. Coordinated entry

CE is "a consistent, community-wide process to match people experiencing homelessness to available community resources that are the best fit for their situation" [3]. In the United States, the Department of Housing and Urban Development (HUD) established guidelines so each Continuum of Care (CoC) may manage CE operations [3]. To some extent, counties follow a first-in-first-out system so that those who have been waiting longest should be served first. However, prioritization is also an important part of allocating housing. HUD has defined four main parts of CE: accessibility, a standardized assessment approach, prioritization, and a referral process to housing.

Upon arrival at CE, the CoC should administer an assessment process to determine the needs of a potential program participant. The assessment method should be uniformly administered, though there are certain type of distinctions allowed by HUD. For example, unaccompanied youths may go through a separate assessment process from adults. Prevention is also an important component of the CoC, so people at

* Corresponding author.
E-mail address: dsingham@nps.edu (D.I. Singham).

https://doi.org/10.1016/j.seps.2025.102212
Received 15 October 2024; Received in revised form 24 March 2025; Accepted 26 March 2025
Available online 4 April 2025
0038-0121/Published by Elsevier Ltd.

imminent risk of becoming homeless may be eligible for prevention assistance. While HUD requires that CoCs use an assessment tool, the tools may vary across regions [3].

Prioritization is explicitly a part of HUD policy, in that "the group of persons with the highest priority is offered housing and supportive services projects first" [3]. In the prioritization step, clients with extreme levels of need may be served first if they are unlikely to survive without housing. As a result, there are often different types of housing allocated for clients with different types of needs. Criteria to evaluate priority could include health challenges and vulnerability to illness and death, high use of crisis or emergency services (emergency departments or jails), and vulnerability to victimization (for example, human trafficking or sexual violence). While some CoCs will maintain separate priority lists according to subpopulations, HUD suggests that CoCs may operate more efficiently using a single list ranked by priority with all persons in the region [3]. There is a potential tradeoff between aligning people exactly with the correct resource, against efficiency in moving people into resources as quickly as possible.

1.2. Research contributions

We present three main research contributions. The first contribution develops a queueing simulation model for CE systems with multiple classes of housing based on priority. This model is inspired by the system used in San Francisco and builds on the literature for routing of participants through homeless care systems (to be discussed in Section 1.3). Our second contribution implements and compares policies which attempt to balance equity (alignment of participants with the correct type of housing resource) and efficiency (reduced waiting times for housing) in allocating resources. These policies assess the effect of changing the thresholds required for participants to receive different tiers of housing, and involve reallocating members of one housing queue to another to potentially increase throughput via jockeying. Because only simple threshold models are amenable to closed-form analytical methods, we employ simulation to estimate how these more-complex policies might be implemented. The third contribution constructs a robust design of experiments to assess how the model performs under a range of uncertain inputs. This analysis reveals how different input factors affect the tradeoffs between equity and efficiency.

1.3. Literature review

The San Francisco Bay Area has been the focus of many streams of research due to high levels of homelessness. Paul et al. [4] conduct a qualitative study on the relationships between race and homelessness on older adults in Oakland, CA. Research in Sacramento, CA, suggests COVID-19 may have had as great or greater economic impact on the homeless population than the effects of the virus itself [5]. There has been much research on the general health outcomes of interventions in the homeless population, for example, with HIV testing [6]. Singham et al. [7] develop simulated queueing models for the flow of people through a CoC in Alameda County, CA, which resides in the East Bay area of San Francisco and includes the city of Oakland. The details of the types of housing options are modeled to test different allocations of resources across the system. This work is motivated by efforts to improve racial equity for support of the homeless population as described in [8]. Direct exploration of the amount of shelter needed as a backstop for a lack of housing was tested in [9] using a quantile field estimation method.

Simulation and statistical research are often used to model healthcare outcomes as they apply to homeless populations. For example, Chapman et al. [10] employs simulation to model the transmission of COVID-19 in homeless shelters, and Ingle et al. [11] uses a probability model to project the amount of shelter beds needed due to COVID-19. Reynolds et al. [12] uses discrete-event simulation to assess the quality of care for homeless patients in a health clinic. Dai and Zhou [13] show the mutual causality between homelessness and poor health outcomes in the United Kingdom.

Optimization models have also been used with success to determine resource allocation. Kaya et al. [14] employ mixed integer linear programming to determine the optimal allocation and expansion of resources in a system modeling youths at risk of being trafficked due to homelessness. Maass et al. [15] solve a mixed integer linear program under different cases to determine the optimal placement of shelters for people at risk of being trafficked. Most recently, Burgess et al. [16] optimize the investment into housing and shelter over time using a fluid model while incorporating policy-based constraints.

Additionally, the importance of using queueing models to align resources with the needs of the population has been analyzed. Rahmattalabi et al. [17] study the effect of matching policies to allocate resources to program participants using a queueing model. Furthermore, there have been a few streams of research on how to manage resources to support populations that are homeless after a natural disaster. Liao et al. [18] uses agent-based simulation to determine the effects of humanitarian logistics structures used to shelter populations, while Souza et al. [19] employs multi-period optimization for allocation of people to shelters and corresponding use of relief items.

Threshold policies have been studied in the queueing literature to determine when optimal policies may exist. Teh and Ward [20] demonstrate the asymptotical optimality of threshold policies for dynamic routing in queueing networks, whereby customers are routed to a particular server as long as its queue is smaller than some threshold. Armony [21] study dynamic routing of customers who are allocated to the faster servers first, i.e., sending the customer at the front of the line to the server which is the fastest. The author demonstrates when this policy is optimal under a regime that balances quality of service with efficiency. Argon et al. [22] explore dynamic routing policy heuristics in systems with multiple different server types, where some customers must be served by dedicated servers, and others can be served by any server. Optimal threshold policies under uncertainty in the arrival and departure rates are studied in [23]. Chen et al. [24] determine optimal threshold routing policies for multi-class server systems with heterogeneous customer types. They are able to show conditions when pairwise dominant policies exist between assigning a customer to the faster server, a slow server, or rejecting the customer from the system.

A recent stream of research has directly focused on routing policies in queueing systems for at-risk youth in New York City. Kaya and Maass [25] prioritize different levels of risk in order to allocate limited shelter beds to runaway and homeless youths at risk of becoming trafficked. Their model manages a global probability of abandonment across all heterogeneous types of clients, and also suggests the number of beds that should be allocated to bound the waiting time of clients. The paper's objective with equity is to increase access for higher-risk clients while remaining fair to clients who may be less vulnerable. A second paper, Kaya and Maass [26], simulates different queueing policies that route runaway youths to different shelters based on their combination of needs. Different shelters have different capabilities, and while ideally each client would be quickly served by a shelter that can address all their needs, in reality limited resources must be allocated in order to best align the right resources with each client. The risk of abandonment from the queue and downstream effects of a mismatch of resources to client needs are also modeled. This paper addresses the tradeoffs between equity and efficiency through a comparison of different queueing policies, and, similar to our approach, relies on simulation to account for complex implementation details that are experienced in practice.

The remainder of the paper is outlined as follows. Section 2 discusses CE with an emphasis on aspects unique to San Francisco, while Section 3 presents the simulation model and threshold policies tested. Section 4 displays the experimental results and Section 5 contains the main conclusions.

2. San Francisco Coordinated Entry system

Although the County of San Francisco is located directly across the San Francisco Bay from Alameda County, each county has its own approach for structuring CE and allocating housing options. General principles must apply across all CE systems according to federal guidelines [3]. These guidelines include the following: the county must conduct a fair assessment of clients entering the system, establish a clear method for prioritization of limited resources, and plan methods of referral of clients to housing resources. However, each county may employ different specific approaches and policies in their CE systems given the particular needs of their population. Singham et al. [7] describes an approach in Alameda County, where there may be up to eight different pathways taken through the system to address different types of need. Some people may require extensive intervention and access to temporary shelter while waiting for housing to become available, while others may only require rental assistance or other financial support to remain housed. This section describes the CE approach taken by San Francisco.

When we refer to a client, person, or program participant, we are generically referring to either an individual, or a family unit treated as a single group to be housed together. Program participants refer to people who are enrolled in the housing system. San Francisco employs a multi-stage approach for clients arriving at a CE access point. There are three general categories of clients. The first is an adult, who is someone over the age of 18, or someone under 18 who has been legally emancipated. The second is a family, meaning adults with minor children or adults who are pregnant. The third group consists of transitional age youths between ages 18–24, ages 25–27 (if they entered CE before the age of 25), or youths under 18 who are legally emancipated. Clients falling into more than one category are eligible for services from each potential system. There may be different CE locations designated for each type of client to enable matching of resources appropriately. Survivors of violence and people who are pregnant are able to enter CE at any access point.

In San Francisco, the CE system facilitates a consistent assessment process and prioritizes individuals based on their Housing Primary Assessment (HPA), which evaluates individuals' needs and vulnerabilities to determine their priority for housing services. The HPA score considers three main factors: chronicity (the length and recurrence of an individual's homelessness), vulnerability (an individual's health, safety and risk of harm), and barriers to housing (legal issues, income, or resource availability). The threshold for HPA scores is set based on the anticipated amount of housing inventory that will be available in 90 days. The county may become aware of new housing being built or made available in order to plan around anticipated inventory. The idea is to assign those with the highest level of need to the housing with the most-intensive services. Those who score above the minimum threshold for their household type are placed in Housing Referral Status, making them eligible for housing queues based on their HPA score. Those below the minimum threshold are not prioritized and do not enter the housing queues. The specific threshold scores can vary from zero to 160, with a score of 160 implying maximum need as evaluated across the assessment. The HPA asks questions about where the client resided over the past year, whether the client is experiencing physical or sexual violence, whether the client has disabling conditions (disability, mental health problems, substance abuse issues), and information about income and arrest records. The responses are then compiled into an HPA score [27].

The model for Alameda County implemented in [7] assumed that each client arrived to the system with a particular type of need, and their homelessness would only be resolved once they received assistance in a particular form suited to them. However, the model for San Francisco takes into account some flexibility among clients who could be supported by multiple types of resources. Clients with extremely high service needs may still require high levels of intervention. However, some clients may benefit from lower levels of intervention that are available sooner, rather than waiting in a long queue for more expensive services. Additionally, San Francisco attempts to adjust the thresholds for different queues based on the anticipated amount of resources available. This means that if additional housing is anticipated to be available in a particular housing category, more people could be directed towards that resource, even if they are ideally suited for another resource with limited capacity.

The overall flow of the San Francisco CE system is represented in Fig. 1. When people enter CE, there are many potential approaches that can be taken to resolve their homelessness. "Problem solving" involves using preventative measures like helping the client find or obtain transportation to housing through their network. Other short term resources may be employed so that the person may quickly return to housing. The goal with problem solving is to enable the person to find their own housing solution by linking them with different types of connections to employment or community services. The housing primary assessment is then collected, and this determines eligibility for housing. Rapid rehousing (RRH) delivers limited rental assistance and services to enable self-sufficiency and keep people housed. Permanent supportive housing (PSH) is "affordable housing designed for people experiencing homelessness with chronic illnesses, disabilities, mental health issues, and/or substance use disorders who have experienced long-term or repeated homelessness". Usually, there are additional services provided along with housing in PSH [29]. We note that transitional living spaces may also be available through separate programs [30] which serve people outside of our considered PSH and RRH resources, but clients may still seek PSH while at these facilities. Transitional living and shelter spaces are an important part of the homeless response system, but are not considered in this queueing model for housing.

There are many other types of specialized resources available, for example temporary shelter may exist while clients are waiting for a housing solution. People are in a "housing referral status" when they are assigned to a queue for either PSH or RRH. The process of housing navigation assists people with collecting documents like identification and income verification so that they are able to move into a housing unit when it is available. Many other solutions may be offered to people seeking services who do not qualify for a housing referral status, though this research focuses exclusively on PSH and RRH.

We now offer more details on how the HPA is used to determine which queue a potential program participant should join. A threshold policy is applied to the HPA score to determine if a participant's score is high enough to qualify for PSH. Fig. 2 shows how a distribution of potential housing scores could determine where a client is routed in the system, with the highest scores referred to PSH (the green area) and moderate need referred to RRH (the blue area). The left-most dotted line shows the minimum threshold required to be placed in a Housing Referral Status, while those scoring below this threshold are placed in a Problem Solving status. The right-most dotted line representing the threshold between PSH and RRH can be modified as time progresses based on anticipated changes to housing inventory with the intention of keeping queue lengths balanced across housing resources. However, this could potentially lead to inequity in allocating the most comprehensive housing to those who need it most. In this paper, we will explore this tradeoff between equity and efficiency using threshold policies for simulated queueing systems.

The San Francisco Department of Homelessness and Supportive Housing determines the ranges of score values which determine the type of housing a person is eligible for. For example, in the fall of 2022, a family with a score between 75 and 160 was eligible to join the queue for PSH, while a score of 50–74 makes the family eligible for RRH. For scores lower than 50, CE will continue to work with them through problem solving or housing referral to ensure each person has a pathway to a housing solution. The ranges for housing referral assignment will be different for adults, veterans, and youths. The

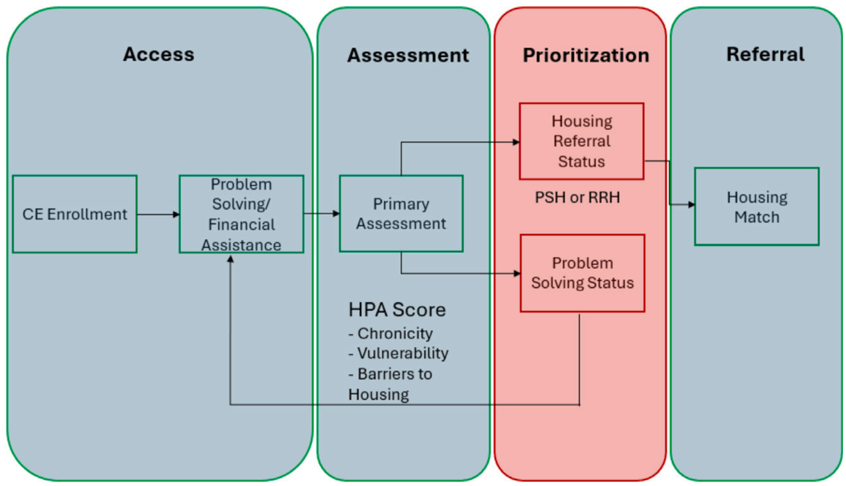

Fig. 1. General flow of San Francisco CE System.
Source: From [28], adapted from [1].

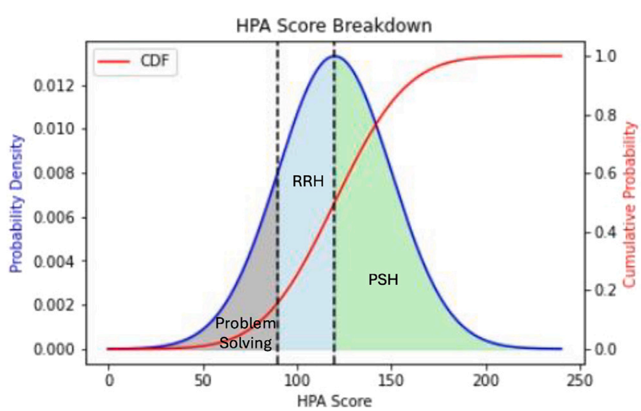

Fig. 2. Illustration of thresholds used to allocate housing resource decisions, with green HPA scores referred to PSH, blue scores to RRH, and gray scores to a Problem Solving status.
Source: From [28].

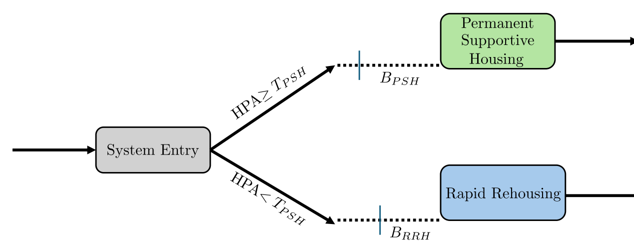

Fig. 3. The queueing model for the CE system.

thresholds that determine which assessment scores are referred to PSH and RRH queues are adjusted over time, depending on the anticipated supplies of housing. This means that if more PSH units will be made available, the lower bound on the threshold for PSH can be reduced to accommodate more people. Next, we construct a simulation model of this system to implement different threshold policies.

3. Simulation model for policy testing

This section presents a simulation model for the CE system described in Section 2. As described in the literature review, housing systems for homeless populations can be thought of as queueing systems. Rather than modeling the details of a particular CE location, we model the aggregate process of people entering the system, waiting for a housing resource, entering housing, then eventually leaving the resource. Singham et al. [7] was the first such approach to model the aggregate flow of people through an entire CoC, with a focus on Alameda County in the East Bay of San Francisco. This model had eight separate pathways for the different housing types through the system. The San Francisco model in this paper only considers two main housing types. We model additional complexity by considering a threshold routing policy based on vulnerability of the client, with the threshold changing when additional housing is anticipated. In both counties, CE is just the first entry point of the CoC, and the correct allocation of resources to people entering at this stage is critical to ensuring they are served appropriately downstream.

3.1. Model layout

In order to model the CE process in San Francisco, we build a simulation model to represent the flow of people through the system. Simulation is an effective tool for analyzing systems with random arrivals and constrained resources. Discrete-event simulation is an effective way of constructing queueing systems because of its efficiency in modeling large numbers of entities flowing through constrained resource systems. It provides a way to track the time people wait for housing in congested systems, and enables fast testing of different housing policies.

We model this system using a process flow paradigm in Simio to track the flow of entities through the system. Simio is a discrete-event simulation software program which enables intelligent objects to be used to model complex queueing systems [31]. When clients arrive to the system, an HPA score is simulated for them. This score, combined with a threshold for eligibility for PSH, determines whether they will receive PSH or RRH. Because we lack specific distributional HPA data, we assume all scores in the lower Problem Solving range will be treated as eligible for RRH housing to simulate a worst-case scenario where RRH is in high demand. This results in a focused model where we focus on changing the threshold between PSH and RRH. Fig. 3 shows the general queueing structure modeled. Thus, we can consider parallel server systems which contain different types of housing to serve different populations. We define some key notation. Let T_{PSH} be the determined threshold of need so that clients with HPA scores higher than T_{PSH} are deemed eligible for PSH. Otherwise, clients are directed to the queue for RRH.

Next, we describe some buffer parameters that determine jockeying in the system. We allow both input buffers for PSH and RRH to have infinite capacity. Let B_{PSH} be a parameter related to the input buffer for the PSH server and B_{RRH} be a parameter related to the input buffer for the RRH server. Let Q_{PSH} and Q_{RRH} be state variables that represent the number of people in the PSH and RRH queues, respectively. We will compare the number in the queue to B_{PSH} and B_{RRH} to determine if jockeying should occur, where a client may move from the queue for RRH to the queue for PSH if they have waited longer than 90 days, and $Q_{RRH} > B_{RRH}$ and $Q_{PSH} < B_{PSH}$. In Fig. 3 we show that these parameters represent some fixed number in the buffer, though they are not the actual buffer limits.

This queueing system faces many unique challenges. The system is often unstable, in that the arrival rate to the system is higher than the service rate. We first discuss what causes a high arrival rate by looking at data collected on the homeless population. Point-In-Time (PIT) counts are conducted to estimate the number of people experiencing homelessness which includes the unsheltered and sheltered population. PIT counts usually occur overnight by having outreach workers walk through the city to record the number of homeless individuals encountered. The following PIT information is found at [32]. While San Francisco has had high levels of people experiencing unshelteredness, the PIT count on January 30, 2024 suggested a 1% decline in the unsheltered population since the 2022 PIT count, and a 16% decrease since 2019. Even though the unsheltered population declined during this time, since 2022 total homelessness increased by 7% because the sheltered population increased by 18%. Overall, the amount of people served through housing and shelter has increased over time, with an estimated 7500 clients served and exiting the homelessness system to successful independent housing between 2022 and 2024. However, it is estimated that three people become homeless for each person housed. While approximately 20,000 people seek homeless services each year, it is not clear how many people are new arrivals versus returns to the system [32]. Despite the increase in the number served by the system, the lack of recent decline in the unsheltered population implies increased arrivals. Thus, this queueing system was likely unstable in recent years because the high arrival rate outpaced the rate that people could be served by limited housing resources. We continue to model the system as unstable to reflect known past conditions, even if recent developments could lead to stability.

Next, we discuss the service rate. Because the intention is for clients to remain housed indefinitely, service rates are very slow. In the queueing literature, many of the routing decisions in heterogeneous server systems are made based on the assumption that one server is faster than the others, and that server should be prioritized in order to maximize the flow of clients through the system. However, in our setting, we want to match clients to servers according to their need, and the more desirable server (PSH) may be much slower, in terms of service speed, due to the needs of its clientele to stay for longer periods. This motivates one main approach to reducing homelessness: increasing the amount of housing resources available. This is accomplished by increasing affordable housing, supportive housing, or shelter to serve the demand for housing. Counties face many challenges in attempting to build housing, including difficulty finding financial resources, locating physical space amid zoning regulations, or obtaining public support for new housing. Given that the quantity of housing resources available to be allocated to homeless populations is limited, counties have developed methods for allocation and prioritization of limited current and future inventory. Our model will explicitly consider the effects of adding additional housing resources to the system.

3.2. Policies

We explore three policies using our simulation model. The intent is to determine whether better results can be achieved by redirecting clients based on available resources, while also trying to keep prioritization aligned with relative need of clients, so that people with higher HPA scores still receive PSH housing as quickly as possible.

1. The first policy is the *baseline* policy. In this policy, we model the HPA threshold, T_{true} above which clients should be directed to PSH based on an assessment of need. T_{true} is meant to reflect the level of client need as measured by the HPA score that would qualify them for PSH, and is defined in more detail in Section 4.1. We set T_{PSH} equal to T_{true}, and this threshold remains constant throughout the model run, even as additional housing inventory is added to PSH. This is the policy displayed in Fig. 3.
2. The second policy is a *dynamic* policy which changes the threshold for PSH when more inventory is added, so the value of T_{PSH} begins with the value T_{true} but can change at only one time period chosen midway through the model run. In practice, San Francisco may lower the value of T_{PSH} to allow more people to enter the queue for PSH in anticipation of increased housing being available. We assume that once someone is placed in a housing referral status so that they are in the queue for a housing resource, they will remain in that queue even if the thresholds change.
3. The third policy employs *jockeying*, whereby if the queue for RRH becomes longer than B_{RRH} and the PSH queue is smaller than B_{PSH}, clients will be redirected from the RRH queue to the PSH queue (i.e., jockeying occurs from RRH to PSH if $Q_{PSH} < B_{PSH}$ and $Q_{RRH} > B_{RRH}$). The client must wait in the RRH queue for 90 days before being allowed to move to PSH.

San Francisco implements a policy similar to our dynamic policy. Housing planners across Bay Area counties also informally consider the effects of jockeying to balance queues and allow clients to be served quickly. Discussions with these planners motivate our policy choices. While the model layout in Fig. 3 for the baseline policy appears straightforward, implementing the logic for the dynamic and jockeying policy involves complex logic using add-on processes in Simio. Simio enables each client to retain its HPA score information as it travels through the housing servers, and add-on processes behind the model layout interface allow for individual jockeying decisions to be made. In all policies, the server capacity for PSH is increased at a certain time midway through the model run to represent an increase in resources. In the dynamic policy, the HPA threshold for PSH eligibility is also changed 90 days before this anticipated increase in inventory. In the jockeying policy, the model is constructed so that clients will jockey from the RRH queue to the PSH queue after 90 days if $Q_{PSH} < B_{PSH}$ and $Q_{RRH} > B_{RRH}$. This means that if the RRH queue is relatively long, clients will move to the PSH queue if it is relatively short. This will allow for balancing of queues if the PSH server system is operating more efficiently than the RRH system.

3.3. Measures of performance

In order to assess the effect of different policies, we evaluate the model according to two types of metrics: equity and efficiency. Equity ensures that people are matched with the resource that is right for them, as one of the goals of prioritization is to give those with the greatest need easier access to housing with the most services. We measure equity as the proportion of people who are matched with the correct resource given their HPA score according to the original threshold T_{true} which is determined by need, rather than anticipated inventory. Equity can be calculated as the proportion of clients who are correctly served by the resource intended for them according to T_{true}:

$$\text{Equity} = \frac{\sum_i \mathbf{1}_i^{PSH} \times \mathbf{1}_{\{\text{HPA}_i \geq T_{true}\}} + \mathbf{1}_i^{RRH} \times \mathbf{1}_{\{\text{HPA}_i < T_{true}\}}}{\#\text{ total clients}} \tag{3.1}$$

where HPA_i is the score of client i and $\mathbf{1}_i^{PSH}$ is 1 if client i is served by PSH and 0 otherwise. Similarly $\mathbf{1}_i^{RRH}$ is 1 if client i is served by RRH and 0 otherwise. We let $\mathbf{1}_{\{A\}}$ be the standard indicator function which is 1 when event A is true and false otherwise. The baseline policy that does not vary the threshold from T_{true} or redirect clients will have an equity score of 1 (equivalently, 100%). In reality, clients may not end up in the exact housing type designed for them due to friction in the system and constraints on resources affecting implementation decisions, so equity is not 100% in practice. However, this research assumes perfect implementation to establish a baseline policy for comparison. Furthermore, we note that we have explicitly chosen to model equity as the proportion of people receiving the correct resource designed for them (according to T_{true}) based on discussions we had about tiered queues with stakeholders. More generally, equity metrics could incorporate ideas of housing as many people as possible, though this was not considered in this paper.

For the dynamic policy, there is some luck in assignment based on when people arrive to the system, as the threshold can change over time. This means that it is possible for someone with a lower assessment score than T_{true} to end up in PSH if they arrive after the threshold is decreased, compared to someone with a higher score who arrived when the threshold was higher and was referred to the RRH queue. Similarly with jockeying, clients originally assigned to one type of housing based on the threshold T_{true} may end up served by a different resource than the one originally assigned to them. Thus, the dynamic and jockeying policies allow for the possibility that equity will be less than 100%.

The second measure of performance is efficiency of the system, which can be measured in different ways. One main queueing metric is the waiting time to receive service (or the time in the queue, denoted by PSH_Wait and RRH_Wait). Note that PSH_Wait refers to waits experienced by clients who should receive PSH housing based on their HPA score compared to T_{true}. RRH_Wait refers to waiting times for clients who should originally have received RRH, but could also have waited for PSH because a dynamic or jockeying policy was used. Higher efficiency implies lower wait times. A second method is to look at the overall throughput of the model, with higher numbers of clients successfully entering housing implying better efficiency. We anticipate that the dynamic and jockeying policies which redirect clients to housing servers with more space may lead to better efficiency, at the cost of reduced equity. We will use our model and associated experiments to evaluate this tradeoff.

4. Experimental results

This section describes the experiments conducted to compare the effectiveness of the three policies using our equity and efficiency metrics. The threshold for PSH may vary for different types of clients. For example, veterans housing may have different thresholds than family housing. Details on thresholds for different populations used are available at [27]. The housing inventory available to the system is updated regularly, and it can be hard to anticipate when and how much housing will be available. Because the threshold T_{true} varies across populations and over time, we conduct experiments on the simulation model varying T_{true} and other key inputs to compare the effects of the three policies under uncertain conditions.

First, we explain the model parameters chosen in Section 4.1. Section 4.2 compares the three policies in a baseline experiment. Section 4.3 describes the experimental design used to vary key policy parameters and presents the main results. Finally, Section 4.4 explores regression results and highlights factors that have large influences on the results.

4.1. Parameter values

This section describes the parameters that were used in the simulation model. First we describe values that were calibrated from data and fixed throughout the model run, and then present parameters that were varied using an experimental design. Table 1 shows the system parameters used in all model runs. The initial inventory and queue values were calibrated using values available from online data from [27]. The arrival rate and service times are estimated from general Bay Area values used in past analysis. The distribution for PSH is chosen to represent long stays in housing, while the distribution for RRH represents shorter-term assistance. The mean HPA score is estimated to be 84 from aggregate data online. Given the scores can range from 0 to 160, we choose a uniform distribution between (0,160) to allow for high variability with a mean near 84. The normal distribution was also tested, but would have to be truncated to limit observations outside the allowable range. However, we note that different distributions could easily be implemented in the model if data were available for calibration. Finally, we note that we need to model the current queue in the system at the start of the model run to avoid initialization bias. After modeling this influx of clients to the system, we run the model with a warmup of 2 years to allow entities to circulate through the system. We change the threshold at 2.5 years in anticipation of a change in inventory 3 months later at 2.75 years, and run the model for an additional year to capture the effects of the change. These values can be easily modified in the model as updated information becomes available.

Table 1
Model baseline parameters. Parameters below the line will be varied in Section 4.3.

Parameter	Value
Initial PSH inventory	11 267 units
Initial PSH queue	2500 people
Initial RRH inventory	2082 units
Initial RRH queue	2000 people
Distribution of PSH housing time	Exponential(mean 6 years)
Distribution of RRH housing time	Exponential(mean 9 months)
Distribution of HPA score	Uniform(0,160)
Model warmup period	2 years
Time of threshold change	2.5 years
Time of inventory change	2.75 years
Total model runtime	3.5 years
Arrival rate	Exponential(rate 10 people/day)
T_{true}	112
New PSH inventory	12 000
$T_{lowered}$ proportion	0.8
B_{PSH}	1300
B_{RRH}	300

Next, we discuss the parameters below the line in Table 1 that will eventually be varied in an experimental design. The arrival rate is the rate of arrivals to the system, often estimated to be 10/day, though this value is highly uncertain. We next define T_{true} as the HPA score threshold that determines if the client is eligible for PSH housing, and set the baseline as an example taken from [27]. The value of T_{true} should be based on the absolute need of the client (ignoring the current queues in the system) so that clients are appropriately aligned with the correct resource. This value may change over time as the nature of the housing resources and HPA scoring method changes, so we wanted to allow for variability in our experiments. In practice, this threshold determining client routing may vary based on current system performance, so $T_{lowered}$ is the new HPA score threshold used by CE in anticipation of new PSH housing, measured as a proportion of T_{true}. Thus, the value of T_{PSH} in Fig. 3 will start as T_{true}, and be changed to $T_{lowered} * T_{true}$ as the system progresses. We start with a baseline 0.8 (80%) to represent a 20% drop in the threshold. Finally, the jockeying policy baseline values are set as $B_{PSH} = 1300$ and $B_{RRH} = 300$, respectively.

4.2. Policy comparison

In order to compare the performance of all three policies, we conduct a one-way Analysis of Variance (ANOVA) F-test to compare the means of the three groups for the Equity, Total_Wait, and Total_Served metrics. Total_Wait is computed as the weighted sum of PSH_Wait and RRH_Wait and is measured in weeks. Total_Served is computed as the total number of clients served by either server. The ANOVA will test the

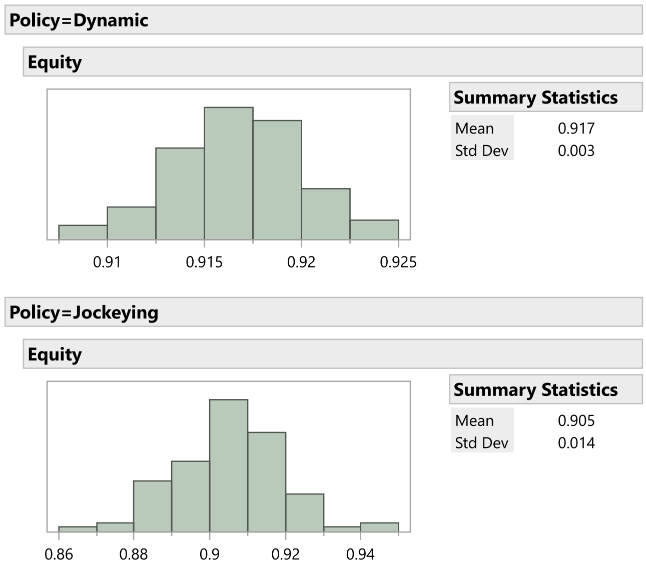

Fig. 4. Empirical distributions of equity by policy.

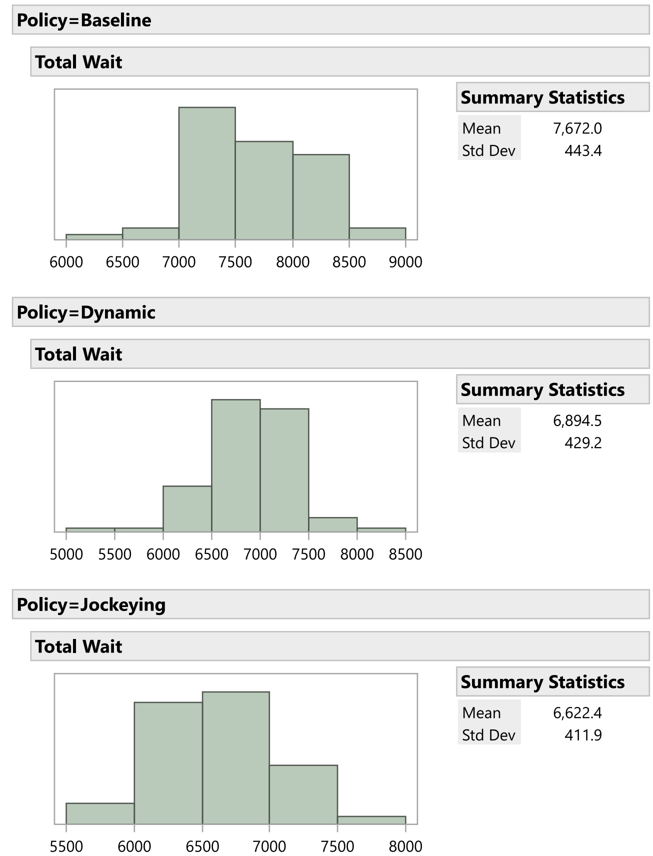

Fig. 5. Empirical distributions of total wait by policy.

hypothesis that the three group means are equal. This is followed by a Student's t-test for each pair of policies, testing if there is a significant difference between them [33].

The first assumption of the ANOVA test, independence of observations within each group, holds because each result was independently simulated. The other assumptions are normality of the results within each group and equal variance across the groups. We note that ANOVA is robust to moderate deviations from the normality and equal variance assumptions, particularly with large and equal sample sizes. One hundred observations within each group meets the large and equal sample size requirement. We next show the empirical distribution of each metric by policy to demonstrate that there is not severe departure from either the normality or equal variance assumptions.

Fig. 4 shows the distribution of Equity for the dynamic and jockeying policies. As mentioned previously, Equity, by design, will always equal 1 for the baseline policy, and a statistical test is not actually needed to demonstrate that it is significantly different from the dynamic or jockeying distributions. Fig. 4 indicates that the dynamic and jockeying distributions do not demonstrate severe departures from the normality or equal variance assumptions. Figs. 5 and 6 show the distribution of Total Wait and Total Served, respectively, by policy. These figures indicate that the distributions do not demonstrate severe departures from the normality or equal variance assumptions. We next discuss the result of the ANOVA test for each metric.

Fig. 7 shows the mean equity with confidence diamonds for each of the policies, and the non-overlapping circles illustrate that there are statistically significant differences between the policies with respect to equity. We see that the baseline policy has the best equity scores of 1 (by construction) relative to the dynamic and jockeying policies which have significantly lower equity scores, with jockeying having the lowest due to more clients being able to move from RRH to PSH.

Fig. 8 shows the mean Total_Wait with confidence diamonds for each of the policies, and the non-overlapping circles illustrate that there are statistically significant differences between the policies with respect to Total_Wait. We see that the dynamic and jockeying policies have significantly lower wait times than the baseline policy, suggesting that these policies improve efficiency at the expense of some loss in equity.

Fig. 9 shows the mean Total_Served with confidence diamonds for each of the policies, and the non-overlapping circles illustrate that there are statistically significant differences between the policies with respect to Total_Served. Again, we see a potential improvement in efficiency using the dynamic and jockeying policies because they significantly increase the total number of clients that are able to be served by the housing resources.

Table 2
Mean performance for each policy.

Policy	Equity	Total_Wait (weeks)	Total_Served
Baseline	1.000	21.8	29 375
Dynamic	0.916	20.0	29 994
Jockeying	0.904	19.3	30 093
ANOVA p-value	<0.0001	<0.0001	<0.0001

Table 2 summarizes these tests, providing the mean of each metric for each policy and the p-value of the ANOVA test. These tests indicate that both the dynamic and jockeying policies provide an improvement over the baseline with respect to both efficiency and throughput, coming at the expense of some loss of equity. Jockeying yielded the best mean efficiency and throughput but also resulted in the lowest mean equity.

The results in this section highlight the broad effects that might be achieved when implementing the three policies, and provide intuition about how the policies affect the tradeoffs between equity and efficiency. However, given these results depend on particular parameter choices, we next explore an experimental design to reduce dependency on the baseline parameters.

4.3. Experimental design

To quantify the impact of uncertainty in key parameters in the simulation model, we design and run three experiments, each tailored to one of the three policies. Leveraging the power of the design of experiments methodology allows us to systematically investigate the effects of multiple factors on outputs of a simulation model, enabling the discovery of broad insights that would otherwise not be possible [34].

Table 3
Experiment factor settings for policies in the CE simulation.

Parameter	Baseline	Low	High	Increment	Policy	Type
T_{true}	112	75	125	5	All	11-level
New_PSH_Inventory	12 000	11 250	14 000	250	All	12-level
Arrival rate (1/day)	10	6	15	1	All	10-level
$T_{lowered}$ proportion	0.8	0.6	0.9	0.05	Dynamic	7-level
B_{PSH}	1300	500	2000	100	Jockeying	16-level
B_{RRH}	300	100	1000	100	Jockeying	10-level

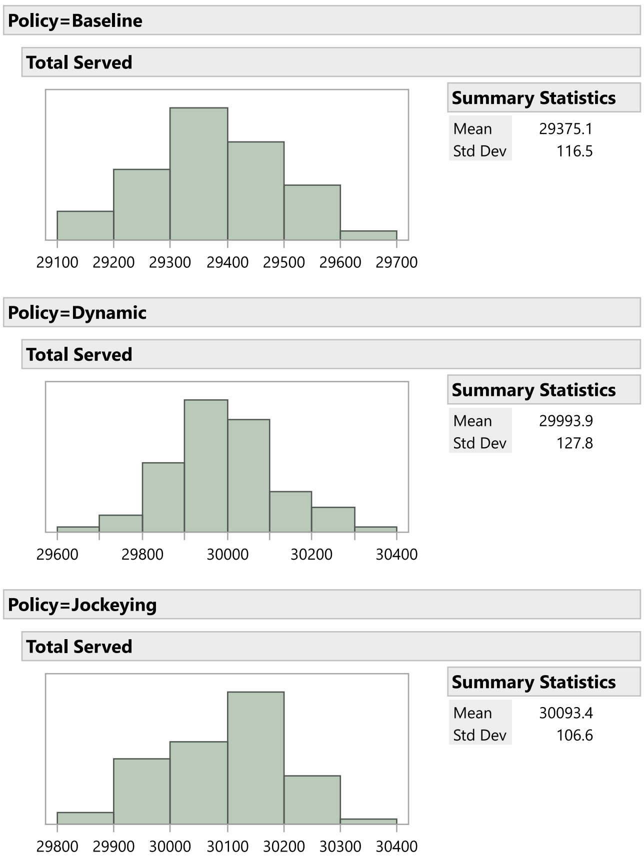

Fig. 6. Empirical distributions of total served by policy.

The first step in designing an experiment is determining the model inputs or parameters to be varied (called factors) and the range or levels over which each will be varied. These factors are provided in Table 3, and were chosen either because they were highly uncertain, or because they are important choices to be made as part of our routing policies. This experiment was conducted with the goal of determining policies around managing queues in the face of a potential increase in PSH housing. Because PSH housing is more intensive, it may make sense to allow more clients to access that resource if extra inventory is available. Thus, we explore routing policies that would enable more clients to enter the PSH queue if additional inventory becomes available. Table 3 outlines each parameter and the factor levels used, including the baseline value, low and high ranges for factor settings, and the increment used to determine factor levels. The ranges for numeric factors were chosen after a series of iterative experiments to induce meaningful and interesting variation in the metrics while also representing a reasonable range of uncertainty in the real system.

There is also great uncertainty around the availability of new housing, so we vary the amount of new total inventory in the system (New_PSH_Inventory in Table 3) to show the new inventory levels varied from the baseline level of 12,000 housing units. We allow for the possibility of total inventory to decrease slightly with a low factor level of 11,250, to allow for unforeseen circumstances where capacity in the system was lost. Because the arrival rate to the system is uncertain and will continue to change, we vary the arrival rate to be between 6/day and 15/day. Prevention policies could result in a decreased arrival rate, though given that the homeless crisis is still facing many challenges, we want to test whether the polices are robust to increased arrival rates from the current baseline estimates of 10/day. Finally, we vary the parameters of the buffer policy, B_{PSH} and B_{RRH} to be the range of queue values allowed for RRH clients to switch to the PSH queue.

There are many choices of design available. A well-known design is the full factorial which tests every possible combination of the factors. Though ideal, as the number of factors and levels grows, it quickly becomes prohibitive in terms of the runs needed, and consequently the time required to run the experiment. Space-filling designs are a popular choice for sampling the interior of a space in an effective and efficient manner [35]. For our experiments, we employ the 2nd Order Nearly Orthogonal and Balanced (NOAB) space-filling design which allows for a mix of factor types (continuous, discrete, or categorical) and provides enough degrees of freedom to fit a wide variety of complex metamodels while minimizing correlations between all terms in a 2nd order regression model [36]. This design ensures that we have good and balanced coverage of the factor space, allowing for the independent assessment of each factor's influence. The space-filling nature of the design allows for identification of thresholds and change points and its efficiency means that we can effectively sample the space using far fewer runs that would be required with a full factorial design. The custom design builder used to construct our design is available publicly for download at https://harvest.nps.edu.

4.4. Analysis of experiment data

We use statistical metamodeling to capture the relationship between experiment factors and the model output, or responses. There are many forms of metamodels including, for example, multiple linear regression, logistic regression, Gaussian process modeling, and tree-based methods [35,37]. By fitting statistical metamodels, we can thereby quantify the effect of the experiment factors on equity and efficiency, capturing the impact of uncertainty. For the analysis presented in this paper, we employed multiple linear regression using stepwise selection, allowing all terms in a second-order regression model to be considered for entry. In each case, we obtained a well-fitting regression model, so there was not a need to consider higher-order terms.

We fit regression models to PSH_Wait and RRH_Wait for the baseline policy experiment. For the dynamic and jockeying experiments, we fit models to Equity, PSH_Wait and RRH_Wait. Thus, we fit eight regression models in total. The tables that follow contain information about the significant terms in two selected regression models, one corresponding to the dynamic experiment and one corresponding to the jockeying experiment. The remaining experiments contained similar insights so the results are excluded here for brevity.

We begin with the regression table for PSH_Wait in the dynamic experiment shown in Table 4. Not including the intercept, this regression model contains four main effects and four two-way interactions and these are listed in the table in decreasing order of impact. Impact is measured by the absolute value of the t-statistic, also known as the t-ratio, computed as the coefficient value divided by its standard

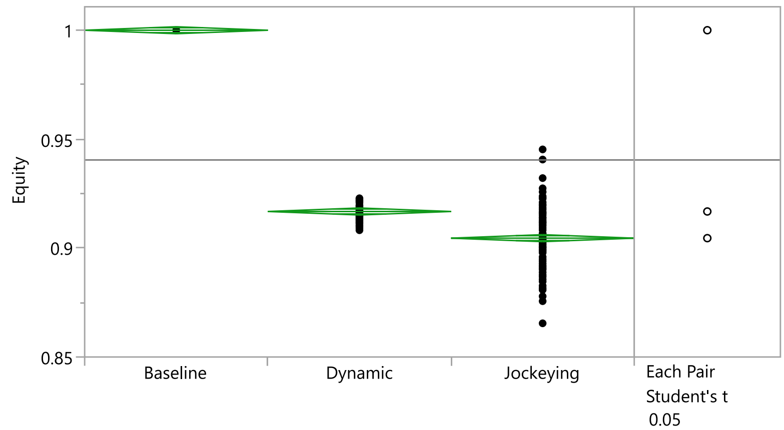

Fig. 7. One-way analysis of variance (Equity by policy).

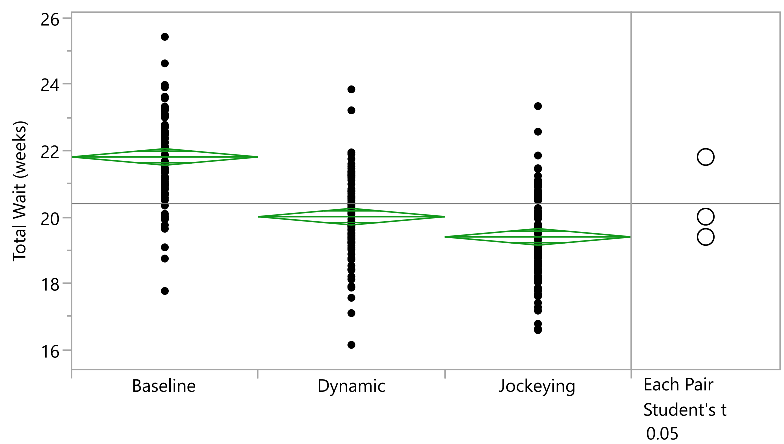

Fig. 8. One-way analysis of variance (Total_Wait in weeks by policy).

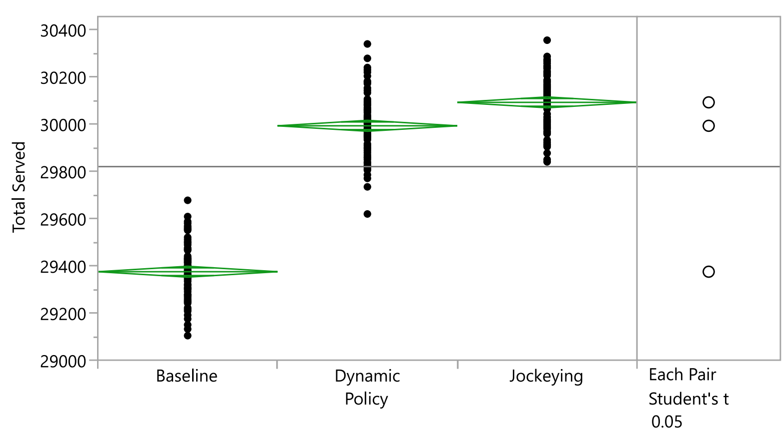

Fig. 9. One-way analysis of variance (Total_Served by policy).

error. Main effects with positive coefficients increase PSH_Wait while those with negative coefficients decrease PSH_Wait. The term with highest impact is the arrival rate, which was highly significant for all regression models. Interaction terms represent a situation where the combined effects of two factors is more or less than the sum of the independent effects. The term with the second highest impact is the interaction between the arrival rate and the amount of New_PSH_Inventory. Further inspection of the interaction result reveals that increased levels of New_PSH_Inventory do not impact PSH_Wait much at lower values of the arrival rate, as the increased inventory is not needed. Increased inventory substantially reduces PSH_Wait for higher arrival rates. Specifically, when the arrival rate is set to its highest value in this experiment, 15/day, then increasing PSH inventory from 11,250 to 14,000 reduces PSH_Wait by approximately 40 percent.

Interpreting the next most impactful interaction term, between the arrival rate and the true threshold, an increased arrival rate has a substantially greater impact when the true threshold is lower and much less of an impact when the true threshold is higher. In other words, the system can handle an increased arrival rate when the true threshold is higher. The interaction between the lowered threshold and New_PSH_Inventory is interpreted as: lowering the threshold increases PSH_Wait for higher levels of New_PSH_Inventory by approximately 20

Table 4
Regression table for PSH_Wait in the dynamic experiment.

Term	Coefficient	Std error	t-statistic	p-value
Arrival_Rate	31.76986	1.80708	17.58	<.0001
Arrival_Rate*New_PSH_Inventory	−0.00143	0.00013	−10.76	<.0001
Arrival_Rate*T_{true}	−0.08306	0.00789	−10.53	<.0001
T_{true}	−2.14939	0.28061	−7.66	<.0001
T_{true}*New_PSH_Inventory	0.00016	0.00002	7.49	<.0001
New_PSH_Inventory*$T_{lowered}$	−0.00763	0.00297	−2.57	0.0122
$T_{lowered}$	90.27962	37.41342	2.41	0.0186
New_PSH_Inventory	−0.00099	0.00071	−1.38	0.1718

Table 5
Regression table for RRH_Wait in the jockeying experiment.

Term	Coefficient	Std error	t-statistic	p-value
Arrival_Rate	6.10914	0.13335	45.81	<0.0001
T_{true}	0.45741	0.02449	18.61	<0.0001
Arrival_Rate*T_{true}	0.09698	0.00850	11.41	<0.0001
Arrival_Rate*Arrival_Rate	0.54689	0.05285	10.35	<0.0001
T_{true}*T_{true}	−0.00845	0.00174	−4.85	<0.0001
New_PSH_Inventory*Arrival_Rate	−0.00041	0.00015	−2.70	0.0083
T_{true}*New_PSH_Inventory	−0.00005	0.00002	−2.09	0.0393
New_PSH_Inventory	−0.00076	0.00045	−1.71	0.0902

percent, because in this case, individuals who would have otherwise been provided with RRH services were instead sent to PSH, increasing total throughput. The term listed last is the New_PSH_Inventory. Though not significant at the 0.05 or 0.10 level of significance as a main effect, it is the convention to retain it in the model because it is contained in several significant interactions with other factors.

We next discuss the regression for RRH_Wait in the jockeying experiment, whose terms are shown in decreasing order of impact in Table 5. This regression model contains three main effects, two quadratic terms, and three two-way interactions and these are listed in the table in decreasing order of impact. Investigation of the quadratic term for the arrival rate reveals that RRH_Wait experiences a polynomial rate of increase as the arrival rate is increased over its range from 6/day to 15/day, with near-zero wait at the lower end and approximately 55 weeks at the upper end. To determine the impact of increasing PSH inventory, we look to the interaction between New_PSH_Inventory and the arrival rate. Inspection of this interaction reveals that when the arrival rate is set to its highest value in this experiment, 15/day, then increasing PSH inventory from 11,250 to 14,000 reduces RRH_Wait by almost ten weeks.

5. Conclusion

We build a discrete-event simulation model to explore complex routing policies associated with allocating limited housing resources to people arriving to a homeless CoC. One goal of a CE system is to effectively align clients with the type of housing that best suits their needs. We approach this goal by developing quantitative metrics of equity (percentage of correct housing assignments according to need) and efficiency (waiting time for housing and total number of clients served). The simulation model allows us to estimate how different routing policies perform according to these metrics. We explore a baseline threshold policy, a dynamic threshold policy that allows the HPA score threshold to change over time, and a jockeying policy that allows clients to move to a queue for a better housing resource.

A comparison of the three policies reveals that the dynamic and jockeying policies are able to improve efficiency in the system through reduced waiting times and increases in the number of clients housed, though at the expense of decreased equity. However, equity levels are still relatively high, where most clients are receiving the housing resource originally intended for them. Because system conditions are constantly changing and uncertain, we conduct a design of experiments to compare the policies while varying inputs to the model. This reveals which input factors are key in influencing the results, and these factors should be carefully considered in implementation planning.

We note that the operations of CE in locations such as San Francisco and Alameda County are continually changing based on local conditions and input from policymakers and stakeholders. While our model was based on general existing conditions at the time of this writing, it can be adapted to consider more than two types of housing resources, or different types of routing policies as the system evolves. Many other systems, such as healthcare systems, may also benefit from this type of exploration of routing policies that are too complex for analytical queueing methods. Future work will address optimization of simulation planning models for multi-tier CE systems.

CRediT authorship contribution statement

Dashi I. Singham: Writing – review & editing, Writing – original draft, Supervision, Conceptualization. **Mary McDonald:** Writing – review & editing, Visualization, Supervision, Investigation, Formal analysis, Data curation. **Robert Elliot:** Writing – review & editing, Software, Investigation, Formal analysis, Data curation.

Acknowledgments

We are extremely grateful to Dr. Jessie Shimmin of the San Francisco Department of Homelessness and Supportive Housing and her colleagues for discussions which motivated our construction of the model. Some introductory text in this paper is replicated from the Master's thesis of the third author, Robert Elliot. However, the model in this paper is modified from the thesis and the experimentation is new. Finally, we thank two referees whose thorough comments greatly improved the presentation of this paper.

Data availability

No data was used for the research described in the article.

References

[1] Focus Strategies. San Francisco coordinated entry report. 2022, URL https://hsh.sfgov.org/wp-content/uploads/2022/06/SF-CE-Evaluation-Report.pdf.

[2] Krivkovich Alexis, Modi Kunal, Pan Eufern, Parthasarathy Ramya, Schiff Robert. The ongoing crisis of homelessness in the bay area: What's working, what's not. 2023, URL https://www.mckinsey.com/industries/public-sector/our-insights/the-ongoing-crisis-of-homelessness-in-the-bay-area-whats-working-whats-not.

[3] US Department of Housing and Urban Development. Coordinated entry core elements. 2024, URL https://www.hudexchange.info/resource/5340/coordinated-entry-core-elements/.

[4] Paul Jr Dereck W, Knight Kelly R, Olsen Pamela, Weeks John, Yen Irene H, Kushel Margot B. Racial discrimination in the life course of older adults experiencing homelessness: Results from the HOPE HOME study. J Soc Distress Homelessness 2020;29(2):184–93.

[5] Finnigan Ryan. Self-reported impacts of the COVID-19 pandemic for people experiencing homelessness in Sacramento, California. J Soc Distress Homelessness 2022;31(1):72–80.

[6] Feld Jamie E, Anaya Henry D, Hoang Tuyen, Knapp Herschel, Asch Steven M. Implementing an HIV rapid testing intervention for homeless veterans in shelter settings within Los Angeles County, USA. J Soc Distress Homeless 2009;19(1–2):17–40.

[7] Singham DI, Lucky J, Reinauer S. Discrete-event simulation modeling for housing of homeless populations. PLoS One 2023.

[8] Oakland-Berkeley-Alameda County CoC. Centering racial equity in homeless system design. 2020, URL https://everyonehome.org/wp-content/uploads/2021/02/2021-Centering-Racial-Equity-in-Homeless-System-Design-Full-Report-FINAL.pdf.

[9] Singham DI. Estimating quantile fields for a simulated model of a homeless care system. In: Proceedings of the 2023 Winter Simulation Conference. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.; 2023.

[10] Chapman Lloyd AC, Kushel Margot, Cox Sarah N, Scarborough Ashley, Cawley Caroline, Nguyen Trang Q, Rodriguez-Barraquer Isabel, Greenhouse Bryan, Imbert Elizabeth, Lo Nathan C. Comparison of infection control strategies to reduce COVID-19 outbreaks in homeless shelters in the United States: a simulation study. BMC Med 2021;19(1):1–13.

[11] Ingle Tanvi A, Morrison Maike, Wang Xutong, Mercer Timothy, Karman Vella, Fox Spencer, Meyers Lauren Ancel. Projecting COVID-19 isolation bed requirements for people experiencing homelessness. PLoS One 2021;16(5):e0251153.

[12] Reynolds Jared, Zeng Zhen, Li Jingshan, Chiang Shu-Yin. Design and analysis of a health care clinic for homeless people using simulations. Int J Heal Care Qual Assur 2010.

[13] Dai Li, Zhou Peng. The health issues of the homeless and the homeless issues of the ill-health. Soc-Econ Plan Sci 2020;69:100677.

[14] Kaya Yaren Bilge, Maass Kayse Lee, Dimas Geri L, Konrad Renata, Trapp Andrew C, Dank Meredith. Improving access to housing and supportive services for runaway and homeless youth: Reducing vulnerability to human trafficking in New York City. IISE Trans 2024;56(3):296–310.

[15] Maass Kayse Lee, Trapp Andrew C, Konrad Renata. Optimizing placement of residential shelters for human trafficking survivors. Soc-Econ Plan Sci 2020;70:100730.

[16] Burgess Graham, Singham Dashi I, Rhodes-Leader Luke. Time-varying capacity planning for designing large-scale homeless care systems. 2025, Under Review.

[17] Rahmattalabi Aida, Vayanos Phebe, Dullerud Kathryn, Rice Eric. Learning resource allocation policies from observational data with an application to homeless services delivery. In: 022 ACM conference on fairness, accountability, and transparency. FAccT '22, New York, NY: ACM; 2022, http://dx.doi.org/10.1145/3531146.3533181.

[18] Liao Haiyan, Holguín-Veras José, Calderón Oriana. Comparative analysis of the performance of humanitarian logistic structures using agent-based simulation. Soc-Econ Plan Sci 2023;90:101751.

[19] Souza Juliano Silva, Lim-Apo Flávio Araújo, Varella Leonardo, Coelho Antônio Sérgio, Souza João Carlos. Multi-period optimization model for planning people allocation in shelters and distributing aid with special constraints. Soc-Econ Plan Sci 2022;79:101087.

[20] Teh Yih-Choung, Ward Amy R. Critical thresholds for dynamic routing in queueing networks. Queueing Syst 2002;42:297–316.

[21] Armony Mor. Dynamic routing in large-scale service systems with heterogeneous servers. Queueing Syst 2005;51:287–329.

[22] Argon Nilay Tanik, Ding Li, Glazebrook Kevin D, Ziya Serhan. Dynamic routing of customers with general delay costs in a multiserver queuing system. Probab Engrg Inform Sci 2009;23(2):175–203.

[23] Jain Ankit, Lim Andrew EB, Shanthikumar J George. On the optimality of threshold control in queues with model uncertainty. Queueing Syst 2010;65(2):157–74.

[24] Chen Sha, Duenyas Izak, Iravani Seyed. Admission and routing control of multiple queues with multiple types of customers. IISE Trans 2023;1–15.

[25] Kaya Yaren Bilge, Maass Kayse Lee. Leveraging priority thresholds to improve equitable housing access for unhoused-at-risk youth. 2023, arXiv preprint arXiv:2212.03777.

[26] Kaya Yaren Bilge, Maass Kayse Lee. Queue routing strategies to improve equitable housing coordination in New York City. 2023, arXiv preprint arXiv:2304.12487.

[27] San Francisco Department of Homelessness and Supportive Housing. Housing referral status range. 2023, URL https://hsh.sfgov.org/wp-content/uploads/2023/10/CE-Written-Standards_9.2023_Clean.pdf.

[28] Elliot Robert. Equity and efficiency tradeoffs in multi-tier simulated queueing of homeless care systems [Master's thesis], Monterey, CA: Naval Postgraduate School; 2024.

[29] San Francisco Department of Homelessness and Supportive Housing. Coordinated entry written standards. 2023, URL https://hsh.sfgov.org/wp-content/uploads/2023/10/CE-Written-Standards_9.2023_Clean.pdf.

[30] San Francisco Department of Homelessness and Supportive Housing. Reentry transitional and supportive housing. 2025, URL https://www.sf.gov/reentry-transitional-and-supportive-housing.

[31] Smith Jeffery S, Sturrock Kenneth T. Simio and simulation: modeling, analysis, applications. 7th ed.. 2024.

[32] San Francisco Department of Homelessness and Supportive Housing. San francisco point-in-time counts. 2024, URL https://hsh.sfgov.org/about/research-and-reports/pit/.

[33] DeGroot Morris H, Schervish Mark J. Probability and statistics. 4th ed.. Addison-Wesley; 2012.

[34] Sanchez Susan M, Sanchez Paul J, Wan Hong. Work smarter, not harder: A tutorial on designing and conducting simulation experiments. In: Proceedings of the 2020 Winter Simulation Conference. Institute of Electrical and Electronics Engineers, Inc.; 2020, p. 1128–42.

[35] Kleijnen Jack PC. Design and analysis of simulation experiments. Springer; 2018.

[36] MacCalman Alex D, Vieira H, Lucas T. Second-order nearly orthogonal latin hypercubes for exploring stochastic simulations. J Simul 2017;11(2):137–50.

[37] Barton Russell R. Tutorial: Simulation metamodeling. In: Proceedings of the 2015 Winter Simulation Conference. Institute of Electrical and Electronics Engineers, Inc.; 2015, p. 1765–79.

Prof. Dashi Singham is a Research Associate Professor in the Operations Research Department at the Naval Postgraduate School, where she has taught since 2010. She is the Director of the SEED Center (Simulation Experiments & Efficient Designs) where she manages multiple research projects on large-scale simulation modeling and experimentation. Dr. Singham earned a Bachelors degree in Operations Research & Financial Engineering from Princeton University, and two Masters (Operations Research & Statistics) and a Ph.D. (Industrial Engineering & Operations Research) from the University of California, Berkeley. She received an NSF graduate research fellowship for her work at Berkeley. Dr. Singham is the recipient of the 2020 Menneken Award for Highly Meritorious Research at NPS. She is a twice-elected officer of the INFORMS Simulation Society (Treasurer, and Council Member), and has served as editorial members of IISE Transactions and Proceedings of the Winter Simulation Conference. Her primary theoretical research interests are simulation modeling and applied probability with an emphasis on sequential procedures and standardized time series for confidence interval estimation. She has had multiple multi-year research projects from sponsors such as the NSF, DTRA, ARP, as well as intelligence and naval research communities.

Mary McDonald graduated from Northwestern University, where she received a B.A. in Mathematics and graduated as a member of the Phi Beta Kappa Honor Society. She received her Marine Corps commission through the Naval Reserve Officers Training Corps program in 1991. Her active tours of duty included being the Fiscal Officer for the 1st Supply Battalion, a Naval Postgraduate School student in the Applied Mathematics Department, where she graduated with an M.S. in Applied Mathematics and the Navy League Award for Highest Academic Achievement, and an Operations Research Analyst in the Office of Science Technology and Integration, Marine Corps Warfighting Laboratory. After 10 years of active service, she continued serving in the Marine Corps Reserve where she was activated and served as the Peacetime Wartime Support Team OIC for the Reserve Training Center (RTC) Naval District Washington (NDW) and Mathematics Instructor for the Marine Enlisted to Officer Commissioning Program (MECEP), ultimately retiring as a Lieutenant Colonel. Mary has been with the Naval Postgraduate School (NPS) as a Lecturer and Faculty Associate for Research, within the Simulation Experiments and Efficient Designs (SEED) Center for Data Farming, since 2008. She assists and advises SEED Center master's thesis students, teaches courses in probability and statistics and simulation analysis, and develops modules for other NPS courses. She is a member of the Military Operations Research Society and the Institute for Operations Research and Management Science.

Robert Elliot received an M.S. in Operations Research from the Naval Postgraduate School in 2024 where he completed his thesis "Equity and Efficiency Tradeoffs in Multi-Tier Simulated Queueing of Homeless Care Systems". He previously received a B.S.E. from Arizona State University in Industrial Engineering in 2016. He currently serves as a Lieutenant in the U.S. Navy.