

# Pricing and Capability Planning of the Referral System Considering Medical Quality and Delay-Sensitive Patients—Based on the Chinese Medical System

Zhenzhong Guan, Hezhi Chen, Na Zhao, Aifeng Zhang

School of Economics and Management, Southwest Jiaotong University, Chengdu, China

Email: [zzguan@swjtu.cn](mailto:zzguan@swjtu.cn), [hezhi.chen@qq.com](mailto:hezhi.chen@qq.com), [729203432@qq.com](mailto:729203432@qq.com), [18213464149@163.com](mailto:18213464149@163.com)

**How to cite this paper:** Guan, Z.Z., Chen, H.Z., Zhao, N. and Zhang, A.F. (2020) Pricing and Capability Planning of the Referral System Considering Medical Quality and Delay-Sensitive Patients—Based on the Chinese Medical System. *Journal of Mathematical Finance*, 10, 96-131.

<https://doi.org/10.4236/jmf.2020.101008>

**Received:** December 12, 2019

**Accepted:** January 19, 2020

**Published:** January 22, 2020

Copyright © 2020 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

Combined with the development of Chinese medical industry, we have analyzed the characteristics of the referral system with Chinese characteristics and the possible problems in it. We combine the two main factors that affect patient care: medical quality and waiting costs, to study a most representative referral system made up of a tertiary hospital and a community hospital for its pricing and capacity planning issues. We build an integrated model of queuing theory and game theory to analyze if all patients make rational choices, what the community hospital should do to maximize the service capacity. Meanwhile, how the tertiary hospital makes pricing to achieve its profit maximization in the case of government taxation. The results show that the service capacity of community hospital is affected by government subsidies, cure rate, and the medical expenses of the tertiary hospital. And those of tertiary hospital are influenced most by government subsidies. Therefore, the results show that the smooth implementation of the government's macro-control plays a crucial role in the two-way referral system.

## Keywords

Referral System, Medical Quality, Waiting Cost, Service Capacity, Pricing Strategy

## 1. Introduction

With the improvement of residents' living standards, the demands and requirements for medical care are getting higher and higher, but the gaps and contradictions between the people's growing medical and health needs and the current

level of medical services are getting wider. In order to alleviate this contradiction, the new medical reform proposes hierarchical diagnosis and two-way referrals, which aims to coordinate various medical resources between large-scale comprehensive medical institutions and primary care through macro-control to achieve rational use of limited resources. At the same time, more patients are encouraged to go to the primary medical institutions for the first visit, so as to change the medical habits and alleviate the pressure of large-scale comprehensive medical institutions. However, for patients, the level of primary medical care in the traditional concept is limited and it is difficult to ensure the quality of medical care. This problem has become the key to hinder the implementation of this policy.

In addition, the patient's focus on medical services is not limited to the medical level, but also includes factors such as service quality, waiting time (waiting cost), medical environment, and costs of diagnosis and treatment. Therefore, considering the above issues, how to adjust the operation strategy of the hospital and promote the development of tiered diagnosis and two-way referral while achieving its own goals has become an urgent and practical problem in the field of medical operation management.

There are great differences between the two in terms of strength or reputation, leading to a sharp contrast between them in admissibility. Tertiary hospitals in the large-scale comprehensive hospitals receive far more patients than they can every day, while the community hospital in the primary hospitals are less visited, resulting in a serious waste of medical resources. In addition, due to the large number of patients in comprehensive hospitals, they have to wait for too long a time for treatment. The treatment experience is getting worse, and the domestic medical satisfaction has plummeted for more and more people.

Due to the uneven distribution of medical service resources (including: outpatient resources, medical technology resources, operating room resources, and ward resources, etc.), the quality of medical care at different levels of hospitals varies greatly.

When selecting the first hospital, patients often prefer the higher-level hospitals, that is, the hospitals with higher medical quality, which has directly become the primary reason for the high number of doctors in the tertiary hospital. Therefore, in the context of medical reform, how to take advantage of the difference in the reimbursement ratio between the first consultation in community hospital and the first consultation in tertiary hospital, and rationally set the price of consultations at different levels of hospitals in the referral system to attract more patients to the community hospital has become an indispensable part of the research on the referral system.

In addition, due to the large number of large general hospitals, the waiting time for patients is too long, and the patient's experience is worse. Some patients will choose other hospitals with shorter waiting time for the first consultation. Therefore, how to scientifically set up the hospital's service capacity, that is, the

number of first-patient reception, has become an important factor affecting patient waiting time and experience.

Based on the above points, only comprehensive consideration of the hospital's medical quality, patient waiting time, and medical pricing and service capabilities can create better opportunities for the implementation of referrals, while helping the hospital to obtain maximum benefits, while helping the hospital to obtain the maximum benefit, the patient's perceived value is maximized.

We consider a two-way referral medical service system consisting of a community hospital with the goal of maximizing service capacity and the tertiary hospital with the goal of maximizing profit. We consider a two-way referral medical service system consisting of a community hospital with the goal of maximizing service capacity and a tertiary hospital with the goal of maximizing profit. We consider the quality of medical care, introduce the choice of medical treatment for delay-sensitive patients, study the pricing issues and capacity planning of referral systems, find the best balance point for tertiary hospital and community hospital, and help medical units in the referral system to obtain the optimal solution they want, so as to maximize the profits of the tertiary hospital and maximize the service capacity of community hospital, while helping patients make better medical treatment choices.

In this paper, we start from the choice of patients in the two-way referral system, taking medical quality and delay-sensitive patients as influential factors, to discuss how to optimize referral systems and achieve effective convergence between the tertiary and community hospital. We then construct mathematical models to discuss the optimal solution of each subject in the referral system under different prices and preferences of delay-sensitive patients. We can then maximize the profits of the tertiary hospital and the service capacity of community hospital.

The paper is divided into three parts. The first part is to introduce the background and significance of the research, summarize the relevant literature and construct the research ideas of this paper. The second part is to describe the patients' choices of medical treatment based on the referral system. We build utility functions according to the choosing behaviors of patients to analyze medical treatment behaviors of patients under different circumstances. Based on this, we analyze how community hospital set their service capacity to accommodate and serve more patients. Meanwhile, we introduce the government subsidies to analyze its impact on system performance. We then build the profit model of the tertiary hospital to study the pricing strategy of them. Finally, we will discuss the impact of medical quality on patient arrival rate, medical service price and waiting cost respectively, in combination with the medical quality and the waiting cost of delayed-sensitive patients. The third part is about the numerical simulation and the prospect of the conclusion. We validate the conclusion and the correctness of the paper through numerical simulation to summarize the research and put forward some suggestions on the operation and management of the tertiary hospital and community hospital at present. Finally, comparing the research status of

domestic and foreign scholars and the actual operation situations of the hospitals with the research in this paper, we point out the defects and shortcomings in this study and the research direction of future referral system.

## 2. Literature

The two-way referral system originated in the early twentieth century in British, aiming at alleviating the tense use of medical resources in large and medium-sized hospitals and the phenomenon of service vacuum in community hospital. The system has achieved good results, so other countries began to follow suit, learning and using this method. In the study of referral systems, Shortell *et al.* [1] focused on the exploration of referral behaviors and referral doctors' attitude. Mackey *et al.* [2] analyzed the 3 years' referral data of a community health center and recommended to improve the medical process and the quality of primary health care services. Wilkin *et al.* [3] focused on the impact of medical expenses and medical service level on medical referral. Evans *et al.* [4] focused on the impact of general practitioners' medical technology and hospital services on referral rates. Harvey *et al.* [5] believed that improving the medical service level of primary medical institutions is the key to ensure the healthy operation of the referral system. In addition, Goddard [6] and Vardy [7] *et al.* also studied the paying ways of public hospitals. For the resource allocation research of the referral system, we referred to the two review articles of Fomundam [8] and Soni [9]. Bhattacharjee and Ray [10], Meli *et al.* [11], Song *et al.* [12] studied the medical cooperation model for upward referrals, mainly discussed the current situation of the allocation of medical resources in the United States. They found that medical resources in high-level hospitals are very tight, but many medical resources in ordinary hospitals are mostly idle or wasted. From this, they found that for ordinary hospitals, there are problems in upward referrals. Then they analyzed the importance of upward referrals by establishing a mathematical model of the cooperation model between high-level hospitals and ordinary hospitals, and proposed measures and policies to promote community medical service centers to refer patients upwards to large hospitals.

For the research object, China, in this paper, the study of the referral system originated in the 1980s and 1990s. During the same period, Chinese scholars began to combine the national conditions, to explore a two-way referral system with Chinese characteristics. Lai and Chen [13] proposed that the main problem facing with the Chinese two-way referral system was the conflict between institutional design and reality. Wu and Li [14] have found, through field research, that the two-way referral bottleneck mainly originated from the regional referral standard confusion, the lagging medical technology and hardware facilities of primary medical institutions, as well as the lack of government public finance investment. Wang and Qiu [15] analyzed the 56 related articles on domestic two-way referral in recent years and discussed the related problems in the implementation of the two-way referral system in China. Tang [16] suggested optimizing the

development of two-way referral system from the aspects of raising the level of primary medical care and expanding the clinical path. Mei [17] introduced the two-way referral system experience in Taiwan and Singapore. Cao and Wu [18], Liu and Liao [19] analyzed Hong Kong's public health services. Ren and Li [20], Ma and He [21], Su and Wang [22] and so on also put forward development methods of the two-way referral system suitable for China's national conditions, combining with foreign experience.

After understanding the referral system, we need to look at the two influencing factors that affect the stability of the referral system, namely, delay sensitiveness and medical quality.

At present, the research results on delay sensitive are more abundant. There is theoretical research as well as empirical research by constructing models. Naor [23] introduced, for the first time, the queuing model into the study of delay sensitiveness and service pricing. This idea was referred to by many scholars later [24]-[29]. Pangburn [30] explores the profit level of different service strategies through the study of heterogeneous customers with different delay sensitivity and price sensitivity. Boyaci and Ray [31] shifted the research focus to the relationship between time delay cost and service perceived value. Yu and Zhao [32] explored the different attitudes of the customers to the service product and the delay cost perception under the monopoly and the duopoly market, by constructing the Nash equilibrium model under the service competition condition. Afeche and Mendelson [33] analyzed the important role of delay sensitiveness in the service industry. In addition, Lederer and Li [34], Sharma and Mehrotra [35], Taylor [36] have also stressed in their research that the service industry should not only pay attention to enhancing the internal service quality, but also the waiting time of the customers in the service process, to increase the competitiveness of enterprises.

Chinese scholars also conducted more delay-sensitive research. Chen and Zhang [37] explored the impact of the choices of time-sensitive customers on corporate marketing strategies. Tang and Feng [38] analyzed the impact of people's patience on the road in the traffic jam. Wang *et al.* [39] analyzed the consumers who were more sensitive to the waiting cost. Chen *et al.* [40] analyzed the different choice behaviors of delayed sensitive patients based on the queuing theory and the game theory. Qi [41] constructed the resource allocation model to discuss the rescue resource allocation scheme in the disaster environment and obtain the multi-objective programming equation based on the delay time, rescue costs and other elements. Tian and Huang [42] brought the travelers' delay costs into crowded costs to study the congestion pricing and travel planning of the urban traffic. Ni and Zhao [43] studied the medical distribution problems of the hospitals for patients with different injury through the construction of the function model.

The study of medical quality has always been a hot issue for scholars around the world. Vargas [44] analyzed methods and results of medical quality evaluation and monitoring by graphs. Headley and Miller [45] regarded medical quality as a

means of healthcare companies to improve their competitiveness and obtain resources. TS Dagger and Sweeney [46] introduced cohort analysis into the study of medical quality and health care services. Wong *et al.* [47] discussed the relationship between medical service perception and medical cost in different health conditions taking the patients with acute symptoms and chronic diseases. Bansal and Taylor [48] focused on the evaluation indicators of the service quality. In general, the study of medical quality focuses on the interaction between the influencing factors of medical quality and the customers' perception [49] [50] [51]. In addition, there are also many scholars analyzing the gap between the medical quality of different hospitals by constructing models [52]-[58].

With the development of Chinese medical cause, the study of medical quality has gradually improved. Zhao and Jiao [59] divided the medical service system development into three stages through the discussion of development and medical evaluation indicators of foreign medical quality management. Xu and Li [60] focused on the medical service evaluation of the rural medical institutions. Cheng [61] discussed the quality supervision in the medical quality system. Tan and Ou [62] believed that the key to improving and enhancing the medical quality was to meet the medical needs of patients and improve the safety of medical services. Dong [63] built a patient trust evaluation index system based on medical quality and analyzed patient satisfaction based on this. Tian and Niu [64] discussed the current phenomenon in medical industry that the large and medium-sized hospitals were overcrowding, but the primary community hospitals were inadequate in service needs. Tu and Wang [65] emphasized that the medical quality was the key factor in the sustainable development of medical organizations combining with the current experience in their development.

It is not difficult to find that, through the reading and analysis of the above literature, the study of referral system is more perfect in Western countries and that in China is relatively backward. If we copy the oversea methods to China, it won't get very good results, but also cause greater trouble because it's not suitable for the national conditions. Therefore, in this paper, I combined with Chinese national conditions to study the influence of the patients' choices on the referral system from the patients' points of view—the medical subject.

In previous studies, both domestic scholars and foreign scholars, the research on the pricing and capacity planning of referral systems was scarce. The few papers were only from a single perspective for analysis. However, we comprehensively considered two topics of high patient attention during the visit: medical quality and waiting time, combined the two to study the perceived value of the patient's visit.

We take the patient selection behavior in the two-way referral system as the starting point, and discuss how to optimize the referral system by adding medical quality and delay-sensitive patients as influencing factors, so as to achieve effective connection between the tertiary hospital and community hospital, and discussed the optimal solution for each subject in the two-way referral system under different pricing and delay-sensitive patient preferences to maximize the

profits of the tertiary hospital and maximize the service capacity of community hospital, thereby promoting hierarchical diagnosis and two-way referrals. The promotion and implementation of the diagnosis system make the optimal use of medical resources, and provide some reference opinions for the rational distribution of patients and the construction of an efficient referral system.

### 3. Basic Model

In the modern society with rapid economic development, there will be a number of hospitals in every region. When people need to get medical treatment, they will consider a variety of factors including the distance from the hospitals, the medical and health level, the service attitude, the medic are reimbursement ratio, the convenience of transportation, the hospital environment. However, for most patients, the final choice can only be judged by the feelings. And there is no accurate data or an answer to tell them exactly what hospital to choose. In such cases, a study of patient behaviors has occurred.

The article deals with the following assumptions:

1) Suppose there are only two hospitals on the market: one is  $A$ , a tertiary hospital, and the other is  $C$ , a community hospital. When people get sick, they can only choose to go to  $A$  or  $C$  for medical treatment, or go to neither to wait for self-cure.

2) For the actual treatment behavior, patients will take into account a number of factors when choosing hospitals, including medical quality, attitude of the medical staff, queuing time, medical expenses and the distance. But to facilitate research, in this paper, we mainly consider the medical quality, medical expenses and waiting time. Among them, the two factors of service quality and treatment environment are included in medical quality, and medical quality can be quantified by the medical data of the hospital in recent years.

3) For the two hospitals in the system, the community hospital represents the primary medical and health institution with ordinary medical quality, while the tertiary hospital represents authority and high quality in the eyes of the residents for its high ranks and scores. Therefore, in the study of the medical quality of the two, we take the tertiary hospital as a reference for medical quality, assuming that patients in the tertiary hospital can all be cured, namely, the cure rate of the tertiary hospital is 100% and the medical quality of the community hospital should be judged by this reference.

4) In actual medical treatment, emergency patients and special patients are also included apart from outpatient appointment patients. To study patients' selection behaviors, it is assumed that, in this paper, all patients are able to wait in non-emergency situations and they all make rational choices.

5) As there are only two hospitals in the system, patients either choose  $H_A$  or  $H_C$  or neither. So we assume that the patients arriving at the two hospitals are independent of each other. In other words, the queuing system in each hospital in the treatment is also independent of each other.

6) Since each patient that arrives needs to queue in the queuing system, to facilitate queuing, we assume that the patients all arrive individually and the patient arrives in a steady period of time and is subject to the Poisson distribution.

7) In the outpatient treatment process, we assume that the service time of each hospital for the patients is subject to the exponential distribution, and the service hours of the two are subject to independent index distribution.

8) In order to maximize the service to all patients, we also assume that the total service capacity of the two hospitals is infinite in the system. In other words, as long as there are patients to seek medical treatment, they will be received. Regardless of the length of the waiting time, they will eventually get the service.

We first study how patients would choose the first hospital for diagnosis considering the medical quality and the waiting costs at the same time. Let's mark the tertiary hospital as  $H_A$  and the community hospital as  $H_C$ .

Two patients arrive to form an M/M/1 queuing system in the two hospitals respectively. We assume the service capacity of the tertiary hospital is larger and can serve more patients, marked by  $\mu_A$ . And that of the community hospital smaller comparatively, marked by  $\mu_C$ . Each hospital needs to pay a certain cost when serving patients. We mark the serving cost of the tertiary hospital as  $c_1$  and that of the community hospital as  $c_2$ .

But as there are differences in medical level and medical quality between  $H_A$  and  $H_C$ , we further assume that the medical quality of  $H_A$  is  $q_A$  and that of  $H_C$  is  $q_C$ . We assume that all patients in  $H_A$  can be cured, namely, the cure rate of  $H_A$  is 1, the cure rate of  $H_C$  will be  $\frac{q_C}{q_A}$  and the non-cure rate of it will be  $1 - \frac{q_C}{q_A}$ .

With the Chinese government's efforts for hierarchical diagnosis and treatment, we assume that the patients transferred from the community hospital to the tertiary hospital all can enjoy certain percentage of the benefits and the reimbursement ratio is assumed as  $1 - \theta$  ( $0 < \theta < 1$ ). We mark the medical expenses of the tertiary hospital is  $p_A$  and those of the community hospital as  $p_C$ . Then the patients who are not cured in the community hospital will actually pay  $\theta p_A$  for the medical expenses.

We also assume that the process of patient arriving at the hospital obeys the Poisson distribution and in the system, the total patient arrival rate is  $\Lambda$ . We assume that all patients don't know the number of people queuing up before arriving at the hospital and will not leave after arrival. And then we mark the arrival rate of  $H_A$  as  $\lambda_A$  and that of  $H_C$  as  $\lambda_C$  ( $\lambda_A + \lambda_C \leq \Lambda$ ). Under normal circumstances, namely the non-emergency situations, the patients have two choices. The first one is to go directly to  $H_A$  for first diagnosis. And the second one is to  $H_C$  firstly and to transfer to  $H_A$  if they are not cured. Therefore, the treatment process of the patients in sick who need to select a hospital is shown in **Figure 1**.



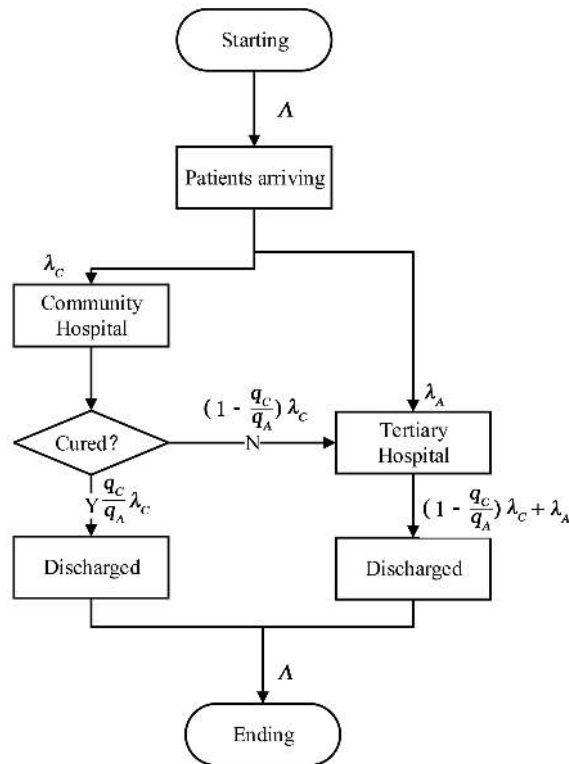


Figure 1. Treatment process of patients choosing hospitals.

Patients need to wait during the medical treatment. The longer the waiting time, the greater the waiting cost. The delay cost for all delay-sensitive patients is  $h$ . Assuming that the patient’s delay waiting cost is about the waiting time linear, we mark  $W(\lambda_i, \mu_i)$  as the waiting time for a patient and then the delay cost is  $hW(\lambda_i, \mu_i)$ .

For all delayed-sensitive patients, whether to enter the queue depends on the “net utility value” of the service offered by the providers. When the patient chooses to enter a hospital, its net utility value  $U$  must be non-negative, and  $q - p - hW(\lambda_i, \mu_i) \geq 0$ .

As the purpose of health care reform is to make everyone afford to get cured, in this paper, we assume that all patients in the system will go to the hospital when they get sick.

**Assumption 1.** When  $p \leq q - \frac{h}{\mu - \lambda}$ , all patients will go to the hospital, then  $\lambda_A + \lambda_C = \Lambda$ .

For the patients, whether they will choose to go to  $H_A$  or  $H_C$  depends on the net utility values of the two. The patients will give priority to the one with a large net utility value.

If the patients choose to go to  $H_A$  for medical treatment, the net utility value will be marked as  $U_A$ .

$$U_A = q_A - p_A - \frac{h}{\mu_A - \left(1 - \frac{q_C}{q_A}\right)\lambda_C - \lambda_A} \tag{1}$$

If the patients choose to go to  $H_C$ , the net utility value will be marked as  $U_C$ , it is made up of two parts. The one part is net utility value cured in  $U_C$ . And the other part is the one not cured and transferred to  $H_A$ . Among them, because the patients are transferred to  $H_A$  through the referral system, when they receive medical treatment in  $H_A$ , they only need to pay  $\theta p_A$  for the medical expenses. Then:

$$U_C = \frac{q_C}{q_A} \left( q_C - p_C - \frac{h}{\mu_C - \lambda_C} \right) + \left( 1 - \frac{q_C}{q_A} \right) \left( q_A - \theta p_A - \frac{h}{\mu_C - \lambda_C} - \frac{h}{\mu_A - \left( 1 - \frac{q_C}{q_A} \right) \lambda_C - \lambda_A} \right) \quad (2)$$

We can learn easily from the customer perceived value theory and utility model that when there are only the two hospitals available for selection, the patients will choose the one that bring them the larger utility value (perceived value). By observing the two utility function models for the patients to the community hospital and the tertiary hospital for first diagnosis, we can easily find that medical quality, medical expenses, waiting cost as well as the reimbursement proportion will all influence the patients on choosing the first hospital to go.

We classify the patients as quality-oriented patients, waiting cost-oriented patients and price-oriented patients. For patients of different types when choosing the first hospital, we would focus on the more important factors they care and take them as important basis for their selection of the first hospital.

## 4. Extension to the Basic Model

### 4.1. Research for the Service Capability Design of Community Hospital

In the previous section, we mainly study how the patients rationally choose the first hospital taking into account the medical quality, the waiting costs and other factors. Although the results show the patients' choice behaviors under different circumstances, if the service capacity of the community hospital is not sufficient to support patients who come for medical treatment, the assumptions in the previous section will not be established and later analysis and discussion will be meaningless. Therefore, this section will focus on the design of service capabilities of community hospital. And we make the following supplementary assumptions:

- 1) Assume that the medical expenses of the tertiary hospital is  $p_A$ .
- 2) Assume that the service capacity of the tertiary hospital is large enough.

For the community hospital, the purpose is not to get much profit, but rather on how to improve their own service capacity to accommodate more patients for the first diagnosis. Therefore, we need to set the maximum service capacity  $\mu_C$  to serve more patients and increase the first diagnosis rate. Therefore, the service

capacity design of community hospital is transformed into seeking the value of the following formula:

$$\begin{aligned} & \max_{\mu_C \geq 0} \lambda_C(\mu_C) \\ & \text{s.t. } 0 < \lambda_C < \mu_C \end{aligned}$$

However, before solving the maximum service capacity of community hospital, we must ensure that there are patients to choose the community hospital. Or the community hospital will lose its existing meaning without patients.

When  $U_A = U_C$ , there was no difference in patient choice. That is, the arrival rate of the two hospitals are  $\lambda_A^*$  and  $\lambda_C^*$ . So the  $\lambda_A^*$  and  $\lambda_C^*$  should fit in the following formula:

$$\begin{aligned} q_A - p_A - \frac{h}{\mu_A - \left(1 - \frac{q_C}{q_A}\right)\lambda_C - \lambda_A^*} &= \frac{q_C}{q_A} \left( q_C - p_C - \frac{h}{\mu_C - \lambda_C^*} \right) \\ &+ \left(1 - \frac{q_C}{q_A}\right) \left( q_A - \theta p_A - \frac{h}{\mu_C - \lambda_C^*} - \frac{h}{\mu_A - \left(1 - \frac{q_C}{q_A}\right)\lambda_C^* - \lambda_A^*} \right) \end{aligned}$$

To facilitate the computerizing, we set

$$\begin{aligned} \alpha &= 1 - \frac{q_C}{q_A}, \quad \left( \frac{q_C}{q_A} = 1 - \alpha \right) \\ A_0 &= (1 - \alpha)(q_C - q_A - p_C) + (1 - \alpha\theta)p_A \\ A_1 &= (1 - \alpha)h \\ A_2 &= \mu_A - \Lambda \end{aligned}$$

**Proposition 1.** When  $\mu_C \geq \frac{\mu_A - \Lambda}{1 - \alpha}$ ,  $\lambda_C^* \geq 0$ ,  $H_C$  will have patients for medical treatment, and the first diagnosis rate is

$$\lambda_C^* = \frac{A_0\mu_C(1 - \alpha) - A_0A_2 - 2A_1 + \sqrt{[A_0A_2 + A_0\mu_C(1 - \alpha)]^2 + 4A_1^2}}{2A_0(1 - \alpha)} \quad (3)$$

Proposition 1 shows that the community hospital must reach a certain size. In other words, the service capacity  $\mu_C$  should not be lower than the value of  $\frac{q_C}{q_A}$ , the ratio of the residual service capability of  $H_A$  and the cure rate of  $H_C$ .

Otherwise, the patients will go directly to the tertiary hospital  $H_A$  because the net utility value is too low.

From **Proposition 1**, we know that whether patients choose to go to the community hospital is, to a large extent, influenced by the cure rate of the community hospital, the cure rate is affected by the medical quality. One of the purposes of graded diagnosis and treatment is to make it easier for patients to seek medical treatment. That is, if the patient only has some simple and minor illnesses, it

can be resolved through a community hospital. You can go directly to a large general hospital for treatment of a disease, and it will not cause the long waiting time of the large hospital because everyone goes to the large hospital. Therefore, to study this issue, we discuss the different situations and make further analysis for the cure rate in them.

**Case 1.** Common diseases, frequently-occurring diseases and chronic diseases.

Community hospital are an important and integral part of primary medical and health treatment. Their comprehensive strength, especially the diagnosis and treatment level is also increasing year by year. So we need to focus on the study on the treatment methods and effects of the common diseases, frequently-occurring disease and chronic diseases. At present, the vast majority of community hospitals have been able to basically prevent and guide the treatment for these diseases. In other words, for the common diseases, frequently-occurring diseases and chronic diseases, the treatment level of the community hospital (medical quality) is equal to that of the tertiary hospital. At this time,  $q_C = q_A$ ,  $\frac{q_C}{q_A} = 1$ , and  $\alpha = 0$ .

As described in Proposition 1, when  $\mu_C \geq \mu_A - \Lambda$  patients will go to  $H_C$ . That is, when the service capability of  $H_C$  is better than the residual service capability of  $H_A$ , the patients will go to the community hospital for the first diagnosis. Or they will go directly to  $H_A$  because the net utility value is too low.

**Case 2.** Difficult and complicated miscellaneous diseases, acute and serious illness.

However, due to the collective effects of medical and health treatment level, medical team and equipment, coupled with uneven distribution of medical resources for many years, the community hospital have great difficulty in dealing with the difficult and complicated miscellaneous diseases, acute and serious illness. Especially for cancer and other incurable diseases, community hospital have almost no solution. At this time, the  $q_C$  of  $H_C$  is far more less than the  $q_A$  of  $H_A$ . When  $q_C \ll q_A$  and  $\frac{q_C}{q_A} = 0$ , then  $\alpha = 1$ .

Because  $\alpha = 1$ , then in the Formula (3), the denominator is 0. The formula is meaningless. So it has to be computerized again and we get the following value.

$$\lambda_C^* = \mu_C - \frac{h}{(1-\theta)p_A}$$

Therefore, when  $\mu_C - \frac{h}{(1-\theta)p_A} \geq 0$ ,  $\lambda_C^* \geq 0$ . And there are patients going to the community hospital. Namely, the  $\mu_C$  of  $H_C$  should not be lower than the ratio of  $h$  and  $(1-\theta)p_A$ , in which  $h$  is the unit delay cost and  $(1-\theta)p_A$  is the additional reimbursement of the patients transferred to the tertiary hospital. Otherwise, the patients will go directly to  $H_A$  for a maximum net utility value because that of the community hospital is too low.

Through the analysis for the above two cases, we can easily find that in the face of different types of diseases, the design for the hospital's service capacity is also different. Therefore, when the community hospital designs further the service capacity, it should make full use of large data and the knowledge of mathematical statistics to analyze the disease distribution in the hospital patients, whether there are more patients with common, frequently-occurring and chronic diseases or with difficult and miscellaneous diseases, acute and serious illness. They need to re-plan the design according to the distribution of the proportions. But no matter what kind of choice they will choose, they shouldn't make an arbitrary one, because in the actual scenario, there cannot be only one class of patients for treatment. And they need to make an overall thinking and judging to achieve the optimal allocation of resources in the hospital, and also to reduce the medical burden for the tertiary hospital, provide more convenience for community residents and make actual contribution to the hierarchical diagnosis and treatment in the new medical reform.

Therefore, if we want to truly realize the purpose of "cure all small disease within the community", the medical service capability, the patient accommodating scenarios are also crucial apart from the great support of the country and the improvement in the medical and health treatment of the community hospital.

**Lemma 1.** When  $\mu_C \geq \frac{\mu_A - \Lambda}{1 - \alpha}$ , the  $\lambda_C^*$  of  $H_C$  will increase with the increase of  $\mu_C$ , then  $\frac{\partial \lambda_C^*}{\partial \mu_C} > 0$ .

Therefore, the design for the service capability of  $H_C$  is transformed to calculate the value of the following formula:

$$\max_{\mu_C \geq 0} \lambda_C^*(\mu_C) = \frac{A_0 \mu_C (1 - \alpha) - A_0 A_2 - 2A_1 + \sqrt{[A_0 A_2 + A_0 \mu_C (1 - \alpha)]^2 + 4A_1^2}}{2A_0 (1 - \alpha)}$$

$$s.t. \mu_C \geq \frac{\mu_A - \Lambda}{1 - \alpha}$$

In addition, in order to encourage residents to visit the community hospital for first diagnosis, the Government provides the patients with a lot of benefits, including medical expenses and medical charges. They are mainly embodied in reducing the medical expenses and enjoying the outpatient treatment when transferring to other hospitals. However, if we blindly provide benefits for the patients, the community hospital income is bound to reduce. In response to this scenario, the Government will provide a total of  $B$  subsidies for the community hospital to balance the total income. Therefore, from an economic point of view, only when the community hospital have access to government subsidies and the total income is more than or equal to the total cost can the community hospital survive in the market regardless of the excessive loss, namely,  $B + p_C \mu_C \geq c_2 \mu_C$ .

And because when  $\mu_C \geq \frac{\mu_A - \Lambda}{1 - \alpha}$ , the first diagnosis rate of  $H_C$  will exist. We can get the following proposition.

**Proposition 2.** In the scenario of given  $p_C$  of  $H_C$ , if  $B \geq (c_2 - p_C) \frac{\mu_A - \Lambda}{1 - \alpha}$ , the best service of  $H_C$  will be  $\mu_C^* = \frac{B}{c_2 - p_C}$ , and at this time  $\lambda_C^* > 0$ . Otherwise the community hospital will be out of the market because there is no patient to come, then  $\lambda_C^* = 0$ .

Proposition 2 shows the effect of government subsidies on the first diagnosis rate of community hospital. It is only when the government subsidy  $B$  reaches the value or above  $(B \geq (c_2 - p_C) \frac{\mu_A - \Lambda}{1 - \alpha})$  will there be patients to  $H_C$  for medical treatment. This also proves the speculation on the impact of the government's macro-control on primary health care institutions (community hospital) in the previous part of the paper. In addition to policy guidance, the government also needs to use the appropriate "invisible hand" to intervene and provide them with some subsidies and funds, helping them improve the overall level and at the same time, ensuring that the operation will not be in a loss state. And this is precisely very important for the community hospital in the fierce competition. It can save them from getting out of the market because of the loss, but also ensure the effective implementation of hierarchical diagnosis and treatment and the two-way referral policy.

In short, only when the community hospital reaches a certain size or more, that is, the service capacity is not less than the ratio of the remaining service capacity of the tertiary hospital to the cure rate of the community hospital, the patient will choose the community hospital for the first consultation. Otherwise, patients will choose a tertiary hospital because their net utility value is too low. However, the cure rate of community hospital is also affected by the type of disease. After researching the classification of different types of diseases, it was found that the service capacity of community hospital without the cure rate was analyzed, and two special cases were discussed. The service capacity design of community hospital is under different circumstances.

## 4.2. Pricing Strategy Research on Tertiary Hospital

For the referral system in the study made up of a tertiary hospital and a community hospital, the main purpose of the community hospital is to maximize the service capacity to serve more community residents. And for a tertiary hospital that needs to pay a certain tax to the government every year, and undertakes the treatment for major social diseases and medical events, its main purpose is also to ensure maximum profits to maintain the normal operation of the hospital in addition to the completion of the above tasks. Therefore, this section will focus on the pricing strategies of the tertiary hospital to maximize its profits through the optimal pricing strategy.

In the previous section, we have solved the followers, that is, the community hospital's objective function value, that is, the optimal solution of its service ca-

capacity. So in this section, we will make further analysis based on this, to obtain the optimal pricing strategy of the tertiary hospital. We make the following supplementary assumptions:

- 1) The community hospital's medical expenses are not subject to fluctuations in the market and will only be affected by government regulation.
- 2) The tertiary hospital needs to pay the government a certain amount of tax.
- 3) The total income of the hospital is the total medical expenses. The total cost is the cost of the service incurred during the course of treatment and the tax costs that need to be paid to the government.

Compared with community hospital, the tertiary hospital is much higher in the medical quality than the general hospitals, especially in the cure rate, which can almost reach 100%. Relatively speaking, the cure rate of community hospital is affected by the types of diseases, the patients' scenarios and other factors. There will be some patients transferred to the tertiary hospital for medical treatment through the referral system.

In addition, due to the government's macro-control, the tertiary hospitals need to pay an additional tax of a total amount of  $T$ . Therefore, for the tertiary hospital, it is necessary to assume the needs of the referral patients in the referral system of the medical reform, and also to ensure that their operations are not in a loss. This requires the tertiary hospital to seek an optimal solution in the medical expense pricing, ensuring the arrival rate of patients and also their maximum profits. Namely, they maximize the profit value  $\pi(p_A)$  by setting the  $p_A$  of the  $H_A$ .

From the previous assumptions and market demand, we can know that the service capacity  $\mu_A$  of the tertiary hospital satisfies the condition of  $\mu_A > \Lambda$ . Therefore, its decision-making is not limited by service capabilities. We make the service cost of the tertiary hospital within a unit time as  $c_1$ , then the profit model can be expressed as follows:

$$\pi(p_A) = p_A \lambda_A + \theta p_A \alpha \lambda_c - c_1 \mu_A - T \tag{4}$$

So the decision-making model of the tertiary hospital can be expressed as follows:

$$\begin{aligned} \max_{p_A \geq 0} \pi(p_A) &= p_A \lambda_A + \theta p_A \alpha \lambda_c - c_1 \mu_A - T \\ \text{s.t. } \frac{c_1 \mu_A + T}{\theta \alpha \Lambda} &\leq p_A \leq q_A - \frac{h}{\mu_A - \Lambda} \end{aligned}$$

**Proposition 3.**  $\pi(p_A)$  of the tertiary hospital is a concave function about the medical expenses of  $p_A$ .

We can see from the nature of the function that when the medical expenses of the tertiary hospital gradually increase, the profit of the tertiary hospital become firstly bigger and then smaller. In other words, in the actual scenario, if the tertiary hospitals want to achieve profitability, they cannot blindly improve the medical expenses. Otherwise the hospital will suffer a lower rate of patients due to high costs, and the hospital will be in a loss state.

We calculate the profit function of the tertiary hospital. We set the reciprocal as zero, and then the medical expenses of the tertiary hospital will be  $p_A^*$ :

$$p_A^* = \frac{4}{3} \cdot \left( \sqrt{\frac{24(1-\alpha)(q_C - q_A - p_C)}{15\mu_C - 32\Lambda}} - (1-\alpha)(q_C - q_A - p_C) \right)$$

**Proposition 4.** When  $\mu_A$  satisfies the condition of  $\Lambda < \mu_A < \left(\frac{47}{15} - \frac{32}{15}\alpha\right) \cdot \Lambda$ , there will be the optimal pricing strategy:

1) When  $\frac{\mu_A - \Lambda}{1 - \alpha} \leq \mu_C < \frac{32}{15}\Lambda$ ,

$$p_A = q_A - \frac{h}{\mu_A - \Lambda}$$

2) When  $\mu_C \geq \frac{32}{15}\Lambda$ ,

a) If  $\frac{c_1\mu_A + T}{\theta\alpha\Lambda} \leq p_A^* \leq q_A - \frac{h}{\mu_A - \Lambda}$ , then  $p_A = p_A^*$ ;

b) If  $p_A^* > q_A - \frac{h}{\mu_A - \Lambda}$ , then  $p_A = q_A - \frac{h}{\mu_A - \Lambda}$ .

Proposition 4 shows how the service capacity  $\mu_C$  of the community hospital affects the medical expenses  $p_A$  of the tertiary hospital  $H_A$ .

1) When the service capability of  $H_C$  is small and ensures that it won't be too small to lead to no patients ( $\frac{\mu_A - \Lambda}{1 - \alpha} \leq \mu_C < \frac{32}{15}\Lambda$ ), the profits of  $H_A$  increase as the medical expenses increase, as shown in:

1) Of Proposition 4. However, for the current Chinese medical units, not all units have the power of self-pricing, therefore:

a) If the hospitals in the region have achieved the power of self-pricing, then the optimal price of the tertiary hospital  $H_A$  should be as follows:

$$p_A = q_A - \frac{h}{\mu_A - \Lambda}$$

b) If the hospitals in the region haven't achieved the power of self-pricing, then the price of  $H_A$  should refer to the government limit and take the government limit  $P$  as the optimal price, that is,  $p_A = P$ .

2) When the service capacity of the community hospital is large enough ( $\mu_C \geq \frac{32}{15}\Lambda$ ), the optimal pricing strategy of  $H_A$  should be discussed according to the following scenarios:

a) If the hospital can make self-pricing, and when the  $p_A^*$  could ensure that the two hospitals are neither in loss nor losing their patients, when

$\frac{c_1\mu_A + T}{\theta\alpha\Lambda} \leq p_A^* \leq q_A - \frac{h}{\mu_A - \Lambda}$ , the price of  $H_A$  should be as follows:

$$p_A = p_A^* = \frac{4}{3} \cdot \left( \sqrt{\frac{24(1-\alpha)(q_C - q_A - p_C)}{15\mu_C - 32\Lambda}} - (1-\alpha)(q_C - q_A - p_C) \right)$$



b) If the hospital can make self-pricing, and  $p_A^*$  has gone beyond  $p_A$ , namely,  $p_A^* > q_A - \frac{h}{\mu_A - \Lambda}$ , then the optimal price of  $H_A$  should be the upper limit of the price range, that is,

$$p_A = q_A - \frac{h}{\mu_A - \Lambda}$$

c) If the hospital cannot make self-pricing, and  $P$  goes beyond the upper limit of  $p_A$  ( $P \geq p_A$ ), then the optimal price of  $H_A$  should be the upper limit of the price range, that is,

$$p_A = q_A - \frac{h}{\mu_A - \Lambda}$$

d) If the hospital cannot make self-pricing, and  $P$  is between  $p_A^*$  and the upper limit ( $p_A^* \leq P < p_A$ ), then the optimal price of  $H_A$  should be the optimal solution of the price, then,

$$p_A = p_A^* = \frac{4}{3} \cdot \left( \sqrt{\frac{24(1-\alpha)(q_C - q_A - p_C)}{15\mu_C - 32\Lambda}} - (1-\alpha)(q_C - q_A - p_C) \right)$$

e) If the hospital cannot make self-pricing, and  $P$  is smaller than  $p_A^*$  ( $P < p_A^*$ ), then the optimal price of  $H_A$  should be the upper limit of  $P$ , namely,  $p_A = P$ .

We can see from Proposition 2 that it is only when the government subsidy  $B$  reaches a certain value ( $B \geq (c_2 - p_C) \frac{\mu_A - \Lambda}{1 - \alpha}$ ), there will be patients coming to  $H_C$ . And also in the scenario of a given  $p_C$  of  $H_C$ , if  $B \geq (c_2 - p_C) \frac{\mu_A - \Lambda}{1 - \alpha}$ , then the service capacity of  $H_C$  will be  $\mu_C^* = \frac{B}{c_2 - p_C}$ . And at this time,  $\lambda_C^* > 0$ . Otherwise the community hospital will be out of the market due no patients,  $\lambda_C^* = 0$ . Therefore, when  $\frac{\mu_A - \Lambda}{1 - \alpha} \leq \mu_C < \frac{32}{15} \Lambda$ ,  $B$  should satisfies the condition of:

$$(c_2 - p_C) \cdot \frac{\mu_A - \Lambda}{1 - \alpha} \leq B < (c_2 - p_C) \cdot \frac{32}{15} \Lambda$$

When  $\mu_C \geq \frac{32}{15} \Lambda$ ,  $B$  should satisfies the condition of

$$B \geq (c_2 - p_C) \cdot \frac{32}{15} \Lambda$$

Therefore, in the *Stackelberg* game model in this study, we take  $H_A$  as the leader and  $H_C$  as the follower. When  $\Lambda < \mu_A < \left(\frac{47}{15} - \frac{32}{15}\alpha\right) \cdot \Lambda$  and

$B \geq (c_2 - p_C) \cdot \frac{\mu_A - \Lambda}{1 - \alpha}$ ,  $H_A$  get the maximum profit by designing an optimal pricing strategy, that is, the best strategy.

Based on the above conclusions, we can also draw the following Lemmas:

**Lemma 2.**  $\mu_C^*$  of  $H_C$  is related to  $B$  and  $c_2$ , and it increases with the increase of  $B$  and decreases with the increase of  $c_2$ .

Lemma 2 shows that, for community hospital, when  $c_2$  is constant, the bigger  $B$  is, the higher the service capacity. However, when  $B$  is constant, the increasing service cost per unit time will reduce its service capacity. Therefore, in order to avoid the decline in service capacity of the community hospital due to the increase in service costs, community hospital should improve service efficiency and other means to control service costs, so as to ensure its service capacity so that there will be enough patients to come for treatment.

**Lemma 3.** When  $(c_2 - p_c) \cdot \frac{\mu_A - \Lambda}{1 - \alpha} \leq B < (c_2 - p_c) \cdot \frac{32}{15} \Lambda$ , the price of the tertiary hospital  $p_A$  will not be affected by it. In contrast, it increases with the increase of  $q_A$  and the increase of  $\mu_A$  of the tertiary hospital, and it decreases with the increase of  $\Lambda$  in the system.

Lemma 3 shows that when  $B$  is comparatively less, the community hospitals are difficult to contend and compete with the tertiary hospital. At this time, the tertiary hospitals have sufficient autonomy to make self-pricing to achieve its profit maximization. Meanwhile, to improve the medical quality and service capacity of the tertiary hospital will increase its pricing level, thereby increasing its profit values. But with the increase in the total number of patients in the system, patients will have greater autonomy, which will also force the tertiary hospital to reduce their optimal pricing, resulting in lower profits.

**Lemma 4.** When  $B$  is comparatively higher and satisfies  $B \geq (c_2 - p_c) \cdot \frac{32}{15} \Lambda$ , if the optimal price of the tertiary hospital is  $p_A = p_A^*$ , the optimal price of the medical expenses in the tertiary hospital will decrease with the increase of  $B$ .

Lemma 4 shows that when the subsidy  $B$  is higher, the service capacity of the community hospital can be improved. In such a case, the community hospital will have the ability to compete with the tertiary hospital. Therefore, at this time, the tertiary hospital's blind increase in medical expenses will lead to the loss of patients. So they have to lower  $p_A$  to attract more patients.

**Lemma 5.**  $\lambda_C^*$  of the community hospital increases with the increase of  $p_A^*$  of the tertiary hospital.

Lemma 5 shows that when the medical expenses of the tertiary hospital continue to increase, the patients will pay a too high price, which will lead to a lower utility value. And at this time, they will consider visiting the community hospital for the first diagnosis, increasing to the first diagnosis rate of them.

**Lemma 6.**  $\lambda_C^*$  of the community hospital increases with the increase of  $q_C$ .

Lemma 6 shows that with the improvement of the medical quality of the community hospital, the non-cure rate due to the too low medical quality will decline and more patients begin to favor the convenient community hospital with relatively low medical expenses. Patients who go to community hospital for first diagnosis will receive a higher utility value relative to the tertiary hospital. Meanwhile, it will also force the tertiary hospital to adjust their pricing strategies to

better participate in the market competition.

**Lemma 7.** When patients go to the community hospital or the tertiary hospital for first diagnosis, their waiting time and waiting costs will increase with the reduction of  $q_c$  of the community hospital.

Lemma 7 shows that with the improvement of the medical quality of the community hospital, the non-cure rate will decline. For the patients who go to the community hospital for first diagnosis, the probability of requiring referral will be reduced and the waiting time and waiting costs caused by referral will decrease. Similarly, for patients who go to the tertiary hospital for first diagnosis, their waiting time will be reduced due to the reduction in the number of referral patients so that the waiting costs will also be reduced.

It is not difficult to find that the profit value of the tertiary hospital does not simply increase with the increase of medical treatment fees, but increases first and then decreases. This also reflects that, in actual situations, reasonable adjustments must be made according to market competition, so as to choose the optimal pricing to achieve the maximum benefits. However, according to the state's relevant regulations on medical pricing, not all hospitals have the power to make independent pricing, and the tertiary hospital should not only consider their own interests, but also weigh the relationship with community hospital. Therefore, its pricing cannot be too low, so that patients in community hospital are lost to the tertiary hospital, which puts greater pressure on the development of medical work in the tertiary hospital

## 5. Numerical Analysis

### 5.1. The Influence of the Non-Cure Rate $\alpha$ of the Community Hospital on the Arrival Rate of Patients $\lambda_c$ of Them

Assume that the model parameters are:

$$q_a = 180, q_c = 90, p_c = 90, \Lambda = 1, h = 1, \theta = 0.5$$

We draw the relationship between  $\lambda_c$  and  $\alpha$  when  $\mu_c = 0.5, 0.8, 1$ , as shown in **Figure 2**.

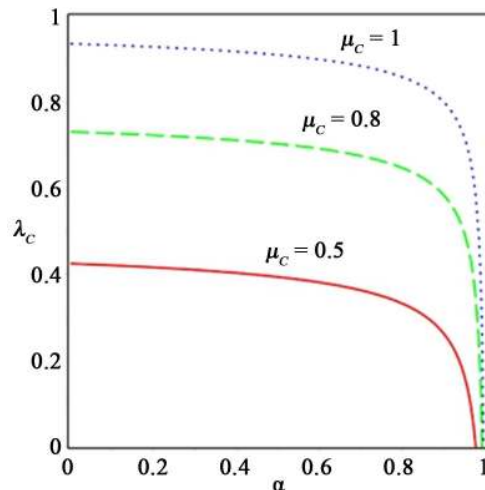
We can see from **Figure 2** that with the increase of  $\alpha$ ,  $\lambda_c$  is decreasing. In other words, when the medical quality of the community hospital decreases (non-cure rate increases), the patient will choose not to go there for the first diagnosis due to the too low net utility value, which is consistent with routine thinking patterns and judging criteria of the majority of patients.

In addition, when the non-cure rate does not change, the arrival rate of patients will also increase with the improvement of  $\mu_c$ , which proves the conclusion of Lemma 1, the  $\lambda_c^*$  of  $H_c$  will increase with the increase of  $\mu_c$ .

### 5.2. The Impact of Government Subsidy $B$ on Community Hospitals' Service Ability of $\mu_c$

Assume that the model parameters are

$$\alpha = 0.5, \mu_A = 1.6, q_a = 180, q_c = 80, c_2 = 90, h = 1, \theta = 0.5$$



**Figure 2.** The influence of  $\alpha$  on  $\lambda_c$ .

We draw the relationship between  $\mu_c$  and  $B$  when  $p_c = 60, 70, 100$  respectively, as shown in **Figure 3**.

We can see from **Figure 3**, when the medical expenses of the community hospital are fixed, the service capacity changes with the change of government subsidies. From the foregoing assumptions, we can see that community hospital as an important part of primary medical institutions, the medical expenses are comparatively lower. They often need government subsidies to ensure that they will not be out of the competitive market because of losses. Therefore, in most cases,  $p_c$  is lower than  $c_2$ . We can easily find through numerical analysis that when  $p_c$  is lower than  $c_2$ , the service capacity of community hospital increases with the increase of government subsidy. When  $p_c$  is higher than  $c_2$ , the service capacity of community hospital decreases with the increase of government subsidy. And through **Figure 1** and the analysis, we can also see that the service capacity of community hospital will affect the first diagnosis rate. Therefore, we can easily infer that the government subsidy to community hospital not only will affect their service capacity, but also their first diagnosis rate. In other words, it will affect the patients' treatment options, thus affecting the implementation of hierarchical diagnosis and treatment.

### 5.3. The Effect of the Total Amount of Patients $\Lambda$ on System Performance When Government Subsidy $B$ Is Comparatively Less

Assume that the model parameters are

$$\alpha = 0.5, \mu_A = 1.6, q_a = 180, q_c = 80, p_c = 90$$

$$h = 1, \theta = 0.5, T = 1, c_1 = 1.2, c_2 = 1.1$$

We draw respectively the relationship between  $\Lambda$  and  $p^*$  as well as that between  $\Lambda$  and  $\pi$ , when  $B$  is comparatively less. As shown in **Figure 4(a)** and **Figure 4(b)**.

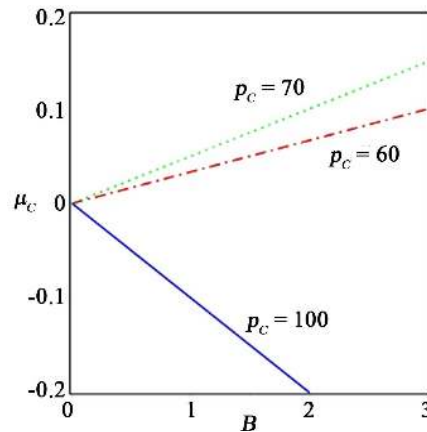


Figure 3. The impact of  $B$  on community hospital service ability of  $\mu_c$ .

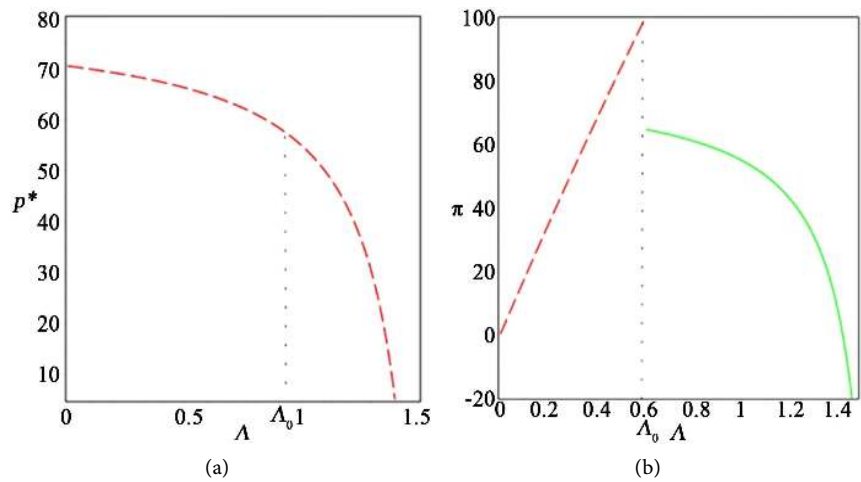


Figure 4. (a) The effect of  $\Lambda$  on  $p^*$  when the government subsidy is comparatively less; (b) The effect of  $\Lambda$  on  $\pi$  when the government subsidy is comparatively less.

When  $B$  is comparatively less, the service capacity of community hospital will be comparatively smaller and the patient arrival rate will be comparatively lower. At this time, most patients will choose to go to tertiary hospital for first diagnosis. We can see from Figure 4(a) that when  $B$  is comparatively less, the total arrival quantity of patients will increase and the optimal price of the tertiary hospital shows a decreasing trend. It decreases gradually in  $(0, \Lambda_0]$  and rapidly in  $[\Lambda_0, \Lambda)$ . That is to say, when the total patient volume in the system gradually increases to  $\Lambda_0$ , the optimal price of the tertiary hospital decreases gradually. But because the number of patients at this time is showing a growth trend and the price decline is slow, so in general, the profit of the tertiary hospital still shows a state of growth. However, when the total amount of patients is more than  $\Lambda_0$ , the patients will choose to go directly to the community hospital for they will wait too long time if going to the tertiary hospital for first diagnosis. At this time, due to the sudden leaving of the patients, the price of the tertiary hospital will decline greatly. Then the first diagnosis rate and the price decrease at the same time, the profit of the tertiary hospital will also have a “diving” decrease.

### 5.4. The Effect of the Total Amount of Patients $\Lambda$ on System Performance When Government Subsidy $B$ Is Comparatively More

Assume that the model parameters are

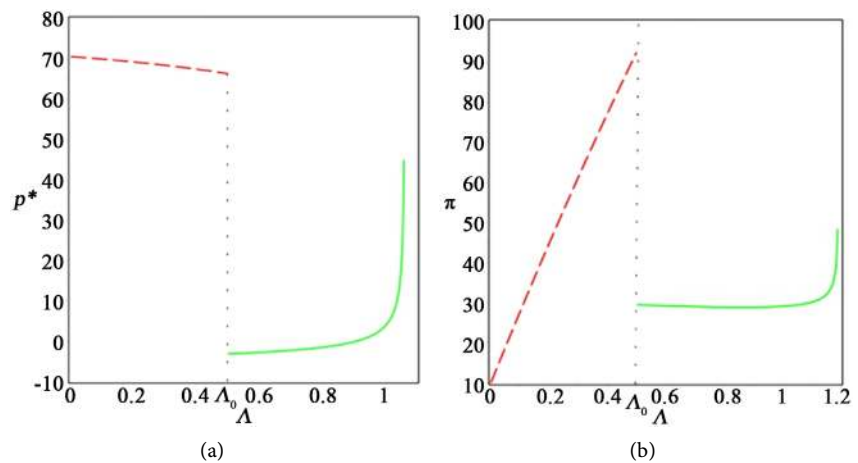
$$\alpha = 0.5, \mu_A = 1.6, q_a = 180, q_c = 80, p_c = 90$$

$$h = 1, \theta = 0.5, T = 1, c_1 = 1.2, c_2 = 1.1$$

We draw respectively the relationship between  $\Lambda$  and  $p^*$  as well as that between  $\Lambda$  and  $\pi$ , when  $B$  is comparatively more. As shown in **Figure 5(a)** and **Figure 5(b)**.

When the government subsidy  $B$  is comparatively more, the service capacity of community hospital will be comparatively larger and the patient arrival rate will be comparatively higher. At this time, most patients will choose to go to community hospital for first diagnosis.

We can see from **Figure 5(a)** that when  $B$  is comparatively more, although the optimal price of the tertiary hospital is higher, the profit value will also be at a low level as fewer patients come for the first diagnosis. With the increase in the overall arrival rate of patients, the optimal price will show a slight decline in the trend, and at this time, with the increase in the number of patients, the profit value will be improved. However, when the total amount of patients reaches  $\Lambda_0$ , due to service capacity constraints of community hospital, the patients have to wait for too long and the waiting cost increases, the patient begin to consider going to the tertiary hospital for the first diagnosis. At this time, to attract the patients waiting in community hospital, the tertiary hospital must lower the price. So the optimal price of the tertiary hospital at this time will suffer a substantial “diving” decrease and the total profit will reduce significantly due to the price reduction. When the total number of patients is getting bigger and bigger, the arrival rate of patients in the tertiary hospital is also growing rapidly. At this time, in order to coordinate the community hospital and the tertiary hospital on the arrival rate of



**Figure 5.** (a) The effect of  $\Lambda$  on  $p^*$  when the government subsidy is comparatively more; (b) The effect of  $\Lambda$  on  $\pi$  when the government subsidy is comparatively more.

patients, the tertiary hospital must raise prices to force the patients to choose community hospital for first diagnosis. But because of the impact of the total amount of patients, the profit will still show a dramatic increase.

## 6. Conclusions

In this paper, the two influencing factors of medical quality and waiting cost that have been focused more are selected as the research object. We combine with the existing related research of the domestic and foreign scholars and build a two-way referral system taking a tertiary hospital and a community hospital as the main body. We study the pricing of the tertiary hospital and the service capacity design issues of the community hospital through theoretical research and numerical simulation. The main work and conclusions are as follows.

1) We construct a utility function model of the patients' first diagnosis to the community hospital and the tertiary hospital. We compare the utility values of the two and determine what the first choices of the patients are to provide reference for community hospital and tertiary hospital on attracting patients better and improving the medical and health level comprehensively.

2) We analyze how community hospital can set their own service capabilities to accommodate and serve more patients and improving the patients' utility values.

3) We construct the profit function model of the tertiary hospital, analyze the nature of the function and obtain the maximum profit in what kind of pricing scenario.

4) We take into account the effect of government subsidies on the system performance and analyze the operation conditions of the system under different subsidies

However, in this study, we conduct the research based on assumptions in many places. However, in reality, the situations are often a lot more complicated. Therefore, there are some shortcomings in this paper needed to be discussed and studied in the future research. Some patients feel that medical quality is more important than waiting cost, and vice versa. Therefore, in the future study, we can divide the two with a certain balance. In this paper, we only consider the referral to the upper hospitals, not to the lower ones. We only consider the two-level system, not the three-level or multi-level systems. We can explore the more complex patterns in future research.

## Acknowledgements

This research was supported by the National Natural Science Foundation of China under Grants No. 71572154, the Fundamental Research Funds for the Central Universities under Grants No. 26816WTD25, and the Service Science and Innovation Key Laboratory of Sichuan Province under Grants No. KL1705.

## Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

## References

- [1] Shortell, S.M. and Anderson, O.W. (1970) The Physician Referral Process: A Theoretical Perspective. *Health Services Research*, **6**, 39-48.
- [2] Mackey, R.A., Harvey, D.S., Taschman, A., et al. (1967) Periodic Surveys of Community Resources: A Project to Improve Referrals for Direct Services. *Community Mental Health Journal*, **3**, 331-334. <https://doi.org/10.1007/BF02349231>
- [3] Wilkin, D. and Smith, A.G. (1987) Variation in General Practitioners' Referral Rates to Consultants. *Journal of the Royal College of General Practitioners*, **37**, 350-353.
- [4] Evans, E., Aiking, H. and Edwards, A. (2011) Reducing Variation in General Practitioner Referral Rates through Clinical Engagement and Peer Review of Referrals: A Service Improvement Project. *Quality in Primary Care*, **19**, 263-272.
- [5] Mackey, R.A., Harvey, D.S.W. and Taschman, A. (1967) Periodic Surveys of Community Resources: A Project to Improve Referrals for Direct Services. *Community Mental Health Journal*, **3**, 331-334. <https://doi.org/10.1007/BF02349231>
- [6] Goddard, J. and Tavakoli, M. (2008) Efficiency and Welfare Implications of Managed Public Sector Hospital Waiting Lists. *European Journal of Operational Research*, **184**, 778-792. <https://doi.org/10.1016/j.ejor.2006.12.003>
- [7] Vardy, D.A., Freud, T., Sherf, M., et al. (2008) A Co-Payment for Consultant Services: Primary Care Physicians' Referral Actualization. *Journal of Medical Systems*, **32**, 37-41. <https://doi.org/10.1007/s10916-007-9105-9>
- [8] Fomundam, S. and Herrmann, J.W. (2007) A Survey of Queuing Theory Applications in Healthcare. *Institute for Systems Research*, **24**, 1-22.
- [9] Soni, K. and Saxena, K. (2011) A Study of Applicability of Waiting Line Model in Health Care: A Systematic Review. *Operation Research*, **19**, 75-91.
- [10] Bhattacharjee, P. and Ray, P.K. (2014) Patient Flow Modelling and Performance Analysis of Healthcare Delivery Processes in Hospitals: A Review and Reflections. *Computers & Industrial Engineering*, **78**, 299-312. <https://doi.org/10.1016/j.cie.2014.04.016>
- [11] Meli, C.L., Khalil, I. and Tari, Z. (2014) Load-Sensitive Dynamic Workflow Re-Orchestration and Optimisation for Faster Patient Healthcare. *Computer Methods and Programs in Biomedicine*, **113**, 1-14. <https://doi.org/10.1016/j.cmpb.2013.06.019>
- [12] Song, J., Chen, W. and Wang, L. (2012) A Block Queueing Network Model for Control Patients Flow Congestion in Urban Healthcare System. *International Journal of Services Operations and Informatics*, **7**, 82-95. <https://doi.org/10.1504/IJSOI.2012.051394>
- [13] Lai, W. and Chen, M.S. (2010) Policy Thinking on Further Perfecting Two-Way Referral System. *The Chinese Modern Hospital Management*, **8**, 1-3.
- [14] Wu, H.M. and Li, Y.P. (2010) Discussing the Status of Two-Way Referral System and Experiences from It. *The Chinese Health Service Management*, **27**, 441-442.
- [15] Wang, N.X., Qiu, W.Y. and Zhang, Y.F. (2014) Problems and Solutions on Dual Referral. *Chinese Hospitals*, No. 5, 66-68.
- [16] Tang, Y.G. (2014) Status Quo and Progress of Dual Referral between Hospital and Community Health Center. *Chinese Hospitals*, No. 7, 44-45.
- [17] Mei, S.Y. (2013) The Dual Referral System of Taiwan and Singapore and Its Enlightenment. *Medicine and Philosophy*, **34**, 70-72. <https://doi.org/10.5406/amerjtheophil.34.1.0070>



- [18] Cao, H.L. and Wu, G.Z. (2008) Investigation and Reflection on Hong Kong Public Health Service System. *Shanghai Journal of Preventive Medicine*, **20**, 134-135.
- [19] Liu, Y.Y. and Liao, C.J. (2001) Training and Medical Continuing Education of Surgeons in Foreign and Hong Kong. *Chinese Journal of General Surgery*, **16**, 629-632.
- [20] Ren, X. and Li, J.-R. (2010) Current Problems of Dual Referral and Countermeasures in New Medical Service Care Reformations. *Chinese Health Economics*, **29**, 44-46.
- [21] Ma, Y.N. and He, Q.C. (2007) Study on Two-Way Transfer Treatment System in Community in China Based on Management Medicine of America. *Chinese Health Quality Management*, **14**, 76-78.
- [22] Su, L.L. and Wang, J. (2008) The Characteristics of Dual-Referral Model in Other Countries. *Soft Science of Health*, **22**, 160-162.
- [23] Naor, P. (1969) The Regulation of Queue Size by Levying Tolls. *Econometrica*, **37**, 15-24. <https://doi.org/10.2307/1909200>
- [24] Chen, H. and Wan, Y.W. (2003) Capacity Competition of Make-to-Order Firms. *IIE Transactions*, **35**, 817-832. <https://doi.org/10.1080/07408170304412>
- [25] Griffiths, J.D., Pricelloyd, N., Smithies, M., et al. (2006) A Queueing Model of Activities in an Intensive Care Unit. *IMA Journal of Management Mathematics*, **17**, 277-288. <https://doi.org/10.1093/imaman/dpi042>
- [26] Mendelson, H. and Whang, S. (1990) Optimal Incentive-Compatible Priority Pricing for the M/M/1 Queue. *Operations Research*, **38**, 870-883. <https://doi.org/10.1287/opre.38.5.870>
- [27] Maister, D. (1984) Psychology of Waiting Lines. Harvard Business School Cases. Harvard University, Boston, 71-78.
- [28] Yankovic, N. and Green, L.V. (2011) Identifying Good Nursing Levels: A Queuing Approach. *Operations Research*, **59**, 942-955. <https://doi.org/10.1287/opre.1110.0943>
- [29] Frank, M. (2004) Monopoly Pricing When Customers Queue. *IIE Transactions*, **36**, 569-581. <https://doi.org/10.1080/07408170490438690>
- [30] Pangburn, M.S. and Stavoulaki, E. (2008) Capacity and Price Setting for Dispersed, Time-Sensitive Customer Segments. *European Journal of Operational Research*, **184**, 1100-1121. <https://doi.org/10.1016/j.ejor.2006.11.044>
- [31] Boyaci, T. and Ray, S. (2003) Product Differentiation and Capacity Cost Interaction in Time and Price Sensitive Markets. *Manufacturing & Service Operations Management*, **5**, 18-36. <https://doi.org/10.1287/msom.5.1.18.12757>
- [32] Yu, D.Z., Zhao, X. and Sun, D. (2013) Optimal Pricing and Capacity Investment for Delay-Sensitive Demand. *IEEE Transactions on Engineering Management*, **60**, 124-136. <https://doi.org/10.1109/TEM.2012.2194288>
- [33] Afèche, P. and Mendelson, H. (2004) Pricing and Priority Auctions in Queueing Systems with a Generalized Delay Cost Structure. *Management Science*, **50**, 869-882. <https://doi.org/10.1287/mnsc.1030.0156>
- [34] Lederer, P.J. and Li, L. (1997) Pricing, Production, Scheduling, and Delivery-Time Competition. *Operations Research*, **45**, 407-420. <https://doi.org/10.1287/opre.45.3.407>
- [35] Sharma, A. and Mehrotra, A. (2007) Choosing an Optimal Channel Mix in Multi-channel Environments. *Industrial Marketing Management*, **36**, 21-28. <https://doi.org/10.1016/j.indmarman.2006.06.012>

- [36] Taylor, S. (1994) Waiting for Service: The Relationship between Delays and Evaluations of Service. *Journal of Marketing*, **58**, 56-69.  
<https://doi.org/10.1177/002224299405800205>
- [37] Chen, J. and Zhang, N. (2008) Study on Backorder Incentives and Inventory Control Policies with Time-Based Customer-Choice Behavior. *Journal of Management in China*, **11**, 53-63.
- [38] Tang, Y.M. and Feng, S.W. (2008) Model of Traffic Behavior Based on Game Theory under Road-Pricing Regulation of Government. *Journal of Management in China*, **11**, 76-82.
- [39] Wang, C., Li, Y.J. and Wu, J.B. (2012) Decision-Making Analysis and Empirical Study on Consumers' Shopping Channels Based on Perceptive Utility. *Management Review*, **10**, 85-93.
- [40] Chen, Y., Zhou, W.H., Hua, Z.S. and Dan, G.Y. (2015) Pricing and Capacity Planning of the Referral System with Delay-Sensitive Patients. *Journal of Management Sciences in China*, **18**, 73-81.
- [41] Qi, J.J. (2010) Research on the Allocation Model of Emergency Resources Based on Delay Time. Shandong University, Jinan.
- [42] Tian, Q. and Huang, H.J. (2008) Dynamic Model for Passengers' Equilibrium Commuting in Urban Transit Systems. *Journal of Management in China*, **12**, 1-8.
- [43] Ni, Y. and Zhao, L.D. (2015) Study on Deployment of Emergency Medical Resources with Injury Condition Consideration. *Logistics Technology*, No. 9, 115-118.
- [44] Vargas, H. (1986) The Methods and Findings of Quality Assessment and Monitoring: An Illustrated Analysis. *Journal of Public Health Policy*, **7**, 134.  
<https://doi.org/10.2307/3342133>
- [45] Headley, D.E. and Miller, S.J. (1993) Measuring Service Quality and Its Relationship to Future Consumer Behavior. *Journal of Health Care Marketing*, **13**, 32.
- [46] Dagger, T.S. and Sweeney, J.C. (2007) Service Quality Attribute Weights: How Do Novice and Longer-Term Customers Construct Service Quality Perceptions. *Journal of Service Research*, **10**, 22-42. <https://doi.org/10.1177/1094670507303010>
- [47] Wong, M.D., Andersen, R., Sherbourne, C.D., et al. (2001) Effects of Cost Sharing on Care Seeking and Health Status: Results from the Medical Outcomes Study. *American Journal of Public Health*, **91**, 1889-1894.  
<https://doi.org/10.2105/AJPH.91.11.1889>
- [48] Bansal, H.S. and Taylor, S. (2015) Investigating the Relationship between Service Quality, Satisfaction and Switching Intentions. In: *Proceedings of the 1997 Academy of Marketing Science (AMS) Annual Conference*, Springer International Publishing, Berlin, 304-313. [https://doi.org/10.1007/978-3-319-13141-2\\_107](https://doi.org/10.1007/978-3-319-13141-2_107)
- [49] Mcmanus, M.L., Long, M.C., Cooper, A., et al. (2004) Queuing Theory Accurately Models the Need for Critical Care Resources. *Anesthesiology*, **100**, 1271.  
<https://doi.org/10.1097/00000542-200405000-00032>
- [50] Green, L.V., Kolesar, P.J. and Whitt, W. (2007) Coping with Time-Varying Demand When Setting Staffing Requirements for a Service System. *Production & Operations Management*, **16**, 13-39. <https://doi.org/10.1111/j.1937-5956.2007.tb00164.x>
- [51] Sadat, S. (2013) Theory of Constraints for Publicly Funded Health Systems. *Health Care Management Science*, **16**, 62-74. <https://doi.org/10.1007/s10729-012-9208-9>
- [52] Green, L. (2006) Queueing Analysis in Healthcare. In: Hall, R.W., Ed., *Patient Flow: Reducing Delay in Healthcare Delivery*, Springer, New York, 281-307.  
[https://doi.org/10.1007/978-0-387-33636-7\\_10](https://doi.org/10.1007/978-0-387-33636-7_10)

- [53] Grönroos, C. (1984) A Service Quality Model and Its Marketing Implications. *European Journal of Marketing*, **18**, 36-44. <https://doi.org/10.1108/EUM000000004784>
- [54] Parasuraman, A., Zeithaml, V.A. and Berry, L.L. (1985) A Conceptual Model of Service Quality and Its Implications for Future Research. *Journal of Marketing*, **49**, 41-50. <https://doi.org/10.1177/002224298504900403>
- [55] Thunhurst, C. (2012) Public Health Systems Analysis—The Transfer of Learning between Developed and Developing Countries. *Health Care Management Science*, **15**, 283-291. <https://doi.org/10.1007/s10729-012-9192-0>
- [56] McDougall, G.H.G. and Levesque, T.J. (1995) A Revised View of Service Quality Dimensions. *Journal of Professional Services Marketing*, **11**, 189-210. [https://doi.org/10.1300/J090v11n01\\_13](https://doi.org/10.1300/J090v11n01_13)
- [57] Vany, A.S.D. and Saving, T.R. (1983) The Economics of Quality. *Journal of Political Economy*, **91**, 553-555. <https://doi.org/10.1086/261196>
- [58] Dabholkar, P.A., Shepherd, C.D. and Thorpe, D.I. (2000) A Comprehensive Framework for Service Quality: An Investigation of Critical Conceptual and Measurement Issues through a Longitudinal Study. *Journal of Retailing*, **76**, 139-173. [https://doi.org/10.1016/S0022-4359\(00\)00029-4](https://doi.org/10.1016/S0022-4359(00)00029-4)
- [59] Zhao, M.G. and Jiao, Y.J. (2009) Study on International Medical Quality Management. *Chinese Health Quality Management*, **16**, 1-4.
- [60] Xu, J.Q. and Li, W.C. (2013) Comprehensive Evaluation of Medical Quality Based on Combined Entropy Weight Gray Relational Analysis. *Chinese Journal of Health Statistics*, **30**, 247-248.
- [61] Cheng, F. (2012) Strengthen the Regulatory Role of Government Health Law Enforcement Department for Medical Quality. *Modern Hospital Management*, **10**, 49-51.
- [62] Tan, T.L. and Ou, B.X. (2011) Practice and Experience of the Establishment of Long Term Mechanism of Continuous Quality Improvement. *Chinese Hospital Management*, **31**, 25-27.
- [63] Dong, E.H. (2012) The Development and Studies on Patients' Trust Index System Based Oil Medial Quality Management. Shanghai Jiao Tong University, Shanghai.
- [64] Tian, J. and Niu, Y.M. (2015) Exploration of the Methods of Homogeneity Management of Medical Quality in Medical Alliance. *Chinese Hospital Management*, **35**, 70-72.
- [65] Tu, Z.L. and Wang, Y.G. (2010) Analysis and Countermeasures of Medical Quality Management. *Chinese Health Quality Management*, **17**, 28-30.

**Proof**

**Proposition 1.**

The balanced arrival rate of patients in the tertiary hospital  $H_A$  and community hospital  $H_C$  ( $\lambda_A^*$ ,  $\lambda_C^*$ ) satisfies the following formula:

$$\begin{aligned} q_A - p_A - \frac{h}{\mu_A - \alpha\lambda_C - \lambda_A} - \frac{\alpha}{1-\alpha} p_A (1-\theta) \\ = q_C - p_C - \frac{h}{\mu_C - \lambda_C} - \frac{\alpha}{1-\alpha} \cdot \frac{h}{\mu_C - \lambda_C} \end{aligned} \tag{1-1}$$

To further simplify the formula, assuming that:

$$\begin{aligned} A_0 &= (1-\alpha)(q_C - q_A - p_C) + (1-\alpha\theta) p_A \\ A_1 &= (1-\alpha)h \\ A_2 &= \mu_A - \Lambda \end{aligned}$$

Then:

$$A_0 + \frac{A_1}{A_2 + (1-\alpha)\lambda_C^*} - \frac{A_1}{(1-\alpha)(\mu_C - \lambda_C^*)} = 0 \tag{1-2}$$

Then:

$$\begin{aligned} A_0(1-\alpha)\lambda_C^{*2} + [A_0A_2 + 2A_1 - A_0\mu_C(1-\alpha)]\lambda_C^* \\ + \left( \frac{A_1A_2}{1-\alpha} - A_0A_2\mu_C - A_1\mu_C \right) = 0 \end{aligned} \tag{1-3}$$

$$\lambda_C^* = \frac{A_0\mu_C(1-\alpha) - A_0A_2 - 2A_1 + \sqrt{[A_0A_2 + A_0\mu_C(1-\alpha)]^2 + 4A_1^2}}{2A_0(1-\alpha)}$$

In order to make  $\lambda_C^* \geq 0$ , then:

$$\begin{aligned} A_0\mu_C(1-\alpha) - A_0A_2 - 2A_1 + \sqrt{[A_0A_2 + A_0\mu_C(1-\alpha)]^2 + 4A_1^2} \geq 0 \tag{1-4} \\ \sqrt{[A_0A_2 + A_0\mu_C(1-\alpha)]^2 + 4A_1^2} \geq 2A_1 \end{aligned}$$

Therefore, it only needs to satisfy  $A_0\mu_C(1-\alpha) - A_0A_2 \geq 0$

That is: when  $\mu_C \geq \frac{\mu_A - \Lambda}{1-\alpha}$ , the Formula (1-4) holds, otherwise  $\lambda_C^* = 0$ , at this time:

$$\lambda_A^* = \Lambda - \frac{A_0\mu_C(1-\alpha) - A_0A_2 - 2A_1 + \sqrt{[A_0A_2 + A_0\mu_C(1-\alpha)]^2 + 4A_1^2}}{2A_0(1-\alpha)}$$

$$\lambda_C^* = \frac{A_0\mu_C(1-\alpha) - A_0A_2 - 2A_1 + \sqrt{[A_0A_2 + A_0\mu_C(1-\alpha)]^2 + 4A_1^2}}{2A_0(1-\alpha)}$$

**Proposition 2.**

$$\pi(p_A) = p_A\lambda_A + \theta p_A\alpha\lambda_C - c_1\mu_A - T \tag{2-1}$$

$$U_A = q_A - p_A - \frac{h}{\mu_A - \alpha\lambda_C - \lambda_A}$$

When  $U_A \geq 0$ , the patient will choose the first visit to a tertiary hospital, then:

$$p_A \leq q_A - \frac{h}{\mu_A - \alpha\lambda_C - \lambda_A} \tag{2-2}$$

When the first consultation of all patients in the system chooses to go directly to the tertiary hospital, the revenue of the tertiary hospital will reach the maximum value. At this time, the pricing of the tertiary hospital can also be intentionally lowered because there is no need to worry about no one coming. In other words,  $p_A$  can take the maximum value. Due to  $\lambda_C = 0$ ,  $\lambda_A = \Lambda$ , then, (2-2) becomes:

$$p_A \leq q_A - \frac{h}{\mu_A - \Lambda} \tag{2-3}$$

If you want the hospital to maintain normal operation without loss, the minimum income must be greater than or equal to the total cost of the hospital, that is, the sum of service costs and taxes, can be expressed as:

$$\theta p_A \alpha \Lambda \geq c_1 \mu_A + T \tag{2-4}$$

Then:

$$p_A \geq \frac{c_1 \mu_A + T}{\theta \alpha \Lambda} \tag{2-5}$$

Therefore, by combining (2-3) and (2-5), the value range of  $p_A$  can be obtained:

$$\frac{c_1 \mu_A + T}{\theta \alpha \Lambda} \leq p_A \leq q_A - \frac{h}{\mu_A - \Lambda} \tag{2-6}$$

Therefore, the decision model of the tertiary hospital can be expressed as:

$$\max_{p_A \geq 0} \pi(p_A) = p_A \lambda_A + \theta p_A \alpha \lambda_C - c_1 \mu_A - T \tag{2-7}$$

$$s.t. \frac{c_1 \mu_A + T}{\theta \alpha \Lambda} \leq p_A \leq q_A - \frac{h}{\mu_A - \Lambda} \tag{2-8}$$

Formula (2-7) can be simplified as:

$$\pi(p_A) = \Lambda p_A - (1 - \alpha \theta) \lambda_C p_A - c_1 \mu_A - T \tag{2-9}$$

Already found:

$$\lambda_C^* = \frac{A_0 \mu_C (1 - \alpha) - A_0 A_2 - 2A_1 + \sqrt{[A_0 A_2 + A_0 \mu_C (1 - \alpha)]^2 + 4A_1^2}}{2A_0 (1 - \alpha)}$$

$$A_0 = (1 - \alpha)(q_C - q_A - p_C) + (1 - \alpha \theta) p_A$$

Let:

$$k = (1 - \alpha)(q_C - q_A - p_C)$$

(2-9) becomes:

$$\begin{aligned} \pi(p_A) = & \Lambda p_A - \frac{1 - \alpha \theta}{2} p_A \mu_C + \frac{1 - \alpha \theta}{2} \cdot \frac{A_2 p_A}{1 - \alpha} + \frac{h(1 - \alpha \theta)}{k + (1 - \alpha \theta) p_A} \cdot p_A - c_1 \mu_A \\ & - T - \frac{(1 - \alpha \theta) p_A \sqrt{[k + (1 - \alpha \theta) p_A]^2 \cdot [A_2 + \mu_C (1 - \alpha)]^2 + 4A_1^2}}{2(1 - \alpha)[k + (1 - \alpha \theta) p_A]} \end{aligned} \tag{2-10}$$

Let:

$$A_3 = \frac{1 - \alpha\theta}{2}$$

Then:

$$\pi(p_A) = \Lambda p_A - A_3 \mu_C p_A + \frac{A_2 A_3}{1 - \alpha} p_A + \frac{2A_3 h}{k + 2A_3 p_A} \cdot p_A - c_1 \mu_A - T - \frac{A_3 p_A \sqrt{(k^2 + 4A_3^2 p_A^2 + 4A_3 k p_A) [A_2 + \mu_C (1 - \alpha)]^2 + 4A_1^2}}{(1 - \alpha)(k + 2A_3 p_A)} \quad (2-11)$$

$$\frac{\partial \pi}{\partial p_A} = \Lambda - A_3 \mu_C + \frac{A_2 A_3}{1 - \alpha} + \left( \frac{2A_3 h}{k + 2A_3 p_A} \cdot p_A \right)' - \left\{ \frac{A_3 p_A \sqrt{(k^2 + 4A_3^2 p_A^2 + 4A_3 k p_A) [A_2 + \mu_C (1 - \alpha)]^2 + 4A_1^2}}{(1 - \alpha)(k + 2A_3 p_A)} \right\}'$$

Let:

$$m = \frac{2A_3 h}{k + 2A_3 p_A} \cdot p_A$$

$$n = \frac{A_3 p_A \sqrt{(k^2 + 4A_3^2 p_A^2 + 4A_3 k p_A) [A_2 + \mu_C (1 - \alpha)]^2 + 4A_1^2}}{(1 - \alpha)(k + 2A_3 p_A)}$$

Then:

$$\frac{\partial m}{\partial p_A} = \left( 2A_3 h \cdot \frac{p_A}{k + 2A_3 p_A} \right)' = 2A_3 h \cdot \frac{k + 2A_3 p_A - p_A \cdot 2A_3}{(k + 2A_3 p_A)^2} = \frac{2A_3 h k}{(k + 2A_3 p_A)^2}$$

Let:

$$\sqrt{G} = \sqrt{(k^2 + 4A_3^2 p_A^2 + 4A_3 k p_A) [A_2 + \mu_C (1 - \alpha)]^2 + 4A_1^2}$$

$$\frac{\partial n}{\partial p_A} = \frac{A_3}{1 - \alpha} \cdot \frac{(p_A \sqrt{G})' (k + 2A_3 p_A) - (p_A \sqrt{G}) \cdot 2A_3}{(k + 2A_3 p_A)^2}$$

$$= \frac{A_3}{1 - \alpha} \cdot \frac{[\sqrt{G} + p_A (\sqrt{G})'] \cdot (k + 2A_3 p_A) - 2A_3 p_A \sqrt{G}}{(k + 2A_3 p_A)^2}$$

$$= \frac{A_3}{1 - \alpha} \cdot \frac{k \sqrt{G} + p_A (\sqrt{G})' \cdot (k + 2A_3 p_A)}{(k + 2A_3 p_A)^2}$$

$$= \frac{A_3}{1 - \alpha} \cdot \frac{k \sqrt{G} + (k + 2A_3 p_A) \cdot p_A \cdot \frac{1}{2\sqrt{G}} \cdot [A_2 + \mu_C (1 - \alpha)]^2 \cdot (8A_3^2 p_A + 4A_3 k)}{(k + 2A_3 p_A)^2}$$

$$= \frac{A_3}{1 - \alpha} \cdot \left\{ \frac{k \sqrt{G}}{(k + 2A_3 p_A)^2} + \frac{p_A \cdot [A_2 + \mu_C (1 - \alpha)]^2 \cdot (8A_3^2 p_A + 4A_3 k)}{2(k + 2A_3 p_A) \sqrt{G}} \right\}$$

Then:

$$\frac{\partial \pi}{\partial p_A} = \Lambda - A_3 \mu_C + \frac{A_2 A_3}{1 - \alpha} + \frac{2 A_3 h k}{(k + 2 A_3 p_A)^2} - \frac{A_3}{1 - \alpha} \cdot \left\{ \frac{k \sqrt{G}}{(k + 2 A_3 p_A)^2} + \frac{p_A \cdot [A_2 + \mu_C (1 - \alpha)]^2 \cdot (8 A_3^2 p_A + 4 A_3 k)}{2(k + 2 A_3 p_A) \sqrt{G}} \right\} \quad (2-12)$$

Further derivate the Formula (2-12).

Let:

$$M = \frac{\partial m}{\partial p_A} = \frac{2 A_3 h k}{(k + 2 A_3 p_A)^2}$$

$$N = \frac{\partial n}{\partial p_A} = \frac{A_3}{1 - \alpha} \cdot \left\{ \frac{k \sqrt{G}}{(k + 2 A_3 p_A)^2} + \frac{p_A \cdot [A_2 + \mu_C (1 - \alpha)]^2 \cdot (8 A_3^2 p_A + 4 A_3 k)}{2(k + 2 A_3 p_A) \sqrt{G}} \right\}$$

Then:

$$\frac{\partial^2 \pi}{\partial^2 p_A} = M' - N'$$

And:

$$M' = \frac{\partial M}{\partial p_A} = \frac{-2 A_3 h k (8 A_3 k + 8 A_3^2)}{(k + 2 A_3 p_A)^4} < 0$$

Then:

$$\frac{\partial^2 \pi}{\partial^2 p_A} = M' - N' < 0$$

Therefore, according to the nature of the function,  $\pi(p_A)$  is a concave function about  $p_A$ .

**Proposition 3 & Proposition 4**

$$\begin{aligned} \frac{\partial \pi}{\partial p_A} &= \Lambda - A_3 \mu_C + \frac{A_2 A_3}{1 - \alpha} + \frac{2 A_3 h k}{(k + 2 A_3 p_A)^2} \\ &\quad - \frac{A_3}{1 - \alpha} \cdot \left\{ \frac{k \sqrt{G}}{(k + 2 A_3 p_A)^2} + \frac{p_A \cdot [A_2 + \mu_C (1 - \alpha)]^2 \cdot (8 A_3^2 p_A + 4 A_3 k)}{2(k + 2 A_3 p_A) \sqrt{G}} \right\} \\ &= 0 \end{aligned}$$

We can find a zero boundary point of the function.

Because there are too many variables and parameters involved in this formula, we will optimize the solution during the calculation, otherwise it will affect the final profit function because the formula is too complicated.

Let:

$$\alpha = 0.5$$

$$\theta = 0.5$$

$$h = 1$$

Then:

$$A_0 = \frac{1}{2}(q_C - q_A - p_C) + \frac{3}{4}p_A$$

$$A_1 = \frac{1}{2}$$

$$A_2 = \mu_A - \Lambda$$

$$A_3 = \frac{3}{8}$$

$$k = k_1 = \frac{1}{2}(q_C - q_A - p_C)$$

Then:

$$\begin{aligned} \pi(p_A) = & \Lambda p_A - \frac{3}{8}\mu_C p_A + \frac{3}{16}(\mu_A - \Lambda)p_A + \frac{3}{4} \cdot \frac{p_A}{k_1 + \frac{3}{4}p_A} \\ & - \frac{3}{16} \cdot p_A \cdot \frac{\sqrt{\left(k_1 + \frac{3}{4}p_A\right)^2 \left[\mu_A - \Lambda + \frac{1}{2}\mu_C\right]^2 + 1}}{k_1 + \frac{3}{4}p_A} - c_1 \mu_A - T \end{aligned}$$

Let:

$$\sqrt{G_1} = \sqrt{\left(k_1 + \frac{3}{4}p_A\right)^2 \left[\mu_A - \Lambda + \frac{1}{2}\mu_C\right]^2 + 1} \approx \left(k_1 + \frac{3}{4}p_A\right) \cdot \left(\mu_A - \Lambda + \frac{1}{2}\mu_C\right)$$

Then:

$$\begin{aligned} \frac{\partial \pi}{\partial p_A} = & \Lambda - \frac{3}{8}\mu_C + \frac{3}{16}(\mu_A - \Lambda) + \frac{\frac{3}{4}k}{\left(k_1 + \frac{3}{4}p_A\right)^2} \\ & - \frac{3}{16} \left[ \frac{k_1 \sqrt{G_1}}{\left(k_1 + \frac{3}{4}p_A\right)^2} + \frac{p_A \left(\mu_A - \Lambda + \frac{1}{2}\mu_C\right)^2 \cdot \frac{3}{2} \left(k_1 + \frac{3}{4}p_A\right)}{2 \left(k_1 + \frac{3}{4}p_A\right) \sqrt{G_1}} \right] \end{aligned}$$

When:

$$\begin{aligned} k_2 = & \mu_A - \Lambda + \frac{1}{2}\mu_C \\ \beta = & k_1 + \frac{3}{4}p_A \quad \left( p_A = \frac{4}{3}(\beta - k_1) \right) \end{aligned}$$

Then:

$$\frac{\partial \pi}{\partial p_A} = \Lambda - \frac{3}{8}\mu_C + \frac{3}{16}(\mu_A - \Lambda) + \frac{3}{4} \frac{k_1}{\beta^2} - \frac{3}{16} \cdot \frac{k_1 k_2}{\beta} - \frac{3}{16} \cdot (\beta - k_1) \cdot \frac{k_2}{\beta}$$

If  $\frac{\partial \pi}{\partial p_A} = 0$ , then:

$$\Lambda - \frac{3}{8}\mu_C + \frac{3}{16}(\mu_A - \Lambda) + \frac{3}{4} \frac{k_1}{\beta^2} - \frac{3}{16} \cdot \frac{k_1 k_2}{\beta} - \frac{3}{16} \cdot (\beta - k_1) \cdot \frac{k_2}{\beta} = 0$$



And:

$$\beta^* = \sqrt{\frac{12k_1}{6\mu_C - 13\Lambda - 3\mu_A + 3k_2}} = \sqrt{\frac{12k_1}{\frac{15}{2}\mu_C - 16\Lambda}} = \sqrt{\frac{24k_1}{15\mu_C - 32\Lambda}}$$

Because of  $p_A = \frac{4}{3}(\beta - k_1)$

$$p_A^* = \frac{4}{3} \cdot \left( \sqrt{\frac{24k_1}{15\mu_C - 32\Lambda}} - k_1 \right)$$

Then:

$$\begin{aligned} \lambda_C^* &= \frac{A_0\mu_C(1-\alpha) - A_0A_2 - 2A_1 + \sqrt{[A_0A_2 + A_0\mu_C(1-\alpha)]^2 + 4A_1^2}}{2A_0(1-\alpha)} \\ &\approx \mu_C - \frac{1}{A_0} = \mu_C - \frac{1}{k_1 + \frac{3}{4}p_A^*} \end{aligned}$$

If  $p_A = \frac{4}{3} \cdot \left( \sqrt{\frac{24k_1}{15\mu_C - 32\Lambda}} - k_1 \right)$

$$\begin{aligned} \pi(p_A^*) &= p_A^*\lambda_A + \theta p_A^*\alpha\lambda_C^* - c_1\mu_A - T \\ &= p_A^*\left(\Lambda - \frac{3}{4}\lambda_C^*\right) - c_1\mu_A - T \\ &= p_A^*\left(\Lambda - \frac{3}{4}\mu_C + \frac{3}{4k_1 + 3p_A^*}\right) - c_1\mu_A - T \\ &= \left(\Lambda - \frac{3}{4}\mu_C\right)\left(\sqrt{\frac{24k_1}{15\mu_C - 32\Lambda}} - k_1\right) + \frac{3\left(\sqrt{\frac{24k_1}{15\mu_C - 32\Lambda}} - k_1\right)}{k_1 + 3\sqrt{\frac{24k_1}{15\mu_C - 32\Lambda}}} - c_1\mu_A - T \end{aligned}$$

It can be known from the above solution process that the value of  $k_1$  changes with the change of  $\alpha$ . Therefore, when  $k_1$  is reduced to  $k$  and when

$\frac{\partial \pi}{\partial p_A} = 0$ ,  $p_A^*$  is:

$$p_A^* = \frac{4}{3} \cdot \left( \sqrt{\frac{24(1-\alpha)(q_C - q_A - p_C)}{15\mu_C - 32\Lambda}} - (1-\alpha)(q_C - q_A - p_C) \right)$$

Then:

$$\begin{aligned} \pi(p_A^*) &= \left(\Lambda - \frac{3}{4}\mu_C\right)\left(\sqrt{\frac{24(1-\alpha)(q_C - q_A - p_C)}{15\mu_C - 32\Lambda}} - (1-\alpha)(q_C - q_A - p_C)\right) \\ &\quad + \frac{3\sqrt{\frac{24(1-\alpha)(q_C - q_A - p_C)}{15\mu_C - 32\Lambda}} - 3(1-\alpha)(q_C - q_A - p_C)}{(1-\alpha)(q_C - q_A - p_C) + 3\sqrt{\frac{24(1-\alpha)(q_C - q_A - p_C)}{15\mu_C - 32\Lambda}}} - c_1\mu_A - T \end{aligned}$$

To ensure that  $p_A^*$  exists, it must satisfy:

$$\frac{24(1-\alpha)(q_C - q_A - p_C)}{15\mu_C - 32\Lambda} \geq 0$$

The previous article has been obtained,  $0 \leq \alpha \leq 1$ ,  $q_C \leq q_A$ ,  $q_C - q_A - p_C < 0$ , therefore, to satisfy (5-18), then  $15\mu_C - 32\Lambda < 0$ , that is:

$$\mu_C < \frac{32}{15}\Lambda$$

If and only if  $\mu_C \geq \frac{\mu_A - \Lambda}{1 - \alpha}$ ,  $H_C$  will have patients come to the clinic:

$$\begin{aligned} \frac{\mu_A - \Lambda}{1 - \alpha} &< \frac{32}{15}\Lambda \\ \Lambda &< \mu_A < \left(\frac{47}{15} - \frac{32}{15}\alpha\right) \cdot \Lambda, \quad \frac{\mu_A - \Lambda}{1 - \alpha} \leq \mu_C < \frac{32}{15}\Lambda \\ \frac{c_1\mu_A + T}{\theta\alpha\Lambda} &\leq p_A \leq q_A - \frac{h}{\mu_A - \Lambda} \end{aligned}$$

**Lemma 2.**

When  $B \geq (c_2 - p_C) \frac{\mu_A - \Lambda}{1 - \alpha}$ , the optimal service capability of  $H_C$  is:

$$\mu_C^* = \frac{B}{c_2 - p_C}$$

Then:

$$\frac{\partial \mu_C^*}{\partial B} > 0, \quad \frac{\partial \mu_C^*}{\partial c_2} < 0$$

**Lemma 3.**

When

$$\begin{aligned} \frac{\mu_A - \Lambda}{1 - \alpha} &\leq \mu_C < \frac{32}{15}\Lambda, \\ (c_2 - p_C) \cdot \frac{\mu_A - \Lambda}{1 - \alpha} &\leq B < (c_2 - p_C) \cdot \frac{32}{15}\Lambda \\ p_A &= q_A - \frac{h}{\mu_A - \Lambda} \end{aligned}$$

Then:

$$\frac{\partial p_A}{\partial q_A} > 0, \quad \frac{\partial p_A}{\partial \mu_A} > 0, \quad \frac{\partial p_A}{\partial \Lambda} < 0$$

**Lemma 4.**

When  $p_A = p_A^*$

$$p_A = p_A^* = \frac{4}{3} \cdot \left( \sqrt{\frac{24(1-\alpha)(q_C - q_A - p_C)}{15\mu_C^* - 32\Lambda}} - (1-\alpha)(q_C - q_A - p_C) \right)$$

$$\frac{\partial p_A^*}{\partial \mu_C^*} = -10\mu_C^{\frac{-3}{2}} < 0$$

$$\frac{\partial \mu_C^*}{\partial B} > 0$$

Then:

$$\frac{\partial p_A^*}{\partial B} = \frac{\partial p_A^*}{\partial \mu_C^*} \cdot \frac{\partial \mu_C^*}{\partial B} < 0$$

**Lemma 5.**

Because

$$\lambda_C^* = \mu_C - \frac{1}{k + \frac{3}{4} p_A^*}$$

We can find that:

$$\frac{\partial \lambda_C^*}{\partial p_A^*} > 0$$

**Lemma 6.**

$$\lambda_C^* = \mu_C - \frac{1}{k + \frac{3}{4} p_A^*}$$

Then

$$p_A^* = \frac{4}{3} \cdot \left( \sqrt{\frac{24(1-\alpha)(q_C - q_A - p_C)}{15\mu_C - 32\Lambda}} - (1-\alpha)(q_C - q_A - p_C) \right)$$

$$\lambda_C^* = \mu_C - \frac{1}{\sqrt{\frac{24(1-\alpha)(q_C - q_A - p_C)}{15\mu_C - 32\Lambda}}}$$

$$\frac{\partial \lambda_C^*}{\partial \alpha} < 0$$

Therefore,  $\lambda_C^*$  decreases as  $\alpha$  increases.

Due to

$$\alpha = 1 - \frac{q_C}{q_A},$$

$$\frac{\partial \alpha}{\partial q_C} < 0$$

Then:

$$\frac{\partial \lambda_C^*}{\partial q_C} = \frac{\partial \lambda_C^*}{\partial \alpha} \cdot \frac{\partial \alpha}{\partial q_C} > 0$$

**Lemma 7.**

The waiting costs for patients are:

$$h_A = \frac{h}{\mu_A - \left(1 - \frac{q_C}{q_A}\right) \lambda_C - \lambda_A}$$

$$h_C = \frac{h}{\mu_C - \lambda_C} + \left(1 - \frac{q_C}{q_A}\right) \cdot \frac{h}{\mu_A - \left(1 - \frac{q_C}{q_A}\right) \lambda_C - \lambda_A}$$

$$\lambda_C + \lambda_A = 1$$

Then:

$$h_A = \frac{h}{\mu_A - 1 + (1 - \alpha) \cdot \left[ \mu_C - \frac{1}{\sqrt{\frac{24(1 - \alpha)(q_C - q_A - p_C)}{15\mu_C - 32\Lambda}}} \right]}$$

$$h_C = h \cdot \sqrt{\frac{24(1 - \alpha)(q_C - q_A - p_C)}{15\mu_C - 32\Lambda}}$$

$$+ \alpha \cdot \frac{h}{\mu_A - 1 + (1 - \alpha) \cdot \left[ \mu_C - \frac{1}{\sqrt{\frac{24(1 - \alpha)(q_C - q_A - p_C)}{15\mu_C - 32\Lambda}}} \right]}$$

Can be calculated by:

$$\frac{\partial h_A}{\partial \alpha} > 0, \quad \frac{\partial h_C}{\partial \alpha} > 0$$

Due to  $\frac{\partial \alpha}{\partial q_C} < 0$ , then:

$$\frac{\partial h_A}{\partial q_C} < 0, \quad \frac{\partial h_C}{\partial q_C} < 0$$