Optimal Two-Tier Outpatient Care Network Redesign With a Real-World Case Study of Shanghai

Yewen Deng, Na Li, Nan Kong, Zhibin Jiang, and Xiaoqing Xie

***Abstract*— Healthcare capacity shortage contributes to poor access in many countries. Moreover, rapid urbanization often occurring in these countries has exacerbated the imbalance between healthcare capacity and need. One way to address the above challenge is expanding the total capacity and redistributing the capacity spatially. In this research, we studied the problem of locating new hospitals in a two-tier outpatient care system comprising multiple central and district hospitals, and upgrading existing district hospitals to central hospitals. We formulated the problem with a discrete location optimization model. To parameterize the optimization model, we used a multinomial logit model to characterize individual patients' diverse hospital choice and to quantify the patient arrival rates at each hospital accordingly. To solve the hard nonlinear combinatorial optimization problem, we developed a queueing network model to approximate the impact of hospital locations on patient flows. We then proposed a multi-fidelity optimization approach, which involves both the aforementioned queuing network model as a surrogate and a self-developed stochastic simulation as the high-fidelity model. With a real-world case study of Shanghai, we demonstrated the changes in the care network and examined the impacts on the network design by population center emergence, governmental budget change and considering patients with different age groups or income levels.**

***Note to Practitioners*—Our work focuses on improving system-wide care access in a two-tier care network. We believe that our work can lead to effective development of a location analytics tool for city-wide healthcare system planners. We also think the importance of this study is further strengthened by the case studies based on real-world hospital choice experimental data from Shanghai, China, a region suffering from the imbalance between healthcare capacity and need. Our case studies are expected to make recommendations on care facility expansion and dispersion to better align with the spatial distribution of residential communities and patient hospital choice behavior.**

***Index Terms*— Network design, location optimization, patient choice, queueing network, multi-fidelity optimization.**

Manuscript received 19 November 2021; revised 4 March 2022 and 8 June 2022; accepted 25 July 2022. This article was recommended for publication by Associate Editor X. Xie and Editor J. Li upon evaluation of the reviewers' comments. This work was supported in part by the National Natural Science Foundation of China under Grant 72171144, Grant 71432006, and Grant 71871138; in part by the China Europe International Business School (CEIBS) Healthcare Research Fund; in part by the Major International (Regional) Joint Research Project of the National Natural Science Foundation of China (NSFC) under Grant 71420107024; and in part by the National Science Foundation of U.S. under Grant 1761022. *(Corresponding author: Na Li.)*

Yewen Deng and Na Li are with the Department of Industrial Engineering and Management, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: ywdeng@sjtu.edu.cn; na-li03@sjtu.edu.cn).

Nan Kong is with the Weldon School of Biomedical Engineering, Purdue University, West Lafayette, IN 47907 USA.

Zhibin Jiang is with the Antai College of Economics and Management, Shanghai Jiao Tong University, Shanghai 200240, China.

Xiaoqing Xie is with the Department of Economics and Decision Sciences, China Europe International Business School (CEIBS), Shanghai 201206, China.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TASE.2022.3198484.

Digital Object Identifier 10.1109/TASE.2022.3198484

I. Introduction

THERE is a significant discrepancy between patient need and provider capacity in many major metropolises worldwide, including Shanghai, China. The 2017 Shanghai Statistics Year Book (available from www.tjj.sh.gov.cn/) shows that the annual volume of hospital visits in Shanghai is as high as 273 million. By contrast, the number of licensed and assistant physicians is only 2.31 per thousand residents in Shanghai (available from www.spcsc.sh.cn/). Healthcare capacity shortage becomes a major social issue in many metropolises in China and around the world. Moreover, with rapid urbanization, major metropolises in China expand significantly and resident populations disperse with new residential communities sprouting in inter-city and suburban areas. However, high-quality healthcare resource remains at city centers. As a result, the demand-supply imbalance is further exacerbated.

Like in many other countries, the hospital network in China is a two-tier system, consisting of central hospitals (CHs) and district hospitals (DHs). CHs are typically staffed with the most qualified physicians and clustered at city centers. On the other hand, DHs provide basic medical services, and they are located in various residential communities. For outpatient care, Chinese patients typically choose hospitals based on their preferences, and CHs are often more attractive to them due to their good reputation. However, due to limited capacities, it is often overcrowded at CHs and patients have to wait long. Similar to China, in South Africa, many patients directly visit general hospitals for some minor issues without consulting local primary care centers [1]. In Japan, many patients needing basic care prefer to choose the emergency departments of regional/national public hospitals instead of receiving more appropriate primary care services in a community clinic [2]. Therefore, the limited resources in regional/national public hospitals inevitably lead to poor care access. In addition, since CHs are mostly located in the central city, patients from inter-city or suburban areas, have to travel long distance for outpatient services. Overall capacity shortage and capacity maldistribution motivated a series of reforms in recent years to address the challenge on care access, one of the fundamental challenges in the Chinese healthcare system. The use of our study is not limited to outpatient care network redesign. Besides outpatient care, preventive care is also greatly influenced by poor access [3], [4].

This paper presents our mathematical modeling and optimization work that aims to increase care access through optimal redesign of a two-tier hospital network subject to government budget. For this research, we conduct a choice experiment on hospital visit for some outpatient service, through which we record respondents' choices when presented with hospital type and distance to the hospital. We develop a multinomial logit model to characterize hospital choice behavior. We formulate a nonlinear optimization problem to select locations to build new care facilities and identify some of the existing DHs to upgrade to CHs. We consider a care access measure that combines times spent in traveling to the hospital and waiting for the outpatient service. Finally, we design a multi-fidelity optimization algorithm involving a low-fidelity queueing network based surrogate and a high-fidelity simulation model. This optimization approach can balance efficient identification of promising solutions and accurate comparison among them.

The contribution of this paper is three-fold. First, we illustrate a way to embed discrete choice modeling into location optimization through quantifying the demand diversion. Second, we demonstrate the use of a multi-fidelity optimization approach to solve nonlinear combinatorial optimization problems whose objective is computationally expensive to evaluate precisely even for one candidate solution. Third, we base our case study on an outpatient care facility choice survey and additional census data collected from Shanghai; thus offering potentially meaningful suggestions to the Shanghai municipality's health department.

The remainder of this paper is organized as follows. In Section II, we review the relevant literature. In Section III, we present our model and optimization approach, which includes a location optimization model for network redesign, a multinominal logit model for outpatient choice characterization, and a multi-fidelity optimization algorithm for solution tractability. In Section IV, we illustrate our approach with a Shanghai-based case study and analyze the results for network redesign effectiveness and robustness. Finally we draw conclusions and outline future research in Section V.

II. Literature Review

There are numerous location analysis studies for healthcare planning. According to the classification of Daskin [5], discrete location models in the healthcare OR literature are divided into three categories: covering-based models, median-based models, and other models such as the p-dispersion model. Covering-based and median-based models are classic models used in medical facility location analysis. For review of these classical problems, we refer the readers to [6], [7], and [8]. For example, many researchers addressed the objective of maximizing the total demand covered by a given number of facilities e.g. [9], [10], and [11] or minimizing the distance or travel time when demand locations are assigned to facilities within a certain distance e.g. [12], [13], [14], and [15]

A critical aspect of medical facility location analysis is modeling patient choice and its impact on location decisions. Studies on patient choice are divided into two categories. In the first category, studies assume a *directed-choice* (or *system-optimal*) mechanism, i.e., each patient is assigned to a medical facility by the decision maker, rather than being allowed to choose a facility. These models determine the optimal facility location and patient assignment for maximization of some system outcome. For example, in [10], the demand at each node is assigned to only a facility within some allowable distance. Other studies in this category include [3], [14], [16], and [17].

In the other category, studies assume a *patient-choice* mechanism, i.e., patients are free to choose a medical facility to visit. This category can be further categorized as *deterministic-choice* and *probabilistic-choice* models. For *deterministic-choice* models, the patient choice behavior is simplistic based on some utility-based criteria, i.e., each patient is assumed to visit the most attractive facility. These models often assume patients are rational and fully informed, and they visit the facility they think is most attractive at all time. For example, Kim and Kim [11] studied the problem of locating public health facilities to maximize the number of served patients of two types (low- and high-income). The authors only considered the preferences of high-income patients and assumed that those patients are only assigned to their most preferred facilities. The authors modeled the preference of high-income patients on each facility to be dependent on several factors, including distance to travel and amount of service charge. Instead of analyzing the influence of each factor, the authors simply used a given parameter between 0 and 1 to represent patients' preferences, and assumed patients from the same region have identical preference. Some other studies assumed that patients visit the closest facility [18], [19] or the facility with the minimal total time [20], [21].

For *probabilistic-choice* models, a patient is assumed to visit each facility with a certain probability. Huff [22] proposed the first probabilistic-choice model, for a spatial interaction analysis. In the model, client utility is represented by a gravity formula to estimate market shares of the facilities. Subsequently, [22] was extended in the location analysis literature, including one based on the multiplicative competitive interaction model by [23]. Discrete choice models, relevant to our work, are based on random utility theory in marketing and econometrics, and have been incorporated in location analysis and optimization. For example, to maximize the total participation of a preventive care program, [4] assumed that the only attractiveness attribute in their probabilistic-choice model is the proximity to a facility. The authors used a multinomial logit (MNL) model to characterize the probability that a patient chooses each facility. The parameters in the MNL model are set hypothetically for the study. Reference [24] also modeled the patient choice behavior with an MNL model. In addition to distance (travel time), the authors considered the influence of waiting time on patient choice. The authors estimated the coefficients in their model using actual patient flow data. Reference [25] proposed a location-allocation model to improve the accessibility in medically under-served areas. A discrete choice model is incorporated to describe care consumers' choice decision for obstetrics care. The attribute variables are travel time, level of a hospital, number of obstetrics specialists, the deprivation index of the location of a hospital and whether care consumer i and hospital j belong to the same hospital service area. Other location analysis papers that incorporate patient choice with a discrete choice model include [26], [27], [28], and [29].

The existing literature on hospital choice model (such as [30], [31], [32], [33], [34], and [35]) believes that the factors that affect patients' hospital choice behavior are: medical convenience (e.g., distance to healthcare facility), patient attributes (age, income, etc.), and hospital attributes (price, quality of providers, level of the hospital, etc.). Most of the location studies considering patient choice often assume that distance (or proximity) is a major or the only influencing factor, and a few studies also consider waiting time as an influencing factor. In this paper, we did not use waiting time, because in reality it is difficult for patients to have information about the detailed waiting time of each hospital in advance. Instead, we considered the preference of patients on CHs versus DHs, considered by fewer in the existing studies, as well as the effect of travel distance. In addition, in most studies that considered patient behavior or preference, there were no distinction on patient types, whereas we distinguished patients by their age and analyzed their respective hospital choice behavior. Furthermore, the aforementioned MNL models were either constructed based on expert judgement and scattered evidence, or fitted against patient flow data. On the other hand, we fitted an MNL model against first-hand behavior data collected from a survey of Shanghai residents. Further, we examined whether distance and type of the hospital were influencing factors and estimated the influence of each significant factor. We believe this has made our work more realistic and comprehensive.

Another aspect of the literature related to our work concerns incorporating congestion in facility design models. The most common way to do it is to represent each facility as a queue (e.g., M/M/1 or M/G/1) and consider a constraint on the congestion level. For example, [4], [36], and [37] used M/M/1 (or M/M/c) queues to model service facilities and considered an upper bound on the average waiting time of customers (or equivalent queue length) in a constraint. References [38] and [39] modeled facilities as M/M/1 (or M/M/c) queueing systems and introduced a constraint to ensure the probability that a client enters the queue at a facility with at most b waiting clients is at least α in a maximal covering location-allocation model. Reference [40] considered a re/design of a congested multi-service network with M/M/1 servers. Patient waiting time is added to the objective function to improve service quality. Alternatively, there are studies that assumed a decay function between demand and waiting time, and included this function as a constraint in their models. Reference [20] captured the level of congestion at each facility with an M/M/1 queue in a preventive care facility location problem. They assumed the fraction of customers from each population node to each facility is a decreasing function of the expected total (travel, waiting and service) time and a client chooses the facility with the least total time. The authors then provided a heuristic solution method to determine the number of facilities and the location of each facility so as to maximize the population-level participation. Reference [24] modeled each facility as an M/M/c queue. The authors assumed that the mean system waiting time and travel time are main determinants on customer choice, and used an MNL model to characterize the relationship between the equilibrium flow from each population node to each facility. Reference [3] used spatially distributed M/G/1 queues to model a preventive care network, and captured the congestion due to waiting and service delays. Then to minimize the weighted sum of total time, the authors presented a nonlinear mixed integer program to determine the facility locations, service capacity of each facility, and customer allocation to each facility. Through a search of the above relevant literature that incorporates congestion into facility design, the arrival are usually assumed to be Poisson distributed. Different from the above studies that assume every patient arriving at a facility will join the queue, we modeled each hospital as an M/G/1 queue and considered multiple balks made by some patients, i.e., a patient arriving to a hospital will decide whether or not to join the queue according to his/her estimated waiting time. If he/she goes to a hospital where the waiting time is longer than expected, he will balk and choose another hospital until the waiting time is acceptable, or he will leave the system.

III. Methods

The optimization model presented in this section concerns the government's need to alleviate the current situation of poor care access among patients. Studies suggest that the total time spent on transportation and waiting can be used as a proxy to measure care accessibility [3], [20]. The objective is to improve care access under certain government budget for the capacity expansion and spatial redistribution of care facilities.

Let J_H and J_L be the sets of current sites of CHs and DHs, respectively. Let J be the set of candidate sites for new hospitals, either CHs or DHs. Let I be the set of residential sites (considered as demand nodes). We also denote three sets of decision variables:

$$x_j^H = \begin{cases} 1 & \text{if a CH is located at site } j,\ j \in J, \\ 0 & \text{otherwise.} \end{cases}$$

$$x_j^L = \begin{cases} 1 & \text{if a DH is located at site } j,\ j \in J, \\ 0 & \text{otherwise.} \end{cases}$$

$$y_j = \begin{cases} 1 & \text{if an existing DH at site } j \text{ is upgraded, } j \in J_L, \\ 0 & \text{otherwise.} \end{cases}$$

Thus, the binary decision vector on locating new CHs and DHs is represented by $\mathbf{x}^H = (x_1^H, \ldots x_j^H, \ldots x_{|J|}^H)$ and $\mathbf{x}^L = (x_1^L, \ldots x_j^L, \ldots x_{|J|}^L)$, respectively. The decision vector on whether to upgrade a DH is represented by $\mathbf{y} = (y_1, \ldots y_j, \ldots y_{|J_L|})$. Without loss of generality, denote R to be the set of patient classes.

Given decision vectors $\mathbf{x}^H$, $\mathbf{x}^L$ and $\mathbf{y}$, we calculate the rectilinear distance between each residential site i and each hospital site j, denoted by d_{ij}. We then identify the binary label, denoted by q_j, of a hospital at site j, i.e., whether the hospital at site j is CH or DH. We finally estimate the likelihood of a class r patient at site i choosing a hospital at site j, denoted by $p_{ij}^r(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y})$, $r \in R$, $i \in I$, $j \in J \cup J_H \cup J_L$, specified by the underlying choice model. By converting the population quantity of class r patients at site i into the demand estimate for outpatient care, denoted by λ_i^r, and combining the demand estimate with patient choice preferences $p_{ij}^r(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y})$, we obtain the arrival rate of class r patients from site i to hospital at site j and the total arrival rate of hospital at site j, denoted by $\lambda_{ij}^r(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y})$. Thus, $\lambda_{ij}^r(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y}) = \lambda_i^r p_{ij}^r(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y})$. We further denote $\lambda_j(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y})$ to be the aggregate arrival rate of hospital at site j over all residential sites. Thus, $\lambda_j(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y}) = \sum_{r\in R}\sum_{i\in I} \lambda_i^r p_{ij}^r(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y}))$. Given $\lambda_j(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y})$, we derive the mean waiting time and balking probability at site j, denoted by $W_j^H(\lambda_j(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y}))$ and $p_j^{B_H}(\lambda_j(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y}))$, if the hospital at site j is CH; or those quantities at site j, denoted by $W_j^L(\lambda_j(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y}))$ and $p_j^{B_L}(\lambda_j(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y}))$, if the hospital at site j is DH.

Next, we present a nonlinear optimization model, i.e., (1) – (6), as shown at the bottom of the page, that aims to select optimal sites for new CHs or DHs and upgrade some of the existing DHs. The objective of the optimization model comprises two parts: (i) the cost of waiting for care at hospital site j; and (ii) the cost of traveling to hospital site j from residential site i. In the model, for notational simplicity, we use λ_{ij}, λ_j, d_{ij}, W_j^H, W_j^L, $p_j^{B_H}$ and $p_j^{B_L}$ to replace $\lambda_{ij}(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y})$, $\lambda_j(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y})$, $d_{ij}(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y})$, $W_j^H(\lambda_j(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y}))$, $W_j^L(\lambda_j(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y}))$, $p_j^{B_H}(\lambda_j(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y}))$ and $p_j^{B_L}(\lambda_j(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y}))$, respectively.

In objective (1), C_{d_H} and C_{d_L} are the unit costs of traveling to CH and DH, respectively; C_{W_H} and C_{W_L} are the unit-time costs of waiting at CH and DH, respectively; α_d and α_W are assigned weights on the cost of traveling and cost of waiting, respectively. Constraints (2) state that a DH cannot be built when a CH has been built at some site, and vice versa. Constraint (3) ensures that the proportion of CHs at which the balking probability does not exceed θ_H is no less than ϕ_H. Note that $\mathbf{1}\left(p_j^{B_H} \le \theta_H\right)$, an indicator function, equals 1 if $p_j^{B_H} \le \theta_H$, and equals 0 otherwise. Constraint (4) is similar to constraint (3) and $\mathbf{1}\left(p_j^{B_L} \le \theta_L\right)$ is the other indicator function. Constraint (5) represents that the total government spending on the network redesign is no more than budget T_B, and κ_x^H and κ_x^L are the costs of land acquisition and facility construction for locating a CH and a DH, respectively, κ_y is the cost of upgrading a DH. Constraints (6) ensure non-negativity and binary restrictions.

To quantify the arrival rates, we developed a discrete choice model and characterized patient choice preferences (Section III-A). To compute the system performance measures, such as patient waiting time and balking probability, one can formulate a two-level network model with an M/G/1 queue at each node, which can be used to capture potentially multiple balks made by some patient. However, such a model does not lend itself to the availability of exact closed-form expressions

$$
\begin{aligned}
\min_{\mathbf{x}^H,\mathbf{x}^L,\mathbf{y}} \quad & \alpha_d\Bigg[C_{d_H}\Bigg(\sum_{i\in I}\sum_{j\in J} x_j^H\cdot\lambda_{ij}\cdot d_{ij} + \sum_{i\in I}\sum_{j\in J_L} y_j\cdot\lambda_{ij}\cdot d_{ij} + \sum_{i\in I}\sum_{j\in J_H}\lambda_{ij}\cdot d_{ij}\Bigg)\\
& + C_{d_L}\Bigg(\sum_{i\in I}\sum_{j\in J} x_j^L\cdot\lambda_{ij}\cdot d_{ij} + \sum_{i\in I}\sum_{j\in J_L}\Big(1-y_j\Big)\cdot\lambda_{ij}\cdot d_{ij}\Bigg)\Bigg]\\
& +\alpha_W\Bigg[C_{W_H}\Bigg(\sum_{j\in J} x_j^H\cdot\lambda_j\cdot W_j^H + \sum_{j\in J_L} y_j\cdot\lambda_j\cdot W_j^H + \sum_{j\in J_H}\lambda_j\cdot W_j^H\Bigg)\\
& +C_{W_L}\Bigg(\sum_{j\in J} x_j^L\cdot\lambda_j\cdot W_j^L + \sum_{j\in J_L}\Big(1-y_j\Big)\cdot\lambda_j\cdot W_j^L\Bigg)\Bigg] \qquad (1)\\
\text{s.t.} \quad & x_j^H + x_j^L \le 1, \quad \forall j; \qquad (2)\\
& \frac{\sum_{j\in J} x_j^H\cdot\mathbf{1}\Big(p_j^{B_H}\le\theta_H\Big) + \sum_{j\in J_L} y_j\cdot\mathbf{1}\Big(p_j^{B_H}\le\theta_H\Big) + \sum_{j\in J_H}\mathbf{1}\Big(p_j^{B_H}\le\theta_H\Big)}{\sum_{j\in J} x_j^H + \sum_{j\in J_L} y_j + |J_H|} \ge \phi_H; \qquad (3)\\
& \frac{\sum_{j\in J} x_j^L\cdot\mathbf{1}\Big(p_j^{B_L}\le\theta_L\Big) + \sum_{j\in J_L}(1-y_j)\cdot\mathbf{1}\Big(p_j^{B_L}\le\theta_L\Big)}{\sum_{j\in J} x_j^L + \sum_{j\in J_L}(1-y_j)} \ge \phi_L; \qquad (4)\\
& \sum_{j\in J}\left(\kappa_x^H\cdot x_j^H + \kappa_x^L\cdot x_j^L\right) + \sum_{j\in J_L}\kappa_y^H\cdot y_j \le T_B; \qquad (5)\\
& x_j^H, x_j^L, y_j \in \{0,1\}, \quad \forall j. \qquad (6)
\end{aligned}
$$

for the performance measures. We thus simplified the above queueing network model to a two-level network model with an M/M/1 queue at each node, where patients leave the system once they balk (Section III-B). On the other hand, while high-fidelity computer simulation can be used to achieve a high level of precise estimation, it is time consuming even for one candidate solution, needless to say we need to solve a highly nonlinear hospital location optimization model. Hence, we elected to combine the advantages of the above two ideas intelligently (Section III-C). We refer to the queuing network model as the low-fidelity model and the simulation model as the high-fidelity model.

A. The Probabilistic-Choice Model of Hospital Visit Behavior

As described in the literature review, it is common to consider distance as an influencing factor. Studies have suggested the so-called distance decay on the use of hospital, i.e., the inverse association between hospital attractiveness and travel distance [41], [42]. Reference [13] compared the magnitude of distance decay on general hospitals and local health centers. In addition, [43] and [44] examined the influence of provider expertise (or being knowledgeable, experienced, and capable) on provider choice. References [45], [46], and [47] examined the influence of specialists on hospital choice.

To characterize how patient choice is affected by the distance and hospital type (i.e., CH or DH), we surveyed a cohort of online respondents residing in Shanghai with a self-designed questionnaire (see Appendix A). With the questionnaire, we acquired each respondent's age, and asked him/her to make a selection among a set of hospital alternatives. To each respondent, a scenario where CH is more crowded than DH is prompted. And each hospital alternative in the choice set is identified by a randomly generated pair of attributes $(d_{ij},\ q_j)$, where d_{ij} is the distance between residential site i and hospital site j, and q_j is the indicator of hospital type at location j. After completing the survey, we used an MNL model to examine whether the distance and hospital type would play an important role on the hospital choice behavior, and how significant the effect would be.

The utility that a respondent chooses a hospital at site j is given by $U_{ij} = V_{ij} + \varepsilon_{ij},\ i \in I,\ j \in J \cup J_H \cup J_L$. Note that the logit model arises from the assumption that the random error component are drawn from an extreme value (GEV) distribution, i.e., Gumbel distribution. In this formula, V_{ij} is a deterministic component and ε_{ij} is a random error component under a Gumbel distribution. Further, for a hospital at site j with distance d_{ij} and hospital type q_j, we have $V_{ij} = \beta_d d_{ij} + \beta_q q_j$, where β_d and β_q are parameters that capture the preferences of a patient at site i on distance and hospital type. These two parameters need to be estimated. Additionally, d_{ij} is a continuous variable, and q_j is an indicator variable for two hospital types; 1 if the hospital at site j is a CH, 0 otherwise.

We denote the preference parameters for distance and hospital type of class r patients are denoted by β_d^r and β_q^r, respectively. Then the probability a class r patient at site i chooses to visit a hospital at site j is expressed as

$$p_{ij}^r = \frac{e^{(\beta_d^r d_{ij} + \beta_q^r q_j)}}{\sum\limits_{k \in J \cup J_H \cup J_L} e^{(\beta_d^r d_{ik} + \beta_q^r q_k)}}, \quad r \in R, i \in I,\ j \in J \cup J_H \cup J_L. \tag{7}$$

With the definition in equation (7), we derive the arrival rate of class r patients from site i to hospital at site j as shown in equations (8)–(10).

$$\lambda_{ij}^r(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y}) =$$

$$\begin{cases} \lambda_i^r \dfrac{\left(x_j^H + x_j^L\right) e^{(\beta_d^r d_{ij} + \beta_q^r x_j^H)}}{U^r}, & r \in R, i \in I, j \in J \quad (8) \\ \lambda_i^r \dfrac{e^{(\beta_d^r d_{ij} + \beta_q^r)}}{U^r}, & r \in R, i \in I, j \in J_H \quad (9) \\ \lambda_i^r \dfrac{e^{(\beta_d^r d_{ij} + \beta_q^r y_j)}}{U^r}, & r \in R, i \in I, j \in J_L \quad (10) \end{cases}$$

where

$$U^r = \sum_{k \in J} \left(x_k^H + x_k^L\right) e^{(\beta_d^r d_{ik} + \beta_q^r x_j^H)} + \sum_{m \in J_H} e^{(\beta_t^r d_{im} + \beta_q^r)} + \sum_{l \in J_L} e^{(\beta_d^r d_{il} + \beta_q^r y_l)}.$$

Equation (8) represents the arrival rate of class r patients from residential site $i \in I$ to a new hospital at site j. The hospital type q_j can be jointly determined by the decision variables x_j^H and x_j^L. Notice that if the newly built hospital at site j is CH (i.e., $x_j^H = 1,\ x_j^L = 0,\ j \in J$), its hospital type is set to 1 (i.e., $q_j = x_j^H = 1,\ j \in J$); otherwise, if the newly built hospital at site $j \in J$ is DH (i.e., $x_j^H = 0, x_j^L = 1$, $j \in J$) or there is no hospital built at site $j \in J$ (i.e., $x_j^H = 0$, $x_j^L = 0,\ j \in J$), the corresponding hospital type is 0 (i.e., $q_j = x_j^H = 0,\ j \in J$). Equation (9) represents the arrival rate of class r patients from site $i \in I$ to the existing CH at site $j \in J_H$. Accordingly, the hospital type of an existing CH at site $j \in J_H$ is set to 1 (i.e., $q_j = 1,\ j \in J_H$). Similarly, equation (10) represents the arrival rate of class r patients from site $i \in I$ to the existing DH at site $j \in J_L$. When the existing DH at site $j \in J_L$ is upgraded to a CH (i.e., $y = 1$, $j \in J_L$), its hospital type indicator is set to 1 (i.e., $q_j = y_j = 1,\ j \in J_L$), otherwise, its hospital type indicator is 0 (i.e., $q_j = y_j = 0,\ j \in J_L$). Finally, the arrival rate at a hospital at site j is expressed as $\lambda_j(\mathbf{x}, \mathbf{y}) = \sum_{r \in R} \sum_{i \in I} \lambda_{ij}^r(\mathbf{x}, \mathbf{y})$, for any $j \in J \cup J_H \cup J_L$. With these arrival rates, we next use them as inputs to the hospital location optimization model. Note that when a patient arrives at CH or DH according to his/her choice preference, he/she will decide whether or not to join the queue according to his/her estimated waiting time. If the waiting time is longer than expected, he will balk and choose another hospital. When the patient chooses another hospital, his/her choice is still affected by the hospital type and distance to other hospitals, as modeled in the choice model.

B. Low-Fidelity Model Based Approximation

For low-fidelity approximation of the system performance measures, we consider a two-level open network model with an M/M/1 queue at each node, where patients leave the system once they balk. We assume the arrival process from residential site i to hospital site j follows a Poisson distribution with mean arrival rate $\lambda_{ij}\left(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y}\right)$. Thus the aggregate arrival process at hospital site j also follows a Poisson distribution with mean arrival rate $\lambda_j\left(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y}\right)$. In reality, when a patient arrives at a hospital, a long queue discourages the patient from joining the queue (i.e., balking). We follow Liu and Kulkarni [48] to introduce a notion of virtual queueing (vqt). The vqt is the waiting time estimated by a patient once s/he arrives at the hospital. We assume that each patient can estimate the time s/he will wait from multiple sources. For example, in many countries, such as Australia, the estimated waiting time is sometimes available to patients [49], or they may estimate the waiting time based on their previous visits [24].

Next we introduce the balking rule used in the system performance evaluation. We assume when an arriving patient believed that his/her vqt is no more than a fixed amount, s/he would decide to join the queue. The fixed amount is considered essentially the maximum waiting time, and denoted by b_H for CH and b_L for DH, respectively. We thus term the probability that vqt is more than the fixed amount b_H and b_L, as the balking probability. We denote $\hat{J}_H$ to be the set of CH locations (i.e.,$\hat{J}_H = J_H \cup \left\{ j \middle| x_j^H = 1, j \in J \right\} \cup \left\{ j \middle| y_j = 1, j \in J_L \right\}$), and denote $\hat{J}_L$ to be the set of DH locations (i.e.,$\hat{J}_L = \left\{ j \middle| x_j^L = 1, j \in J \right\} \cup \left\{ j \middle| y_j = 0, j \in J_L \right\}$). We thus denote the balking probability at CH $j \in \hat{J}_H$ and DH $j \in \hat{J}_L$ to be $p_j^{B_H}$ and $p_j^{B_L}$, respectively. We also assume the service duration at each CH and each DH to be exponentially distributed. We denote the mean service rate at CH $j \in \hat{J}_H$ and DH $j \in \hat{J}_L$ to be μ_j^H and μ_j^L, respectively. For CH $j \in \hat{J}_H$, an arriving patient leaves the two-level system with balking probability $p_j^{B_H}$, otherwise, the patient waits for the service in an infinite capacity FCFS (first come, first served) queue with probability $1 - p_j^{B_H}$, and leaves the system when the service is completed. For DH $j \in \hat{J}_L$, a patient leaves the system with balking probability $p_j^{B_L}$, otherwise, the patient waits for the service in a queue with probability $1 - p_j^{B_L}$. With these assumptions, we formulate a two-level queueing network model with each node being an M/M/1 queue with balking.

Given $\lambda_j\left(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y}\right)$, $j \in J \cup J_H \cup J_L$, which can be computed via equations (8)-(10), together with μ_j^H, μ_j^L, b_H and b_L, we follow Theorems 1 to 4 in [48] to derive the following closed-form expressions for relevant queueing performance measures, i.e., CH/DH utilization rate, balking probability, and mean waiting time. The system parameters such as the outpatient service demand, the threshold on the wait time tolerance and service rate are exogenous. The parameters such as the arrival rate, mean waiting time, and the balking probability of each hospital are endogenous.

- The utilization rate at CH $j \in \hat{J}_H$:

$$\rho_j^H(\lambda_j\left(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y}\right)) = \frac{\lambda_j\left(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y}\right)}{\mu_H}.$$

- The utilization rate at DH $j \in \hat{J}_L$:

$$\rho_j^L(\lambda_j\left(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y}\right)) = \frac{\lambda_j\left(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y}\right)}{\mu_L}.$$

- The balking probability at CH $j \in \hat{J}_H$:

$$p_j^{B_H}\left(\lambda_j\left(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y}\right)\right) = \omega_j^H \frac{\mu_j^H p_j^H}{1 - p_j^H} \frac{e^{-\left(\mu_j^H - \lambda_j\left(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y}\right)\right) b_H}}{\mu_j^H}, \quad j \in \hat{J}_H,$$

where ω_j^H, as shown at the bottom of the next page, and $p_j^H = \frac{\lambda_j\left(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y}\right)}{\lambda_j\left(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y}\right) + \mu_j^H}$.

- The balking probability at DH $j \in \hat{J}_L$:

$$p_j^{B_L}\left(\lambda_j\left(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y}\right)\right) = \omega_j^L \frac{\mu_j^L p_j^L}{1 - p_j^L} \frac{e^{-\left(\mu_j^L - \lambda_j\left(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y}\right)\right) b_L}}{\mu_j^L}, \quad j \in \hat{J}_L,$$

where ω_j^L, as shown at the bottom of the next page, and $p_j^L = \frac{\lambda_j\left(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y}\right)}{\lambda_j\left(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y}\right) + \mu_j^L}$.

- The mean waiting time at CH $j \in \hat{J}_H$: $W_j^H\left(\lambda_j\left(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y}\right)\right)$, as shown at the bottom of the next page.
- The mean waiting time at DH $j \in \hat{J}_L$: $W_j^L\left(\lambda_j\left(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y}\right)\right)$, as shown at the bottom of the next page.

C. Multi-Fidelity Optimization

In this section, we propose a multi-fidelity optimization approach, for which we borrowed ideas from [50] and [51]. Our discrete location optimization problem has a very large solution space ($3^{230} \times 2^{246}$), thus the computational cost cannot be ignored even with the use of the low-fidelity model. Instead of exhaustive search of all feasible solutions, we applied genetic algorithm to efficiently select a set of promising candidate solutions based on the low-fidelity model. We then sorted the low-fidelity estimates to form a one-dimensional ordinal space, and equally partition the candidate solutions into groups. We next iteratively sampled candidate solutions from each group, ran the high-fidelity simulation to assess the performance measures for the sampled solutions, and computed group-specific sample statistics (i.e., sample mean and sample standard deviation) for each group. To efficiently establish the stochastic domain of candidate solutions, we adaptively allotted the sampling budget to each solution group based on it sample statistics until we used up all the sampling budget.

- Exploration based on the low-fidelity model

Step 1: Set the number of generations in genetic algorithm T, and the number of chromosomes in each generation N.

Step 2: Run the genetic algorithm to generate N chromosome (i.e., N feasible solutions) at each generation: first, generate a chromosome that satisfies constraints (2), (5) and (6), and then examine whether the chromosome satisfies constraints (3) and (4). If so, the chromosome is retained, and so on, until N distinct chromosomes are generated, forming an initial population satisfying constraints (2) - (6). Then, generate new chromosomes for the descendant generation. The feasibility of each new chromosome (whether it meets the constraints (2)-(6)) is examined. If feasible, it is retained into the descendant population to form N feasible chromosome at each generation. The fitness of each chromosome (i.e. objective (1)) is computed via the low-fidelity model. Upon the termination of the algorithm, only retain the distinct solutions obtained from the T generations.

Step 3: Set the group number K. Sort in an ascending order the fitness values of those solutions retained in Step 2. Then partition them equally into K groups, following an ascending order.

- Exploitation based on the high-fidelity simulation

Step 4: Set the initial sampling budget B_0 (i.e., number of simulation runs). Assign each candidate solution group with B_0/K sampling budget. Randomly sample B_0/K solutions from each group and run the high-fidelity simulation exactly one time on each of the selected solutions. Collect the simulation results and compute the initial group-specific sample statistics information (i.e., sample mean $\hat{\upsilon}^0 := (\hat{\upsilon}_1^0, \ldots, \hat{\upsilon}_K^0)$ and sample standard deviation $\hat{\sigma}^0 := (\hat{\sigma}_1^0, \ldots, \hat{\sigma}_K^0)$).

Step 5: Set the total budget B, and the sampling budget in iteration 1, B_1. Set the decrement of sampling budget over two consecutive iterations Δ. Set the iteration index $i = 1$.

Step 6: Select the best group b^{i-1} based on the sample mean $\hat{\upsilon}^{i-1}$ in iteration $i-1$. For any group $k = 1, \ldots, K$, allot its sampling budget in iteration i, i.e., B_k^i, according to the following relationship:

$$\frac{B_\kappa^i}{B_{\kappa'}^i} = \left(\frac{\hat{\delta}_{b^{i-1},\kappa'}^{i-1}}{\hat{\sigma}_{\kappa'}^{i-1}} \cdot \frac{\hat{\sigma}_\kappa^{i-1}}{\hat{\delta}_{b^{i-1},\kappa}^{i-1}}\right)^2, \quad \forall \kappa, \kappa' = 1, \ldots K, \tag{11}$$

$$B_{b^{i-1}}^i = \hat{\sigma}_{b^{i-1}}^{i-1} \sqrt{\sum_{k=1,k\neq b^{i-1}}^{K} \frac{(B_k^i)^2}{(\hat{\sigma}_k^{i-1})^2}}, \tag{12}$$

$$\sum_{k=1}^{K} B_k^i = B_i, \tag{13}$$

where $\hat{\delta}_{b^{i-1},k}^{i-1}$ represents the sample mean difference between groups b^{i-1} and k in iteration $i-1$ (i.e., $\hat{\delta}_{b^{i-1},k}^{i-1} = \hat{\upsilon}_k^{i-1} - \hat{\upsilon}_{b^{i-1}}^{i-1}$).

Step 7: Randomly sample B_k^i solutions from each group $k = 1, \ldots, K$. Run the high-fidelity simulation exactly one time on each of the sample solutions. Collect the simulation results and update the group-specific sample statistics information (i.e., sample mean $\hat{\upsilon}^i := (\hat{\upsilon}_1^i, \ldots, \hat{\upsilon}_K^i)$ and sample standard deviation $\hat{\sigma}^i := (\hat{\sigma}_1^i, \ldots, \hat{\sigma}_K^i)$).

Step 8: Set $B_{i+1} = B_i - \Delta$. If there is remaining budget (i.e., $B - \sum_{a=1}^{i} B_a \geq B_{i+1}$), set $i = i + 1$ and go to Step 6. Otherwise, STOP and output the best solution over the past i iterations.

$$\omega_j^H = \begin{cases} \left[\dfrac{\mu_j^H p_j^H}{1-p_j^H}\left(\dfrac{1}{\mu_j^H - \lambda_j(\mathbf{x},\mathbf{y})} - \dfrac{e^{-\left(\mu_j^H - \lambda_j(\mathbf{x}^H,\mathbf{x}^L,\mathbf{y})\right)b_H}\lambda_j(\mathbf{x}^H,\mathbf{x}^L,\mathbf{y})}{\left(\mu_j^H - \lambda_j(\mathbf{x}^H,\mathbf{x}^L,\mathbf{y})\right)\mu_j^H}\right)+1\right]^{-1}, & \text{if } \rho_j^H \neq 1, \\ \dfrac{\lambda_j(\mathbf{x}^H,\mathbf{x}^L,\mathbf{y})}{\lambda_j(\mathbf{x}^H,\mathbf{x}^L,\mathbf{y}) + \frac{\mu_j^H p_j^H}{1-p_j^H}(1+\lambda_j(\mathbf{x}^H,\mathbf{x}^L,\mathbf{y})b_H)}, & \text{if } \rho_j^H = 1, \end{cases}$$

$$\omega_j^L = \begin{cases} \left[\dfrac{\mu_j^L p_j^L}{1-p_j^L}\left(\dfrac{1}{\mu_j^L - \lambda_j(\mathbf{x}^H,\mathbf{x}^L,\mathbf{y})} - \dfrac{e^{-\left(\mu_j^L - \lambda_j(\mathbf{x}^H,\mathbf{x}^L,\mathbf{y})\right)b_L}\lambda_j(\mathbf{x}^H,\mathbf{x}^L,\mathbf{y})}{\left(\mu_j^L - \lambda_j(\mathbf{x}^H,\mathbf{x}^L,\mathbf{y})\right)\mu_j^L}\right)+1\right]^{-1} & \text{if } \rho_j^L \neq 1, \\ \dfrac{\lambda_j(\mathbf{x}^H,\mathbf{x}^L,\mathbf{y})}{\lambda_j(\mathbf{x}^H,\mathbf{x}^L,\mathbf{y}) + \frac{\mu_j^H p_j^L}{1-p_j^L}(1+\lambda_j(\mathbf{x}^H,\mathbf{x}^L,\mathbf{y})b_L)} & \text{if } \rho_j^L = 1, \end{cases}$$

$$W_j^H\left(\lambda_j(\mathbf{x}^H,\mathbf{x}^L,\mathbf{y})\right) = \begin{cases} \dfrac{\omega_j^H \mu_j^H p_j^H\left[1 - \left(\mu_j^H - \lambda_j(\mathbf{x}^H,\mathbf{x}^L,\mathbf{y})\right)b_H \cdot e^{-\left(\mu_j^H - \lambda_j(\mathbf{x}^H,\mathbf{x}^L,\mathbf{y})\right)b_H} - e^{-\left(\mu_j^H - \lambda_j(\mathbf{x}^H,\mathbf{x}^L,\mathbf{y})\right)b_H}\right]}{\left(1-p_j^{B_H}\right)\left(1-p_j^H\right)\left(\mu_j^H - \lambda_j(\mathbf{x}^H,\mathbf{x}^L,\mathbf{y})\right)^2} & \text{if } \rho_j^H \neq 1, \\ \dfrac{\omega_j^H b_H^2}{2\left(1-p_j^{B_H}\right)}\dfrac{\mu_j^H p_j^H}{1-p_j^H} & \text{if } \rho_j^H = 1. \end{cases}$$

$$W_j^L\left(\lambda_j(\mathbf{x}^H,\mathbf{x}^L,\mathbf{y})\right) = \begin{cases} \dfrac{\omega_j^L \mu_j^L p_j^L\left[1 - \left(\mu_j^L - \lambda_j(\mathbf{x}^H,\mathbf{x}^L,\mathbf{y})\right)b_L \cdot e^{-\left(\mu_j^L - \lambda_j(\mathbf{x}^H,\mathbf{x}^L,\mathbf{y})\right)b_L} - e^{-\left(\mu_j^L - \lambda_j(\mathbf{x}^H,\mathbf{x}^L,\mathbf{y})\right)b_L}\right]}{\left(1-p_j^{B_L}\right)\left(1-p_j^L\right)\left(\mu_j^L - \lambda_j(\mathbf{x}^H,\mathbf{x}^L,\mathbf{y})\right)^2} & \text{if } \rho_j^L \neq 1, \\ \dfrac{\omega_j^L b_L^2}{2\left(1-p_j^{B_L}\right)}\dfrac{\mu_j^L p_j^L}{1-p_j^L} & \text{if } \rho_j^L = 1. \end{cases}$$

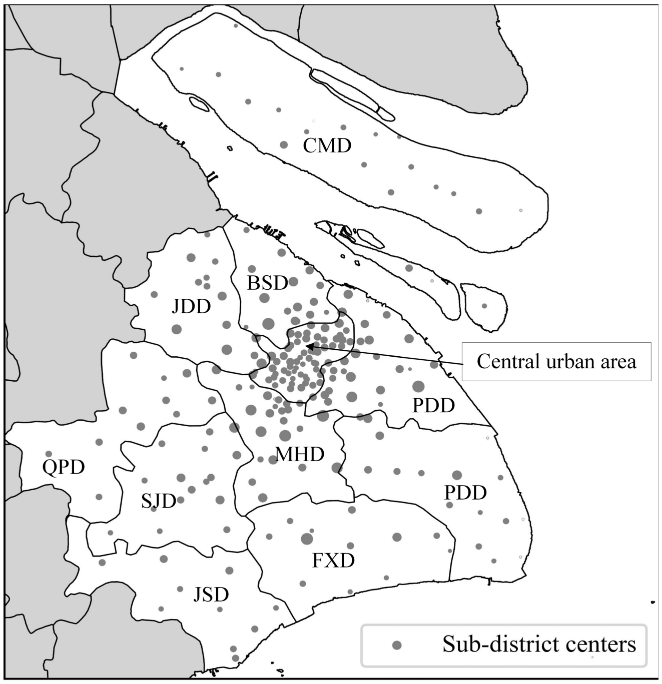


Fig. 1. The spatial distribution of sub-districts[1].

IV. Case Study

In this section, we present a real-world case study. For the case study, we used the multi-hospital system in Shanghai as the real-world context. We aimed to answer the following questions: (1) where new hospitals of either type should be built and what existing DHs should be upgraded; (2) what happens to the design of the hospital system with the emergence of population center; and (3) what happens to the design of the hospital system if the redesign budget increases. Through this real-world case study, we expect to offer system redesign recommendations to the Shanghai municipal government.

A. Background

In Shanghai, there are 16 administrative districts, or equivalently 230 administrative sub-districts (available from http://www.shanghai.gov.cn/). These sub-districts can be labeled as central urban area, semi-central semi-suburban area, and suburban area; see Figure 1. Table X in Appendix B lists the district names and the corresponding population distribution. In our model, we used the 230 sub-districts in Shanghai to identify the care demand and provision sites. The current system comprises 285 hospitals, 39 of which are CHs, 246 are DHs; see Figure 2–3. Tabulation on the 2010 Population Census of the People's Republic of China by Township (available from www.stats.gov.cn/) provides the population quantity of permanent residents at each of the 230 sub-districts in Shanghai; see Table XI and Figure 1. Figure 2 shows that most of the CHs are located in central urban areas, where the population is only 30.35% of the total population of Shanghai.

The two-week consultation rate for outpatient services is reported to differ by age [52]. In our case study, we divided the age group into two groups: 0-64 years old and over 65 years old. From the Shanghai Statistics Year Book (available from www.tjj.sh.gov.cn/), we extracted the population of either age group, and derived the outpatient service demand in two weeks based on the consultation rate. By assuming seven work days per week and 8 work hours per day, we converted the population quantity into the unit-time outpatient service demand from either age group, to calculate the mean arrival rate. Finally, based on the real data of patient service time at partnering CHs (Ruijin hospital and Shanghai No. 6 People Hospital) and DHs (Xujiahui Street Community Health Service Center, Longhua Street Community Health Service Center, and Hongmei Street Community Health Service Center) between 2015–17, we estimated the mean service rates at CHs and DHs.

[1]Sub-districts with more population correspond to larger points.

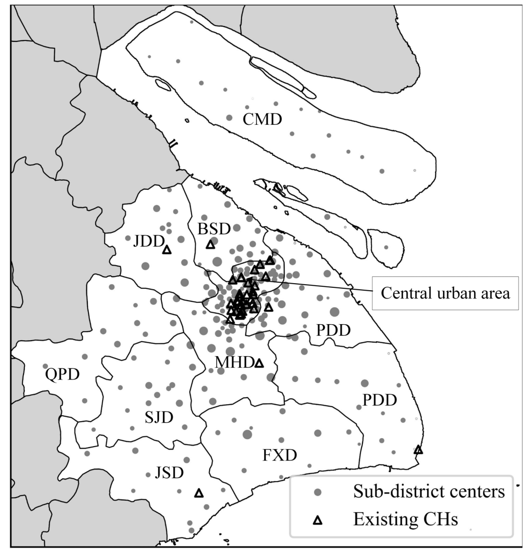


Fig. 2. Current location of CHs.

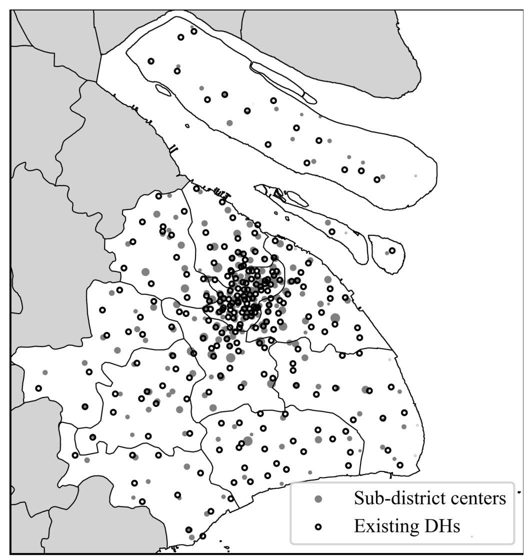


Fig. 3. Current location of DHs.

TABLE I
CHOICE MODEL COEFFICIENT ESTIMATION RESULTS

Age Group	Variable Name	Parameter Estimate	Standard Error	p –Value
0-64	q_j	1.099	0.235	0.000
	d_{ij}	-0.085	0.014	0.000
	Constant	0.108	0.109	0.326
65+	q_j	0.541	0.283	0.056
	d_{ij}	-0.042	0.016	0.008
	Constant	0.150	0.141	0.286

B. Choice Modeling Results

To characterize how patient hospital visit behaviors are influenced by the hospital type and distance to hospital, we designed a questionnaire (see Appendix A). We administered an online survey with the questionnaire in September 2021 on a Chinese internet survey platform (www.wenjuan.com). A total of 1018 respondents in Shanghai participated in our study, and 997 of which were deemed valid samples. The respondents are anonymous and the data source is reliable.

In the questionnaire, we presented the following scenario. We asked each respondent to imagine that she experienced fever and cough, accompanied by chest tightness, shortness of breath and other symptoms. This is likely a respiratory disease. As a result, the respondent would like to make an outpatient visit. We then provided each respondent a choice card containing two hospitals identified by distance and type. The hospital attributes were randomly generated. We set a plausible range of distance on each hospital, from which we drew uniform samples. The lower bound of the distance range on DH is 0 and the upper bound is 3 kilometers. The lower bound of the distance range on CH is 1 and the upper bound is 30 kilometers.

With the behavior experiment data, we parameterized a choice model for either age group. Table I presents our choice model results, i.e., respective estimates of parameters β_d^r and β_q^r from equation (7). We verified that for both age groups, the attribute *hospital type* is significant and positively correlates to the hospital choice, suggesting that patients prefer a CH than a DH when the distances to the two hospitals being equal. Our results also verified that for both age groups, the attribute *distance to hospital* is a factor as significant as *hospital type*. That is, when hospital type being equal, the shorter the distance is, the more likely the patient is to visit it. In addition, by comparing the probabilities of choosing CH between the two age groups (see Appendix C), we found that when the differences between traveling to DH and to CH are the same for the two age groups, (1) patients aged 0-64 are less likely to choose to visit CH than patients aged 65+ if the distance between CH and DH is relatively large, and (2) patients aged 0-64 are more likely to choose to visit CH than patients aged 65+ if the distance between CH and DH is relatively small.

C. Network Redesign Results

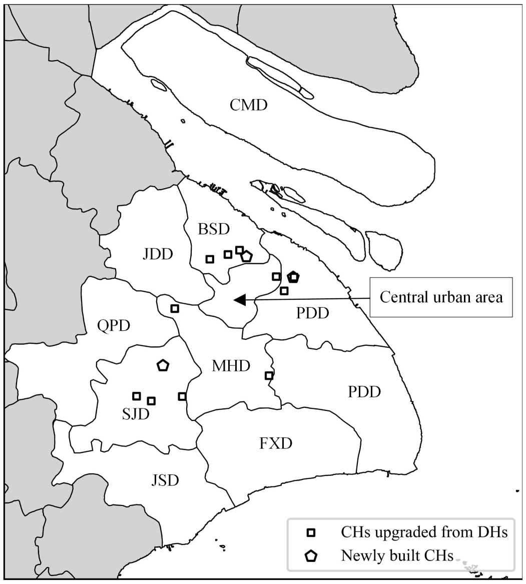
In this section, we report three Shanghai-based case studies. Through interviews with managers of some hospitals in Shanghai, we are able to obtain reasonable weight coefficient and each cost in the objective function. We set C_{d_H}, C_{d_L}, C_{W_H}, and C_{W_L}, the per kilometer travel cost of CH and DH, and the per hour waiting cost of CH and DH, to be 1.8 RMB/kilometer, 2 RMB/kilometer, 12 RMB/hour, 8 RMB/hour, respectively. We also set α_d, the weighting coefficient of distance, to be 0.4, and α_W, the weighting coefficient of waiting times, to be 0.6. We set the upper bounds of the balking probability at CH and DH, i.e., θ_H and θ_L, to be 0.2 and 0.3, respectively. The cost of land acquisition and facility construction for locating a CH and DH is from literature [53] and [54] and website (www.sohu.com/). We set κ_x^H, κ_x^L, the cost of land acquisition and facility construction for locating a CH and a DH, to be 1.4×10^8 RMB and 3×10^6 RMB, respectively, κ_y, the costs of upgrading a DH, to be 1.37×10^8 RMB, and T_B, the redesign budget, to be 2×10^9 RMB. Through field investigation at hospitals, we estimated the service rates of a CH and a DH to be $\mu_j^H = 100$ patients/hour and $\mu_j^L = 10$ patients/hour, and by conducting small-scale investigations in several hospitals into the threshold on the wait time tolerance, we specified the threshold at CH to be $b_H = 2$ hour and the threshold at DH to be $b_L = 0.5$ hour. We then set the lower bounds for the proportion of CHs and DHs at which the balking probability does not exceed b_H and b_L according to experience, i.e., ϕ_H and ϕ_L, to be 0.8 and 0.7, respectively. Finally, we assumed each patient balks at most twice in the simulation model.

For the implementation of our multi-fidelity optimization approach, we set the number of groups $K = 6$, initial sampling budget $B_0 = 30$, total iteration budget $B = 90$, budget of iteration 1 $B_1 = 26$, and budget decrement over two consecutive iterations $\Delta = 4$. In the GA algorithm, we set the population size $N = 40$, and the number of generations $T = 600$.

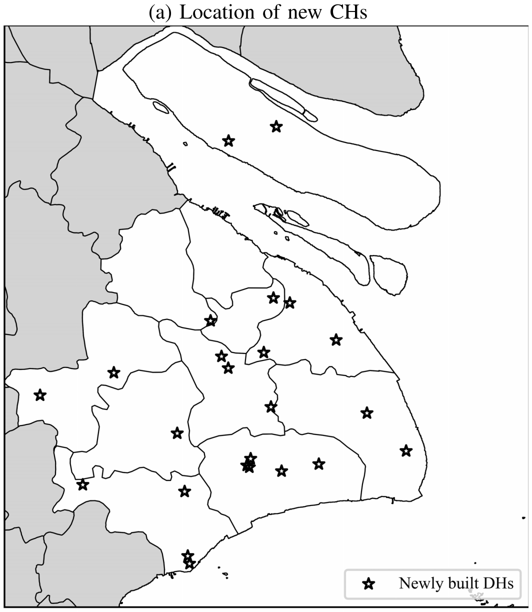
Study 1: What are the changes of the care network through the proposed redesign?

Figure 4 displays that 3 new CH and 24 new DHs are built and 11 existing DHs are upgraded in the optimal redesign. CHs in the current network are mainly clustered in central city (see Figure 2), whereas the location of new CHs in the optimal redesign is more dispersed (see Figure 4a). Taking a close look at the optimal location, we noticed that some suburbs with large populations, such as Baoshan District, Songjiang District, and Minhang District, add CHs. Upgraded DHs are also distributed in suburban areas that previously had no CHs or few CHs, such as Songjiang District and Baoshan District. This can be explained as follows. Relatively scattered CHs in the optimally redesigned network would be conveniently accessible to surrounding residents that still primarily concern the distance to hospital when choosing which hospital to visit. Moreover, new DHs in the optimal redesign are dispersed in each district. These results imply that the hospital network redesign is aligned with the expansion and dispersion of residential communities.

We compared the waiting time and traveling distance for accessing care between the optimally redesigned network and the current network. For the interest of space, we term two outcome measures as *wait time* and *distance*. Table II indicates that the optimal redesign, mostly with upgrading existing DHs



(a) Location of new CHs



(b) Location of new DHs

Fig. 4. Location for new CHs and DHs in the optimal redesign.

and building new DHs, reduces mean CH and DH wait times and mean DH distance to some extent. On the other hand, the mean CH distance slightly increases. Most of the observations are intuitive and only the last piece deserves a bit explanation. Unlike the locations of existing CHs, new CHs are often built far away from central city to cover recently appeared residential communities. Given that residential communities are still mostly concentrated at central city, some of those central-city dwellers may have to travel to suburban CHs with the redesigned network. Such slight inconvenience is largely offset by the increase in care access across the board with much reduced in-facility wait time. Note that Chinese patients have traditionally the tendency of visiting CHs without minding much the travel distance or even the waiting time. For more information, please refer to Appendix D.

TABLE II
PERFORMANCE COMPARISON BETWEEN THE OPTIMALLY REDESIGNED CARE NETWORK AND THE CURRENT NETWORK

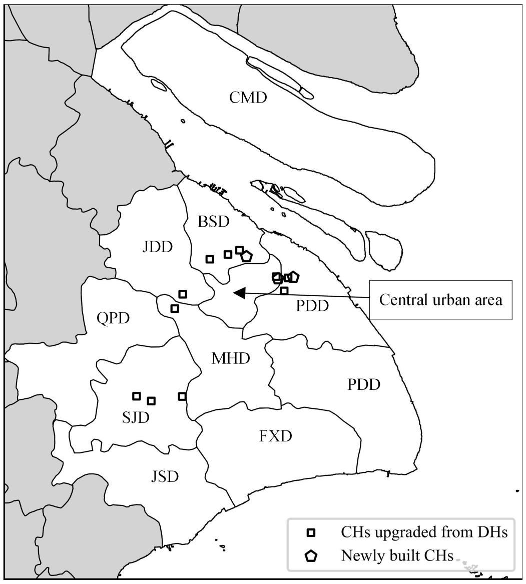
Hospital type	Performance		Study 1	Current	p-value
CH	Wait time (h)	mean	1.386	1.920	0.000
		min	1.361	1.919	\
		max	1.455	1.921	\
	Distance (Km)	mean	18.628	16.494	0.000
		min	18.562	16.468	\
		max	18.886	16.524	\
DH	Wait time (h)	mean	0.065	0.140	0.000
		min	0.064	0.138	\
		max	0.067	0.142	\
	Distance (Km)	mean	0.983	1.098	0.000
		min	0.977	1.093	\
		max	0.989	1.100	\

Furthermore, to discuss the benefit of Exploitation of the multi-fidelity optimization, we compare the inputs and outputs of Exploration and Exploitation respectively. We take the best solution obtained by low-fidelity model (Exploration) as the input (shown in Figure 5), and run the high-fidelity simulation to obtain the performance measures, such as wait time and distance as outputs of Exploration procedure. As for the Exploitation procedure, its input is the best solution obtained through the multi-fidelity optimization approach (shown in Figure 4). And the performance measures obtained from the high-fidelity simulation, such as wait time and distance, are the outputs of Exploitation procedure. Figure 5a displays the optimal 14 new CHs, including 3 newly built and 11 upgraded. Comparing Figure 4a and Figure 5a, the main difference is the locations of some CHs in the three suburbs of Minhang District, Jiading District, and Songjiang District. The location of new CHs in the Exploitation procedure is more dispersed, so as to meet the care needs of suburban residents as much as possible. By comparing the outputs of Exploration and Exploitation (see Table III), we find the objective function value, the mean CH and DH wait times, and the mean CH and DH distance obtained from Exploitation are all smaller than those from Exploration, which verifies the effectiveness of Exploitation.

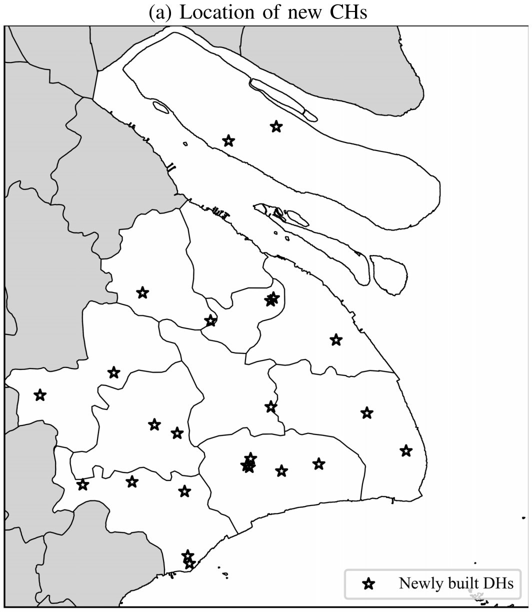
Study 2: What is the impact on the optimal network redesign from the emergence of population center?

Next we assumed the emergence of a new population center somewhere in Shanghai (say e.g., Qingpu District–QPD, a northwest suburb) and compared the optimal network redesign before (i.e., optimal design from Study 1) and after. This study was inspired by the trend of Shanghai's rapid changing spatial distribution of its population. In this study, we increased the population quantity of some sub-district far from the city center (see Table IV) and kept other model parameters the same as before.

Figures 6 displays that 4 new CHs and 23 new DHs are built and 10 existing DHs are upgraded. Compared with Study 1, we noticed that 2 new CHs in Songjiang District are located towards the new population center. For two adjacent districts of Qingpu District, two CHs in Songjiang District are removed and one CH in Minhang District are moved to the remote



(a) Location of new CHs



(b) Location of new DHs

Fig. 5. Location for new CHs and DHs in the low-fidelity model.

TABLE III

PERFORMANCE COMPARISON BETWEEN THE LOW-FIDELITY MODEL (EXPLORATION) AND THE HIGH-FIDELITY MODEL (EXPLOITATION)

Hospital type	Performance		Outputs: Exploitation	Outputs: Exploration	p-value
	Objective function value	mean	2.03×10^7	2.08×10^7	0.000
		min	2.01×10^7	2.03×10^7	\
		max	2.04×10^7	2.10×10^7	\
CH	Wait time (h)	mean	1.386	1.435	0.000
		min	1.361	1.354	\
		max	1.455	1.462	\
	Distance (Km)	mean	18.628	18.753	0.000
		min	18.562	18.698	\
		max	18.886	18.832	\
DH	Wait time (h)	mean	0.065	0.069	0.000
		min	0.064	0.068	\
		max	0.067	0.071	\
	Distance (Km)	mean	0.983	0.985	0.000
		min	0.977	0.965	\
		max	0.989	0.944	\

TABLE IV

POPULATION SIZE OF THE SELECTED SUB-DISTRICTS

	Current		Alternative	
Residential site	QPD5	QPD6	QPD5	QPD6
Population quantity	59600	57960	238400	173880

TABLE V

PERFORMANCE COMPARISON WITH EMERGENCY OF A NEW POPULATION CENTER

Hospital type	Performance		Study 2	Study 1	p-value
CH	wait time (h)	mean	1.519	1.386	0.000
		min	1.502	1.361	\
		max	1.530	1.455	\
	Distance (km)	mean	18.413	18.628	0.000
		min	18.376	18.562	\
		max	18.450	18.886	\
DH	wait time (h)	mean	0.062	0.065	0.000
		min	0.061	0.064	\
		max	0.064	0.067	\
	Distance (km)	mean	0.981	0.983	0.000
		min	0.977	0.977	\
		max	0.985	0.989	\

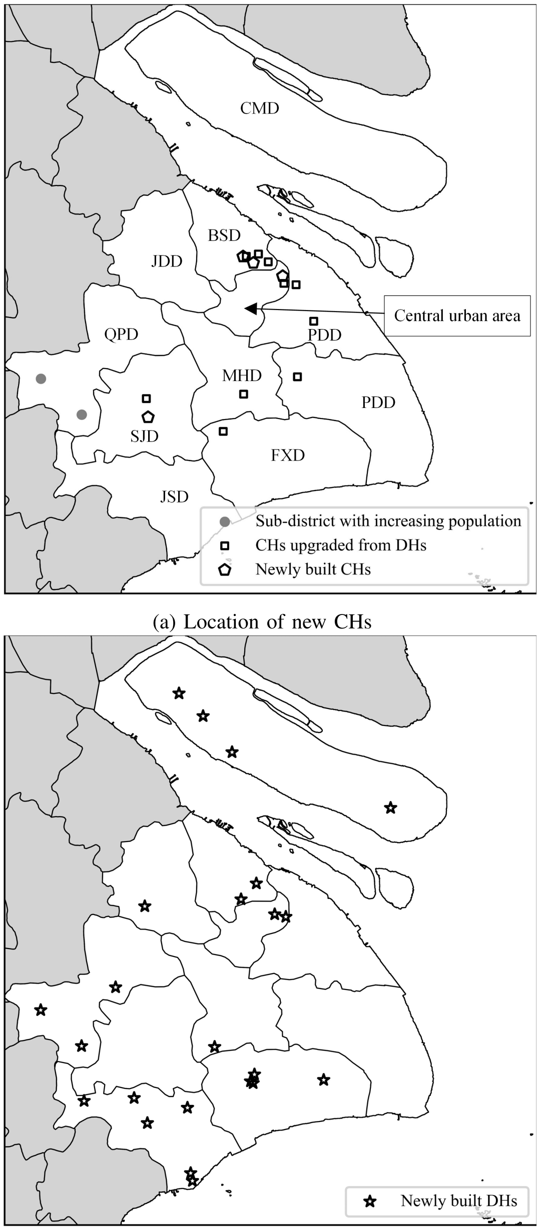
area. On the other hand, one CH are built in Fengxian District and one CH are located in remote area of Pudong District, which intuitively is used to meet the care access requirement by residents in suburban area. The rest of the network is similar to Study 1.

Further, we compared the optimal redesign of this study with that from Study 1, in terms of the wait time and distance. Table V indicates that the mean CH wait time has a slight increase compared to before, whereas the other three outcomes hardly change. The performance comparison suggests reasonable robustness with the optimal network redesign in response to modest changes in the spatial population distribution.

Study 3: What is the impact on the optimal network redesign from the increase of redesign budget?

In this study, we increased the government budget for the network redesign. We kept the other model parameters the same as in Study 1. First, we increased the budget from 2×10^9 RMB to 3×10^9 RMB.

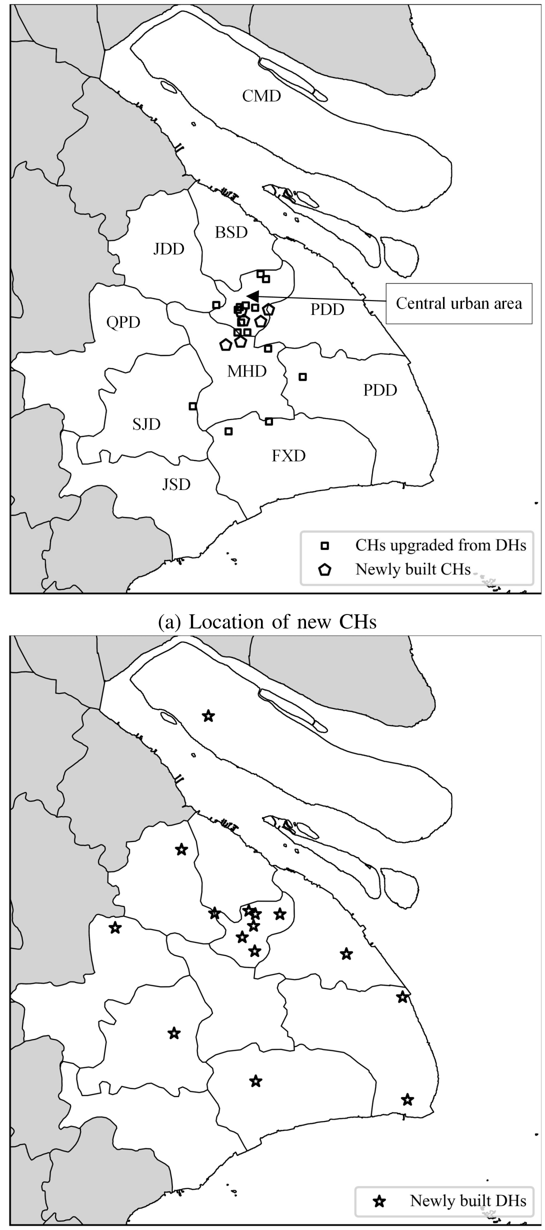
Figure 7a displays the optimal location of 21 new CHs, including 6 newly built and 15 upgraded, with the increased budget. Compared with the baseline optimal redesign, new CHs appear in central city and on the periphery of the suburbs. This can be explained as follows. The central urban area in Shanghai have relatively densely populated residential communities. Thus, when the budget is sufficient, adding CHs in central city has a better chance to achieve the system-wide requirement on care access. Figure 7b displays the optimal location of 15 newly built DHs. Compared with Study 1, the newly built DHs in Study 3 are decreased for the reason that the construction budget should be allocated to the scarce CHs as much as possible. Table VI shows that the mean CH and DH wait times decrease significantly, but the mean CH distance increases slightly. This is also because Chinese patients have



(a) Location of new CHs

(b) Location of new DHs

Fig. 6. Location for new CHs and DHs after the emergence of the new population center.



(a) Location of new CHs

(b) Location of new DHs

Fig. 7. Location for new CHs and DHs with $B = 3 \times 10^9$.

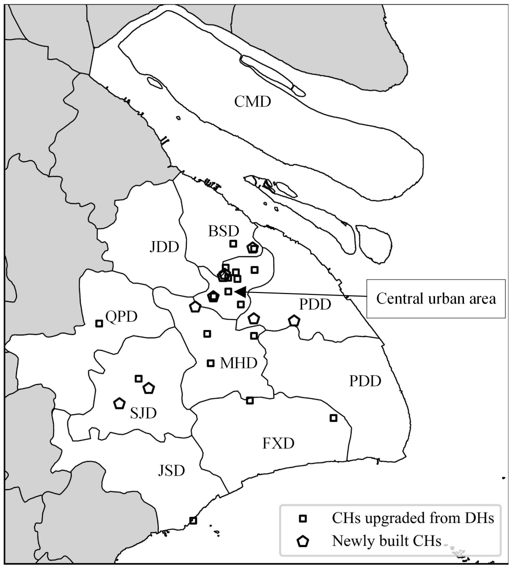
the tendency of visiting CHs without minding much of the travel distance. On the other hand, the mean DH distance decreases slightly.

We further increased the budget to $B = 4 \times 10^9$. Figure 8a displays that the optimal location of 28 CHs, including 9 new built and 19 upgraded. Compared to previous optimal location (i.e., $B = 3 \times 10^9$), the number of new CHs and DHs increases and the distribution of CHs and DHs is more scattered. Further, the suburban area that previously had no CHs, such as Qingpu District and Jingshan District, has a new CH. Table VIII shows that the mean CD and DH wait time further decreases as the number of new CHs and DHs increases, and the mean CH distance decreases and mean DH distance increases slightly.

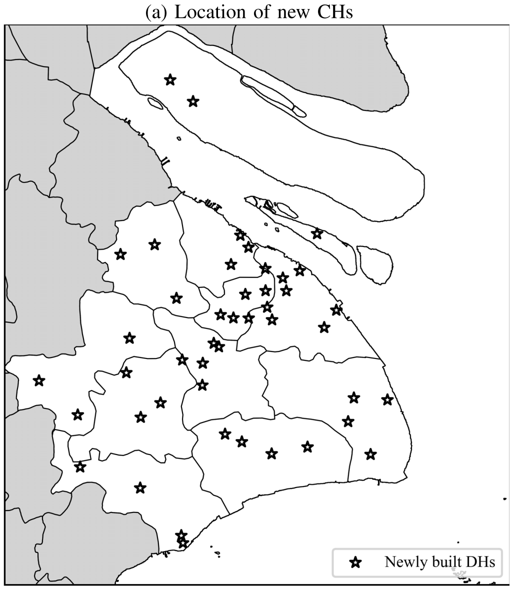
In addition to analyzing the impact of redesign budget on optimal network redesign, we next briefly analyze the impact of other parameters on system design. When decision makers pay more attention to the proximity of patients' visit to hospitals, i.e. $\alpha_d > \alpha_W$, the locations of CHs and DHs tend to be located in communities with large populations. When the threshold on the wait time tolerance at CH (i.e.,b_H) decreases, more CHs will be built to meet patients' care need. And conversely, when the threshold at CH b_H increases, less CHs need be added. Similarly, for b_L, when the threshold at DH decreases, more DHs will be built to meet patients' care need. And when the threshold at DH increases, less DHs need be added.

Study 4: What is the impact on the optimal network redesign considering patients with different income levels?

In this case study, we categorized patients by income level. With the behavior experiment data, we parameterized

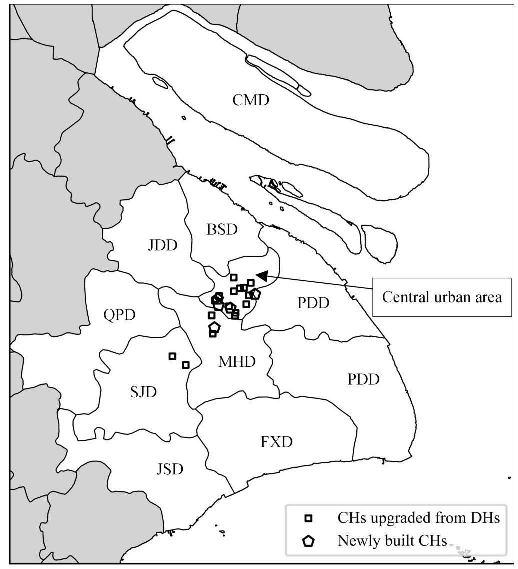


(a) Location of new CHs

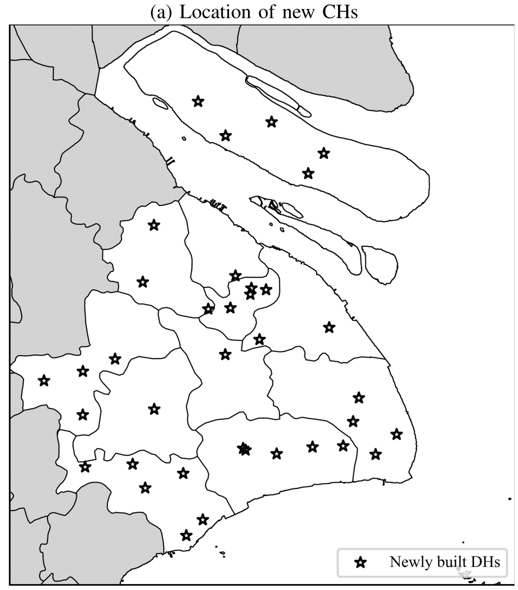


(b) Location of new DHs

Fig. 8. Location for new CHs and DHs with $B = 4 \times 10^9$.



(a) Location of new CHs



(b) Location of new DHs

Fig. 9. Location for new CHs and DHs when considering patients with different income levels.

TABLE VI
PERFORMANCE COMPARISON WHEN THE BUDGET INCREASES FROM 2×10^9 TO 3×10^9

Hospital type	Performance		$B = 3 \times 10^9$	$B = 2 \times 10^9$	p-value
CH	wait time (h)	mean	0.092	1.386	0.000
		min	0.089	1.361	\
		max	0.094	1.455	\
	Distance (Km)	mean	19.135	18.628	0.000
		min	19.111	18.562	\
		max	19.157	18.886	\
DH	wait time (h)	mean	0.051	0.065	0.000
		min	0.050	0.064	\
		max	0.053	0.067	\
	Distance (Km)	mean	0.803	0.983	0.000
		min	0.797	0.977	\
		max	0.811	0.989	\

a choice model for two income groups (0-10000 RMB and over 10000 RMB). Table VII presents the respective estimates of parameters β_d^r and β_q^r. From the result, we verified that for both income groups, the attribute hospital type is significant and the positive estimate of the parameter β_q^r suggests that patients prefer a CH than a DH when the distances to the two hospitals being equal. Our results also verified that for both income groups, the attribute distance to hospital is a significant factor. That is, when hospital type being equal, the shorter the distance is, the more likely the patient is to visit it.

Figure 9 displays that 5 new CHs and 36 new DHs are built and 16 existing DHs are upgraded, with the budget $B = 3 \times 10^9$. CHs are mainly clustered in central city, and five new CHs is located in suburban area with large populations, i.e., Minhang District and Songjiang District (see Figure 9a). Similar to Study 3 with $B = 3 \times 10^9$, because of the relatively dense population in the central urban area, it is more likely to meet the system-wide care access needs by adding CHs in the central urban area when the budget is sufficient. And new DHs

TABLE VII

CHOICE MODEL COEFFICIENT ESTIMATION RESULTS FOR DIFFERENT INCOME GROUPS

Income Group	Variable Name	Parameter Estimate	Standard Error	p –Value
0-10000	q_j	0.974	0.243	0.000
	d_{ij}	-0.075	0.014	0.000
	Constant	0.125	0.115	0.279
10000+	q_j	0.742	0.269	0.006
	d_{ij}	-0.057	0.015	0.000
	Constant	0.116	0.131	0.375

TABLE VIII

PERFORMANCE COMPARISON WHEN THE BUDGET FURTHER INCREASES FROM $B = 3 \times 10^9$ AND STUDY WITH $B = 4 \times 10^9$

Hospital type	Performance		$B = 4 \times 10^9$	$B = 3 \times 10^9$	p-value
CH	wait time (h)	mean	0.025	0.092	0.000
		min	0.024	0.089	\
		max	0.025	0.094	\
	Distance (km)	mean	18.595	19.135	0.000
		min	18.576	19.111	\
		max	18.632	19.157	\
DH	wait time (h)	mean	0.046	0.051	0.000
		min	0.045	0.050	\
		max	0.047	0.053	\
	Distance (km)	mean	0.977	0.803	0.000
		min	0.970	0.797	\
		max	0.984	0.811	\

TABLE IX

CHANGES IN PERFORMANCE WHEN CONSIDERING PATIENTS WITH DIFFERENT INCOME LEVELS

Hospital type	Performance		Study4	Study3	p-value
CH	wait time (h)	mean	0.103	0.092	0.000
		min	0.100	0.089	\
		max	0.106	0.094	\
	Distance (km)	mean	20.030	19.135	0.000
		min	20.007	19.111	\
		max	20.044	19.157	\
DH	wait time (h)	mean	0.052	0.051	0.009
		min	0.050	0.050	\
		max	0.054	0.053	\
	Distance (km)	mean	0.898	0.803	0.000
		min	0.891	0.797	\
		max	0.906	0.811	\

in the optimal redesign are dispersed in each district, as shown in Figure 9b. The hospital network redesign is aligned with the expansion and dispersion of residential communities.

Table IX indicates that the mean CH distance has a slight increase compared to Study 3 with budget $B = 3 \times 10^9$, whereas the other three outcomes change less. This can be explained as follows. Compared with Figure 7a, CHs in Figure 9b are more concentrated in central city, which leads to a longer distance for suburban patients to CH. The slight inconvenience to visit CH is largely offset by the increase in care access across the board with much reduced in-facility wait time at CH.

V. CONCLUSION AND FUTURE RESEARCH

Healthcare capacity expansion and spatial redistribution needs to coordinate with care-seeking behavior, population change, and government budget. A well redesigned hospital network is expected to improve care access for the entire population without incurring significant spending on establishing new hospitals and upgrading the existing ones.

In this paper, we studied the optimal network redesign problem for a two-tier hospital network consisting of central hospitals and district hospitals, which was inspired by real challenges in Chinese metropolitan areas. In our study, we developed (1) a multinomial logit choice model to characterize hospital visit behaviors of different patient groups; (2) a queuing network model with consideration of patient balking and choice model derived arrival rates; (3) a multi-fidelity optimization approach with the queuing network model being low-fidelity and a self-developed simulation being high-fidelity; and (4) a case study based on the hospital network of Shanghai municipality. Our study makes two contributions: (1) integrating choice models of multiple patient types into multi-type facility network redesign optimization with queuing performance measures in the objective; (2) embedding genetic algorithm into the multi-fidelity optimization framework, which speeds up the execution of discrete location optimization based on the low-fidelity queuing network model; and (3) conducting a proof-of-the-concept case study based on real-world data for Shanghai, a place of imminent need, with anticipation that we will make meaningful recommendation to Shanghai's healthcare system. However, limitation still exist. In the choice model, the current work only analyzes the impact of distance and hospital type on patients' choice behavior in the situation where CH is more crowded than DH, but do not analyze the impact of specific waiting time of each hospital as one of the attributes in the choice model.

The key findings of the patient hospital visit choice modeling are (1) hospital type (central vs. district) and distance to hospital are influential factors in patient behavior; (2) younger patients are less likely to choose CH than older patients if the distance between CH and DH is relatively large, and younger patients are more likely to choose CH than older patients if the distance between CH and DH is relatively small. Through the location analysis in our case study, we can make recommendations on care facility expansion and dispersion to better align with the spatial distribution of residential communities.

We plan to pursue further research in the following directions. First, we plan to incorporate additional factors such as the detailed waiting time into the choice model. Thus, arrival rate and waiting time will affect each other, and then need to solve the equilibrium problem between them, which will make the model more complex and interesting. Second, we will introduce additional heterogeneity among hospitals such as differentiated pricing in medical services, and further set the multi-service pricing decision in a model extension. Moreover, we will investigate the robustness of the optimal network design in face of further population aging in the coming decades.

APPENDIX A
THE HOSPITAL VISIT CHOICE BEHAVIOR QUESTIONNAIRE

In this appendix, we present the questionnaire we used to survey a cohort of online respondents and model the hospital visit choice behavior in Shanghai. The questionnaire is originally written in Chinese. We provide its English translation here.

A. Introduction

You are being invited to take part in a research study about hospital visit choices. Please note that there are no right or wrong answers to any questions in this questionnaire. We are only interested in your opinions and feedback. Your kind and valid response will help the Shanghai municipal government develop better hospital network and will help make you feel more satisfied about the accessibility of hospitals in the future. This questionnaire should take approximately 3-5 minutes to complete.

We assure you that the responses you provide will not be linked to any personal identifiable information. Your participation in this study is on a voluntary basis and you are free to withdraw from the study at any time without penalty. We thank you again for your willingness to participate in this study. Please feel free to contact us if you need any additional information about this project.

B. Section 1: Basic Information

The first section of the questionnaire includes questions about your demographics and other related information. We will only use your responses to these questions to compare across survey participants. We assure you that your privacy is protected.

1) What is your gender?
   a) Female b) Male
2) Which of the following categories does your age falls into?
   a) 0-18 b) 19-34 c) 35-49
   d) 50-64 e) 65-69 f) over 70
4) What's your occupation?
   a) Unemployed b) Student
   c) Employees of state-owned enterprises and institutions d) Self-employed or private owners
   e) Employees of private or foreign companies f) Peasant
   g) Worker h) Retired
   i) Other (Please specify)
5) Which of the following income groups includes your monthly individual income
   a) less than 3000 RMB b) 3000-5000 RMB
   c) 5000-10000 RMB d) 10000-30000 RMB
   e) over 30000 RMB

C. Section 2: Choice Scenario

In the following, we will present a scenario where we would like you to choose whether to go to a central hospital or a community-based hospital. Please note that there are no correct or incorrect responses, and your choice should be based on your own preferences, experiences, and specific needs.

Suppose you had fever and cough with headache, muscle pain and other symptoms, you would go to a hospital in need of basic medical service, e.g., an outpatient consultation. Imagine you have two options, either going to a central hospital, i.e., a CH, or a community-based hospital, i.e., a DH. Please note that in Shanghai, CH is more crowded than DH, and the access time (indirect waiting time to get an appointment) to CH is usually longer than that to DH.

	Alternatives	
	Hospital 1	Hospital 2
Type	Central hospital	Community-based hospital
Distance	X_1	X_2
Which one would you choose?		

TABLE X
DIVISION OF SHANGHAI AND THE CORRESPONDING PROPORTION OF POPULATION

Type	Administrative divisions	Percentage of population	
		each district	each type
Central urban area	HPD (Huangpu)	2.95%	30.35%
	JAD (Jingan)	4.68%	
	XHD (Xuhui)	4.71%	
	CND (Changning)	3.00%	
	YPD (Yangpu)	5.70%	
	HKD (Hongkou)	3.70%	
	PTD (Putuo)	5.60%	
Semi central area and semi suburban area	PDD (Pudong)	21.91%	21.91%
Suburban area	BSD (Baoshan)	8.28%	47.74%
	JDD (Jiading)	6.39%	
	MHD (Minhang)	10.55%	
	SJD (Songjiang)	6.87%	
	QPD (Qingpu)	4.70%	
	FXD (Fengxian)	4.71%	
	JSD (Jinshan)	3.18%	
	CMD (Chongming)	3.06%	

APPENDIX B
ADDITIONAL INFORMATION ON DISTRICTS AND SUB-DISTRICTS OF SHANGHAI

Table X and XI list the relevant data on districts and sub-districts of Shanghai, respectively.

APPENDIX C
EXPLANATIONS ABOUT THE CONCLUSION ON SECTION IV-B

Without loss of generality, we assume that there are a CH and a DH, and patients choose one hospital according to their preference modeled by equation (7). Based on the coefficient estimation results in Table I, the probabilities of choosing the CH for the two types of patients are given by equation (14) and (15).

$$p_C^{0\text{-}64} = \frac{e^{(-0.085*d_C^{0\text{-}64}+1.099*q_C+0.108)}}{e^{(-0.085*d_C^{0\text{-}64}+1.099*q_C+0.108)}+e^{(-0.085*d_D^{0-64}+1.099q_D+0.108)}} = \frac{1}{1+e^{(-0.085*(d_D^{0\text{-}64}-d_C^{0\text{-}64})+1.099*(q_D-q_C))}}; \quad (14)$$

TABLE XI
POPULATION QUANTITY OF EACH SUB-DISTRICT IN SHANGHAI

HPD1	**HPD2**	**HPD3**	**HPD4**	**HPD5**	**HPD6**	**LWD1**	**LWD2**	**LWD3**	**LWD4**
66285	64896	89776	74994	61042	72898	82403	59085	57931	49360
XHD1	**XHD2**	**XHD3**	**XHD4**	**XHD5**	**XHD6**	**XHD7**	**XHD8**	**XHD9**	**XHD10**
60533	36281	69710	112400	118872	97171	34877	100444	92915	108582
XHD11	**XHD12**	**XHD13**	**XHD14**	**CND1**	**CND2**	**CND3**	**CND4**	**CND5**	**CND6**
85769	97917	67415	2244	72730	51883	73230	56628	73757	84664
CND7	**CND8**	**CND9**	**CND10**	**JAD1**	**JAD2**	**JAD3**	**JAD4**	**JAD5**	**PTD1**
59551	24487	46865	146776	75272	34288	36544	29173	71511	98267
PTD2	**PTD3**	**PTD4**	**PTD5**	**PTD6**	**PTD7**	**PTD8**	**PTD9**	**ZBD1**	**ZBD2**
120920	128647	112498	120217	111185	172397	229925	194825	34749	77968
ZBD3	**ZBD4**	**ZBD5**	**ZBD6**	**ZBD7**	**ZBD8**	**ZBD9**	**HKD1**	**HKD2**	**HKD3**
80726	97630	77710	156276	78079	74633	152725	73328	102564	122669
HKD4	**HKD5**	**HKD6**	**HKD7**	**HKD8**	**YPD1**	**YPD2**	**YPD3**	**YPD4**	**YPD5**
125634	98094	87401	113751	129035	100480	85870	95382	92505	105613
YPD6	**YPD7**	**YPD8**	**YPD9**	**YPD10**	**YPD11**	**YPD12**	**MHD1**	**MHD2**	**MHD3**
70195	90334	192554	124954	149090	27251	178994	185991	149141	65256
MHD4	**MHD5**	**MHD6**	**MHD7**	**MHD8**	**MHD9**	**MHD10**	**MHD11**	**MHD12**	**MHD13**
277934	283352	189604	193777	165877	344434	121164	103989	292750	56103
BSD1	**BSD2**	**BSD3**	**BSD4**	**BSD5**	**BSD6**	**BSD7**	**BSD8**	**BSD9**	**BSD10**
136814	104162	172284	118323	371856	204564	139328	54329	240185	127512
BSD11	**BSD12**	**BSD13**	**JDD1**	**JDD2**	**JDD3**	**JDD4**	**JDD5**	**JDD6**	**JDD7**
89615	127347	18567	55223	106164	60924	81854	139845	232503	172864
JDD8	**JDD9**	**JDD10**	**JDD11**	**JDD12**	**PDD1**	**PDD2**	**PDD3**	**PDD4**	**PDD5**
165452	46355	80896	256218	72933	100548	112507	144668	76916	104932
PDD6	**PDD7**	**PDD8**	**PDD9**	**PDD10**	**PDD11**	**PDD12**	**PDD13**	**PDD14**	**PDD15**
107130	112031	206017	146237	177468	121449	221327	20219	369032	184486
PDD16	**PDD17**	**PDD18**	**PDD19**	**PDD20**	**PDD21**	**PDD22**	**PDD23**	**PDD24**	**PDD25**
276547	132038	129267	186012	81537	137625	110552	165297	360516	213845
PDD26	**PDD27**	**PDD28**	**PDD29**	**PDD30**	**PDD31**	**PDD32**	**PDD33**	**PDD34**	**PDD35**
147329	84183	71162	27162	174672	110060	51013	104945	62519	59567
PDD36	**PDD37**	**PDD38**	**PDD39**	**PDD40**	**PDD41**	**PDD42**	**PDD43**	**PDD44**	**JSD1**
59323	24346	37408	688	508	862	1349	5514	23617	87901
JSD2	**JSD3**	**JSD4**	**JSD5**	**JSD6**	**JSD7**	**JSD8**	**JSD9**	**JSD10**	**SJD1**
120084	82477	37057	122272	52808	33658	70819	40722	84640	112671
SJD2	**SJD3**	**SJD4**	**SJD5**	**SJD6**	**SJD7**	**SJD8**	**SJD9**	**SJD10**	**SJD11**
93330	161438	98888	94279	75507	167687	155856	57861	253110	41626
SJD12	**SJD13**	**SJD14**	**SJD15**	**SJD16**	**QPD1**	**QPD2**	**QPD3**	**QPD4**	**QPD5**
44011	33627	80104	51606	60797	137321	118708	106830	94351	68485
QPD6	**QPD7**	**QPD8**	**QPD9**	**QPD10**	**QPD11**	**FXD1**	**FXD2**	**FXD3**	**FXD4**
67735	74409	127936	153203	39756	92288	361185	176938	62388	108264
FXD5	**FXD6**	**FXD7**	**FXD8**	**FXD9**	**FXD10**	**FXD11**	**FXD12**	**FXD13**	**CMD1**
65389	89163	62589	28457	57341	15413	29151	16710	10475	113442
CMD2	**CMD3**	**CMD4**	**CMD5**	**CMD6**	**CMD7**	**CMD8**	**CMD9**	**CMD10**	**CMD11**
60111	42737	45926	40823	26265	29894	40741	25274	53996	7061
CMD12	**CMD13**	**CMD14**	**CMD15**	**CMD16**	**CMD17**	**CMD18**	**CMD19**	**CMD20**	**CMD21**
23416	27466	11646	15112	99134	9581	27916	1695	35	1451

$$p_C^{65+} = \frac{e^{(-0.042*d_C^{65+}+0.541*q_C+0.150)}}{e^{(-0.042*d_C^{65+}+0.541*q_C+0.150)} + e^{(-0.042*d_D^{65+}+0.541*q_D+0.150)}} = \frac{1}{1+e^{(-0.042*(d_D^{65+}-d_C^{65+})+0.541*(q_D-q_C))}}. \quad (15)$$

Note that q_C (q_D) represents the indicator variable for the CH (DH), and $q_C = 1$, $q_D = 0$. $d_C^{0\text{-}64}$($d_D^{0\text{-}64}$) is the distance for a patient aged 0-64 to travel from his/her residential site to a CH (DH), and d_C^{65+} (d_D^{65+}) is the distance for a patient aged 65+ to travel. The probability of choosing the CH can thus be further derived as:

$$p_C^{0\text{-}64} = \frac{1}{1+e^{(-0.085*(d_D^{0\text{-}64}-d_C^{0\text{-}64})-1.099)}};$$

$$p_C^{65+} = \frac{1}{1+e^{(-0.042*(d_D^{65+}-d_C^{65+})-0.541)}}.$$

When the differences between traveling to the DH and traveling to the CH are the same for the two age groups of patients (i.e., $d_D^{0\text{-}64} - d_C^{0\text{-}64} = d_D^{65+} - d_C^{65+} = \Delta d$), we have $-0.085 * \Delta d - 1.099 - (-0.042 * \Delta d - 0.541) = -0.043 * \Delta d - 0.558$.

If $\Delta d < -12.98$ km, then $-0.043 * \Delta d - 0.558 > 0$, and we have $p_C^{0\text{-}64} < p_C^{65+}$, which means that younger patients are less willing to choose a CH than older patients when a CH is far away from a DH. On the contrary, If $\Delta d > -12.98$ km, then $-0.043 * \Delta d - 0.558 < 0$, and we have $p_C^{0\text{-}64} > p_C^{65+}$, which means that younger patients are more willing to choose a CH than older patients when a CH is not two far away from a DH.

Appendix D
Mathematical Explanations for the Conclusion of Study 1

Before the optimization of the hospital system, for patients at any residential site i who choose CH, we denote their probability of going to an existing CH j as $p_{ij}^H, i \in I, j \in J_H$. Thus we have $\sum_{j \in J_H} p_{ij}^H = 1, \forall i$. The mean distance from site i to these existing CHs, denoted as $\bar{d}_i^H$, is expressed as

$$\bar{d}_i^H = \sum_{j \in J_H} d_{ij} p_{ij}^H, \quad \forall i \in I. \quad (16)$$

Without loss of generality, when a new CH is located at site $j' \in J \cup J_L$ after the optimization, the probability of patients at residential site i visiting CH at site $j \in J_H$ or the new CH at site $j' \in J \cup J_L$ is denoted by $p_{ij}^{H'}$ and $p_{ij'}^{H'}$, respectively. We can derive the expression as follows.

$$\sum_{j \in J_H} p_{ij}^{H'} + p_{ij'}^{H'} = 1, \quad \forall j' \in J \cup J_L \quad (17)$$

The mean distance from site i to all CHs after the reconfiguration is represented as

$$\bar{d}_i^{H'} = \sum_{j \in J_H} d_{ij} p_{ij}^{H'} + d_{ij'} p_{ij'}^{H'} \quad (18)$$

The MNL model we used to quantify patient choice behavior has the property of independence of irrelevant alternatives (IIA). That is, the ratio of choice probabilities of any two alternatives is independent of the systematic utilities of any other alternatives [55]. For example, $p_{i1}^H / p_{i2}^H = p_{i1}^{H'} / p_{i2}^{H'}$. Therefore, we can obtain the following equation.

$$p_{ij}^{H'} = p_{ij}^H \cdot \left(1 - p_{ij'}^{H'}\right), \quad \forall i \in I, j \in J_H, j' \in J \cup J_L. \quad (19)$$

Thus, the mean distance $\bar{d}_i^{H'}$ can be expressed as

$$\begin{aligned} \bar{d}_i^{H'} &= \sum_{j \in J_H} d_{ij} p_{ij}^{H'} + d_{ij'} p_{ij'}^{H'} \\ &= \sum_{j \in J_H} d_{ij} p_{ij}^H \left(1 - p_{ij'}^{H'}\right) + d_{ij'} p_{ij'}^{H'} \\ &= \bar{d}_i^H \left(1 - p_{ij'}^{H'}\right) + d_{ij'} p_{ij'}^{H'} \\ &= \bar{d}_i^H + p_{ij'}^{H'} \left(d_{ij'} - \bar{d}_i^H\right), \quad \forall i \in I, j \in J_H, j' \in J \cup J_L \end{aligned} \quad (20)$$

When $d_{ij'} > \bar{d}_i^H$, we will get $\bar{d}_i^{H'} > \bar{d}_i^H$. In other words, if the distance from the new CH j' to residential site i is larger than the mean distance from existing CHs to residential site i, the mean distance from site i to all CHs in the reconfigured network will increase. In contrast, when $d_{ij'} < \bar{d}_i^H$, then $\bar{d}_i^{H'} < \bar{d}_i^H$, implying that if the distance from the new CH j' to residential site i is smaller than the mean distance from existing CHs to residential site i, the mean distance from site i to all CHs in the reconfigured network will decrease.

References

[1] C. Visser, G. Marincowitz, I. Govender, and G. Ogunbanjo, "Reasons for and perceptions of patients with minor ailments bypassing local primary health care facilities," *South Afr. Family Pract.*, vol. 57, no. 6, pp. 333–336, Nov. 2015.

[2] Y. Kadooka, A. Asai, A. Enzo, and T. Okita, "Misuse of emergent healthcare in contemporary Japan," *BMC Emergency Med.*, vol. 17, no. 1, p. 23, Dec. 2017.

[3] N. Vidyarthi and O. Kuzgunkaya, "The impact of directed choice on the design of preventive healthcare facility network under congestion," *Health Care Manage. Sci.*, vol. 18, no. 4, pp. 459–474, Dec. 2015.

[4] Y. Zhang, O. Berman, and V. Verter, "The impact of client choice on preventive healthcare facility network design," *OR Spectr.*, vol. 34, no. 2, pp. 349–370, Apr. 2012.

[5] M. S. Daskin, "What you should know about location modeling," *Nav. Res. Logistics*, vol. 55, no. 4, pp. 283–294, 2008.

[6] M. S. Daskin and L. K. Dean, "Location of health care facilities," in *Operations Research and Health Care*. Cham, Switzerland: Springer, 2005, pp. 43–76.

[7] B. P. Berg, "Location models in healthcare," in *Handbook of Healthcare Operations Management*. Cham, Switzerland: Springer, 2013, pp. 387–402.

[8] A. Ahmadi-Javid, P. Seyedi, and S. S. Syam, "A survey of healthcare facility location," *Comput. Oper. Res.*, vol. 79, pp. 223–263, Mar. 2017.

[9] P. M. Griffin, C. R. Scherrer, and J. L. Swann, "Optimization of community health center locations and service offerings with statistical need estimation," *IIE Trans.*, vol. 40, no. 9, pp. 880–892, Jul. 2008.

[10] S. S. R. Shariff, N. H. Moin, and M. Omar, "Location allocation modeling for healthcare facility planning in Malaysia," *Comput. Ind. Eng.*, vol. 62, no. 4, pp. 1000–1010, May 2012.

[11] D.-G. Kim and Y.-D. Kim, "A Lagrangian heuristic algorithm for a public healthcare facility location problem," *Ann. Oper. Res.*, vol. 206, no. 1, pp. 221–240, Jul. 2013.

[12] C. Stummer, K. Doerner, A. Focke, and K. Heidenberger, "Determining location and size of medical departments in a hospital network: A multiobjective decision support approach," *Health Care Manage. Sci.*, vol. 7, no. 1, pp. 63–71, Feb. 2004.

[13] P. Mitropoulos, I. Mitropoulos, I. Giannikos, and A. Sissouras, "A biobjective model for the locational planning of hospitals and health centers," *Health Care Manage. Sci.*, vol. 9, no. 2, pp. 171–179, May 2006.

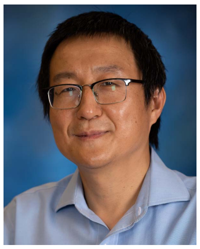
[14] S. Beheshtifar and A. Alimoahmmadi, "A multiobjective optimization approach for location-allocation of clinics," *Int. Trans. Oper. Res.*, vol. 22, no. 2, pp. 313–328, Mar. 2015.
[15] A. M. Mestre, M. D. Oliveira, and A. P. Barbosa-Póvoa, "Location–allocation approaches for hospital network planning under uncertainty," *Eur. J. Oper. Res.*, vol. 240, no. 3, pp. 791–806, Feb. 2015.
[16] R. Rahmaniani, G. Rahmaniani, and A. Jabbarzadeh, "Variable neighborhood search based evolutionary algorithm and several approximations for balanced location–allocation design problem," *Int. J. Adv. Manuf. Technol.*, vol. 72, nos. 1–4, pp. 145–159, Apr. 2014.
[17] V. Hajipour, P. Fattahi, M. Tavana, and D. Di Caprio, "Multi-objective multi-layer congested facility location-allocation problem optimization with Pareto-based meta-heuristics," *Appl. Math. Model.*, vol. 40, nos. 7–8, pp. 4948–4969, Apr. 2016.
[18] V. Verter and S. D. Lapierre, "Location of preventive health care facilities," *Ann. Oper. Res.*, vol. 110, nos. 1–4, pp. 123–132, 2002.
[19] K. Dogan, M. Karatas, and E. Yakici, "A model for locating preventive health care facilities," *Central Eur. J. Oper. Res.*, vol. 28, no. 3, pp. 1091–1121, Sep. 2020.
[20] Y. Zhang, O. Berman, and V. Verter, "Incorporating congestion in preventive healthcare facility network design," *Eur. J. Oper. Res.*, vol. 198, no. 3, pp. 922–935, Nov. 2009.
[21] Y. Zhang, O. Berman, P. Marcotte, and V. Verter, "A bilevel model for preventive healthcare facility network design with congestion," *IIE Trans.*, vol. 42, no. 12, pp. 865–880, Sep. 2010.
[22] D. L. Huff, "Defining and estimating a trading area," *J. Marketing*, vol. 28, no. 3, pp. 34–38, Jul. 1964.
[23] R. Aboolian, O. Berman, and D. Krass, "Competitive facility location and design problem," *Eur. J. Oper. Res.*, vol. 182, no. 1, pp. 40–62, Oct. 2007.
[24] Y. Zhang and D. Atkins, "Medical facility network design: User-choice and system-optimal models," *Eur. J. Oper. Res.*, vol. 273, no. 1, pp. 305–319, Feb. 2019.
[25] K. Hwang, T. B. Asif, and T. Lee, "Choice-driven location-allocation model for healthcare facility location problem," *Flexible Services Manuf. J.*, pp. 1–26, 2021, doi: 10.1007/s10696-021-09441-8.
[26] V. Marianov, M. Ríos, and M. J. Icaza, "Facility location for market capture when users rank facilities by shorter travel and waiting times," *Eur. J. Oper. Res.*, vol. 191, no. 1, pp. 32–44, Nov. 2008.
[27] K. Haase and S. Müller, "A comparison of linear reformulations for multinomial logit choice probabilities in facility location models," *Eur. J. Oper. Res.*, vol. 232, no. 3, pp. 689–691, Feb. 2014.
[28] R. Krohn, S. Müller, and K. Haase, "Preventive healthcare facility location planning with quality-conscious clients," *OR Spectr.*, vol. 43, no. 1, pp. 59–87, Mar. 2021.
[29] B. Kucukyazici, Y. Zhang, A. Ardestani-Jaafari, and L. Song, "Incorporating patient preferences in the design and operation of cancer screening facility networks," *Eur. J. Oper. Res.*, vol. 287, no. 2, pp. 616–632, 2020.
[30] J. M. Bronstein and M. A. Morrisey, "Bypassing rural hospitals for obstetrics care," *J. Health Politics, Policy Law*, vol. 16, no. 1, pp. 87–118, Feb. 1991.
[31] A. Coulter, N. L. Maistre, and L. Henderson, *Evaluation of London Patient Choice: Patients' Experience of Choosing Where to Undergo Surgical Treatment*. Oxford, U.K.: Picker Institute, Jan. 2005.
[32] B. J. Borah, "A mixed logit model of health care provider choice: Analysis of NSS data for rural India," *Health Econ.*, vol. 15, no. 9, pp. 915–932, 2006.
[33] W.-T.-C. Tai, F. W. Porell, and E. K. Adams, "Hospital choice of rural medicare beneficiaries: Patient, hospital attributes, and the patient-physician relationship," *Health Services Res.*, vol. 39, pp. 1903–1922, Dec. 2004.
[34] A. Victoor, D. M. Delnoij, R. D. Friele, and J. J. Rademakers, "Determinants of patient choice of healthcare providers: A scoping review," *BMC Health Services Res.*, vol. 12, no. 1, p. 272, Dec. 2012.
[35] M. Varkevisser and S. A. van der Geest, "Why do patients bypass the nearest hospital? An empirical analysis for orthopaedic care and neurosurgery in The Netherlands," *Eur. J. Health Econ.*, vol. 8, no. 3, pp. 287–295, Sep. 2007.
[36] Q. Wang, R. Batta, and C. M. Rump, "Algorithms for a facility location problem with stochastic customer demand and immobile servers," *Ann. Oper. Res.*, vol. 111, nos. 1–4, pp. 17–34, 2002.
[37] O. Berman and Z. Drezner, "Location of congested capacitated facilities with distance-sensitive demand," *IIE Trans.*, vol. 38, no. 3, pp. 213–221, Mar. 2006.
[38] V. Marianov and D. Serra, "Probabilistic, maximal covering location—Allocation models for congested systems," *J. Regional Sci.*, vol. 38, no. 3, pp. 401–424, Aug. 1998.
[39] V. Marianov and D. Serra, "Location-allocation of multiple-server service centers with constrained queues or waiting times," *Ann. Oper. Res.*, vol. 111, nos. 1–4, pp. 35–50, 2002.
[40] N. Ramshe and A. Ahmadi-Javid, "A portfolio optimization approach to design of congested multi-service networks with application in community healthcare: Fractional 0–1 programming," *Int. Trans. Oper. Res.*, vol. 28, no. 5, pp. 2464–2495, Sep. 2021.
[41] M. A. McGuirk and F. W. Porell, "Spatial patterns of hospital utilization: The impact of distance and time," *Inquiry*, vol. 21, no. 1, pp. 84–95, Apr. 1984.
[42] B. Farhan and A. T. Murray, "Distance decay and coverage in facility location planning," *Ann. Regional Sci.*, vol. 40, no. 2, pp. 279–295, Jun. 2006.
[43] A. Victoor, D. M. Delnoij, R. D. Friele, and J. J. Rademakers, "Determinants of patient choice of healthcare providers: A scoping review," *BMC Health Services Res.*, vol. 12, no. 1, p. 272, Dec. 2012.
[44] P. F. Schnatz, J. L. Murphy, D. M. O'Sullivan, and J. I. Sorosky, "Patient choice: Comparing criteria for selecting an obstetrician-gynecologist based on image, gender, and professional attributes," *Amer. J. Obstetrics Gynecol.*, vol. 197, no. 5, p. 548, 2007.
[45] M. Wei, "The anaysis of the influence factors on selecting hospital," M.S. thesis, School Health Manage., Anhui Medical Univ., Hefei, China, 2014.
[46] Y. Wei, Y. Yao, and X. He, "Analysis on influencing factors of outpatients' behaviors in large general hospitals," *Practical J. Clin. Med.*, vol. 15, no. 4, pp. 240–243, 2018.
[47] X. Kang, "Anaysis on the influencing factors of hospital selecting behavior in Wuhan," M.S. thesis, School Econ. Bus. Admin., Central China Normal Univ., Wuhan, China, 2016.
[48] L. Liu and V. G. Kulkarni, "Balking and reneging in M/G/s systems exact analysis and approximations," *Probab. Eng. Informational Sci.*, vol. 22, no. 3, pp. 355–371, Jul. 2008.
[49] P. Guo, C. S. Tang, Y. Wang, and M. Zhao, "The impact of reimbursement policy on social welfare, revisit rate and waiting time in a public healthcare system: Fee-for-service vs. bundled payment," *Manuf. Service Oper. Manage.*, vol. 21, no. 1, pp. 154–170, 2019.
[50] S. C. Horng, F. Y. Yang, and S. S. Lin, "Applying PSO and OCBA to minimize the overkills and re-probes in wafer probe testing," *IEEE Trans. Semicond. Manuf.*, vol. 25, no. 3, pp. 531–540, Aug. 2012.
[51] Y. Qiu, J. Song, and Z. Liu, "A simulation optimisation on the hierarchical health care delivery system patient flow based on multi-fidelity models," *Int. J. Prod. Res.*, vol. 54, no. 21, pp. 6478–6493, Jul. 2016.
[52] H. Qiu, "Health serivces need, demand and utilization of residents in Xuhui district in Shanghai," M.S. thesis, School Public Health, Fudan Univ., Shanghai, China, 2012.
[53] G. Wu, F. Lu, Y. Bao, F. Yan, and E. A. Yue Li, "Condition analysis of building up station cost, houses and equipment in community health service station," *Chin. Primary Health Care*, vol. 13, no. 6, pp. 14–16, 1999.
[54] J. Bing, "How much does it cost to build a community hospital?" *Window Northeast*, no. 17, pp. 16–17, 2008.
[55] M. E. Ben-Akiva, S. R. Lerman, and S. R. Lerman, *Discrete Choice Analysis: Theory and Application to Travel Demand*, vol. 9. Cambridge, MA, USA: MIT Press, 1985.

Yewen Deng received the B.S. degree in industrial engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 2014. She is currently pursuing the Ph.D. degree with the Department of Industrial Engineering and Management, Shanghai Jiao Tong University, Shanghai, China.

Her current research interests include operation management, performance evaluation, and optimization in healthcare delivery systems.

Na Li received the B.S. degree in mechanical engineering and automation from Xi'an Jiaotong University, Xi'an, China, in 2003, and the Ph.D. degree in industrial engineering from Tsinghua University, Beijing, China, in 2008.

She is currently an Associate Professor with the Department of Industrial Engineering and Management, Shanghai Jiao Tong University, Shanghai, China. Her research interests include stochastic modeling in the application areas of production and the healthcare delivery systems.



Nan Kong received the B.S. degree in automation engineering from Tsinghua University, Beijing, China, in 1999, the M.S. degree in operations research and industrial engineering from Cornell University, Ithaca, NY, USA, in 2000, and the Ph.D. degree in industrial engineering from the University of Pittsburgh, Pittsburgh, PA, USA, in 2006.

His current research interests include applying operations systems engineering principles and techniques to improving healthcare delivery.

Zhibin Jiang received the Ph.D. degree in manufacturing engineering and engineering management from the City University of Hong Kong, Hong Kong, China, in 1999.

He is currently the Changjiang Scholar Chair Professor of the Ministry of Education (MOE), China, and a Distinguished Professor with the Antai College of Economics and Management and the Dean of the Sino-US Global Logistics Institute, Shanghai Jiao Tong University, Shanghai, China. He has published four books and authored/coauthored 250 publications. His research interests include discrete-event modeling and simulation and operation management in manufacturing and health care systems.

Xiaoqing Xie received the B.S. degree in mathematics and applied mathematics from Hunan University, Hunan, China, the M.S. degree in mathematics from the University of Western Ontario, Canada, and the Ph.D. degree in operations management from Cornell University, USA.

Her current research primarily focuses on service operations, revenue management, dynamic pricing, and operations/marketing interface.