



No. 036/2020/POM

Two-tier hospital system redesign optimization based on consumer
choice and queueing network modeling

Yewen Deng

Department of Industrial Engineering & Management
Shanghai Jiao Tong University

Na Li *

Department of Industrial Engineering & Management
Shanghai Jiao Tong University

Nan Kong

Weldon School of Biomedical Engineering
Purdue University

Zhibin Jiang

Department of Industrial Engineering & Management
Shanghai Jiao Tong University

Xiaoqing Xie

Department of Economics and Decision Sciences
China Europe International Business School (CEIBS)

August 2020

* Corresponding author: Na Li (na-li03@sjtu.edu.cn). Address: Na Li, PhD Associate Professor Industrial Engineering and Management Shanghai JiaoTong University 800 E. Dongchuan Road Minhang. Shanghai, China +86-15000630829

Two-tier hospital system redesign optimization based on consumer choice and queueing network modeling

Yewen Deng^a, Na Li ^a, Nan Kong^b, Zhibin Jiang^c, Xiaoqing Xie^d

^aDepartment of Industrial Engineering & Management, Shanghai Jiao Tong University, China; ^bWeldon School of Biomedical Engineering, Purdue University, United States; ^cAntai College of Economics and Management, Shanghai Jiao Tong University, China; ^dDepartment of Economics and Decision Sciences, China Europe International Business School (CEIBS), China

ARTICLE HISTORY

Compiled February 2, 2020

ABSTRACT

The hospital system in many countries experiences capacity shortage, which results in poor access to healthcare in certain parts of the country. Moreover, rapid urbanization, which creates many new population centers, has exacerbated the imbalance between care need and service capacity. One way to address the challenge is to conduct capacity expansion and spatial redistribution. In this research, we studied the problem of optimal decision making for locating new hospitals in a two-tier hospital system comprising both central and district hospitals, and upgrading existing district hospitals to central hospitals, with incorporation of patient preferences on seeking care. We first formulated the problem with a discrete location optimization model to minimize the total cost (i.e., a weighted sum of travel cost, waiting cost, and government spending). Then we constructed a multinomial logit model with real-world data to characterize hospital choice behaviors, and quantify patient arrival rates at each hospital accordingly. We also developed a multi-hospital queueing network model to analyze the impact of hospital locations on patient flows. By solving the resultant nonlinear combinatorial optimization problem via a genetic algorithm, we verified the effectiveness of hospital location reconfiguration and confirmed the influence of individual-specific attributes (e.g., insurance type and balking tendency).

KEYWORDS

Healthcare service system; Network redesign; Location; Patient choice; Queueing

1. Introduction

There is a significant discrepancy between patient need and provider capacity in many major metropolises worldwide such as Shanghai. 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 273.423 million. On the other hand, the number of licensed physicians and assistant physicians in Shanghai is only 2.31 per thousand people, ranking behind most municipalities and provincial capitals in China (available from www.spccsc.sh.cn/). Healthcare resource scarcity has become a major social issue

in Shanghai. Furthermore, with rapid development of industrialization and urbanization, major metropolises in China have expanded significantly and urban populations have dispersed with new residential communities sprouting in inter-city and suburban areas. This contributes to the imbalance of healthcare resource distribution among sections of the metropolises. In these places, hospitals are mainly located in central cities whereas inter-city and suburban areas lack healthcare resources.

Like in many other countries, the healthcare system in China consists of central hospitals (CHs) and district hospitals (DHs). CHs have a high reputation with majority of the most qualified specialists, and are mainly concentrated in the central urban area. On the other hand, DHs provide basic medical services, and they are located at various communities. For outpatient care, Chinese patients typically choose hospitals based on their preferences, and CHs are often more attractive to them due to high reputation. However, for its limited capacities, CHs are always overcrowded, resulting in long waiting times. In addition, since CHs are mostly located in central city, patients from inter-city or suburban areas, which can be far from the city center, have to travel long distances. Overall resource scarcity and regional capacity maldistribution have motivated a series of reforms in recent years to address the challenge in care access, one of the fundamental challenges to the Chinese healthcare system. Besides outpatient services, effectiveness of preventive care is also greatly influenced by poor access (Vidyarthi and Kuzgunkaya 2015; Zhang, Berman, and Verter 2012).

To combat the above challenge, one viable way is to finance necessary care capacity expansion especially in new population centers. This can be achieved by simply opening more hospitals or by converting DHs to CHs (i.e., upgrading). In this paper, we develop a framework for mathematically analyzing the decision issue related to reconfiguring a network of CHs and DHs. We thus consider the following government investment decisions: (1) locate more CHs and DHs to increase care capacity; and (2) upgrade DHs to CHs for the same purpose. With increased and better distributed care capacity in the hospital network, we expect an overall improvement in care access (e.g., reduced waiting time and travel distance). Our objective is to redesign the hospital network such that the resultant access improvement and government spending can reach some desired compromise, e.g., their sum can be minimized. Furthermore, in the hospital system, patients seeking outpatient services have choices to different points of access to care as they like, thus demand from the same location will split between medical facilities. Therefore, patient choice behavior in the system with two types of hospitals is a critical part of our modeling. We consider hospital choice behavior modeling of two classes of patients identified by their insurance type.

The following questions were answered in our research. One, what is the effect of hospital types and distance to hospitals on hospital choice behavior of either class of patients? Two, what is the influence of hospital choice on the performances of a hospital system consisting of CHs and DHs? Three, what is the optimal design of the hospital system? To answer these questions, we first modeled the hospital choice behavior under the influence from locations of CHs and DHs. We conducted a survey which includes a choice experiment on hospital visit for some outpatient service by assessing respondents' reactions to hospital types and distances to hospitals, and then developed a multinomial logit model (MNL). To answer the second question, we built a queuing network considering patient balking behavior, and then applied queue theory to derive closed-form expressions of operational performance measures (i.e., balking probability, utilization rate, and average waiting time) with respect to the location of CHs and DHs and upgrade decision of DHs. Finally, we formulated a nonlinear optimization model to select appropriate locations for new CHs and DHs and upgrade

some of the existing DHs, so as to achieve optimality in minimization of the weighted sum of patient travel, waiting, and government spending on the system reconfiguration.

This paper makes two main contributions: (1) We embedded a discrete choice model into an optimal location problem with different streams of service requests and different types of service facilities; and (2) We obtained real-world experimental data from Shanghai, the target catchment area, for the quantification of two known influential factors, hospital type and distance to hospital, on patient choice of hospital visitation for outpatient services. To the best of our knowledge, both contributions above are rare practice in the OR/MS literature. In addition, we considered not only building new hospitals but also upgrading existing DHs, which coincides with the expectation set by the Shanghai municipal government.

The remainder of this paper is organized as follows. In Section 2, we review the relevant literature. In Section 3, we present an optimization model for redesign of the hospital network. To support this, we present a patient hospital choice model and a multi-hospital queueing network model in this section. In Section 4, we report our optimal hospital network design of real-world cases based on Shanghai. Finally we draw conclusions and outline future research in Section 5.

2. Literature review

There have been numerous location analysis studies in healthcare planning. According to the classification of Daskin (2008), the discrete location models appear in the healthcare systems literature can be 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 the basis of models used in healthcare applications. For review of these classical problems, we refer readers to Daskin and Dean (2005), Berg (2013) and Ahmadi-Javid, Seyedi, and Syam (2017). For example, many researchers addressed the objective of maximizing the total demand covered by a given number of facilities (e.g. Griffin, Scherrer, and Swann 2008; Shariff, Moin, and Omar 2012; Kim and Kim 2013), or minimizing the distance or travel time when demand locations are assigned to facilities within a certain distance (e.g. Stummer et al. 2004; Mitropoulos et al. 2006; Beheshtifar and Alimoahmadi 2015; Mestre, Oliveira, and Barbosa-Póvoa 2015).

A critical aspect of medical facility location analysis is modeling patient choice and its impact on the 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., patients are assigned to a medical facility by the decision maker, rather than users choosing the facility to access. These models determine the optimal location of facilities and assign patients to selected facilities. For example, in Shariff, Moin, and Omar (2012), patient demand is assigned to a facility within some allowable distance, and is assumed to be allocated to one facility at most. Other studies that take the allocation of demand as a decision variable include Rahmaniani, Rahmaniani, and Jabbarzadeh (2014); Beheshtifar and Alimoahmadi (2015); Vidyarthi and Kuzgunkaya (2015); Hajipour et al. (2016).

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 simplified, namely each patient is assumed to visit the most *attractive* facility. This requires patients to be rational and fully informed, and

they will visit the most attractive facility they think at all times. For example, Kim and Kim (2013) studied the problem of locating public healthcare facilities to maximize the number of served patients of two types (low- or high-income). The authors only considered the preferences of high-income patients and assumed that they are only allocated to 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 service spending. 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 to have identical preference. Some other studies assume that patients visit the facility closest to them (e.g. Verter and Lapierre 2002), and some studies assume patients visit the facility with the minimal expected total time, including the travel time plus the expected time spent at the facility (e.g. Zhang, Berman, and Verter 2009; Zhang et al. 2010).

For *probabilistic-choice* models, a patient is assumed to visit each facility with a certain probability, which is based on the *attractiveness* of the facility. Huff (1964) presented the first probabilistic-choice model, which is for a spatial interaction analysis. In the model, client utility is represented by a gravity formula to estimate market shares of the facilities. Later, extensions of Huff (1964) appeared in location analysis such as the multiplicative competitive interaction model by Aboolian, Berman, and Krass (2007). In addition, there are other ways of incorporating probabilistic-choice model. Related to our work, discrete choice models, based on random utility theory common in marketing and econometrics, have been incorporated into location models. For example, to maximize the total participation of preventive healthcare (PH), Zhang, Berman, and Verter (2012) 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 estimate the probability that a patient chooses each facility. In that work, the parameter representing the sensitivity to the attractiveness determinant is given directly in the case study. Zhang and Atkins (2019) 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 patients' choice. The authors estimated the coefficients in the MNL model using actual patient flow data. Other location analysis papers that incorporate patient choice with a discrete choice model include Marianov, Ríos, and Icaza (2008); Haase and Müller (2014); Zhang and Atkins (2019).

Most of the above 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 the influencing factor. In this paper, we did not use waiting time as an influencing factor, because in reality it is difficult for patients to have information about the waiting time of each hospital in advance. Instead, in our choice model, we considered the preference of patients for CHs versus DHs, and the effect of distance to these two types of hospitals on patients, which is less considered by healthcare facility location studies in the literature. In addition, in most studies that considered patient behavior or preference, there are no differences on patient types, whereas we differentiated patients by their insurance types and analyzed their respective hospital choice behaviors. Furthermore, in contrast to the above studies using MNL models, which directly predetermined or used patient flow data in hospitals to fit the values of parameters related to patient choice preference, we fitted an MNL model against first-hand behavior data collected from a survey of 358 Shanghai residents. Then, we examined whether distance and type of the hospital were influencing determinants and estimated the influence of significant factors. We believe this has made our work more realistic and comprehensive.

Another aspect of research related to our work concerns incorporating congestion in facility design models, in which the most common way 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, to incorporate service congestion, several studies considered a constraint on wait time. Wang, Batta, and Rump (2002); Berman and Drezner (2006); Zhang, Berman, and Verter (2012) used M/M/1 (or M/M/C) queues to model facilities and considered the upper bound on the average waiting time of clients (or equivalent queue length) in a constraint. Marianov and Serra (1998, 2002) modeled service facilities as M/M/1 (or M/M/C) queueing systems and introduced a constraint to ensure the probability that a client enters a queue at a facility with at most b waiting clients is at least α in a maximal covering location-allocation model. Alternatively, there are studies that establish a decay function between demand and waiting time in constraints. Zhang, Berman, and Verter (2009) captured the level of congestion at each facility with an M/M/1 queue in a preventive health facility location problem. They assumed the fraction of clients from each population node to each facility is a decreasing function of the expected total (travel, waiting and service) time and assumed clients choose the facility with a minimum 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. Zhang and Atkins (2019) modeled each medical facility as an M/M/c queue. The authors assumed that the mean system waiting time and travel time are main determinants for client choice, and used an MNL model to establish the relationship between the equilibrium flow from each population node to each facility. Vidyarthi and Kuzgunkaya (2015) used spatially distributed M/G/1 queues to model a preventive health facility network, and captured congestion due to waiting and service delays. Then to minimize the weighted sum of total travel time and waiting and service delays, the authors presented a model to determine the location of the facilities, the service capacity of each facility, and the allocation of clients to each facility. Different from the above studies that assume every patient arriving at a facility will join the queue, we modeled each hospital as an M/M/1 queue and considered balking. That is, a patient arriving to a hospital will decide whether or not to join the queue according to his/her estimated waiting time.

3. Model formulation

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 combined time spent on transport and waiting to access care can be used as a proxy to measuring care accessibility of hospitals (Zhang, Berman, and Verter 2009; Vidyarthi and Kuzgunkaya 2015). The objective is to tradeoff between care accessibility and government spending for the capacity expansion.

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 vector of decision variables for locating new CHs and DHs is represented by $\mathbf{x}^H = (x_1^H, \dots, x_j^H, \dots, x_{|J|}^H)$ and $\mathbf{x}^L = (x_1^L, \dots, x_j^L, \dots, x_{|J|}^L)$, respectively. The decision vector on whether to upgrade a DH is represented by $\mathbf{y} = (y_1, \dots, y_j, \dots, y_{|J_L|})$. Based on the questionnaire used for the choice model, the cohort of respondents was divided into two classes distinguished by insurance types. Without loss of generality, denote R to be the set of patient classes.

Given decision vector \mathbf{x}^H , \mathbf{x}^L and \mathbf{y} , one can calculate the rectilinear distance between residential site i and hospital site j , denoted by d_{ij} , 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, and then 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 at site i into the demand estimate for outpatient care, which is denoted by d_i^r , and combining the demand estimate with patient choice preferences $p_{ij}^r(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y})$, one can obtain the arrival rate of class r patients from site i to hospital at site j and the total arrival rate at a hospital at site j , which are denoted by $\lambda_{ij}^r(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y})$. Thus, $\lambda_{ij}^r(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y}) = d_i^r p_{ij}^r(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y})$. One can further denote $\lambda_j(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y})$ to be the combined arrival rate of the hospital at site j over all residential sites. Thus, $\lambda_j(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y}) = \sum_{r \in R} \sum_{i \in I} d_i^r p_{ij}^r(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y})$.

Given $\lambda_j(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y})$, one can further 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 the model aiming to selecting optimal sites for new CHs or DHs and selecting existing DHs to upgrade so as to minimize an objective, which comprises three components: (i) the cost of weighted mean wait times for service at hospital site j , i.e., $W_j^H(\lambda_j(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y}))$ or $W_j^L(\lambda_j(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y}))$; (ii) the cost of traveling to hospital site j for care from residential site i , i.e., $d_{ij}(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y})$; and (iii) the spending of the government for expanding the capacity of the hospital network.

The nonlinear program for the location optimization problem is presented as in (1) – (5). In the optimization 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 present $\lambda_{ij}(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y})$, $\lambda_j(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y})$, $d_{ij}(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y})$, $W_j^H(\lambda_j(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y}))$, $W_j^L(\lambda_j(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y}))$, $p_j^{B_H}(\lambda_j(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y}))$ and

$p_j^{B_L}(\lambda_j(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y}))$, respectively.

$$\begin{aligned}
\min_{\mathbf{x}^H, \mathbf{x}^L, \mathbf{y}} \quad & \alpha_d \left[C_{d_H} \left(\sum_{i \in I} \sum_{j \in J} x_j^H \cdot \lambda_{ij} \cdot d_{ij} + \sum_{i \in I} \sum_{j \in J_L} y_j \cdot \lambda_{ij} \cdot d_{ij} + \sum_{i \in I} \sum_{j' \in J_H} \lambda_{ij'} \cdot d_{ij'} \right) + \right. \\
& C_{d_L} \left(\sum_{i \in I} \sum_{j \in J} x_j^L \cdot \lambda_{ij} \cdot d_{ij} + \sum_{i \in I} \sum_{j \in J_L} (1 - y_j) \cdot \lambda_{ij} \cdot d_{ij} \right) \left. \right] + \\
& \alpha_W \left[C_{W_H} \left(\sum_{j \in J} x_j^H \cdot \lambda_j \cdot W_j^H + \sum_{j \in J_L} y_j \cdot \lambda_j \cdot W_j^H + \sum_{j \in J_H} \lambda_j \cdot W_j^H \right) + \right. \\
& C_{W_L} \left(\sum_{j \in J} x_j^L \cdot \lambda_j \cdot W_j^L + \sum_{j \in J_L} (1 - y_j) \cdot \lambda_j \cdot W_j^L \right) \left. \right] + \\
& \alpha_I \left[\sum_{j \in J} \left(\kappa_x^H \cdot x_j^H + \kappa_x^L \cdot x_j^L \right) + \sum_{j \in J_L} \kappa_y^H \cdot y_j \right]
\end{aligned} \tag{1}$$

$$\text{s.t.} \quad x_j^H + x_j^L \leq 1, \forall j \tag{2}$$

$$p_j^{B_H} \leq \theta_H, \forall j \tag{3}$$

$$p_j^{B_L} \leq \theta_L, \forall j \tag{4}$$

$$x_j^H, x_j^L, y_j \in \{0, 1\}, \forall j \tag{5}$$

Objective (1) comprises three components: (i) the cost of traveling; (ii) the cost of waiting; and (iii) the government spending. In (1), C_{d_H} and C_{d_L} is an unit cost of traveling to CH and DH; respectively; C_{W_H} and C_{W_L} is the unit-time cost of waiting at CH and DH, respectively; κ_x^H , κ_x^L is the costs of land acquisition and facility building for establishing a CH and a DH, respectively, κ_y is the cost of upgrading a DH; weights α_d , α_W and α_I are assigned to the cost of traveling, cost of waiting, and government spending, respectively. Constraints (2) state that a DH cannot be built when a CH has been built at some site, and vice versa. Constraints (3) and (4) are constraints that ensure the balking probability of patients at CH and DH is capped by some level, denoted by θ_H and θ_L , respectively. Constraints (5) ensure non-negativity and binary restrictions.

The queueing model and the choice models embedded in the above location optimization model are highly nonlinear, which makes the optimization model difficult to solve exactly. Thus, we elect to use a Genetic Algorithm to solving the above program; see Appendix B for detailed information.

3.1. The probabilistic-choice model of patient hospital visit behavior

As described in the literature review, it is common to consider distance as an influential factor, such as Mitropoulos et al. (2006); Zhang, Berman, and Verter (2009). Studies have shown that the reduction in attractiveness with distance, which is called distance decay, is a key determinant of the use of any health facility ((McGuirk and Porell 1984; Farhan and Murray 2006)). Similarly, Mitropoulos et al. (2006) compared the

influence of the distance decay on general hospitals and local health centers. Victoor et al. (2012); Schnatz et al. (2007) examined the influence of provider expertise (or them being knowledgeable, experienced, and capable) is an important determinant of patient's provider choice. Wei (2014); Wei, Yao, and He (2018); Kang (2016) examined the influence of specialists on patient's hospital choice.

To investigate how patient choice is affected by the distance and the type of hospitals (i.e., CH or DH), we designed a questionnaire, recruited and surveyed a cohort of online respondents residing in Shanghai (see Appendix A for the web-based questionnaire). With the questionnaire, we acquired each respondent's insurance type first, and then asked him/her to make a choice among a choice set of hospital alternatives. In this choice set for each respondent, each hospital alternative 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 (i.e., CH or DH) at location j . We consider attribute d_{ij} to be a continuous variable and q_j to be an indicator variable for two hospital types, a 1 if the hospital at site j is a CH, 0 otherwise. After completing the survey, we used an MNL model to test 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 with a distance d_{ij} and a type identifier q_j , is given by $U_{ij}^n = V_{ij}^n + \varepsilon_{ij}^n$, $i \in I$, $j \in J \cup J_H \cup J_L$. In this expression, V_{ij}^n is the deterministic component and ε_{ij}^n is a random error component following a Gumbel distribution. Further for a hospital at site j with a distance d_{ij} and a hospital type q_j , we have $V_{ij}^n = \beta_d d_{ij}^n + \beta_q q_j^n$, where β_d and β_q are parameters that capture the preference of a patient at site i on distance and hospital type. These two parameters need to be estimated.

The probability that a patient n at site i chooses a hospital at site j with distance d_{ij}^n and hospital type q_j^n is given by:

$$p_{ij}^n = \frac{\exp(\beta_d d_{ij}^n + \beta_q q_j^n)}{\sum_{k \in J \cup J_H \cup J_L} \exp(\beta_d d_{ik}^n + \beta_q q_k^n)}, \quad i \in I, j \in J \cup J_H \cup J_L \quad (6)$$

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

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

With the above definition of the choice probability in equation (7), we can derive the arrival rate of class r patients from site i to hospital at site j as

$$\lambda_{ij}^r(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y}) = \begin{cases} d_i^r \frac{(x_j^H + x_j^L) \exp(\beta_d^r d_{ij} + \beta_q^r x_j^H)}{U^r}, & r \in R, i \in I, j \in J \quad (8) \\ d_i^r \frac{\exp(\beta_d^r d_{ij} + \beta_q^r)}{U^r}, & r \in R, i \in I, j \in J_H \quad (9) \\ d_i^r \frac{\exp(\beta_d^r d_{ij} + \beta_q^r y_j)}{U^r}, & r \in R, i \in I, j \in J_L \quad (10) \end{cases}$$

where

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

Equation (8) represents the arrival rate of class r patients from residential site $i \in I$ to a new hospital at site j . Notice that when the newly built hospital at candidate site j is CH (i.e., $x_j^H = 1, x_j^L = 0, j \in J$), its hospital type is set to 1 in equation (8) (i.e., $q_j = x_j^H = 1, j \in J$); otherwise, when the newly built hospital at site $j \in J$ (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 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., $y_j = 1, j \in J_L$), its hospital type is also upgraded to 1 (i.e., $q_j = y_j = 1, j \in J_L$), otherwise, when the existing DH at site $j \in J_L$ is not upgraded to a CH (i.e., $y_j = 0, j \in J_L$), its hospital type is 0 (i.e., $q_j = y_j = 0, j \in J_L$).

Then 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 the above arrival rates, we next use them as input to the multi-hospital system model.

3.2. Performance evaluation for the multi-hospital system

We assume the arrival process from resident site i to hospital site j follows a Poisson distribution with mean arrival rate $\lambda_{ij}(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y})$. Thus the total arrival process at hospital site j also follows a Poisson distribution with mean arrival rate $\lambda_j(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y})$. In reality, once a patient arrives at a hospital, too long a queue discourages him or her from joining the queue (i.e., balking). We thus consider impatience and balking of patients in the system. We first define a virtual queueing time (vqt) as introduced in Liu and Kulkarni (2008). 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 (Guo et al. 2017), or patients may estimate the waiting time based on their previous visits (Zhang and Atkins 2019). Next we introduce the balking rule used in our system performance analysis. We assume when an arriving patient evaluates that his/her vqt is no more than a fixed amount, s/he decides to join the queue. The fixed amount is termed the threshold of tolerable waiting time, and denoted by b_H for CH and b_L for DH, respectively. We thus define the probability that vqt is more than the fixed amount b_H and b_L as the balking probability. In addition, we represent \hat{J}_H as the set where CHs are located (i.e., $\hat{J}_H = J_H \cup \{j \mid x_j^H = 1, j \in J\} \cup \{j \mid y_j = 1, j \in J_L\}$), and represent \hat{J}_L as the set where DHs are located (i.e., $\hat{J}_L = \{j \mid x_j^H = 0, j \in J\} \cup \{j \mid y_j = 0, j \in J_L\}$). We thus represent balking probability at CH $j \in \hat{J}_H$ and DH $j \in \hat{J}_L$ as $p_j^{B_H}, j \in \hat{J}_H$ and $p_j^{B_L}, j \in \hat{J}_L$, respectively.

Next we assume the service duration at each CH and at each DH is exponentially distributed. We denote the mean service rates of each CH $j \in \hat{J}_H$ and each DH $j \in \hat{J}_L$

to be μ_j^H and μ_j^L , respectively. For an CH $j \in \hat{J}_H$, an arriving patient leaves the two-level system with balking probability $p_j^{B_H}$, $j \in \hat{J}_H$, or wait for service in an infinite capacity FCFS (first come, first served) queue with probability $1 - p_j^{B_H}$, $j \in \hat{J}_H$, then leave when the service is completed. For an LWH $j \in \hat{J}_L$, a patient will leave the system with balking probability $p_j^{B_L}$, $j \in \hat{J}_L$ at his or her arrival epoch, or wait for service in a queue with probability $1 - p_j^{B_L}$, $j \in \hat{J}_L$. With these assumptions, we formulate a two-level queueing network with each hospital being an M/M/1 queue with balking.

Given $\lambda_j(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y})$, $j \in J \cup J_H \cup J_L$, which can be obtained from equations (8)-(10), μ_j^H , μ_j^L , b_H and b_L , we can derive closed-form expressions for the relevant queueing performance measures according to Theorems 1 to 4 in Liu and Kulkarni (2008), i.e., CH/DH utilization rate, balking probability, and mean waiting time.

- The HWH utilization rate at a CH $j \in \hat{J}_H$: $\rho_j^H(\lambda_j(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y})) = \frac{\lambda_j(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y})}{\mu_j^H}$;
- The LWH utilization rate at a DH $j \in \hat{J}_L$: $\rho_j^L(\lambda_j(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y})) = \frac{\lambda_j(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y})}{\mu_j^L}$;
- The balking probability at an CH $j \in \hat{J}_H$:

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

where

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

$$\text{and } p_j^H = \frac{\lambda_j(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y})}{\lambda_j(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y}) + \mu_j^H};$$

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

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

where

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

$$\text{and } p_j^L = \frac{\lambda_j(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y})}{\lambda_j(\mathbf{x}^H, \mathbf{x}^L, \mathbf{y}) + \mu_j^L};$$

- The mean waiting time at a CH i :

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

- The mean waiting time at a DH j :

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

4. Case study

In this section, we use the multi-hospital system in Shanghai as the real-world context to present a case study. We used this case study to answer the following three types of 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 if the distribution of patients of the two insurance types changes; and (3) what happens to the design of the hospital system if the threshold of patient’s tolerable waiting time changes. Through this real-world based case study, our study is expected to offer system redesign recommendations to the Shanghai municipal government.

4.1. Background

In Shanghai, there are 16 administrative divisions including 230 sub-districts (available from <http://www.shanghai.gov.cn/>), which can be divided according to central urban area, semi-central semi-suburban area, and suburban area; see Table 7 and Figure 1. Table 7 represents the division of administrative districts in Shanghai and the corresponding proportion of each subpopulation. In our model, the 230 sub-districts in Shanghai represent the spatial distribution of population zones and are also used to represent the candidate sites for new hospitals. The considered system comprises 285 existing hospitals, 39 of which are CHs, 246 are DHs. Please 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 in each of the 230 sub-districts in Shanghai; see Table 8 and Figure 1. Figure 2 shows that most of the CHs are located in central urban areas, where the population is though only 30.35% of the total population of Shanghai.

The two-week consultation rate for outpatient services is reported to be different between the two types of basic medical insurance in Shanghai (Qiu 2012), namely *urban employee basic medical insurance* and *urban and rural resident basic medical insurance*. For the interest of space, we refer to the two types as UE insurance and URR insurance. The two-week consultation rate of residents with UE and URR insurance in one district of Shanghai is 11.73% and 9.89% (Qiu 2012). Further, according to the Shanghai Statistics Year Book (available from www.tjj.sh.gov.cn/), the proportion of residents covered by UE and URR insurance is 74.5% and 25.5%, respectively. By assuming five working days per week and 8 working hours per day, we thus converted the population quantity with the two insurance types into the demand for outpatient care per hour, to calculate the mean arrival rates. Finally, based on real data of patient service time in representative CHs (i.e., Ruijin hospital and Shanghai No. 6 People Hospital) and DHs (i.e., Xujiahui Street Community Health Service Center, Longhua Street Community Health Service Center, and Hongmei Street Community Health Service Center) between 2015–17, we were able to estimate the mean service rates at CHs and DHs, i.e., $\mu_j^H = 600$ patients/h and $\mu_j^L = 20$ patients/h.

4.2. Choice model

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 ran an online survey with the questionnaire in October 2019 on www.wenjuan.com, a Chinese internet survey platform. A total of 362 respondents in Shanghai participated in our

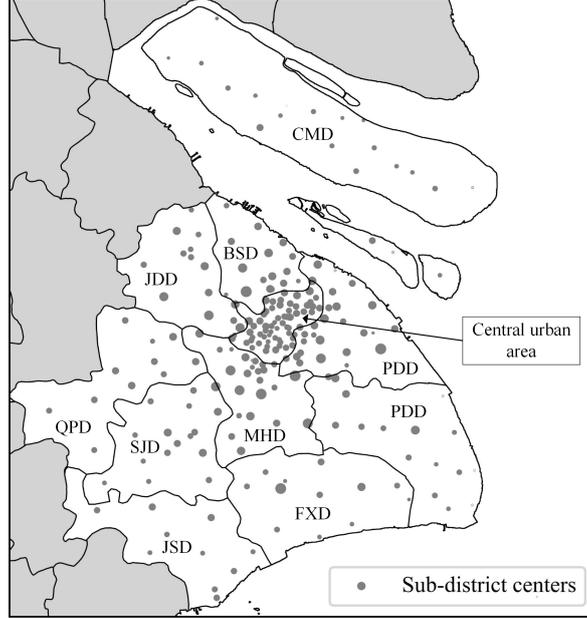


Figure 1.: The spatial distribution of population zones

study, and 358 of which were deemed valid samples. The respondents are anonymous and the data source is reliable.

In the questionnaire, we presented a scenario. We asked each of the respondents to imagine the situation s/he experienced a fever and cough, accompanied by chest tightness, shortness of breath and other symptoms, which is likely to be a respiratory disease. As a result, s/he would visit some outpatient department. We then provided them with a choice card containing two hospitals identified by distance and type. The two pairs of hospital attributes were randomly generated. We set a plausible range of distance on each hospital type from which we drew uniform samples as follows. The lower bound of the distance range on DH is 0 and the upper bound is 3. The lower bound of the distance range on CH is 1 and the upper bound is 30.

With the behavior experiment data, we parameterized a choice model for either insurance type. Table 1 presents our choice model results, i.e., respective estimates of parameters β_d and β_q from equation (6). We verified that for both insurance types, the attribute hospital type is significant and positively correlated to the hospital choice, implying when the distance to hospital being equal, patients prefer CH as opposed to DH. Our results also verified that for both insurance types, the attribute distance to hospital is a factor as significant as hospital type. That is, the results imply when the hospital type being equal, the shorter the distance, the more likely the patient is to visit it. In addition, by comparing the probabilities of choosing CH between the two types of patients (please see Appendix C), we found that when the differences between traveling to the DH and traveling to the CH are the same for the two types of patients, UE patients are more likely to choose to visit the CH than URR patients. In summary, with the choice experiment, we parameterized the two key factors on patient's hospital visit behavior in Shanghai, which paved the way for us to study the multi-hospital system redesign problem of Shanghai.

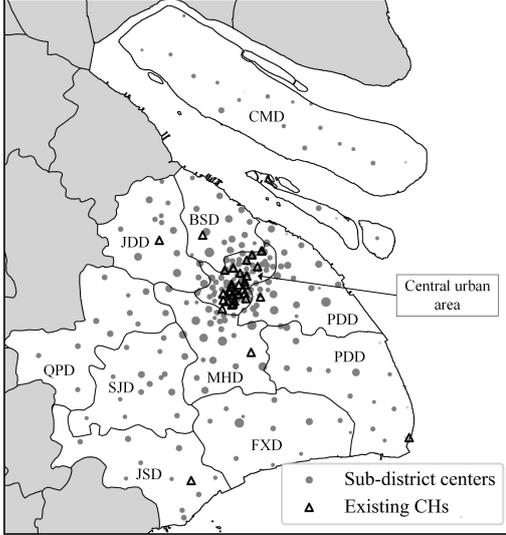


Figure 2.: Current location of CHs

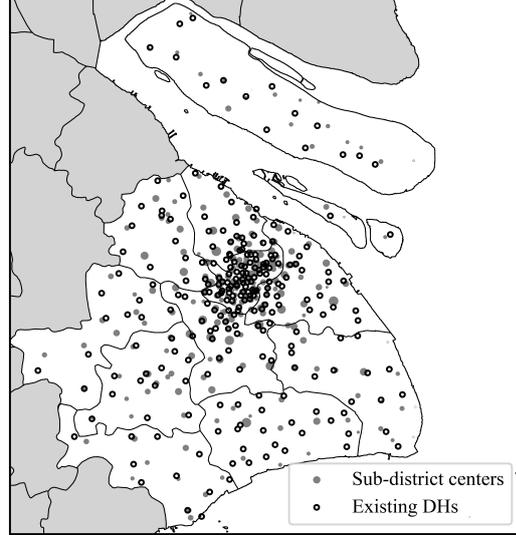


Figure 3.: Current location of DHs

Table 1.: Choice model coefficient estimation results

Insurance Type	Variable Name	Parameter Estimate	Standard Error	p -Value
UE	Type_CH	1.087	0.258	0.000
	Distance	-0.085	0.015	0.000
	Constant	0.148	0.121	0.221
URR	Type_CH	1.027	0.494	0.038
	Distance	-0.093	0.031	0.003
	Constant	-0.150	0.260	0.563

4.3. Results

In this section, we report three Shanghai-based case studies. The three research questions at the beginning of Section 4 have been presented. Our main results are location of new CHs and DHs, and the decision about whether or not to upgrade a DH. These studies involve solving the multi-hospital system redesign optimization model, i.e., Eq. (1)–(5). We set C_{d_H} , C_{d_L} , C_{W_H} , and C_{W_L} , the unit-cost of distance to CH and DH and the unit-time cost of waiting time at CH and DH, to be 2, 1.8, 12, 8, respectively. Then we set α_d , the weighting coefficient of distance, to be 0.4, and set α_W , the weighting coefficient of waiting times to be 0.45, and set α_I , the weighting coefficient of government investment, to be 0.15. The upper bounds of balking probability of patients at CH and DH, i.e., θ_H and θ_L , are set to be 0.2 and 0.3, respectively. κ_x^H , κ_x^L , the cost of land acquisition and establishment of a CH and a DH, is set to be 9000 and 2000, respectively, and the cost of upgrading a DH, i.e., κ_y , is set to be 6000; We estimated the service rates of an CH and an DH to be $\mu_j^H = 600$, $\mu_j^L = 20$, respectively, based on the average number of patients a CH and a DH treats hourly. By conducting field investigation at various hospitals, we specified the threshold on the wait time tolerance at CHs to be $b_H = 2$ hour, and the threshold at DHs to be $b_L = 1$ hour.

Study 1: What is the optimal hospital system design to achieve the best system performance?

Figure 4 displays the optimal location of CHs, including 1 new built and 18 upgraded. CHs in the current system are mainly clustered in the central urban areas (see Figure 2). In contrary, the location of CHs in the optimal design is more dispersed. In particular, we found one new CH would be built at the center of Chongming island (majority of Chongming District), which is far from the central urban area. In addition, most of upgraded hospitals (i.e., DHs upgraded to CHs) would be distributed in suburban areas that previously had no CHs. Taking a closer look at the optimal facility location, we noticed that the suburban area with the most upgraded CHs is Songjiang District, which is located almost at the center of eight Shanghai suburban areas. This result is intuitive that centrally located hospitals would be conveniently accessible to surrounding residents that are still oriented towards the distance to hospital when choosing which hospital to visit.

Figure 5 displays the optimal location of DHs. The optimal location suggests that DHs be upgraded in suburbs relatively close to central city and new DHs be built at more remote suburbs. This result implies that the hospital network redesign is aligned with the expansion and dispersion of residential areas to traditionally remote sites.

To evaluate the performance of the optimal hospital system design, we compared the total costs (i.e., weighted sum of the waiting cost, traveling cost, and government spending). The results show a 41.4% reduction in the total cost under the optimal hospital system design (i.e., 3.43×10^5 under the optimal design vs. 5.86×10^5 under the current system). Further, the results in Table 2 suggest that under the optimal design, the mean waiting times at CHs and DHs are decreased, and the distances to CHs and DHs are decreased as well. These results imply that adding CHs (mostly from upgrading existing DHs) in suburban areas can economically reduce the wait time at CHs, and to some extent, reduce the travel distance.

Table 2.: The weighted¹ average wait time and distance

	Mean waiting time (in hours)		Mean distance (in km)	
	at CH	at DH	to CH	to DH
Current	1.86	0.29	17.0	1.8
Optimal	0.02	0.19	15.4	1.6

¹ The weight of each CH/DH is calculated as the arrival at each CH/DH divided by the total arrival at all CHs/DHs.

Study 2: What is the optimal hospital system design with the different distribution of residents covered by two insurance types?

In this study, we analyzed the would-be impact on the network design if the distribution of patients covered by the two types of insurance changed. This study was inspired by the burden on healthcare capacity management due to rapid urbanization in China. As a result, the proportion of UE patients would increase. In this study, we thus increased its percentage from 74.5% to 95% (see Table 3). We kept other model parameters the same as in Study 1.

Figure 6 displays the optimal location of 19 CHs, including 1 new built and 18 upgraded. Compared with the optimal system design obtained from Study 1, the number of newly built DHs is less when the resident population consists of more UE

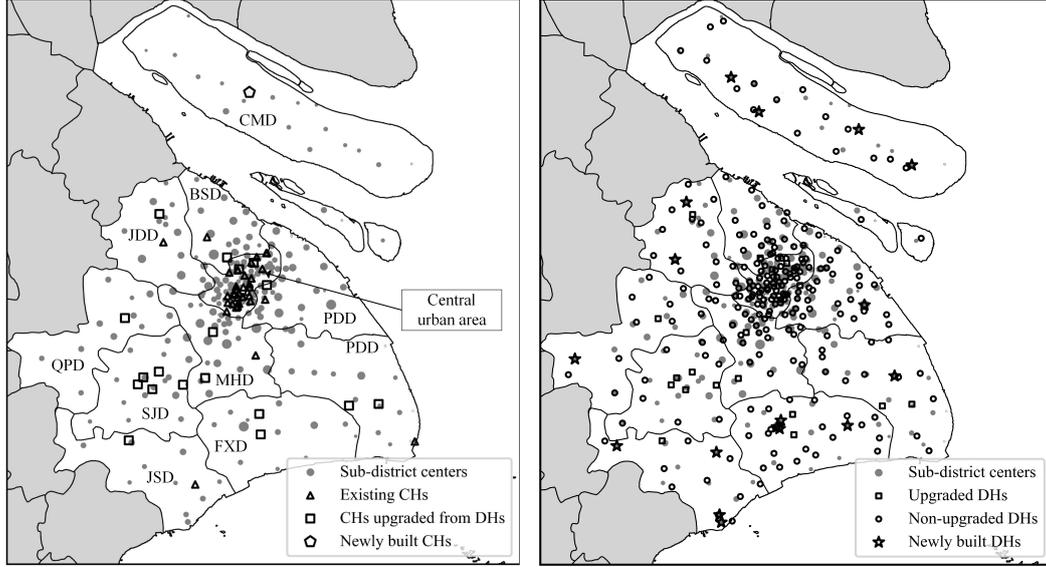


Figure 4.: Optimal location of CHs in Figure 5.: Optimal location of DHs in Study 1

Table 3.: Percentage of patients covered by UE and URR

Insurance type	Current		Alternative	
	UE	URR	UE	URR
Proportion	74.5%	25.5%	95%	5%

patients. This can be explained as follows. UE patients prefer to go to CH compared with URR patients. Thus with increased percentage of UE patients, the service demand for CH would increase and the demand for DH would decrease, which results in fewer DHs to be built. On the other hand, the percentage of UE patients increase does not change the trend of having most of upgraded hospitals in the suburban areas that had no CHs previously. Taking a closer look at the facility location, we found a DH would be upgraded on either side of the Chongming Island (majority of Chongming District), and a new CH would be built in Fengxian District. We noticed that the suburban area with the most upgrades is still Songjiang District, the same as in Study 1. Figure 7 displays the optimal location of newly built DHs, upgraded DHs, and remaining DHs. Similar to Study 1, the optimal location of DHs shows that most of the new DHs are built at more remote suburbs of the city.

To evaluate the performance the optimal hospital system design, we again compared the total costs. The results show a 40.9% reduction in the total cost under the optimal design (i.e., 3.47×10^5 under the optimal design vs. 5.86×10^5 under the current system). Further, the results in Table 4 show that under the optimal design, the mean waiting time at CH and DH are decreased, and the distances to CH and DH are decreased as well. Similar to Study 1, these results imply that adding CHs (mostly from DH upgrades) in suburban areas can economically improve patients' care accessibility. Comparing the two studies, we found in this study the distances to CHs and DHs are slightly higher; the mean waiting time at CHs is increased but that at DHs is decreased. These results further confirm that when the proportion of UE patients increases, the demand for CHs increases accordingly, which also results in a slight increase in the

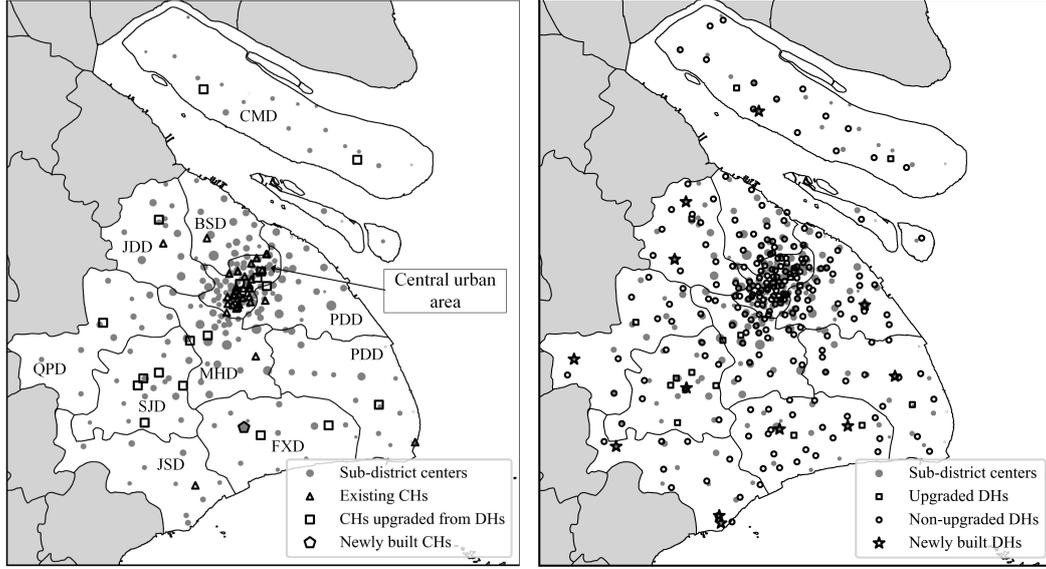


Figure 6.: Optimal location of CHs in Figure 7.: Optimal location of DHs in Study 2

total cost.

Table 4.: The weighted average wait time and distance in Study 2

	Mean waiting time (in hours)		Mean distance(in KM)	
	at CH	at DH	to CH	to DH
The current	1.9	0.29	17.0	1.75
Study 1	0.02	0.19	15.4	1.56
Study 2	0.03	0.17	15.6	1.63

Study 3: What is the optimal hospital system design if the threshold of patient's tolerable waiting time varies?

In this study, we analyzed the impact on the optimal hospital system design when the threshold of patient's tolerable waiting time decreases. We varied the threshold of patient's tolerable waiting time at CH and DH, as shown in Table 5. The baseline column provides the thresholds in Study 1. We kept other model parameters the same as in Study 1.

Table 5.: The threshold of patient's tolerable waiting time (in hours)

Hospital type	Baseline		Alternative	
	CH	DH	CH	DH
Threshold	2	1	0.5	0.1

Figure 8 displays the optimal location of 21 upgraded CHs. Similar to the previous studies, the results suggest most hospital upgraded would appear in suburban areas that had no CHs previously. The suburban area with the highest number of upgrades would still be at the center of suburban areas, i.e., Songjiang District. Taking a closer

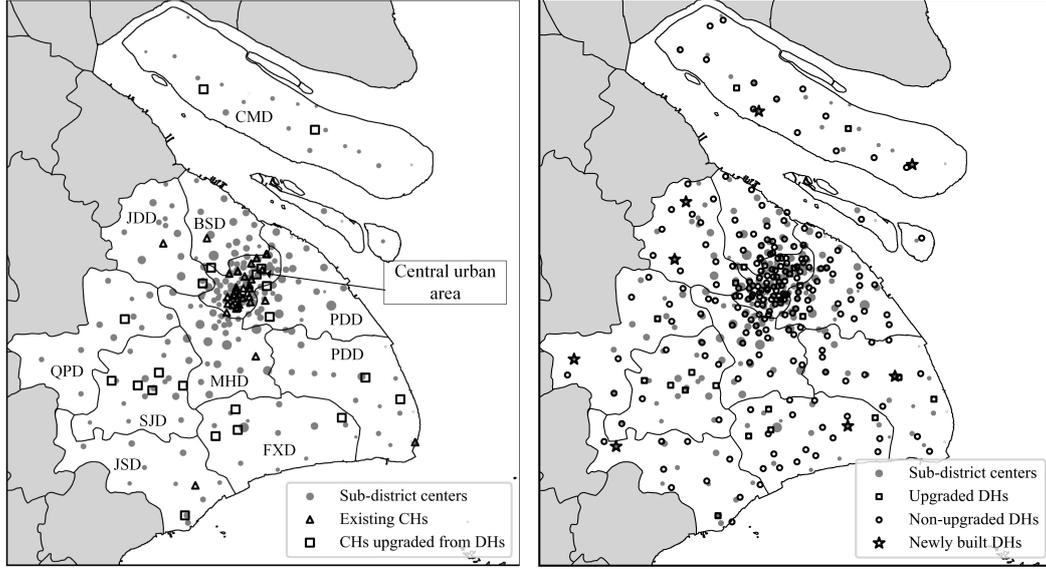


Figure 8.: Optimal location of CHs in Figure 9.: Optimal location of DHs in Study 3

look at, we found that two upgrades would appear on Chongming Island. Figure 9 displays the optimal location of DHs. We again found that most of the DHs would be located at suburban areas. Compared with the baseline optimal design (i.e., from Study 1), we found the number of CHs would increase as the waiting-time tolerance thresholds decrease. This can be explained as follows. With decrease of the thresholds, more CHs would be needed to reduce patient waiting time, to ensure the balking probability to be below a certain level.

To evaluate the performance the optimal hospital system design, we again compared the total costs. The results show a 41.4% reduction in the total cost (i.e., 3.44×10^5 under the optimal design vs. 5.86×10^5 under the current system). Further, the results in Table 6 show that under the optimal design, the mean waiting times and distances are decreased, and the distances to CH and DH are decreased. Comparing the results here and those from Study 1, we found the distances are further increased slightly; the waiting times are further decreased slightly; and a minimal increase in the total cost.

Table 6.: The weighted average wait time and distance in Study 3

	Mean waiting time (in hours)		Mean distance (in KM)	
	at CH	at DH	to CH	to DH
Current	1.86	0.29	17.0	1.75
Study 1	0.022	0.189	15.44	1.56
Study 3	0.020	0.019	15.45	1.66

5. Conclusions and Future Research

Capacity expansion and spatial redistribution is often needed in healthcare to coordinate with population changes and differences on care-seeking behavior. A well-re-

designed hospital system is expected to improve care access for the entire population without incurring significant spending on establishing new hospitals and upgrading the ones in existence.

In this paper, we studied the problem of optimally reconfiguring a two-tier hospital system consisting of central hospitals and district hospitals, which was inspired by real challenges in Chinese metropolitan areas. Our study includes: (1) developing a multinomial logit choice model to characterize hospital visit behaviors of different patient groups; (2) developing a multi-server queuing network model with the arrival rates specified by the choice model; (3) analyzing the performance measures of the queueing network model with consideration of patient balking; (4) solving a discrete location optimization problem for the hospital system redesign. Our study makes two contributions: (1) integration of choice models of multiple customer types into multi-type facility network redesign optimization with queuing performance measures in the objective; (2) real-world case studies for Shanghai, a place of imminent need, with anticipation that we will make direct impact to healthcare service operations there.

The key findings of the patient hospital choice modeling are (1) hospital type (central vs. district) and distance to hospital are influential factors in patient hospital choice behavior; (2) patients with urban employment insurance type are more likely to choose a central hospital than patients with urban rural resident insurance, when other conditions being equal. Our case study recommends hospital system reconfiguration in Shanghai in line with expansion and dispersion of residential areas in remote sites of Shanghai. Our case study also suggests to build fewer district hospitals with more patients of urban employment insurance type, and build more central hospitals when patients become less patient in waiting.

We plan to pursue further research in the following directions. First, 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. Second, we will incorporate individual-specific attributes (such as age) in the choice modeling. Finally, we will consider the effects of demographic changes (e.g., population aging) on the hospital system reconfiguration.

Acknowledgements

This work was supported by the National Natural Science Foundation of China [Grant No. 71432006; 71871138], CEIBS Healthcare Research Fund, and a Major International (Regional) Joint Research Project (71420107024) of the natural science foundation of China (NSFC).

References

- Aboolian, Robert, Oded Berman, and Dmitry Krass. 2007. "Competitive facility location and design problem." *European Journal of Operational Research* 182 (1): 40–62.
- Ahmadi-Javid, Amir, Pardis Seyedi, and Siddhartha S Syam. 2017. "A survey of healthcare facility location." *Computers & Operations Research* 79: 223–263.
- Beheshtifar, Sara, and Abbas Alimoahmadi. 2015. "A multiobjective optimization approach for location-allocation of clinics." *International Transactions in Operational Research* 22 (2): 313–328.
- Berg, Bjorn P. 2013. "Location models in healthcare." In *Handbook of Healthcare Operations Management*, 387–402. Springer.

- Berman, Oded, and Zvi Drezner. 2006. "Location of congested capacitated facilities with distance-sensitive demand." *IIE Transactions* 38 (3): 213–221.
- Daskin, Mark S. 2008. "What you should know about location modeling." *Naval Research Logistics (NRL)* 55 (4): 283–294.
- Daskin, Mark S, and Latoya K Dean. 2005. "Location of health care facilities." In *Operations research and health care*, 43–76. Springer.
- Farhan, Bilal, and Alan T Murray. 2006. "Distance decay and coverage in facility location planning." *The Annals of Regional Science* 40 (2): 279–295.
- Griffin, Paul M, Christina R Scherrer, and Julie L Swann. 2008. "Optimization of community health center locations and service offerings with statistical need estimation." *IIE transactions* 40 (9): 880–892.
- Guo, Pengfei, Christopher S Tang, Yulan Wang, and Ming Zhao. 2017. "The impact of reimbursement policy on social welfare, revisit rate and waiting time in a public healthcare system: fee-for-service vs. bundled payment." *Revisit Rate and Waiting Time in a Public Healthcare System: Fee-for-Service vs. Bundled Payment (October 25, 2017)* .
- Haase, Knut, and Sven Müller. 2014. "A comparison of linear reformulations for multinomial logit choice probabilities in facility location models." *European Journal of Operational Research* 232 (3): 689–691.
- Hajipour, Vahid, Parviz Fattahi, Madjid Tavana, and Debora Di Caprio. 2016. "Multi-objective multi-layer congested facility location-allocation problem optimization with Pareto-based meta-heuristics." *Applied Mathematical Modelling* 40 (7-8): 4948–4969.
- Huff, David L. 1964. "Defining and estimating a trading area." *Journal of marketing* 28 (3): 34–38.
- Kang, xiaolin. 2016. "Anaysis on the influencing factors of hospital selecting behavior in Wuhan." Master's thesis, Central China Normal University.
- Kim, Dong-Guen, and Yeong-Dae Kim. 2013. "A Lagrangian heuristic algorithm for a public healthcare facility location problem." *Annals of Operations Research* 206 (1): 221–240.
- Liu, Liqiang, and Vidyadhar G Kulkarni. 2008. "Balking and renegeing in M/G/s systems exact analysis and approximations." *Probability in the Engineering and Informational Sciences* 22 (3): 355–371.
- Marianov, Vladimir, Miguel Ríos, and Manuel José Icaza. 2008. "Facility location for market capture when users rank facilities by shorter travel and waiting times." *European Journal of Operational Research* 191 (1): 32–44.
- Marianov, Vladimir, and Daniel Serra. 1998. "Probabilistic, maximal covering location—allocation models for congested systems." *Journal of Regional Science* 38 (3): 401–424.
- Marianov, Vladimir, and Daniel Serra. 2002. "Location—allocation of multiple-server service centers with constrained queues or waiting times." *Annals of Operations Research* 111 (1-4): 35–50.
- McGuirk, Marjorie A, and Frank W Porell. 1984. "Spatial patterns of hospital utilization: the impact of distance and time." *Inquiry* 84–95.
- Mestre, Ana Maria, Mónica Duarte Oliveira, and Ana Paula Barbosa-Póvoa. 2015. "Location—allocation approaches for hospital network planning under uncertainty." *European Journal of Operational Research* 240 (3): 791–806.
- Mitropoulos, Panagiotis, Ioannis Mitropoulos, Ioannis Giannikos, and Aris Sissouras. 2006. "A biobjective model for the locational planning of hospitals and health centers." *Health Care Management Science* 9 (2): 171–179.
- Qiu, Huijuan. 2012. "Health Services Need, Demand and utilization of Residents in Xuhui District in Shanghai." PhD diss., Fudan University.
- Rahmaniani, Ragheb, Gona Rahmaniani, and Armin Jabbarzadeh. 2014. "Variable neighborhood search based evolutionary algorithm and several approximations for balanced location—allocation design problem." *The International Journal of Advanced Manufacturing Technology* 72 (1-4): 145–159.
- Schnatz, Peter F, Jessica L Murphy, David M O'Sullivan, and Joel I Sorosky. 2007. "Patient choice: comparing criteria for selecting an obstetrician-gynecologist based on image, gender,

- and professional attributes.” *American journal of obstetrics and gynecology* 197 (5): 548–e1.
- Shariff, SS Radiah, Noor Hasnah Moin, and Mohd Omar. 2012. “Location allocation modeling for healthcare facility planning in Malaysia.” *Computers & Industrial Engineering* 62 (4): 1000–1010.
- Stummer, Christian, Karl Doerner, Axel Focke, and Kurt Heidenberger. 2004. “Determining location and size of medical departments in a hospital network: A multiobjective decision support approach.” *Health Care Management Science* 7 (1): 63–71.
- Verter, Vedat, and Sophie D Lapierre. 2002. “Location of preventive health care facilities.” *Annals of Operations Research* 110 (1-4): 123–132.
- Victoor, Aafke, Diana MJ Delnoij, Roland D Friele, and Jany JDJM Rademakers. 2012. “Determinants of patient choice of healthcare providers: a scoping review.” *BMC health services research* 12 (1): 272.
- Vidyarthi, Navneet, and Onur Kuzgunkaya. 2015. “The impact of directed choice on the design of preventive healthcare facility network under congestion.” *Health care management science* 18 (4): 459–474.
- Wang, Qian, Rajan Batta, and Christopher M Rump. 2002. “Algorithms for a facility location problem with stochastic customer demand and immobile servers.” *Annals of operations Research* 111 (1-4): 17–34.
- Wei, Min. 2014. “The analysis of the influence factors on selecting hospital.” Master’s thesis, Anhui Medical University.
- Wei, Yi, Yingchun Yao, and Xiaoli He. 2018. “Analysis on influencing factors of outpatients’ behaviors in large general hospitals.” *Practical Journal of Clinical Medicine* 15 (4): 240–243.
- Zhang, Yue, and Derek Atkins. 2019. “Medical facility network design: User-choice and system-optimal models.” *European Journal of Operational Research* 273 (1): 305–319.
- Zhang, Yue, Oded Berman, Patrice Marcotte, and Vedat Verter. 2010. “A bilevel model for preventive healthcare facility network design with congestion.” *IIE Transactions* 42 (12): 865–880.
- Zhang, Yue, Oded Berman, and Vedat Verter. 2009. “Incorporating congestion in preventive healthcare facility network design.” *European Journal of Operational Research* 198 (3): 922–935.
- Zhang, Yue, Oded Berman, and Vedat Verter. 2012. “The impact of client choice on preventive healthcare facility network design.” *OR spectrum* 34 (2): 349–370.

Appendices

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.

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.

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) 35-44
 - d) 60-69
 - e) over 70
- 3) What is the highest level of education you have obtained?
 - a) Primary school or below
 - b) Junior high school
 - c) High school or vocational school
 - d) Junior college
 - e) Bachelor degree or above
- 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

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.

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

Appendix B. Location algorithm

In Genetic Algorithm, each chromosome represents a solution. The quality of a solution is represented by a fitness value. In this paper, the integer coding is used to represent a chromosome. Each chromosome contains a number of genes and each gene corresponds to a member in $J \cup J_L$. The actual implementation of the Genetic Algorithm is presented in the following.

Algorithm 1 Genetic Algorithm

Step 0 (Initialization): Randomly generate S_{pop} feasible solutions. Assure that they are all different. For each chromosome in the population, its fitness value is set to the objective function value.

Step 1 (Selection): Sort the chromosomes according to their fitness values. The first N solutions are used as chromosomes in the initila generation.

Step 2 (Generation of new chromosomes): Follow these steps until the descendant population has N chromosomes.

Step 2.1 (Parent selection): Two different chromosomes are randomly selected as parents. For each chromosome, the chance of being selected is equal.

Step 2.2 (Inherit): Copy the gene of the parents to two children. With different probability, they will be treated in one of the following three ways:

Treatment 1 (Crossover): For each pair of chromosomes, randomly select one part on the chromosome of one child, and exchange this part with the corresponding part of the other child.

Treatment 2 (Self Crossover): For each child, randomly select two genes on its chromosome and interchange their values.

Treatment 3 (Mutation): Choose one gene on a chromosome. For each child, randomly re-generate a new gene to replace the old one.

Step 2.3 (Activity): Examine the feasibility of the two children. If feasible, take them into the descendant population.

Step 3 (Replacement): Add the old population to the descendant population. Sort the chromosomes according to their fitness values. The first N chromosomes are used as the new chromosomes for the next generation.

Step 4 (Stopping criterion): If the running limit reaches a predefined limit, then stop. Otherwise go to Step 2.

Appendix C. Explanations about the conclusions on Section 4.2

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 1, the probabilities of choosing the CH for the two types of patients are given by:

$$\begin{aligned}
 p_C^{UE} &= \frac{\exp(-0.085 * d_C^{UE} + 1.087 * q_C + 0.148)}{\exp(-0.085 * d_C^{UE} + 1.087 * q_C + 0.148) + \exp(-0.085 * d_D^{UE} + 1.087 * q_D + 0.148)} \\
 &= \frac{1}{1 + \exp(-0.085 * (d_D^{UE} - d_C^{UE}) + 1.087 * (q_D - q_C))};
 \end{aligned}$$

$$\begin{aligned}
p_C^{\text{URR}} &= \frac{\exp(-0.093 * d_C^{\text{URR}} + 1.027 * q_C - 0.150)}{\exp(-0.093 * d_C^{\text{URR}} + 1.027 * q_C - 0.150) + \exp(-0.093 * d_D^{\text{URR}} + 1.027 * q_D - 0.150)} \\
&= \frac{1}{1 + \exp(-0.093 * (d_D^{\text{URR}} - d_C^{\text{URR}}) + 1.027 * (q_D - q_C))}.
\end{aligned}$$

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

$$p_C^{\text{UE}} = \frac{1}{1 + \exp(-0.085 * (d_D^{\text{UE}} - d_C^{\text{UE}}) - 1.087)};$$

$$p_C^{\text{URR}} = \frac{1}{1 + \exp(-0.093 * (d_D^{\text{URR}} - d_C^{\text{URR}}) - 1.027)}.$$

When the differences between traveling to the DH and traveling to the CH are the same for the two types of patients (i.e., $d_D^{\text{UE}} - d_C^{\text{UE}} = d_D^{\text{URR}} - d_C^{\text{URR}} = \Delta d$), we have $-0.085 * \Delta d - 1.087 - (-0.093 * \Delta d - 1.027) = 0.008 * \Delta d - 0.06$. As we described in Section 4.2, the distance of traveling to the DH is no more than 3km (i.e., $d_D^{\text{UE}} < 3$, and $d_C^{\text{URR}} < 3$), thus $\Delta d < 3$ and $0.008 * \Delta d - 0.06 < -0.036 < 0$. Therefore, when the differences of the travel distances are the same for the two types, we have $p_C^{\text{UE}} > p_C^{\text{URR}}$, which means that UE patients are more willing to choose a CH than URR patients.

Appendix D. Additional information on the district and sub-district of Shanghai

Table 7.: 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%	

Table 8.: 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