1-1-2015

Modeling And Optimization Of Non-Profit Hospital Call Centers With Service Blending

Yanli Zhao
Wayne State University,

MODELING AND OPTIMIZATION OF NON-PROFIT HOSPITAL CALL CENTERS WITH SERVICE BLENDING

by

YANLI ZHAO

DISSERTATION

Submitted to the Graduate School

of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

2015

MAJOR: INDUSTRIAL ENGINEERING

Approved By:

____

Advisor Date

____

____

____

DEDICATION

To my parents,

XUEYAN ZHAO and GUIHUA FENG

for the love and support

ACKNOWLEDGMENTS

I would like to express the deepest appreciation to my advisor, Professor Kai Yang. His inspirational instruction and guidance have been the greatest supports throughout my study; his persistent encouragements and help gave me the strength to overcome the difficulties in dissertation work. It is a pleasurable learning experience and I have benefited a lot from his advice.

I would also like to acknowledge my other committee members. I am particularly grateful to Ms. Susan Yu for offering me the opportunity to observe the practical operations of hospital call centers and taking time from her busy schedule to give me invaluable assistance. My sincere gratitude goes to Dr. Kyoung-Yun Kim for the suggestions on my research. My special appreciation goes to Dr. Qingyu Yang for the tremendous help in learning simulation tools and constant support for my research.

I would like to thank Jihan Wang, Xiaoyu Ma and many other team members and friends for their great company and help during my study.

I also wish to thank my family, a great source of inspiration and energy for me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

The national healthcare expenditure increased fast in recent years and will increase at a higher rate in the following years. The healthcare expenditure as a percent of gross domestic product (GDP) has already approach 18% in 2010. Figure 1.1 shows the trends of national health expenditure, which indicates the great importance of healthcare industry. It will require the design and implementation of novel healthcare systems which can more effectively and efficiently deliver high-quality care while controlling costs.

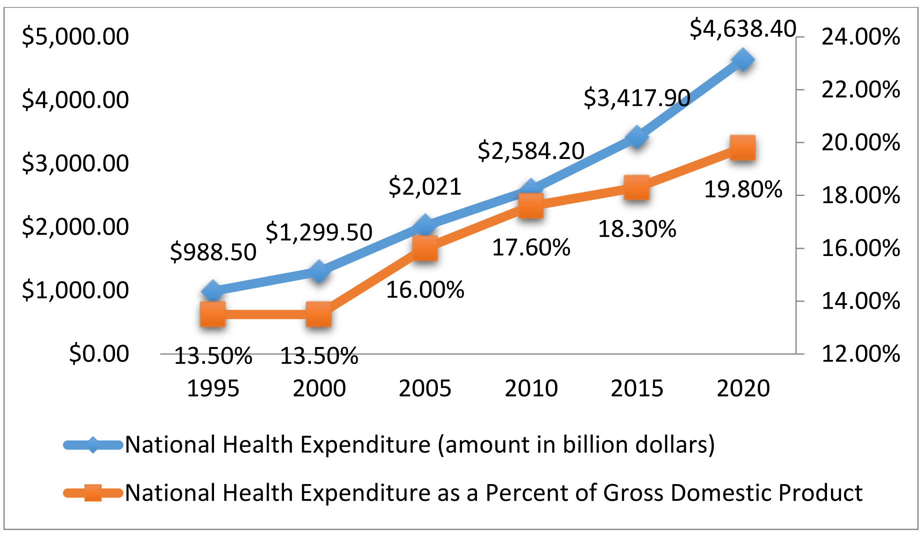

Figure 1.1: Trends of national health expenditure (CMS 1995, 2010)

Call centers are important service channels that organizations provide services to customers through telephones. In recent decades, the falling costs of tele-communications and information technology have made it increasingly economical to interact between service providers and their customers via call center service. A hospital call center functions as an important link between patients and hospitals. From a patient's perspective, telephone is one of the most commonly used alternative venues through which patients seek healthcare needs beyond in-person visits. From the hospital's perspective, call center will well coordinate the care of patients and balance a difficult set of challenges. Hospital call centers are playing an increasingly important role in improving healthcare service quality and decreasing healthcare cost, especially for non-profit hospitals. Figure 1.2 shows the comparison of for-profit hospital and non-profit hospital. In for-profit hospital, patients need to pay for the medical service. So for-profit hospital can increase benefit from increasing quantity of patients and clinic appointments. While patients in non-profit hospitals are specific groups and usually the quantity of patients is constant. Non-profit hospitals don't charge money from patients. So how to saving cost is especially important. Call centers can delivery multiple services, which will decrease the clinic workload effectively. So non-profit hospital call centers are important to decrease cost while improve the service quality.

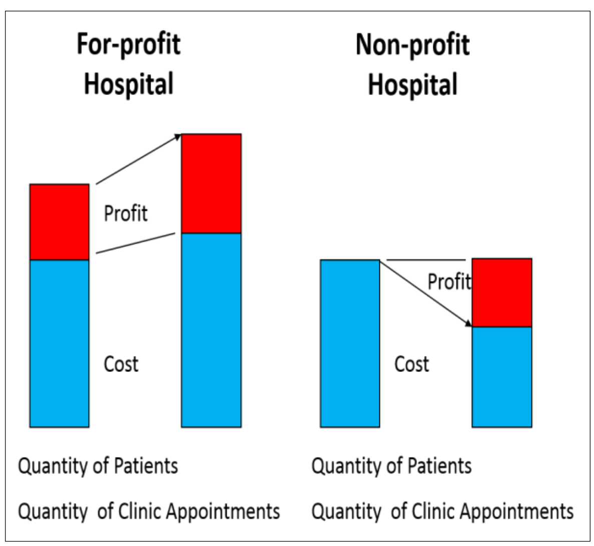

Figure 1.2: Comparison of for-profit hospital and non-profit hospital

A call center has the ability to handle a considerable volume of calls at the same time. Calls may be inbound or outbound, depending on the character of the organization. Inbound calls are initiated by customers. In contrast, outbound calls are initiated by operators. A multi-skill call center in which inbound and outbound calls are combined is known as a "blended" call center. The objectivity of this thesis is modeling and optimization of hospital call centers with service blending.

1.1 Background

In the past decade, there have been many researchers developing new methods to carry out healthcare services totally or partly through telephone in many points of view. Figure 1.3 shows the multiple services of modern hospital call center. The services are delivered through inbound or outbound calls. Inbound calls are initiated by patients while outbound calls are initiated by call center operators. The results indicated strong potential for the use of call center.

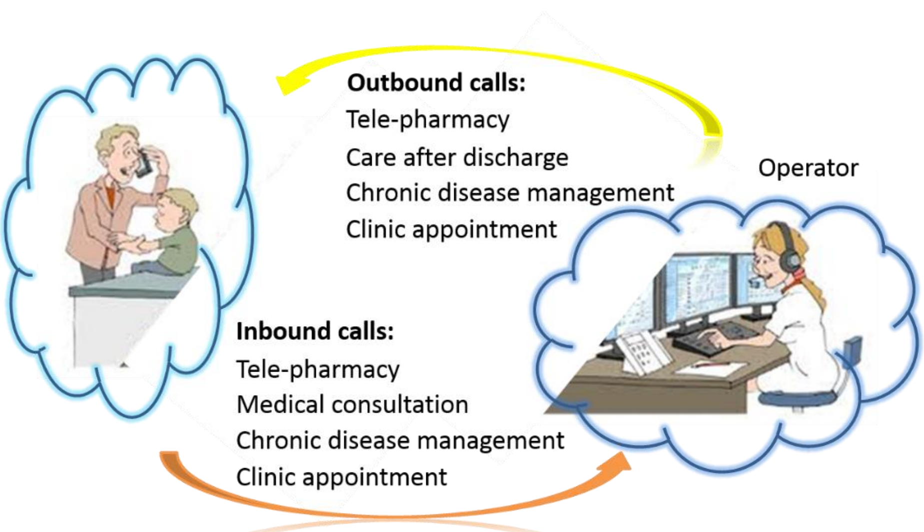

Figure 1.3: Multiple services of modern hospital call centers

1) Appointment Scheduling

Hospital delivers appointment scheduling service to patients through call centers with inbound and outbound calls. Patients call the hospital to scheduling, canceling, changing appointments, for primary care or specialty care, and for lab test and procedures. Hospital calls patients one or two days before patients' clinic appointments to reminder them of their appointments, which is effective to reduce patient no-show rate (Gariti, et al. (1995); Haynes and Sweeney (2006)).

2) Tele-pharmacy

Tele-pharmacy is an innovative solutions using call center to solve the problem of access to pharmacy services. Patients call the hospital for "pharmacy-related issues", refills, renewals, but also medical questions, including side effects, interference between two drugs and efficacy. Tele-pharmacy brings hospital great benefit in avoided clinic cost (Stubbings, et al. (2005); Grabowski and O'Malley (2014)).

3) Medical situation consultation

Medical situation calls are inbound calls. When patients are feeling sick, they may need advice before they want to come in; when patients want results of tests or related medical insurance information, they can get solution through call center.

4) Care after discharge

Hospital readmissions are very frequent. Care after discharge through call center will be an innovative method to decrease hospital readmissions and hospital costs in comparison with the usual care (UC) follow-up program. Every patient who has been discharged from Hospital for 2 to 3 days is tracked through telephone and offered comprehensive care coordination intervention, including medication reconciliation and provision, medical supplies/equipment provision, referral for home health service, patient and caregiver education on post discharge monitoring /follow-ups, and communication of the evaluation findings to primary care providers. The goal of this service is to facilitate seamless transitions between healthcare settings to promote patient safety and reduce unnecessary health care utilization. Result of related projects shows it is necessary to care out the care after discharge via telephone methods (Bostrom, et al. (1996), Kirsch, et al. (2014)).

5) Chronic disease management

Chronic disease is the leading cause of disability and seeking health care, especially for the aged. For chronic disease management, hospital uses telephone to avoid patients' unnecessary admission to institutional care, helping patients self-manage their condition, assessing their bio psychosocial needs and instituting care/case management (Darkins, et al. (2008); Giordano, et al. (2009)). Chronic disease management calls include inbound and outbound calls.

6) Care for some special diseases

The hospital call center also has great potential in care for some special diseases using inbound and outbound calls (van Zyl, et al. (2015)).

1.2 Characteristics of hospital Call Centers

Hospital deliveries increasing kinds of services through call center with inbound and outbound calls. There are different characteristics for inbound and outbound calls. The inbound calls are initialed by patients, so inbound calls are random and have higher priority; while the outbound calls are initialed by hospital call center which means that outbound calls are deterministic for call center and can be planned.

In this research, all the data used to describe the characteristics of non-profit hospital call center are from a sample non-profit hospital call center. Figure 1.4 shows the trend of inbound call volume. Before 2010, the annual inbound call volume increased slowly. From 2010 to 2011, there was a great increase of the annual inbound call volume, which indicated there was a great increase of service offered through hospital call center. Here the scale of hospital call center increases accordingly. It is important to improve the hospital call center's operation. With optimized operation, the hospital call center will effectively play an important role in increasing hospital service for patients and decreasing hospital cost.

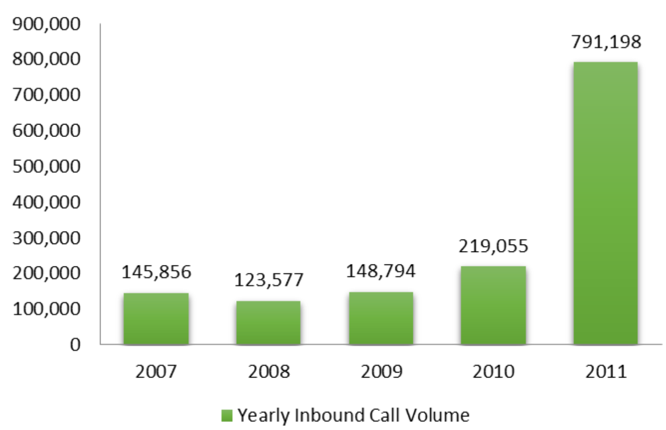

Figure 1.4: Trend of inbound call volume

Figure 1.5 shows the daily inbound calls volume data from a service line of a non-profit hospital call center. The data is for the 36-week period from Feb 28, 2011 to Nov 4, 2011, inclusive. Each week includes five days from Monday to Friday. Six special days are included, two days with missing data due to telephone system and four days as public holidays. The series exhibits no apparent trend and very clear intraweek seasonality: In each week with five working days, Monday is the busiest day with the highest volume of inbound calls. Then the inbound calls volume decreases gradually from Monday to Friday.

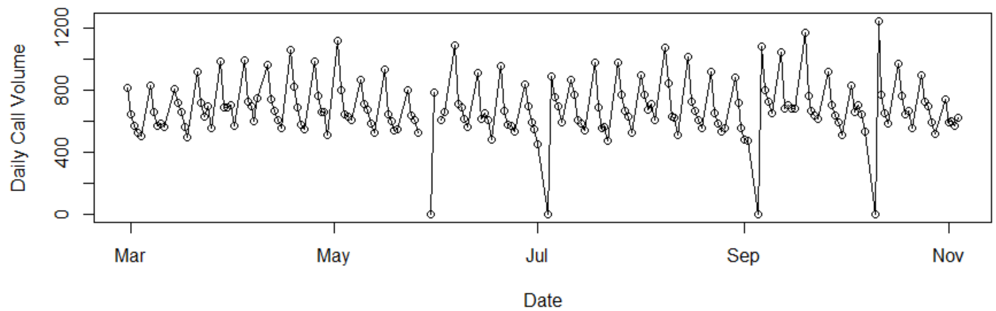

Figure 1.5: Daily Inbound Calls Volume

Figure 1.6 shows the average half-hour call volume data for the same service line. The call center working time is 8:00am-4:30pm, which covers 17 half-hour time intervals. The inbound calls volume varies in each time interval. From 8:00 am, the inbound call arrivals first decreases slightly in around 30 minutes and then increase sharply from 9:00am to 11:00am. After that, the volume decreases linearly as time progresses until 1:00pm. Then the volume slightly increase again until 2:30pm. At around 4:00pm, when the work day is going to end, a sudden decrease in the call volume is shown. There are two peak levels every day, morning peak around 11:00am and afternoon peak around 2:30pm. From Monday to Friday, the inbound call volume shows intraday seasonality.

The large variation of inbound calls cause large variation of needed operators. While the outbound calls are usually planned to be delivered in a certain time period. Operators working for outbound calls can be scheduled with high occupancy rate. In hospital call centers, most of the operators are nurses and full-time employees, which means the quantity of operators is constant. How to assign inbound calls and outbound calls for hospital call center will be an important issue to optimize the occupancy of operators and improve the service quality.

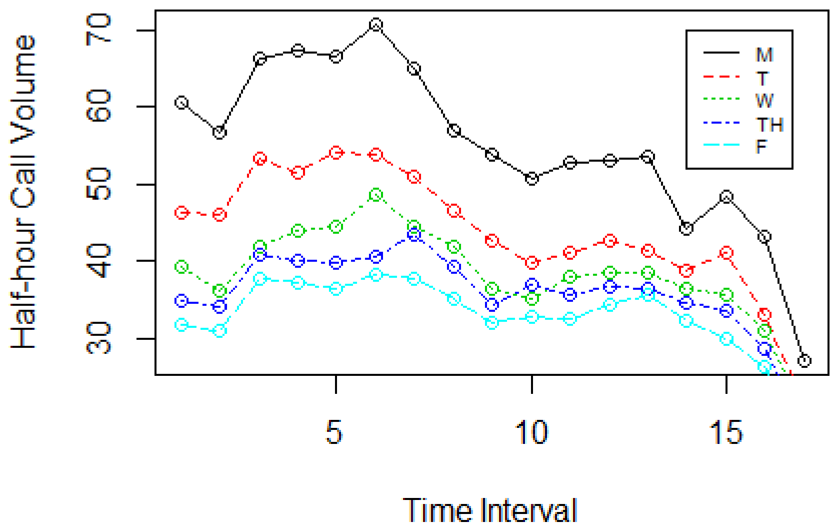

Figure 1.6: Average Half-hour Inbound Calls Volume

There are several performance measures for call centers. Table 1.1 lists the general metrics. There are two categories metrics, product-related metrics and process-related metrics. The product-related metrics includes first call resolution, which is for inbound calls only, and customer satisfaction, for the contents of inbound and outbound calls. The solution to improve product-related service is operators training. The process-related metrics includes several metrics for inbound calls, like probability of blocking, probability of abandonment and service level, which indicate the percentage of calls answered within a specified time limit. This research uses a special process-related metric for outbound calls, on time delivery, which means the outbound calls should be completed in the schedule time interval.

Table 1.1: Call center performance measure

<table>
<tr><th></th><th>Metrics</th><th>Call type</th><th>Solution</th></tr>
<tr><td rowspan="2">Product-related</td><td>First call resolution</td><td>Inbound calls</td><td rowspan="2">Operators training</td></tr>
<tr><td>Customer satisfaction</td><td>Inbound and outbound calls</td></tr>
<tr><td rowspan="4">Process-related</td><td>Probability of blocking</td><td rowspan="3">Inbound calls</td><td rowspan="4">Process optimization</td></tr>
<tr><td>Probability of abandonment</td></tr>
<tr><td>Service level</td></tr>
<tr><td>On time delivery</td><td>Outbound calls</td></tr>
</table>

This research focuses on the process-related metrics. The telecommunication system of the sample hospital call center is able to accommodate enough patients waiting on line so we ignore the probability of blocking metric. The probability of abandonment metric is low compared to call centers in some other industry fields because in hospital call centers, the patient call to get service that they must have. So the probability of abandonment is also ignored. Service level is considered to measure the inbound calls service quality and on time delivery is used to measure the outbound calls service quality in this research.

1.3 Research Objectives

The increasing kinds of service undoubtedly create bottlenecks for the call center operations. Gans, et al. (2003) and Aksin, et al. (2007) summarize comprehensive surveys of call center operation problems. However, most current existing researches focus only on inbound-only call centers. Researches focusing on service-blending call center are rare.

Hence, the novelty of this study is that it provided a generic framework to design a blending call center for hospitals, with the effort to understand the needs of hospital call centers, to facilitate the health care activities arising from patient's requirements, convenience and satisfaction as well as hospital's service quality and cost saving.

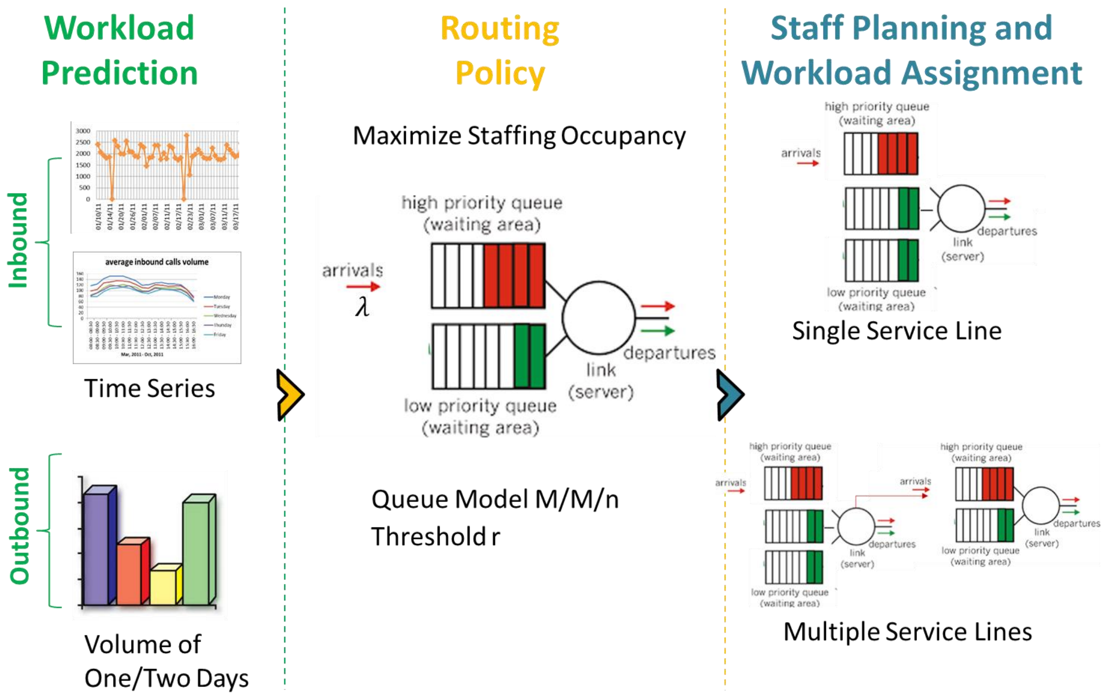

Figure 1.7: Figure Call Center Modeling Stages

Figure 1.7 describes the call center modeling stages. Three stages are included. The first stage is workload prediction. The workload of call centers includes inbound calls workload and outbound calls workload. The inbound calls data are time series data and multiple kinds of methods are studied to predict the inbound calls arrival rate. The outbound calls data are estimated. The second stage is making routing policy for inbound and outbound calls. The third stage is developing model for staffing shift planning and workload assignment.

In this research, we do not focus on the inbound calls workload prediction. We make the assumption that the inbound calls arrival rate can be predicted and the prediction accuracy satisfies the modeling requirements.

The research objective includes three aspects: first, effect inbound and outbound calls routing policies were developed to meet the service quality and improve operators occupancy. Second, based on the calls routing policies, efficient staffing assignment algorithm for inbound and outbound calls were considered. Third, staffing assignment algorithm were adjusted for multiple service lines and multi-skill operators to satisfy the hospitals' need.

1.4 Structure of the Thesis

The reminder of this thesis is organized as follows.

In Chapter 2 the routing policy is considered for a hospital call center with constant number of full-time operators and two kinds of calls, inbound calls and outbound calls. The objective is to improve the system utilization under service quality constraints. This research combines staffing planning, staffing assignment and calls routing. A collection of practical staffing assignment methods, separating staffing policy and mixing staffing policy for different calls are presented and evaluated. Erlang C queuing model is used as a start point to decide the minimum number of required operators and the base for separating staffing policy. Theoretical analysis and numerical experiments illustrate that due to the inbound calls randomness, separate staffing policy just satisfies the inbound calls service quality but fails to take the system utilization into account. Through dynamically assigning the inbound and outbound calls to operators under optimal threshold policy, mixing staffing policy is efficient to balance the system utilization and service quality. Numerical experiments based on real-life data demonstrate how this method can be applied in practice. Chapter 2 is based on Zhao, et al. (2014)

Chapter 3 focus on staffing shift planning model based on the inbound and outbound calls routing policies. Staffing shift planning is an important issue in industry operation, especially for call centers, in which the labor cost constitutes most of the operating cost. A mathematical programming model is developed, based on a hospital call center with one kind of inbound calls and multiple kinds of outbound calls. The objective is to minimize the staffing numbers, by deciding the shift setting and workload allocation. The inbound calls service level and staffing utilization are taken into consideration in the constraints. Numerical experiments based on actual operational data are included. Results show that the model is effective to optimize the shift planning and hence reduce the call centers' cost.

Chapter 4 considers the staffing shift planning problem in a more complex scenario compared to chapter 3. Recently, increasing kinds of service are delivered through call centers so most of call centers have been converted from single skill call centers to multi-skill call centers. Accordingly multi-skill call center staffing shift planning is an increasingly important issue in call centers operations since the labor cost constitutes most of the operating cost. The multi-skill staffing shift planning problem is considered for a hospital call center with two kinds of service lines. Each kind of service is delivered through both inbound calls and outbound calls. Operators on different service line have different labor cost. The inbound calls can be transferred between these two service lines based on the patients' questions. A mathematical programming model is developed. The objective is to minimize the staffing cost, by deciding the shift setting and outbound calls workload allocation. The inbound calls service level and staffing utilization are taken into consideration in the constraints. Numerical experiments are carried out based on actual operational data. Results show that the model is effective to reduce the call centers' labor cost.

CHAPTER 2: EFFICIENT STAFFING FOR HOSPITAL CALL CENTERS WITH SERVICE BLENDING

2.1 Introduction

Call center staffing problem is an important issue in call center operation. Usually up to 60% - 70% of call center operating expenses are capacity cost, which is used to hire the customer service representatives (Gans, et al. (2003)). In this paper, operator is used to name the customer service representative. Hospital call centers provide medical related service to customers. Their operators need have professional medical background, which makes the staffing cost higher than common call centers.

In the very beginning, hospital call centers mainly process inbound calls, which are initialed by patients. They serve patients passively through waiting for patient's calls and offering answers to patients' spontaneous questions. While in recent decades, the falling costs of telecommunications and information technology make it increasingly economical to interact between hospitals and patients via call center service. More and more medical related services are integrated into hospital call centers gradually, such as care after discharge (Bostrom, et al. (1996)), chronic disease management (Bowles, et al. (2009)), etc. Some services need hospital call centers initial outbound calls to contact patients actively. These services increase hospital's benefit and accordingly the hospital call centers become inbound and outbound service blending call centers.

The hospital call centers differ from most other call centers in two aspects. First, the staffing levels are nearly constant. Because of the requirement of medical knowledge background, hospital call centers don't use flexible or outsourcing staffs as some other industry call centers do (Ren and Zhou (2008), Roubos, et al. (2011)). Second, the quantity of outbound calls is comparative to the quantity of inbound calls, not like just a portion of inbound calls as “callback” to e-mail calls or voice message (Koole and Mandelbaum (2002), Armony and Maglaras (2004)). Since new medical services are carried out gradually; we can even consider the outbound calls are infinite in the limited working hours.

The exclusive characteristics of hospital call centers with inbound and outbound calls blending require new staffing methods to optimize the call center operation. The purpose of this chapter is to develop a staffing model for hospital call centers. This model is based on the distinguishing characteristics of hospital call centers and will be more effective to increase the profit of hospital call centers. The model approach has been applied to realistic hospital call centers data. The results show great improvement of hospital call centers performance, which can be seen important insights for hospital call centers staffing problems.

The reminder of this chapter is organized as follows: In section 2.2, a brief literature review related to staffing problems in call centers with calls blending is given. In section 2.3, the call center model is described and the staffing problem is formulated. In section 2.4, the solution approach is introduced. In section 2.5, numerical experiments are used to evaluate the solution approach. Finally, concluding remarks and further research are summarized in section 2.6.

2.2 Literature Review

Over recent years some authors consider multiple types of call centers with inbound and outbound calls blending. Gans, et al. (2003) and Aksin, et al. (2007) provide comprehensive summaries of the state of call center research pertaining to inbound and outbound calls blending. Staffing for blending call centers is more complex since it should consider the inbound and outbound calls simultaneously. Pang and Perry (2014) propose a logarithmic safety staffing rule. Nah and Kim (2013) develop a single mathematical programming model for staffing in call centers with service blending.

We review the literatures through two aspects, which are the main focus of staffing problem in our model. First, how to decide the minimum operators constrained by the inbound calls service level. Second, how to assign the inbound and outbound calls to operators to improve the occupancy.

The inbound calls need real-time service and have higher urgency priority than outbound calls. Thus the first staffing issue should be considered is how to satisfy the inbound calls service requirements. Erlang C (M/M/C) queuing model assumes that all delayed customers wait until they get service (Erlang (1909)). Erlang C system can be used to compute the minimum number of customer services representatives given the expected service level constraint, which is a widely used indicator of the inbound calls service quality measuring the percentage of calls answered within a specified time limit (Jouini, et al. (2013)).

In the real practice, inbound calls arrival rate varies all the time. The most common way is dividing the whole operation time into several short intervals, 15 minutes, 30 minutes or 60 minutes (Aktekin and Soyer (2011), Taylor (2012) , Ibrahim and L'Ecuyer (2013), Goldberg, et al. (2014)). In each short time interval, it is assumed that the arrival rate is constant and the process is considered as Erlang C queuing model.

Accordingly, a series of stationary Erlang C queuing models are set up and the minimum number of operators for each time interval is solved independently. Green, et al. (2001) name this method as stationary independent period by period (SIPP) approach. Some researchers consider that SIPP is a good start to calculate minimum customer service representatives but still not very accurate. Kim and Ha (2010) point out that the SIPP approach does not consider incomplete calls at the end of each time interval; they propose a consecutive staffing using simulation approach. Roubos, et al. (2012) consider the large service-level variability using traditional SIPP approach and develop accurate approximations of service-level distribution through extensive simulation. In this research, the focus is how to improve the capacity utilization under SIPP approach.

Outbound calls in hospital call centers are independent from inbound calls. Computer-based outbound telephone-dialing systems decide whom to call and when to do it, dial automatically and connect the operators and customers, while logging other results (Samuelson (1999)). The less urgency of outbound calls makes it possible to plan whom to call and when to start the call to optimize system performance. Two operational issues generate accordingly, which we call them staffing assignment and mix calls routing problem. The first issue is the assignment of the operators to inbound and outbound calls. Deslauriers, et al. (2007) propose five continuous-time Markov chain (CTMC) models to optimize the blending of inbound-only operators (just process inbound calls) and blend operators (process inbound and outbound calls). The second issue is the mix of inbound and outbound calls. Researchers propose a threshold policy, which assigns outbound calls to operators if the number of available operators exceeds a certain threshold. It has been widely used in the routing problem of inbound and outbound calls (Mehrotra and Fama (2003), Pichitlamken, et al. (2003), Armony and Maglaras (2004)).

This research establishes a framework for real hospital call center optimization. Compared to the related literature, our model is more appropriate for hospital call centers. The main contributions of this paper are the following aspects:

First, the call centers we consider have special features unlike the previous call center models in literature. In hospital context, the call center size is small or medium; the use of flexible operators is rare and operators' turnover rate is low. The staffing quantity can be seen constant. Multiple kinds of medical service can be delivered through outbound calls. So the more outbound calls they make, the more potential benefit the hospital will get, especially for non-profit hospitals.

Second, many former studies of SIPP approach consider the improvement of staffing accuracy under the expected service level constraint for inbound calls. While we consider the staffing problem with the target of finding a balance between service quality and efficiency.

Third, the distinguishing feature of our model is that it shows the way to optimize the threshold value. The threshold value are derived through extensive simulation based on the system parameters.

2.3 Model

We consider a call center with a single type of homogenous operators and two types of customers, inbound and outbound. The number of operators is constant, which is defined as C_{total}. Each operator has ability to process inbound calls and outbound calls. Their ability to process inbound calls is identical. Their ability to process outbound calls is also identical. In order to decrease the staffing cost, we should decide the minimum operators, which is indicated by C_{min}.

We model the inbound process using Erlang C (M/M/C) queuing system. Inbound customers arrive according to a Poisson process with the arrival rate λ. The service times of inbound calls are assumed to be negative exponential distribution with service rate μ_{in}^{-1}. The number of servers (operators) for inbound calls is C_{in}. The inbound calls abandon rate in hospital call centers is low, so we ignore the abandonment in this model.

Inbound customers arriving at the call center will be served immediately if there are operators available. If all the operators are busy, they will wait in a queue for the next available operator. Customers in queue will be served according to first-in first-out rule. W_q is the time that a customer spends waiting in the queue before gets served. However, a waiting customer has a finite patience, which is defined as the acceptable waiting time τ. If the real waiting time exceeds τ, customer may give negative feedback to call center, dissatisfied with the call center's service quality or even abandon the call, which will result in lost to the call center. The service level (SL) is defined as the probability that customers' waiting time is less than the acceptable waiting time, which is shown in the following formula.

$$\mathrm{SL} = \mathrm{P}(W_q \leq \tau)$$

SL is an important metric to measure the inbound call center service quality. In industry the SL has the default standard 80/20. It means that at least 80% of the customers should wait no longer than 20 seconds. We will use this standard as the constraint of the inbound calls service in our model.

The outbound calls are deterministic since they are initialed by the call center. The service times of outbound calls are assumed to be negative exponential distribution with service rate μ_{out}^{-1}. It is assumed that the outbound calls are infinite, so an available operator can always serve an outbound call if desired. The total number of outbound calls that the call center can initial is decided by the number of operators C_{out} which are assigned to process the outbound calls.

There is no staffing cost difference for operators process inbound or outbound calls. We use the operators' utilization U as the system performance measure. U equals the ratio of operators' total busy time to operators' total scheduled time in a certain time interval. The model can be constructed as follows:

Objective: Maximize U

Subject to: $$\mathrm{SL} = \mathrm{P}(W_q \leq \tau) \geq 80\% \tag{2.1}$$

$$C_{in} + C_{out} = C_{total} \tag{2.2}$$

The model is a concept model. The objectivity is to maximize the operators' utilization. Constraint 2.1 indicates that the inbound calls service level should higher than 80%. Constraint 2.2 implies that no matter how to assign operators for inbound calls and outbound calls, the sum of operators is constant.

2.4 Solution Approach

In order to optimize the staffing problem for call centers with inbound and outbound calls service blending, the following functions are needed to be considered.

$C_{min}(\lambda, \mu_{in}, SL)$: The minimum number of operators, constrained by the combination of inbound calls arrival rate λ, inbound calls service time μ_{in} and expected inbound calls service level SL.

$\mathrm{A}(C_{in}, C_{out})$: The assignment policy of inbound and outbound operators, determining the quantities of operators for inbound calls and operators for outbound calls. Two different assignment policies are considered in our model, separating operators for inbound and outbound calls or mixing operators for inbound and outbound calls. In the condition of mixing operators, threshold $\mathrm{r}(C_{total}, \lambda, \mu_{in}, \mu_{out}, \mathrm{SL})$ should be evaluated, which is affected by quantity of total operators C_{total}, inbound calls arrival rate λ, service time of inbound calls μ_{in} and outbound calls μ_{out} as well as inbound calls service rate SL.

2.4.1 Staffing Planning

The inbound calls arrival process is described using M/M/c queuing model, which is shown in figure 2.1. There are c identical servers (operators) working in parallel. The inbound calls have a passion arrival process and exponential service time. The buffer size is infinite and the arrival calls are served first-in-first-out.

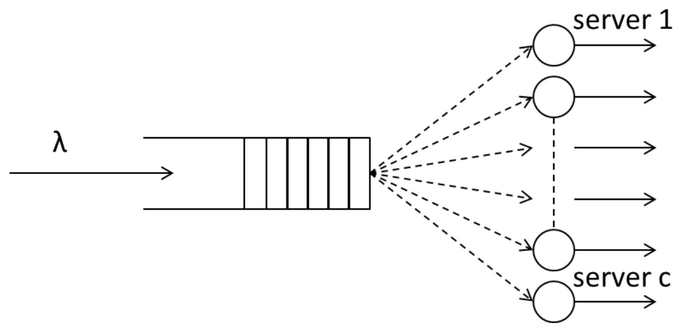

Figure 2.1: M/M/c queuing model

Figure 2.2 shows the transition diagram of the M/M/c queuing model. λ, μ and c indicate the inbound calls arrival rate, inbound calls service time and number of servers respectively.

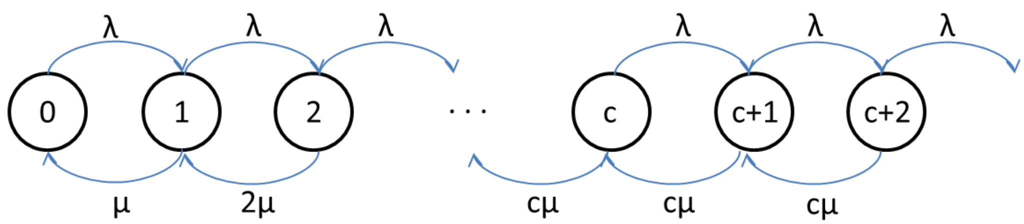

Figure 2.2: Transition diagram of the M/M/c queuing system

We get the minimum number of operators $C_{min}(\lambda, \mu_{in}, SL)$ directly based on the Erlang C (M/M/C) model formula and service level constraint. $a\ (= \lambda/\mu)$ is defined as the offered load and $\rho(= \lambda/c\mu)$ is defined as the offered load per server. When $\rho < 1$, the system has a steady state.

$\mathrm{C}(\mathrm{c}, \alpha)$ defines the steady-state probability that all the servers are occupied.

$$\mathrm{C}(\mathrm{c}, \alpha) = \frac{a^c}{c!(1-(a/c))}\left\{\left[\sum_{n=0}^{c-1} \frac{(c\rho)^n}{n!}\right] + \left[\frac{(c\rho)^c}{c!(1-\rho)}\right]\right\}^{-1} \tag{2.6}$$

When an inbound call arrives, it must wait if all the servers are busy. P(wait > 0) is the delay probability. We have the relation

$$\mathrm{P}\left(W_q > 0\right) = \mathrm{C}(\mathrm{c}, \alpha) \tag{2.7}$$

The expected wait time is given by

$$\mathrm{E}\left(W_q\right) = \frac{C(c,\alpha)}{c\mu-\lambda} \tag{2.8}$$

Define the expected waiting time is τ. The waiting time distribution is given by

$$\mathrm{P}\left(W_q > \tau\right) = \mathrm{C}(\mathrm{c}, \alpha) \cdot e^{-(cu-\lambda)\tau} \tag{2.9}$$

Accordingly, the service level SL is

$$\mathrm{SL} = \mathrm{P}\left(W_q \leq \tau\right) = 1 - \mathrm{C}(\mathrm{c}, \mathrm{a}) \cdot e^{-(cu-\lambda)\tau} \tag{2.10}$$

When the inbound calls arrival rate is time-depending, SIPP approach based on Erlang C model will be used to calculate the minimum inbound operators for each time interval.

2.4.2 Staffing Assignment and Calls Routing

Then we need consider the function $\mathrm{A}(C_{in}, C_{out})$, how to assign the operators for inbound and outbound calls. We use two different policies here and compare their performance in the numerical results section.

Separate staffing policy (S): There are C_{total} operators in total. The operators are classified into two pools, one for inbound calls and one for outbound calls. The inbound operators just process inbound calls while outbound operators just process outbound calls. In order to satisfy the inbound calls service level, the number of inbound operators C_{in} must larger than the minimum value decided using Erlang C formula. The number of operators to handle the outbound calls is C_{out} and $C_{in} + C_{out} \leq C_{total}$. Figure 2.3 shows the separate staffing policy.

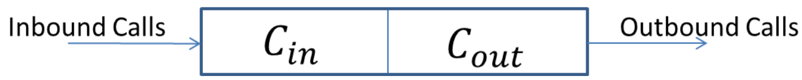

Figure 2.3: Separate staffing policy

Mix staffing policy (M): In the same way, we have C_{total} operators in total. No previous quantified assignment of operators for inbound or outbound calls. When one operator finishes a call, whether his next call is an inbound call or an outbound call, will be decided dynamically according to the system real-time requirement. Since the inbound calls are random with customer waiting, it has higher priority than outbound calls. A threshold function $\mathrm{r}(C_{total}, \lambda, \mu_{in}, \mu_{out}, \mathrm{SL})$ is adopted here to keep inbound calls priority. The inbound calls will be processed immediately when there is idle operator; the outbound calls are initialed only when the idle operators are more than r. Figure 2.4 shows the mix staffing policy.

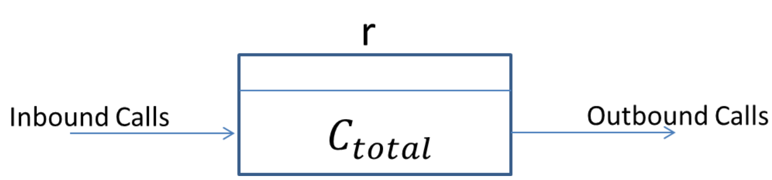

Figure 2.4: Mix staffing policy

2.5 Numerical Experiments and Results Analysis

In this section we evaluate the solution approach through numerical experiments. We start with an experiment with constant inbound arrival rate to optimize the different policies in the solution approach and demonstrate why a certain policy is better. In the following section, an experiment with time-dependent inbound arrival rate is performed and the parameters are historical data in real call center. In the following section, an experiment with time-dependent inbound calls arrival rate is performed and the parameters are historical data in real call center. This experiment shows the appliance process of the solution approach. Three performance measures are used in the simulations: the average service level for inbound calls, the average total outbound calls and the average total utilization for all the operators. The average total outbound calls and the average total utilization for all the operators are correlated actually. Each experiment has 1000 replications, which is large enough for the accuracy requirements.

2.5.1 Constant Arrival Rate

We assume a call center with a fixed number of operators scheduled for the whole time horizon. The inbound calls arrival rate λ is 3 per minute and the service rate μ is 1/3 per minute. The acceptable waiting time τ is 20 seconds. The Erlang C formula tells us that the minimum operators C_{min} for the 80% service-level target is 12. The service rate for outbound calls is the same as inbound calls, 1/3 per minutes. Suppose that the call center operates for a time horizon of T = 510 minutes (8.5 hours). The call center starts empty, and at the end of time horizon T, no new customer will be accepted but the waiting customers will still be served.

Suppose we have 17 operators in total. We first use separate policy to assign task to them. We assign 12 operators to process inbound calls and 5 operators to handle outbound calls. The numerical results show that average inbound calls N_{in}= 1534 arrive at the call center in total and the average service level for inbound calls SL is 81.1%, which is larger than 80%. The average total outbound calls N_{out} is 858 and the total utilization of operators U equals 79.5%.

Next, we use mix staffing policy and the operators will be assigned inbound calls or outbound calls dynamically. We have 17 operators in total. When a operator finishes a call, his state will be changed from busy to idle. If at this moment the total idle operators are less than the threshold, he will stay idle to wait for the next inbound call. Otherwise, if the total idle operators are more than the threshold, he will be assigned an outbound call automatically by the system.

Table 2.1: Performance comparison with constant inbound calls arrival rate

Output	System with S policy	System with M policy					Difference
		r = 0	r = 1	r = 2	r = 3	r = 4	r = 2
N_{in}	1534	1533	1531	1533	1531	1533	——
SL	81.1%	58.9%	78.5%	88.3%	93.3%	96.0%	——
N_{out}	858	1376	1299	1177	1044	901	319
U	79.5%	95.0%	92.4%	88.5%	84.1%	79.5%	9%

The threshold r value is integer since it is the number of idle operators in the system. We set the initial r = 0 with an increment 1. The output with different r value is collected in table 2.1. We can see as r increases, average inbound calls service level (SL) increases monotonously while average total outbound calls N_{out} and average total operators utilization U decrease monotonously. The optimal r should balance the average inbound calls service level SL and average total operators utilization U. The average inbound calls service level SL has the lower constraint limit 80%. So in this experiment r = 2 will be the optimal threshold value.

Let us compare the performance of system with separate staffing policy and system with mix staffing policy (r = 2). Both average inbound calls service levels are large than 80%, which satisfies the service level requirement. The former system generates 858 outbound calls in average while the later system finishes 1177 outbound calls in average, increasing by 319. The average total operators' utilization increases from 79.5% by using separate policy to 88.5% by using mix staffing policy. The comparison shows that system with mix staffing policy performs better than system with separate staffing policy. The mix staffing policy is optimal with higher utilization. We can analyze the reason from two aspects, service level variability and stochastic feature of inbound calls.

First, Erlang queuing method usually considers long-run time average behavior. For each single run, the performance measures shows variability, as figure 2.5 shows. The service level variability indicates uncertainty in the system.

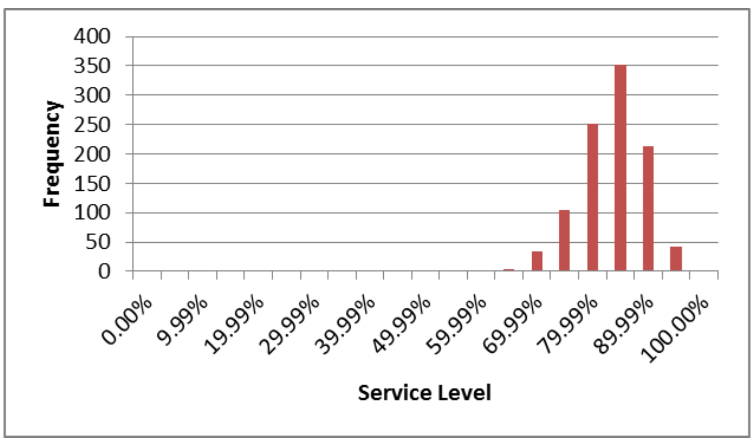

Figure 2.5: Service level variability

Second, the stochastic feature of inbound calls brings uncertainty to the process. We randomly select one simulation from the separate staffing policy experiments and observe the fluctuation of operators' occupancy, as figure 2.6 shows, most of the time the busy operators are less than 12. Because waiting happens not only from the customers' side when no operator is idle, but also from the operators' side, keeping idle for the following inbound calls. Customers' waiting decreases the inbound calls service level while operators' waiting decrease the operators occupancy rate, which is correlated to the system utilization. Customers' waiting and operators' waiting are negative correlated. Decrease customers' waiting will increase operators' waiting, and hence decrease the system utilization. The intrinsic randomness of inbound calls brings uncertainty to the system utilization.

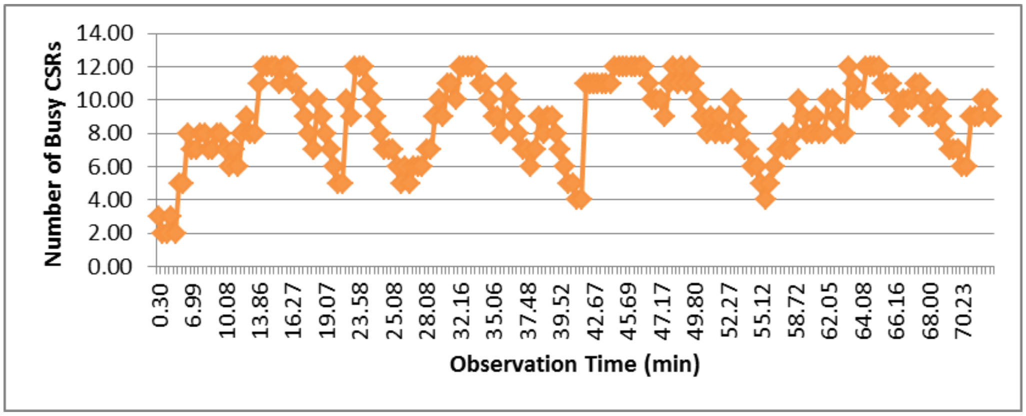

Figure 2.6: Number of busy operators

Figure 2.7 shows the utilization curve for inbound calls only when using Erlang C model. The two curves indicate the utilization value got from Erlang C formula and simulation respectively. They have the same trend and are almost identical. We can see that the utilization increases when the inbound calls arrival rate increases. When the arrival rate is low, the utilization is low. Usually the hospital call center size is small or medium, which means the arrival rate is not very high. Hence, the utilization for inbound calls only will be low. That explains why the capacity utilization is lower when we separate the operators for inbound and outbound calls.

Therefore, the uncertainty in the system is compensated when we use mix staffing policy and dynamically assign the outbound calls according to the system real time state.

Therefore the system utilization increases without decreasing the inbound calls service level.

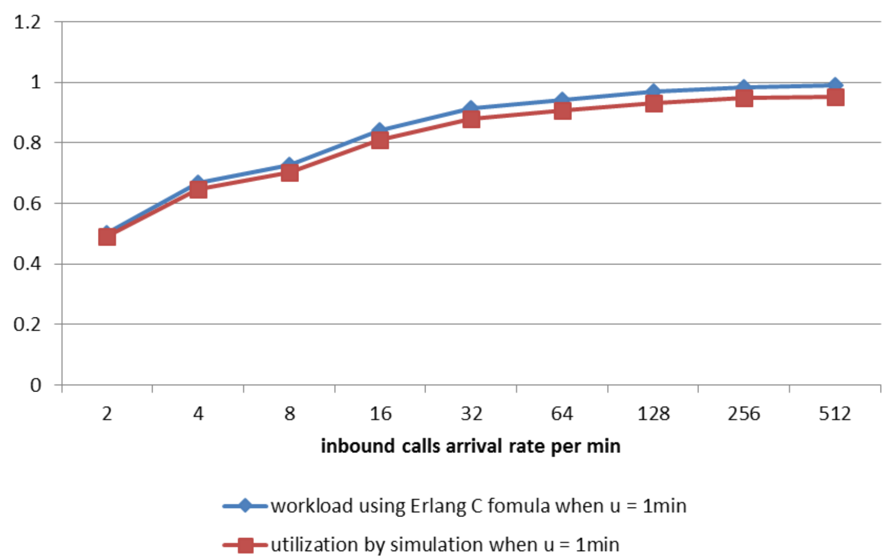

Figure 2.7: Utilization for Inbound calls based on Erlang C model

2.5.2 Time-dependent Arrival Rate

In real call center operation, the inbound calls arrival rates vary as time changes. Table 2.2 shows one sample of Monday arrivals of a hospital call center. We just consider the normal working hours in this hospital call center, from 8:00am to 16: 30pm. There are 8.5 consecutive operating hours in one day. We divided this period into 17 time intervals, each with 30 minutes. All other parameters in this model are the same as the model with constant arrival rate.

We have already demonstrated that the mix staffing policy is better than the separate staffing policy. In this section, we still keep the separate staffing policy performance and use it as a reference to measure the performance improvement by using mix staffing policy.

We use SIPP approach to find the minimum inbound staffing level for each time interval, which are listed in table 2.2. In planning interval 7, the arrival rate reaches the peak value. Accordingly, the minimum inbound operators needed is the most, 22, to satisfy the over 80% inbound calls service level. We define the call center has a constant operators quantity 22. The breaking time of operators here is ignored. In the model with separate staffing policy, for each planning interval, we keep the inbound operators level equals the minimum required inbound operators. The other operators in the total 22 are assigned to work for outbound calls, as shown in table 2.2.

Table 2.2: Inbound calls volume data and minimum inbound operators

Planning interval	Time	Inbound Calls Volume	Inbound Calls Arrival Rate/min	Minimum Inbound operators	Outbound operators using S policy
1	08:00 - 08:30	116	3.87	15	7
2	08:30 - 09:00	95	3.17	13	9
3	09:00 - 09:30	131	4.37	17	5
4	09:30 - 10:00	132	4.40	17	5
5	10:00 - 10:30	136	4.53	17	5
6	10:30 - 11:00	143	4.77	18	4
7	11:00 - 11:30	181	6.03	22	0
8	11:30 - 12:00	135	4.50	17	5
9	12:00 - 12:30	106	3.53	14	8
10	12:30 - 13:00	114	3.80	15	7
11	13:00 - 13:30	129	4.30	16	6
12	13:30 - 14:00	119	3.97	15	7
13	14:00 - 14:30	107	3.57	14	8
14	14:30 - 15:00	101	3.37	14	8
15	15:00 - 15:30	108	3.60	14	8
16	15:30 - 16:00	82	2.73	11	11
17	16:00 - 16:30	53	1.77	8	14

Since the arrival rate and total quantity of operators vary, before the model with blend policy can be performed, we first need to evaluate the threshold r value. We start the evaluation from the arrival rate λ = 1 to λ = 5, which covers the range of the arrival rate in the entire planning intervals, except the peak value. From the model with constant arrival rate, we find that the inbound calls service level increases as the threshold value r increases, so we set the initial r = 0 with an increment 1 for each arrival rate. When r increases to the value that the inbound calls service level increases accordingly to 80%, this r value will be the optimal threshold for the corresponding arrival rate. Each experiment operates 30 minutes. The numerical results are shown in table 2.3.

Table 2.3: Optimal threshold values

Arrival Rate /min	Threshold	Average total inbound calls	Average Service level	Average Outbound Calls	Average operators utilization
1	0	28	85.78%	192	99.7%
	1	28	96.47%	183	95.8%
	2	28	100%	173	91.4%
2	0	55	81.7%	166	99.7%
	1	55	94.6%	158	96.4%
	2	55	98.2%	148	92.2%
	3	56	100%	139	87.9%
3	0	82	74.4%	138	99.7%
	1	82	90.3%	132	97%
	2	82	95.2%	123	93.3%
4	0	109	64.1%	111	99.7%
	1	109	80.8%	106	97.6%
	2	109	89.9%	100	94.5%
5	0	134	50.9%	86	99.7%
	1	135	67.5%	81	98.2%
	2	136	77.9%	76	95.8%
	3	136	85.3%	69	92.9%
6	6	162	79.6%	33	88.1%

	7	162	81.5%	28	86.2%

Figure 2.8 shows the optimal threshold value when the inbound calls arrival rate changes. 80% service level is used the benchmark to evaluate the optimal threshold value. As the inbound calls arrival rate increases from 1/min to 5/min, the optimal threshold value increases from 0 to 3.

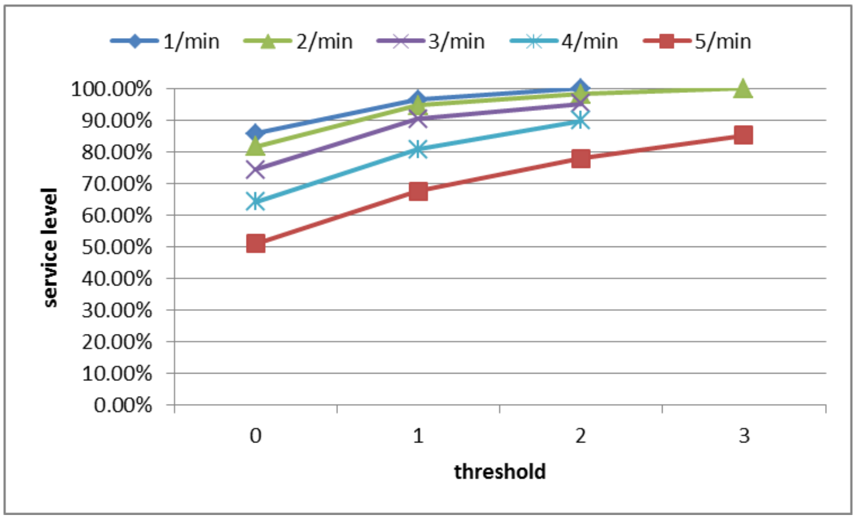

Figure 2.8: Optimal threshold value under different arrival rate

Figure 2.9 shows the influence on utilization when threshold value changes. No matter the value of inbound calls arrival rate, the total utilization will decrease when threshold value decreases. So the optimal threshold value is the smallest value under which the inbound calls service level will keep higher than 80%.

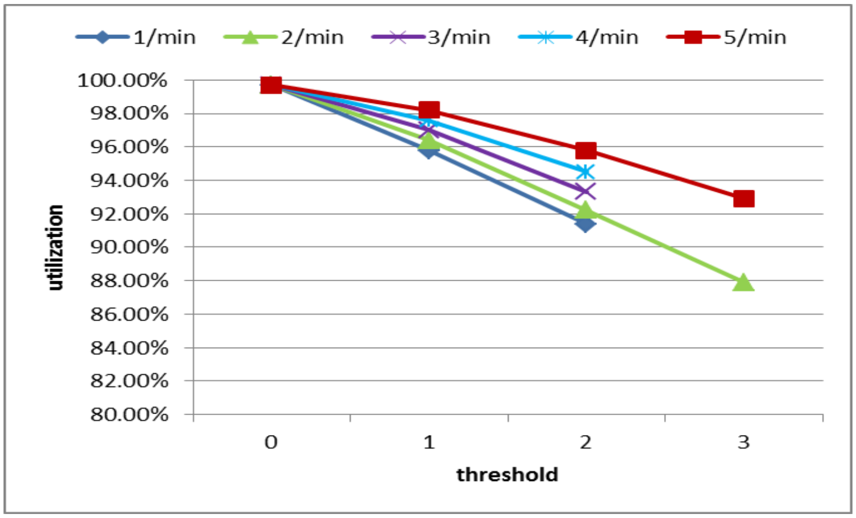

Figure 2.9: Influence on utilization

We compare the system performance using separate staffing policy and mix staffing policy with optimized threshold. Figure 2.10 shows the accumulative time-dependent difference of inbound calls quantity, inbound calls service level, outbound calls quantity and operators utilization.

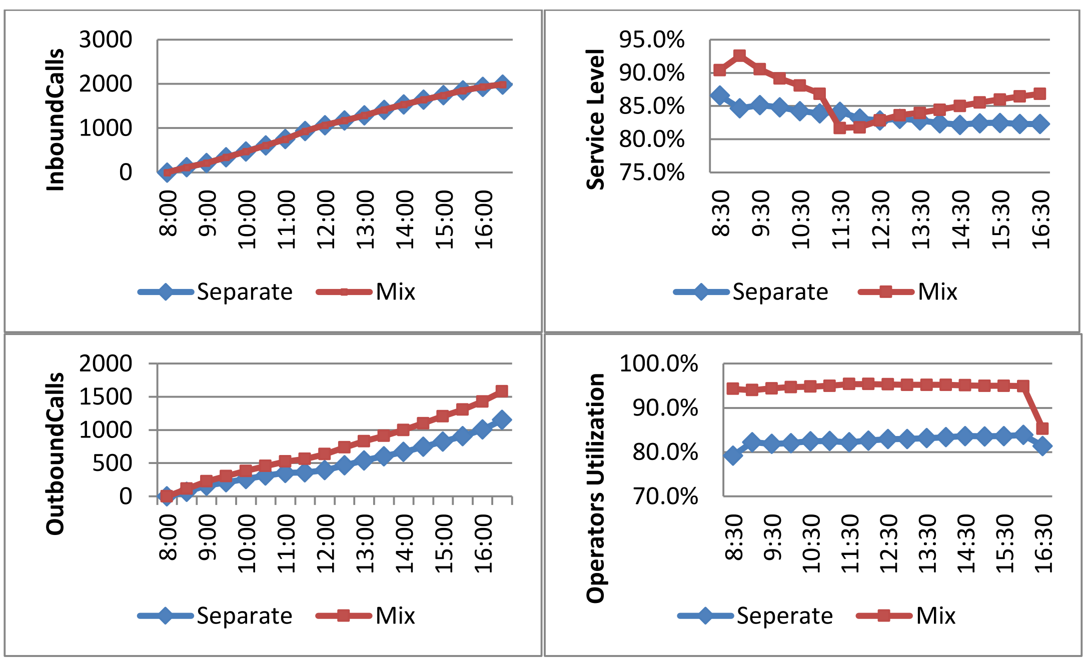

Figure 2.10: System performances comparison

The quantity of inbound calls is related to the inbound calls arrival rate. With the same inbound calls arrival rate, the system receives same quantity of inbound calls, around 1990 inbound calls in the whole operating time interval.

Both of two kinds of staffing policy satisfy the inbound calls service level constraint, higher than 80%. While in most of the whole operating time, the inbound calls service level of system with mix staffing policy is obvious higher, except around 11:00am, which is the peak time of inbound calls arrival rate.

System with separate staffing policy makes average 1153 outbound calls in total. System with mix staffing policy makes average 1580 outbound calls in total, which are 427 more. The operators' utilization reflects the same improvement. In the whole operating time, system with separate staffing policy have average operators' utilization fluctuates between 79% and 84%; while system with mix staffing policy increase the operators utilization higher than 94%. The great decrease at the end of operating time is because the system does not cut off at 16:30; it will finish all the calls start before 16:30.

In summary, the staffing in this paper satisfies the inbound calls service constraint. Compare with the separate staffing policy, the mix staffing policy greatly improves the system service quality and operators utilization. It is an efficient staffing method for hospital call centers.

2.6 Conclusions

In this paper, we have considered an efficient staffing method for hospital call centers with constant number of full-time operators and two kinds of calls, inbound calls and outbound calls.

The main contributions of this paper are the following:

First, the call centers we considered had special features unlike the previous call center models in literature. In hospital context, call centers' size is medium and operators are full time employees with constant quantity. Multiple kinds of medical service can be delivered through outbound calls. The more outbound calls they make, the more potential benefit the hospital will get, especially for non-profit hospitals like Veteran Affairs hospitals. So the outbound calls in our model were assumed infinite. The minimum numbers of operators were set according to the peak value of inbound calls arrival rate, constrained by the service level. Since it was not possible to reduce cost by reducing the operators' quantity, how to improve the system profit became an important issue. In this condition, the objective was to quantify and maximize the operators' utilization, which was equivalent to maximize the system profit.

Second, we considered the staffing problem with the target of finding a balance between service quality and efficiency under SIPP approach. We carried out SIPP approach to determine the minimum number of operators and then used mix staffing policy with optimized threshold to balance the inbound calls service level and operators utilization. The numerical experiments results have showed that our staffing approach not only satisfied the inbound calls service level requirement, but also were efficient to increase the operators' utilization.

Third, for the routing policy, the previous research only demonstrated that the threshold policy was optimal, but failed to show the way to optimize the threshold value under SIPP approach. In our model, the threshold values were derived through extensive simulation based on the system parameters.

With the purpose that our staffing approach can be widely used in multiple kinds of hospital call centers, further research is needed. For one thing, we ignored the inbound calls abandonment when considering the minimum operators in this paper. So one important extension is to determine operators' quantity based on allowing customer abandonment from a practical viewpoint. For another thing, we used numerical experiments to evaluate the staffing policy by considering the same service time distribution of inbound calls and outbound calls. In practice, the distribution of inbound calls and outbound calls service time may be different, which will affect the efficiency of operators utilization. Furthermore, we focused on the staffing policy optimization in this paper and ignored the operators' break time, such as lunch break. In further research, operators shift planning under the optimal staffing policy should be considered.

CHAPTER 3: STAFFING SHIFT PLANNING FOR HOSPITAL CALL CENTERS WITH SERVICE BLENDING

3.1 Introduction

Staffing shift planning is an importance issue in call centers operation because the labor cost constitutes most of the operating cost. In hospital call centers, the operators are required to have medical knowledge background so the labor cost is especially high. The historical inbound call data show that inbound calls volume varies seasonally, which leads to the large variation of number of operators. Usually operators in hospital call centers are permanent staffs; it is hard to hire part-time operators considering the special knowledge requirements and safety issues. So hospital call center has constant number of operators. Staffing shift planning can solve the contradiction between large variation of number of operators required by inbound calls and constant total number of operators.

In recent years, inbound calls are no longer the only workload in hospital call centers. Increasing kinds of medical related services are delivered using outbound calls through hospital call centers gradually. Outbound calls are initialed by the hospital call center. These calls can be classified according to different deadline requirement. For example, outbound calls for chronic disease management should be finished in a planned day; outbound calls for care after discharge should be finished in two days after discharge. Appropriate operators' assignment and planning of outbound calls are able to improve the system efficiency and capacity utilization.

The purpose of this chapter is to develop a staffing planning model for hospital call centers. The reminder of this chapter is organized as follows: In section 3.2, we give a brief literature review related to staffing problems in call centers. In section 3.3, we describe the preliminary work to determine the minimum operators demand constrained by inbound calls. In section 3.4, we present the staffing problem statement. In section 3.5, we develop a mathematical programming model for the staffing planning problem. Section 3.6 shows the numerical results to evaluate the model. Finally, concluding remarks and further research are summarized in section 3.7.

3.2 Literature Review

Staffing planning is very important for call center operations. Gans, et al. (2003) and Aksin, et al. (2007) comprehensively introduce the call center operations research, including call center staffing planning. As the types of call centers increasing, new requirements generate continuingly. Various new studies for call centers' staffing problem are carried out, aiming to solve the new problems and make use of staffing efficiently. Among these call center staffing and scheduling studies, most are focused on the inbound call centers. The inbound calls are stochastic and real-time, which makes the inbound calls have higher priority than outbound calls. So the inbound calls operation is indispensable for blending call centers staffing research.

Various kinds of methods are used to solve the staffing problem in call centers. Roubos and Bhulai (2010) apply approximate dynamic programming to control the time-varying queuing systems. Dietz (2011) develops a practical spreadsheet-based scheduling method to determine the optimal service operators' allocation. Quadratic programming model is used to determine the distribution of operator tours. Mehrotra, et al. (2010) implement a flexible heuristic framework to make intra-day resource adjustment decisions. Erlang formula is used to determine the minimum staffing level. Mandelbaum and Zeltyn (2009), Janssen, et al. (2011) and Zhang, et al. (2012) use diffusion approximation by expanding the Erlang formula to determine the staffing levels. Ásgeirsson (2014) presents an algorithm, which is based on several independent heuristic modules, to create feasible schedule that satisfies both employees' requests and system regulations.

Researchers also carry out studies about staffing operation for inbound calls from different points of view. Service-level is the most important measurement for inbound calls service and it varies when the length of target time interval changes. Different service level definition will lead to different staffing requirements. Baron and Milner (2009), Robbins and Harrison (2010) study the service-level impact on staffing problem. The large variation of inbound calls arrival rate is the root cause for inbound calls operation. Large variation of inbound calls volume causes large variation of operators' requirements, so Keblis and Chen (2006), Batta, et al. (2007), Schrieck, et al. (2014) and Kocaga, et al. (2014) consider flexible operators or outsourcing inbound calls to compensate the large variation.

The inbound calls operation research gives researchers inspirations on the staffing operations for call center with inbound and outbound calls blending. Staffing operations for call centers with inbound and outbound service blending have been carried out. Liao, et al. (2012) present a multi-period staffing problem in a single-shift call center, which handles inbound calls, as well as some alternative back-office jobs. They model the staffing problem as a generalized newsboy-type model under an expected cost criterion. They use stochastic programming formulation and robust programming formulation to solve the problem. Numerical tractability of each formulation is considered in particular. Nah and Kim (2013) consider the workforce planning problem of a hospital reservation call center which handles inbound calls and outbound calls simultaneously. They develop a mathematical programming model with objective to minimize the total cost, including operator labor costs, caller waiting costs and abandonment costs for lost calls. They introduce an inbound-load parameter and use it to get the expected caller waiting time and expected abandonment rate through regression model.

In this chapter, we consider a call center for hospital with one kind of inbound calls service and multiple kinds of outbound calls service. Inbound calls are stochastic and predictable using historical data. Outbound calls can be estimated before planning horizon and have different time to finish. A mathematical programming model is developed to optimize the shift setting and outbound calls workload allocation.

3.3 Preliminary Work

3.3.1 Minimum operators for inbound calls

The inbound calls are real time so they have higher priority than outbound calls. The number of operators should first satisfy the inbound calls requirement. We use service level as the performance measure for the inbound calls, which requires more than 80% customers waiting less than 20 seconds. Steady-state queuing model M/M/s is used to determine the minimum number of operators.

Figure 3.1 shows a sample day's minimum estimated operators based on M/M/s model. For every 30-min time interval, the blue bar indicates the minimum operators under inbound calls service level constraint. The peak value of minimum operators is 22 during 11:00-11:30. So we set 22 as the total number of operators. The red bar indicates the remainder operators which can be assigned to deal with outbound calls.

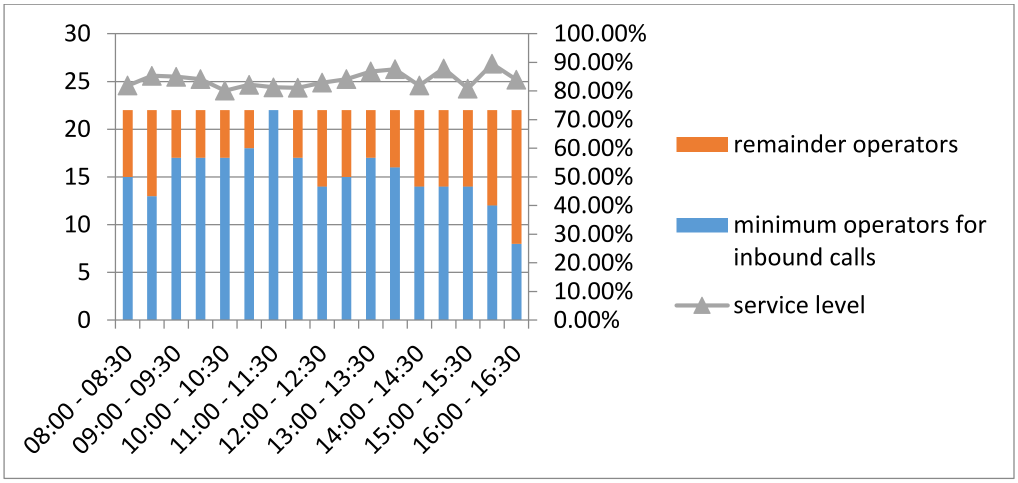

Figure 3.1: Sample of minimum operators under service level constraint

3.3.2 Routing Policy Adjustment

The inbound and outbound calls routing policy is the same policy proposed in chapter 2. From chapter 2 we know that the routing policy which blends inbound and outbound calls under optimal threshold is the optimal routing policy for hospital call centers to improve capacity utilization.

Figure 3.2 shows the operators' occupancy under the parameters quantified in chapter 2. The red line represents the operator occupancy with minimum operators according to the inbound calls arrival rate; the blue line indicates the operators' occupancy for inbound calls only with fixed operators which equals to the minimum operators when the inbound calls arrival rate is 6/min. The green line implies the operators' occupancy when dynamically blending inbound and outbound calls, with the same quantity of operators as the blue line. Operators' Occupancy for inbound calls only is defined as $\rho = \lambda/s\mu$. With optimal threshold policy, the operators' occupancy can be higher than 95%.

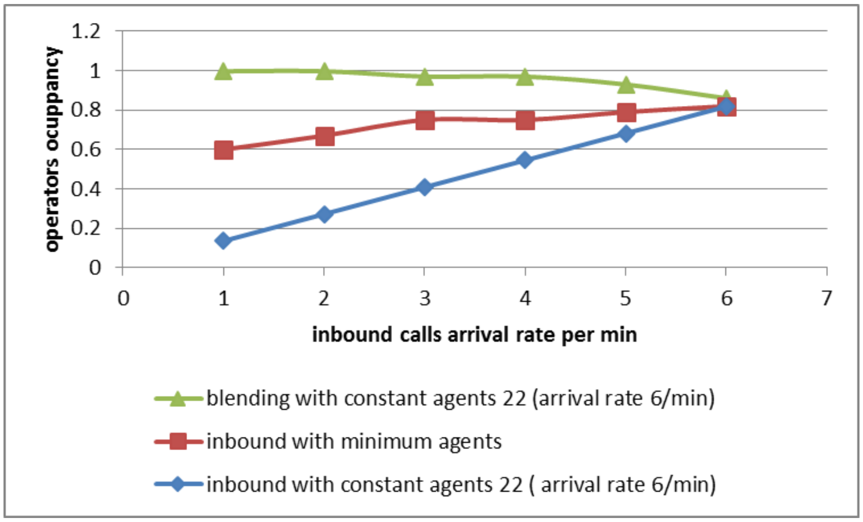

Figure 3.2: Operators' occupancy

3.4 Problem Statement

In this research, a hospital call center is used as a case. The regular working time is from 8:00 to 16:30, 5 days a week. In other time the call center just keep least operators for emergency inbound call, which is out of our scope. Inbound calls from patients are stochastic. Figure 3.3 shows the characters of daily inbound calls volume. A repeating intraweek cycle can be seen in the daily inbound call volume graph: Monday is the busiest day of the week; there is then a gradual decrease in calls from Tuesday to Friday. The inbound calls volume shows intraweek cycle, so the planning horizon in this research is one week. We use D as the set of weekdays indexed by d.

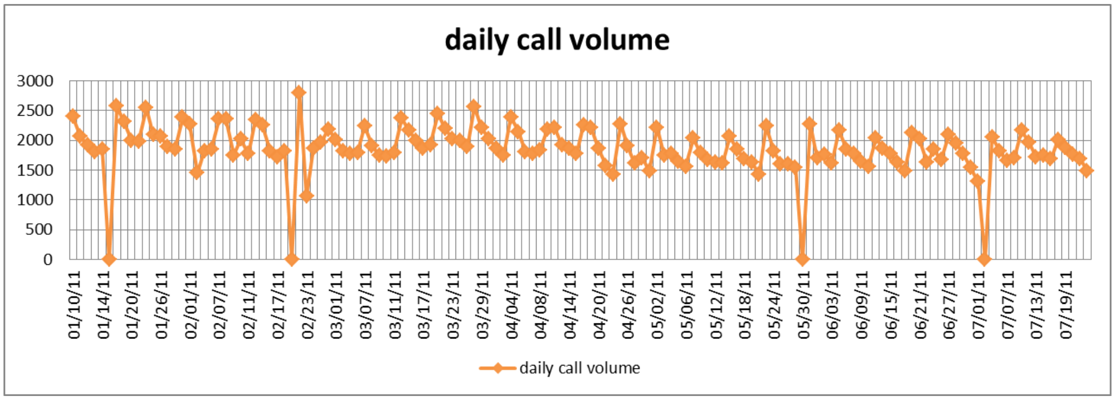

Figure 3.3: Daily call volume

The inbound calls volume also shows within-day cycle. Figure 3.4 reveals the half-hour call volumes for two consecutive weeks. Different color indicates different days, for example red color shows data on Monday. The call center's normal operation time is from 8:00am to 4:30pm. The inbound call volume increases slightly from 8:00am to 9:00am. Then after 9:00am, it increases sharply and reaches the first peak level around 10:30am. The call center is at its busiest for inbound calls between 9:00am to 11:30am. The second peak level generates around 2:00pm. The volume then decreases as the day progresses. The average half hour rate in the morning is higher than in the afternoon. The daily operating time is 8.5 hours. We divide it to 17 time intervals, each with a half hour. T is the time intervals set and t is the time intervals index.

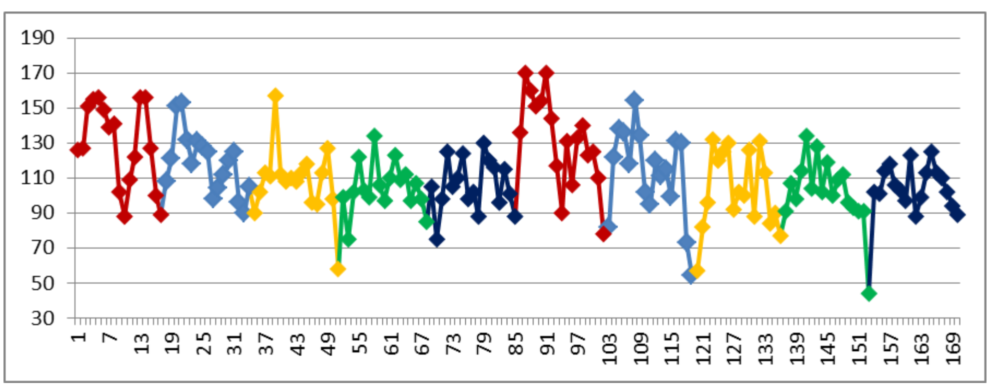

Figure 3.4: Half-hour inbound calls volume

For a certain time interval t on day d, the inbound calls arrival rate λ_{dt} is considered constant in a half hour time interval. λ_{dt} is predicted at least one week ahead of time. S_{dt} represents the minimum number of operators for time interval t on day d. The value of S_{dt} is based on the inbound calls arrival rate λ_{dt}.

Set K denotes the categories of outbound calls. Deferent outbound calls have different deadline requirements. B_{kd} means the total minimum number of outbound calls k which are planned from day *d*. B_{kdt} means the minimum number of outbound calls k which should be assigned in time interval t on day *d*. Two kinds of outbound calls are considered in this model. The first kind of outbound calls should be finished in one day; while the second outbound calls can be finished in two days since it is planned.

The service time of inbound calls and outbound calls is exponentially distributed. The inbound calls service rate is indexed by μ; the outbound calls k service rate is indexed by μ'_k.

Each operator works 8 hours a day, 6 hours for processing the calls, 2 hours for other task like meeting or training and a half hour for lunch break. The operators are assigned to different shifts to make sure that in any time there are operators preparing for the inbound calls. We use I represent the shifts set indexed by i. Parameter a_{it} equals 1 if time interval t is covered by shift i, otherwise it equals 0.

The capacity utilization U is defined as the ratio of operators' total busy time to operators' total scheduled working time in a certain time interval. Parameter U_{dt} denotes the maximum capacity utilization in time interval t on day d.

The operators' busy time includes time used for inbound calls and time used for outbound calls. The queuing system in our model is steady queuing system, so for inbound calls, the system utilization is $\rho = \lambda / s\mu$ (s indicates the total number of operators). For a certain kind of outbound calls, the contents of outbound calls constitutes routing questions, so the variation of service rate is small, which will be ignore in this model.

In this research, two kinds of decision variables are considered: X_{di} presents the number of operators in shift *i* on day *d*; B_{kdt} index the number of outbound calls k assigned in time interval t of day d.

As a summary, we formulate the model with the following notations:

Sets:

D: Set of days in a week indexed by *d* (1, 2,…, 5).

I: Set of shifts indexed by *i* (1,…, *n*).

T: Set of time intervals (half hour) indexed by *t* (1, 2,…, *m*).

K: Set of outbound calls categories indexed by k(1,…,*f*).

Decision variables:

X_{di}: Number of operators working on day d shift *i*.

B_{kdt}: Number of outbound calls k planned in time interval *t* of day *d*.

Parameters:

a_{it}: 1 if time interval *t* is covered by shift *i*, 0 otherwise.

S_{dt}: Minimum number of operators needed in time interval *t* of day *d*.

λ_{dt}: Inbound calls arrival rate in time interval *t* of day *d*.

μ: inbound calls service rate.

μ'_k: outbound calls k service rate.

B_{kd}: Number of outbound calls k on day *d*.

U_{dt}: Maximum capacity utilization

3.5 Model

We develop an integer programming model as the optimization model. Through optimized calculation, the model generates a shift plan with minimum capacity for processing the predicted inbound calls and planned outbound calls for a given finite planning horizon. The model also generates the assignment plan of workload from outbound calls volume, which covers each half-hour time interval during the whole planning horizon.

The integer program is given below:

Objective: Min $\sum_{d=1}^{5} \sum_{i=1}^{n} X_{di}$

Subject to $\sum_{i=1}^{n} a_{it} X_{di} \geq S_{dt} \quad \forall \text{ d, t}$ (3.1)

$$\frac{\lambda_{dt}}{(\sum_{i=1}^{n} a_{ti} X_{di})\mu} + \sum_{k=1}^{f} \left(\frac{B_{kdt}}{30(\sum_{i=1}^{n} a_{ti} X_{di}) u'_k} \right) \leq U_{dt} \quad \forall \text{ d, t} \quad (3.2)$$

$$\sum_{t=1}^{m} b_{1dt} = B_{1d} \quad \forall \text{ d} \quad (3.3)$$

$$\sum_{1}^{d+1} \sum_{t=1}^{m} B_{2dt} \geq \sum_{1}^{d} B_{2d} \quad \text{for d = 1, ..., 4} \quad (3.4)$$

$$\sum_{1}^{d} \sum_{t=1}^{m} B_{2dt} \leq \sum_{1}^{d} B_{2d} \quad \text{for d = 1, ..., 4} \quad (3.5)$$

$$\sum_{1}^{d} \sum_{t=1}^{m} B_{2dt} = \sum_{1}^{d} B_{2d} \quad \text{for d = 5} \quad (3.6)$$

$$\sum_{i=1}^{n} x_{di} = \sum_{i=1}^{n} x_{(d+1)i} \quad \text{for d = 1, ..., 4} \quad (3.7)$$

$X_{di} \geq 0$ and integer (3.8)

$B_{kdt} \geq 0$ and integer (3.9)

The objective function is to reduce the total number of operators used.

Constraint set (3.1) assures that call center has at least the minimum number of operators required every half hour across the whole week. Constraint set (3.2) assures that the total utilization of operators is less than the maximum utilization using inbound and outbound calls service blending policy. Constraint set (3.3) satisfies the number of outbound calls which should be finished in one day. Constraint sets (3.4) - (3.6) include the assignment of outbound calls which should be finished in two days. Constraint set (3.7) indicates that the daily total number of operators is constant. The constraint set (3.8) defines the number of shift operators positive and integer. The constraint set (3.9) implies the outbound calls work load is positive and integer.

LINGO 14.0 (LINDO SYSTEMS INC) is a commercial optimization software. It is used to solve this staffing shift planning model for hospital call center with service blending.

3.6 Numerical Analysis

The planning horizon of the hospital call center in this research is one week. Based on the historical data stored in the computer database system and the prediction model, we can get the predicted inbound calls volume for the coming week, as figure 3.5 shows. The inbound calls service time is exponentially distributed and u^{-1} equals 3 mins.

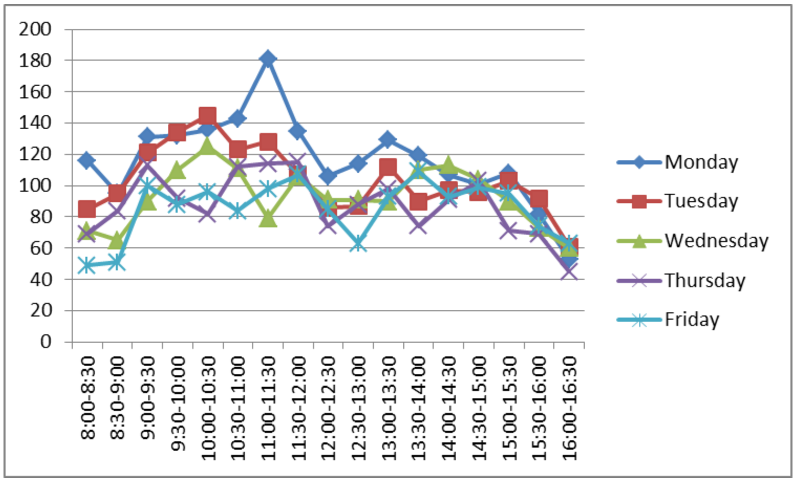

Figure 3.5: Predicted inbound call volumes

The outbound calls volume is estimated from the patient status records, which is shown in figure 3.6. For example, B_{1d} represents the volume of chronic disease management outbound calls that should be operated on day *d*. B_{2d} indicates the volume of care after discharge outbound calls which should be done in two days, either on day *d* or day (*d*+1). All the planned outbound calls should be finished in the same week. The outbound calls service time is also exponentially distributed and $(\mu')^{-1}$ equals 3 mins.

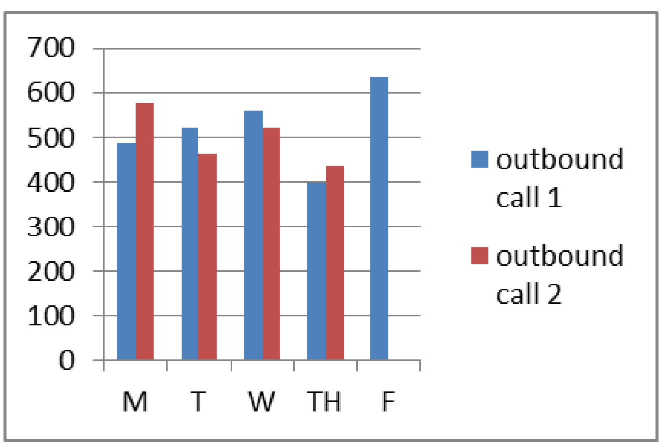

Figure 3.6: Estimated outbound call volumes

The operators were separated into 2 shifts every day. Each shift has 6 hours for calls related work, two hours for meeting or training and a half hour for lunch. More operators should be assigned when inbound calls come to the peak level. And one meeting or training usually lasts one hour. So the meeting or training time should be consecutive. Considering the real situation, the shifts set is shown in table 3.1.

Table 3.1: Shift scheduling set

Shifts	8:00		9:00		10:00		11:00		12:00		13:00		14:00		15:00		16:00	X_i
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
2	1	1	1	1	1	1	1	1	/	1	1	1	1					X_1
			1	1	1	1	1	1	1	1	/			1	1	1	1	X_2

("/" lunch time)

Shift planning results are shown in table 3.2. We test 6 cases under different parameters. The shift sets vary when we choose different maximum system utilization value and outbound calls workload. In each case study, we use constant system utilization value. In real situation, when working consecutively for two or three hours, operators may feel tired and their working efficiency may decrease. So call center manager can adopt different system utilization values according to operators' real working status. The first three cases has deferent capacity utilization value. We also try to change the parameters of outbound calls workload in the last three cases. It shows the way to plan the shift when the workload increases since services delivered through call centers are increasing.

Table 3.2: Shift planning results

Case #	Utilization	Outbound Calls (B_{1d}, B_{2d})	Shifts set (X_1, X_2) M	T	W	TH	F	Total
1	$U_{dt} = 95\%$	(B_{1d}, B_{2d})	(17, 14)	(15, 16)	(15, 16)	(17, 14)	(15, 16)	31
2	$U_{dt} = 92\%$	(B_{1d}, B_{2d})	(17, 14)	(17, 14)	(17, 14)	(17, 14)	(18, 13)	31
3	$U_{dt} = 85\%$	(B_{1d}, B_{2d})	(17, 14)	(17, 14)	(17, 14)	(17, 14)	(15, 16)	31
4	$U_{dt} = 85\%$	$(2B_{1d}, B_{2d})$	(17, 15)	(15, 17)	(18, 14)	(18, 14)	(15, 17)	32

5	$U_{dt} = 85\%$	$(B_{1d}, 2B_{2d})$	(17, 14)	(17, 14)	(17, 14)	(13, 18)	(15, 16)	31
6	$U_{dt} = 85\%$	$(2B_{1d}, 2B_{2d})$	(17, 18)	(15, 20)	(21, 14)	(13, 22)	(15, 20)	35

The outbound calls workload allocation plan is shown in the following graphs. Figure 3.7. Figure shows outbound calls 1 assignment. We can see that there are three kinds of volume assignments: high, medium and zero. In a time interval, if the number of active working operators is much more than the minimum requirement, high volume of outbound calls will be allocated to this time interval; If the number of active working operators is a little more than the minimum requirement, medium volume of outbound calls will be allocated to this time interval; If the time interval is the inbound calls peak time, no outbound calls are allocated to this time interval.

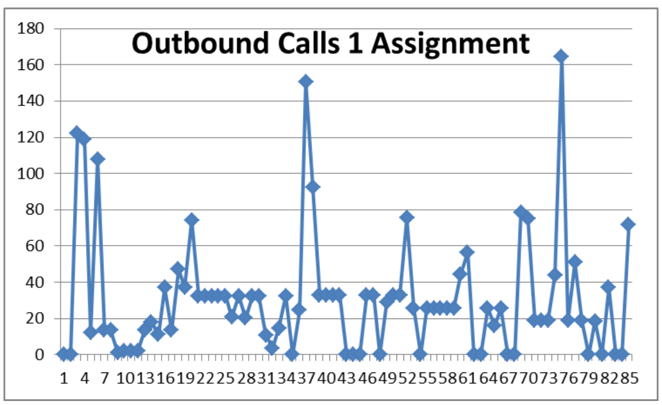

Figure 3.7: Outbound calls 1 workload allocation plan in case 3

Figure 3.8 shows the outbound calls 2 assignment. It shows similar planning results as outbound calls 1. There are three levels of workload assignment, high, medium and zero.

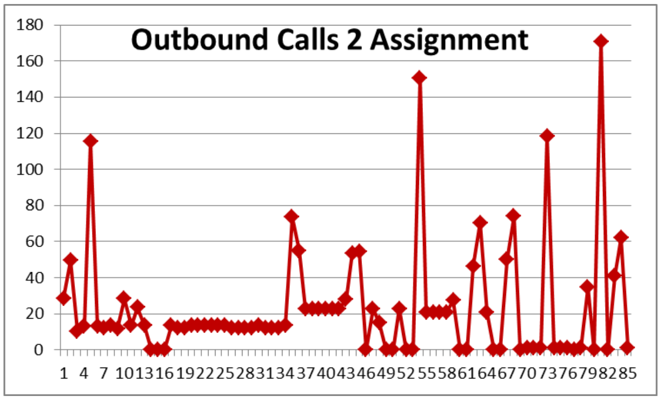

Figure 3.8: Outbound calls 2 workload allocation plan in case 3

In order to see the effect of our model, we compare the performance of this model with the model without routing inbound and outbound calls dynamically, which means the hospital call center separates the operators for inbound and outbound calls: inbound calls operators just process inbound calls and outbound calls operators just process outbound calls. Figure 3.9 shows the performance comparisons. The parameters sets of cases are same as in table 3.2. The red bars indicate the minimum operators using the policies without dynamic routing and the blue bars indicate the minimum operators using the method in this research. The green line shows the percent of operators' decrease. All the cases show great decrease of operators using this model.

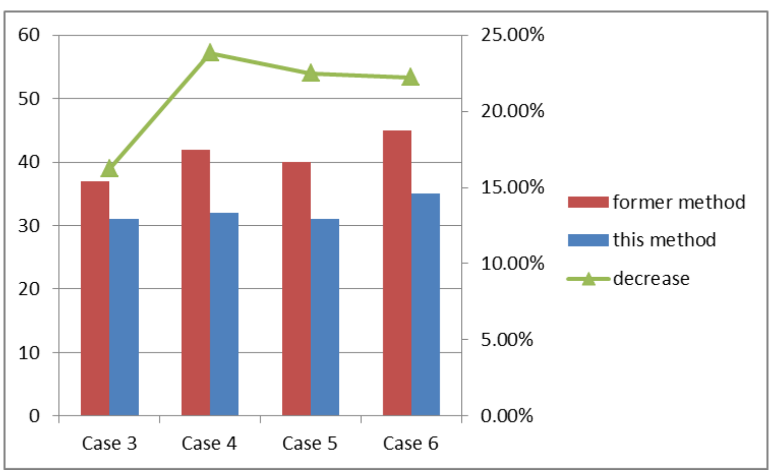

Figure 3.9: Performance comparisons

In summary, this model can effectively help the hospital call center to plan shift settings and outbound calls workload allocation. Parameters can be set adaptively according to the call center's needs.

3.7 Conclusions

In this paper, we have considered an efficient staffing shift planning model for hospital call centers with single kind of inbound calls and multiple kinds of outbound calls. The different kinds of outbound calls have different scheduled time to complete. For example, outbound calls one should be completed in one day while outbound calls two should be completed in two days.

Several preliminary works are carried out before the mathematical model is developed. In order to satisfy the inbound calls service quality, the minimum numbers of operators are set according to the peak value of inbound calls arrival rate. Policy of dynamically routing inbound and outbound calls are evaluated and the system utilization is quantified.

The operators are full time employees and the quantity is constant. The objective of the mathematical model is to minimize the number of total operators. Minimum numbers of operators, system utilization and the time limits for completion of multiple kinds of outbound calls are used as the constraints for outbound calls workload allocation. Numerical results show that the model is effective for staffing shift planning.

This model is based on the predicted inbound calls volume and planned outbound calls volume. In the further research, we should make this model more consistent with the real situation. As the predicted inbound calls volume may update in real-time; the outbound calls may not be get through during the first time dialing. The dynamic process of these factors should be included in the model to make the model more adaptive to the real situation.

CHAPTER 4: MULTI-SKILL STAFFING SHIFT PLANNING FOR A HOSPITAL CALL CENTER WITH SERVICE BLENDING

4.1 Introduction

Increasing kinds of service are delivered through call centers so most of call centers have been converted from single skill call centers to multi-skill call centers. For example, in hospital call centers, nurses are required to handle calls related to medical service; clerks are able to process the appointment schedule or reminder service. The labor cost varies according to the role of operators. Nurses are required to have medical knowledge background so they have higher labor cost than clerks. Accordingly multi-skill call center staffing shift planning is an increasingly important issue in call centers operations since the labor cost constitutes most of the operating cost.

In recent years, inbound calls are no longer the only workload in hospital call centers. Increasing kinds of medical related services are delivered by outbound calls through hospital call centers gradually. For example, outbound calls are used by nurses to offer chronic disease management service while by clerks to offer appointment reminder service. So the staffing planning problem in this research covers the multi-skill staff planning with inbound and outbound calls service blending. Appropriate multi-skill staff planning is able to decrease the labor cost and improve the capacity utilization.

The purpose of this chapter is to develop a multi-skill staffing planning model for hospital call centers. The reminder of this paper is organized as follows: In section 4.2, we give a brief literature review related to multi-skill staffing problems in call centers. In section 4.3, we show the preliminary work to make policy for routing inbound calls between multi-skill operators and determine the minimum operators demand constrained by inbound calls. In section 4.4, we present the multi-skill staffing problem statement. In section 4.5, we develop a mathematical programming model for the multi-skill staffing planning problem. Section 4.6 shows the numerical results to evaluate the model. Finally, concluding remarks and further research are summarized in section 4.7.

4.2 Literature Review

Multi-skill call center problems generate along with increasing characteristics of call centers' service, customer types, and agent skills. Gans, et al. (2003) and Aksin, et al. (2007) summarize comprehensive surveys of call center operation problems, including multi-skill operation problem. As the complexity of call center increases, a stream of research is continuingly carried out to study the multi-skill call center operations from different perspectives.

Tarakci, et al. (2009) study a specialty hospital providing traditional face-to-face consultations by experts and telemedicine services by tele-specialists. The objective is to minimize the cost of staffing, technology investment, incorrect treatment and waiting. A heuristic proposed in queuing theory is used. They find under certain conditions it is not optimal for the hospital to invest in telemedicine and a policy of treating all patients via telemedicine is never optimal.

Cordone, et al. (2011) develop a quantitative approach to organize the call center work shifts and their rest periods to make sure the mix of skills obtained in each time slot is as close as possible to a desired level. The approach is based on a heuristic method which exploits a general purpose linear programming solver.

Perry and Whitt (2011) consider an automatic control between two networked service systems. Each network has its own customers and usually the agents only serve their own customers. With the automatic control mechanism, the network starts serving some of the other customers when an unexpected overload occurs. They propose a queue-ratio control with thresholds to achieve the automatic control of calls routing between the two networked service systems.

Yao and Cassandras (2011) consider an automatic control between two networked service systems. Each network has its own customers and usually the agents only serve their own customers. With the automatic control mechanism, the network starts serving some of the other customers when an unexpected overload occurs. They propose a queue-ratio control with thresholds to achieve the automatic control of calls routing between the two networked service systems.

Roubos and Bhulai (2012) propose an approximate dynamic programming algorithm for multi-skill routing in call centers. The algorithm is one-step policy improvement using a polynomial approximation to relative value functions. These dynamic routing policies are nearly optimal. The policies are also scalable with the problem instance and can be computed online.

Adan and Weiss (2012) investigate the exact first-come first-served matching rates for queues with multi-type servers and multi-type customers. The system is described by multidimensional countable Markov chain. A closed-form formula is given to calculate the matching rate and the chain is ergodic. This model can be connected to the conjectured matching rates for an overloaded multi-skill call center.

Andradóttir, et al. (2013) investigate the design principles for flexible systems with heterogeneous servers and tasks. They consider the system as a network and the bottleneck may span several nodes in the network. Through choosing flexibility structure of the network, they can shift capacity between arbitrary nodes and allow the network to cope with demand fluctuations.

Legros, et al. (2015) propose a flexible architecture for call centers with skill-based routing. Single pooling with only two skills per agent is applied and compared with the most well-known architectures like chaining, which have limited flexibility and fail against asymmetric parameters like unbalanced workload, different service requirements, etc.

In this chapter, we consider a call center for hospital with multiple kinds of inbound calls service and multiple kinds of outbound calls service. Inbound calls are stochastic and predictable using historical data. Outbound calls can be estimated before planning horizon and have different scheduled time to complete. Operators are classified with different skills. A mathematical programming model is developed to optimize the routing problem with multi-skill operators, shift setting and outbound calls workload allocation.

4.3 Preliminary Work

4.3.1 Inbound calls routing policy

The hospital call center operators include nurses and clerks. Nurse operators are expected to process medical related questions through inbound calls and initial outbound calls to deliver service to patient, including chronic disease management, care after discharge and so on. While clerk operators are assigned to handle non-medical related inbound calls and initial outbound calls to deliver service like appointment reminder calls. Nurse and clerk operators are assigned to medical related service line and non-medical related service line separately. When patients initial inbound calls, they can choose medical related service or non-medical related service and the ACD system will automatically rout patients' inbound calls to the right operators.

By analyzing the historical inbound calls records, we find that it is hard to divide all the inbound calls to two kinds. Usually an inbound call includes several questions. A certain portion of inbound calls contains not only medical related questions, but also non-medical related questions. Patients will choose medical related service even if they just have one medical related question. So the inbound calls answered by nurse are not all about medical related questions like appointment, bill or insurance information. The labor cost for nurses is much higher than clerk. So nurse processes non-medical related questions will increase the labor cost.

The call center model with multiple kinds of operators can be described in figure 4.1. There are two different service lines in our model. The operators on the two service lines have different knowledge background, which are represented by S_a and S_b. S_a have higher labor cost than S_b. S_a have the ability to process multiple questions related to content *a* and *b*; while S_b are just able to handle questions related to content *b*. In most cases the inbound calls may not contain only one kind of questions. They may include multiple questions with both kinds of contents. Inbound calls with two kinds of questions usually are routed to S_a. S_a and S_b also process outbound calls related to their knowledge level. The inbound calls have higher priority than outbound calls.

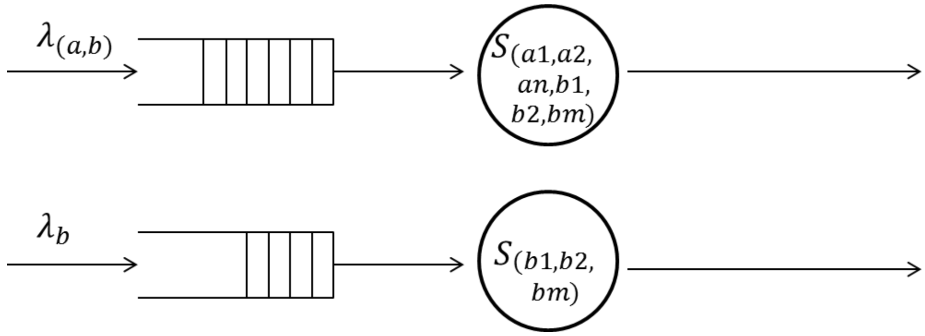

Figure 4.1: A call center model with two service lines

In order to decrease the call center's total labor cost, we try the new inbound calls transfer policy, which is described in figure 2. We allow S_a transfer inbound calls to S_b. When S_a receive an inbound call, they first talk with the patients to know what kinds of questions they have. If the patients have two kinds of questions, they answer the content *a* questions and then transfer these inbound calls to S_b for the remaining content *b* questions. The transfer rate is *p*, which is decided by analyzing the historical inbound calls contents record. Figure 4.2 shows the transferring model. S_a and S_b also process outbound calls related to their knowledge level.

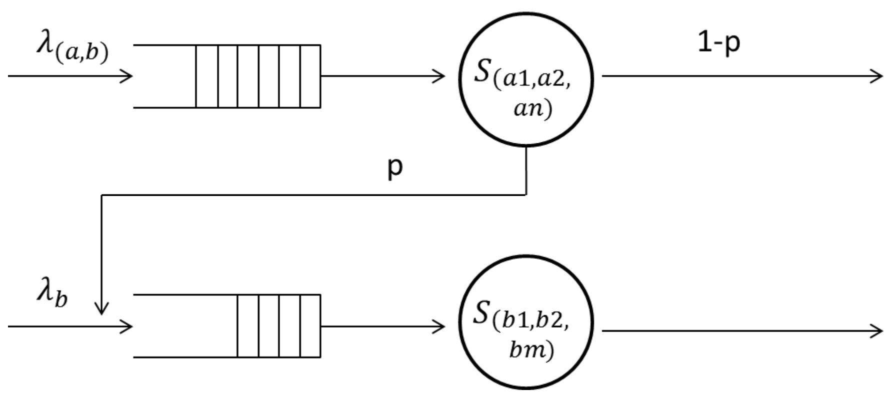

Figure 4.2: Inbound calls transferring model

The service time varies when the call is transferred. The service rate of non-transferred inbound calls is μ_1 for nurse operators and μ_2 for clerk operators; while the transferred inbound calls service rate is μ_1' for nurse operators and μ_2' for clerk operators.

A transferred call will cost a little more time than a non-transferred call. The object of this research is to see if the new routing policy is optimal and how to schedule the staffing and allocate the outbound calls work load under this condition.

4.3.2 Determination of minimum operators

Although both kinds of operators work on inbound and outbound calls. Inbound calls, no matter original or transferred, are real time calls so they have higher priority than outbound calls. The number of operators should first satisfy the inbound calls requirement. We use service level as the performance measure for the inbound calls, which requires more than 80% customers waiting less than 20 seconds. Before transfer policy is adopted, steady-state queuing model M/M/s is used to determine the minimum number of operators. When a portion of inbound calls are transferred between the service lines, each service line processes inbound calls with different service rate, which will lead to increase or decrease of minimum operators demand. There are no closed-form expressions to calculate the minimum operators with multiple service rates. So we adjust the minimum operators demand through extensive simulation.

A sample day's minimum operator estimation is shown in the flowing graphs. Figure 4.3 shows the minimum demand of nurse operators. After transferring a certain percent of inbound calls, the minimum demands of nurse operators decrease in all time intervals while the service rates remain above 80%.

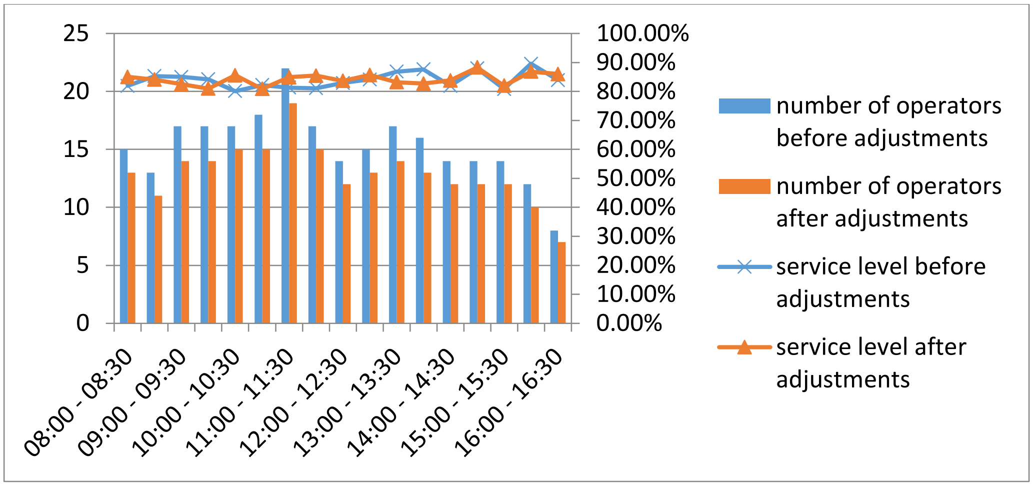

Figure 4.3: Minimum demands of nurse operators

Figure 4.4 shows the minimum demand of clerk operators. After receiving a certain quantity of transferred inbound calls, the minimum demands of nurse operators remain unchanged or increase, constrained by the service level requirement.

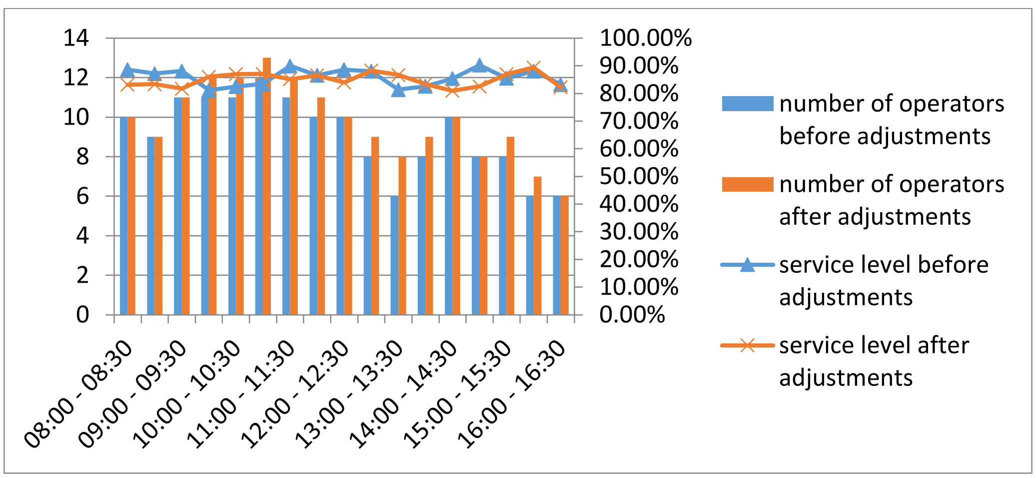

Figure 4.4: Minimum demands of clerk operators

4.3.3 Outbound calls routing policy

The routing policy is the same policy that we propose in chapter 2. Both of the two service lines put inbound calls with higher priority, no matter original inbound calls or transferred inbound calls. The outbound calls on each service line are dynamically assigned to the idle operators of their own service line under optimal threshold value.

4.4 Problem Statement

The regular working time of this call center is from 8:00 to 16:30, 5 days a week. The inbound calls volume shows intraweek cycle, so the planning horizon in this research is one week. We use D as the set of weekdays indexed by d. The inbound calls volume also shows within-day cycle. The daily operating time is 8.5 hours. We divide it to 17 time intervals, each with a half hour. T is the time intervals set and t is the time intervals index.

For a certain time interval t on day d, the arrival rate of inbound calls *o* is considered constant in a half hour time interval. λ_{odt} is predicted at least one week ahead of time. The inbound calls transfer rate is expressed by *p*. S_{odt} represents the minimum number of operators *o* for time interval t on day d. The value of S_{odt} is based on the inbound calls arrival rate λ_{odt} and inbound calls transfer rate.

Set K denotes the categories of outbound calls. Deferent outbound calls have different deadline requirements. B_{okd} means the total minimum number of outbound calls k which are assigned to operators *o* and planned from day *d*. B_{okdt} means the minimum number of outbound calls k which should be assigned in time interval t on day *d* to operators *o*. In this model, for clerk operators, a single kind of outbound calls is considered and these calls must be finished in one day since it is planned. For nurse operators, two kinds of outbound calls are considered. The first kind of outbound calls should be finished in one day; while the second outbound calls can be finished in two days since it is planned.

The service time of inbound calls and outbound calls is exponentially distributed. The service rate of original inbound calls o is indexed by μ_o; the service rate of transferred inbound calls o is indexed by μ'_o; the outbound calls k service rate is indexed by μ''_{ok}.

There are two kinds of operators, nurse and clerk. O is the operator categories set and o is the operator categories index. C_o indicates the daily labor cost of operators *o.* Nurse operators have the skill to process the inbound calls that directly routed to the nurse service line. While clerk operators have the skills to process inbound calls that directly routed to the clerk service line and inbound calls transferred from the nurse service line.

Nurse operators work 8 hours a day, 6 hours for processing the calls, 2 hours for other task like meeting or training and a half hour for lunch break. While clerk operators work 8 hours for calls processing and a half hour for lunch break. The operators are assigned to different shifts to make sure that in any time there are operators preparing for the inbound calls. We use I representing the shifts set indexed by i. Parameter a_{oit} equals 1 if time interval t is covered by shift i of operator o, otherwise it equals 0.

The capacity utilization U is defined as the ratio of operators' total busy time to operators' total scheduled working time in a certain time interval. Parameters U_{odt} denote the maximum capacity utilization in time interval t on day d for operator o.

The operators' busy time includes time used for inbound calls and time used for outbound calls. The queuing system in our model is steady queuing system, so for inbound calls, the system utilization is $\rho = \lambda/s\mu$ (s indicates the total number of operators). For a certain kind of outbound calls, the contents of outbound calls constitutes routing questions, so the variation of service rate is small, which will be ignore in this model.

In this research, two kinds of decision variables are considered: X_{odi} presents the number of operators *o* in shift *i* on day *d*; B_{okdt} indexes the number of outbound calls k assigned in time interval t of day d to operators *o*.

As a summary, we formulate the model with the following notations:

Sets:

D: Set of days in a week indexed by *d* (1, 2,…, 5).

I: Set of shifts indexed by *i* (1,…, *n*).

T: Set of time intervals (half hour) indexed by *t* (1, 2,…, *m*).

O: Set of operators indexed by o (1, 2)

K: Set of outbound calls categories indexed by k(1,…,*f*).

Decision variables:

X_{odi}: Number of operators *o* working on day d shift *i*.

B_{okdt}: Number of outbound calls k assigned to operators *o* in time interval *t* of day *d*.

Parameters:

C_o: Daily salary for operator *o.*

a_{oit}: 1 if time interval *t* is covered by shift *i* of operators *o*, 0 otherwise.

p: inbound calls transfer rate.

S_{odt}: Minimum number of operators *o* needed in time interval *t* of day *d*.

λ_{odt}: Arrival rate of inbound calls for operators *o* in time interval *t* of day *d*.

μ_o: Service rate of original inbound calls for operators *o*.

μ'_o: Service rate of transferred inbound calls for operators *o*.

μ''_{ok}: Service rate of outbound calls k for operators *o*.

B_{okd}: Number of outbound calls k for operators *o* on day *d*.

U_{odt}: Maximum capacity utilization of operators *o.*

4.5 Model

We develop an integer programming model for this multi-skill staffing shift planning problem. Through optimized calculation, the model generates a shift plan with minimum labor cost for processing the predicted inbound calls and planned outbound calls in the planning horizon. The model also generates the assignment plan of workload from outbound calls volume, which covers each half-hour time interval during the whole planning horizon.

The integer program is given in the following part:

Object: Min $\sum_{o=1}^{O} C_o(\sum_{d=1}^{D}\sum_{i=1}^{n} X_{odi})$

Subject to $\sum_{i=1}^{n} a_{oit} X_{odi} \geq S_{odt}$ $\forall$ o, d, t (4.1)

$$\frac{(1-p)\lambda_{1dt}}{(\sum_{i=1}^{n} a_{1ti}X_{1di})u_1} + \frac{p*\lambda_{1dt}}{(\sum_{i=1}^{n} a_{1ti}X_{1di})\mu'_1} + \sum_{k=1}^{f}\left(\frac{B_{1kdt}}{30(\sum_{i=1}^{n} a_{1ti}X_{1di})u''_{1k}}\right) \leq U_{1dt} \quad \forall \text{ d, t} \quad (4.2)$$

$$\frac{\lambda_{2dt}}{(\sum_{i=1}^{n} a_{2ti}X_{2di})u_2} + \frac{p*\lambda_{1dt}}{(\sum_{i=1}^{n} a_{2ti}X_{2di})\mu'_2} + \sum_{k=1}^{f}\left(\frac{B_{2kdt}}{30(\sum_{i=1}^{n} a_{2ti}X_{2di})u''_{2k}}\right) \leq U_{2dt} \quad \forall \text{ d, t} \quad (4.3)$$

$$\sum_{t=1}^{m} b_{o1dt} = B_{o1d} \qquad \forall \text{ o, d} \qquad (4.4)$$

$\sum_{1}^{d+1}\sum_{t=1}^{m} B_{22dt} \geq \sum_{1}^{d} B_{22d}$ for d = 1, …, 4 (4.5)

$\sum_{1}^{d}\sum_{t=1}^{m} B_{22dt} \leq \sum_{1}^{d} B_{22d}$ for d = 1, …, 4 (4.6)

$\sum_{1}^{d}\sum_{t=1}^{m} B_{22dt} = \sum_{1}^{d} B_{22d}$ for d = 5 (4.7)

$\sum_{i=1}^{n} x_{odi} = \sum_{i=1}^{n} x_{o(d+1)i}$ for o, d = 1, …, 4 (4.8)

$X_{odi} \geq 0$ and integer (4.9)

$B_{okdt} \geq 0$ and integer (4.10)

The objective is to reduce the total labor cost.

Constraint set (4.1) assures that call center has at least the minimum number of operators for inbound calls every half hour across the whole week. Constraint set (4.2) assures that the total utilization of nurse operators for all kinds of calls is less than the maximum utilization using inbound and outbound calls service blending policy. Constraint set (4.3) assures that the total utilization of clerk operators on the sum of all kinds of calls is less than the maximum utilization using inbound and outbound calls service blending policy. Constraint set (4.4) satisfies the number of outbound calls which should be finished in one day. Constraint sets (4.5) - (4.7) include the assignment of outbound calls which should be finished in two days. Constraint set (4.8) indicates that the daily total number of operators is constant. The constraint set (4.9) defines the number of shift operators positive and integer. The constraint set (4.10) implies the outbound calls workload is positive and integer.

LINGO 14.0 (LINDO SYSTEMS INC) is a commercial optimization software. It is used to solve this multi-skill staffing shift planning model for hospital call center with service blending.

4.6 Numerical Analysis

The planning horizon of the hospital call center in this research is one week. Based on the historical data stored in the computer database system and the prediction model, we can get the predicted inbound calls volume for the coming week, as figure 4.5 shows. The service time of original inbound calls for nurse and clerk is exponentially distributed with μ_1^{-1} = μ_2^{-1} = 3 mins. The service time of transferred inbound calls for nurse and clerk is also exponentially distributed with $(\mu_1')^{-1}$ = 1 min and $(\mu_2')^{-1}$ = 2.5 mins.

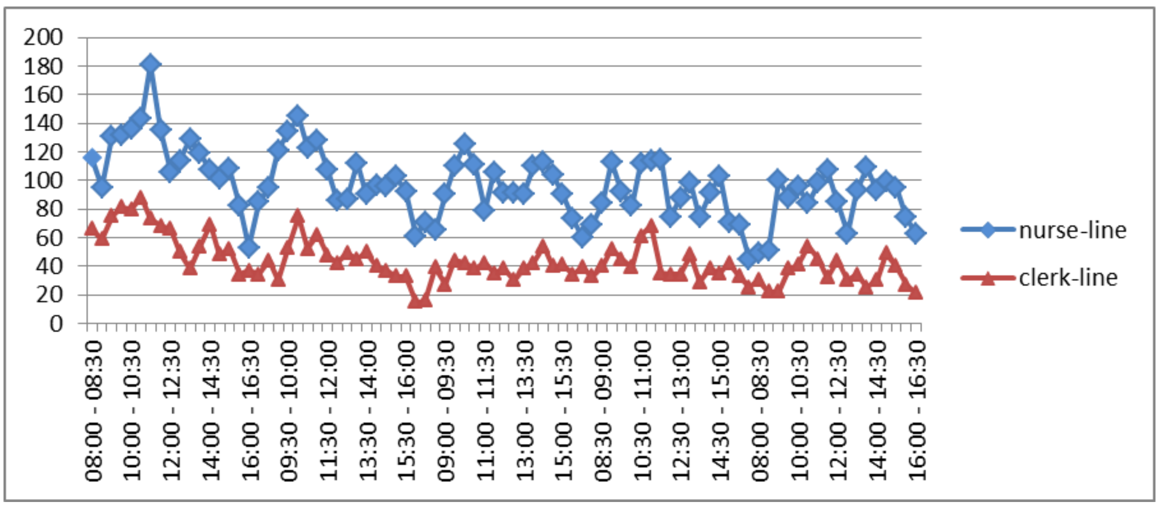

Figure 4.5: Predicted inbound calls volume

The outbound calls volume is estimated from the patient status records. There are three kinds of outbound calls in this model, outbound calls which should be finished by clerk in one day, and outbound calls which should be processed by nurse in one day and two days. Figure 4.6 shows the estimated outbound calls volume in the target week. The outbound calls service time is also exponentially distributed with $(\mu_1'')^{-1}$ = $(\mu_2'')^{-1}$ = 3 mins.

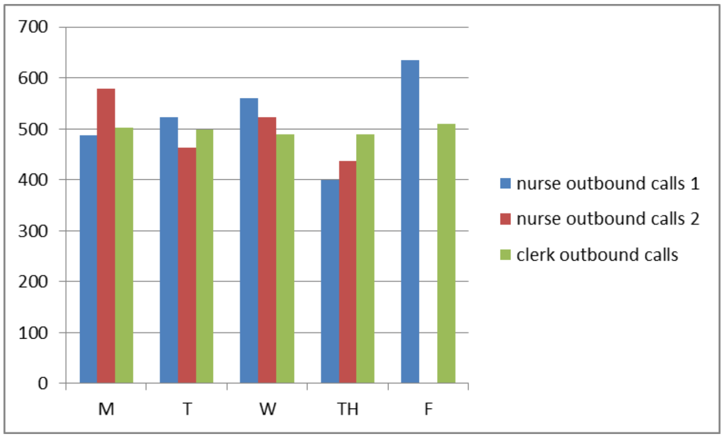

Figure 4.6: Estimated outbound calls volume

Each kind of operators was separated into 2 shifts every day. For nurse operators, each shift has 6 hours for calls related work, two hours for meeting or training and a half hour for lunch. One meeting or training usually last one hour. So the meeting or training time should be consecutive. For clerk operators, each shift works 8 hours for calls related work and a half hour for lunch. More operators should be assigned when the inbound calls come to the peak level. Considering the real situation, the shifts set is shown in table 4.1.

Table 4.1: Shift scheduling set

Operators	Shifts	8:00		9:00		10:00		11:00		12:00		13:00		14:00		15:00		16:00	X_{oi}
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
Nurse	2	1	1	1	1	1	1	1	1	/	1	1	1	1					X_{11}
				1	1	1	1	1	1	1	1	/			1	1	1	1	X_{12}
Clerk	2	1	1	1	1	1	1	1	1	/	1	1	1	1	1	1	1	1	X_{21}
		1	1	1	1	1	1	1	1	1	/	1	1	1	1	1	1	1	X_{22}

("/" lunch time)

The daily labor cost is set identical, \$250/day for nurses and \$150/day for clerks based on Michigan's average wage. Shift planning results under different capacity utilization are shown in table 4.2.

Table 4.2: Shift planning results with inbound calls transfer

Utilization	Operators	Shifts set (X_{o1}, X_{o2})	Total	Labor

	Type	M	T	W	TH	F		Cost
95%	Nurse	(14, 12)	(13, 13)	(13, 13)	(13, 13)	(15, 11)	26	
	Clerk	(9, 10)	(12, 7)	(12, 7)	(13, 6)	(10, 9)	19	
90%	Nurse	(14, 12)	(14, 12)	(14, 12)	(14, 12)	(12, 14)	26	$46750
	Clerk	(9, 10)	(12, 7)	(12, 7)	(13, 6)	(11, 8)	19	
85%	Nurse	(14, 12)	(13, 13)	(14, 12)	(13, 13)	(14, 12)	26	
	Clerk	(9, 10)	(12, 7)	(11, 8)	(13, 6)	(10, 9)	19	

In order to see the efficiency of our model, we also run the model under former policy which we do not transfer the inbound calls between different service lines. The shift planning result is shown in the following table 4.3. We can see that the labor cost is decreased by 10.5% compared to the former policy model. This saving will increase when the call center scales up.

Table 4.3: Shift planning results without inbound calls transfer

Utilization	Operators Type	Shifts set (X_{o1}, X_{o2})					Total	Labor Cost
		M	T	W	TH	F		
95%	Nurse	(17, 14)	(15, 16)	(15, 16)	(17, 14)	(15, 16)	31	
	Clerk	(8, 10)	(11, 7)	(12, 6)	(9, 9)	(5, 13)	18	
90%	Nurse	(17, 14)	(15, 16)	(17, 14)	(17, 14)	(15, 16)	31	$52250
	Clerk	(8, 10)	(11, 7)	(12, 6)	(12, 6)	(11, 7)	18	
85%	Nurse	(17, 14)	(16, 15)	(15, 16)	(17, 14)	(18, 13)	31	
	Clerk	(8, 10)	(11, 7)	(12, 6)	(12, 6)	(11, 7)	18	

4.7 Conclusions

In this paper, we have considered an efficient cost-skill based staffing shift planning model for hospital call centers with multiple kinds of inbound calls and multiple kinds of outbound calls. The inbound calls can be routed to different operators targeting cost savings.

Several preliminary works are carried out before the mathematical model is developed. Routing policy is set up according to the calls content and operators cost difference. Erlang C queuing model and extensive simulation are used to determine the minimum numbers of operators in each time interval according to the inbound calls arrival rate. Routing policy of dynamically blending inbound and outbound calls is evaluated and the system utilization is quantified.

The operators are full time employees and quantity is constant. The objective of the mathematical model is to minimize the total labor cost. Minimum numbers of operators, system utilization and the time limits for completion of multiple kinds of outbound calls are used as the constraints for outbound calls workload allocation. Numerical results show that the model is effective to decrease labor cost.

CHAPTER 5: CONCLUSIONS AND FUTURE RESEARCH

Hospital call center is an increasingly important service channel for hospital. More and more kinds of medical services are planned and delivered through hospital call center centers to substitute or coordinate the traditional clinic activity. There are great potential benefits using hospital call centers to decrease healthcare cost, improve service quality and increase patients' satisfaction. Hospital call centers is especially useful for non-profit hospitals to decrease cost by decreasing clinic workload without service quality reduction.

The new characteristic of operations in hospital call centers includes multiple kinds of service, constant operators as well as inbound and outbound calls blending. The objective is to decrease the operating cost and increase service quality. Labor cost constitutes up to 70% of operating cost, so increasing labor occupancy is the main method to decrease the operating cost.

In the first stage, we consider the routing policy for inbound and outbound calls. The objective is to improve the system utilization and the constraints include the requirements of service quality and operators' quantity. Practical staffing assignment methods, separating and mixing staffing policy are presented and evaluated. Erlang C queuing model is used to decide the minimum number of operators under inbound calls service quality constraints. Theoretical analysis and numerical experiments illustrate that through dynamically assigning the inbound and outbound calls to operators under optimal threshold policy, mixing staffing policy is more efficient to balance the system utilization and service quality. Numerical experiments based on real-life data demonstrate how this method can be applied in practice.

In the following part, we study the staffing shift planning problem based on the inbound and outbound calls routing policies. We developed two mathematical programming models, one for homogenous operators and one for heterogeneous operators. When the operators are homogenous, we dynamically assign inbound or outbound calls. When operators are heterogeneous, we transfer inbound calls between different kinds of operators as well as dynamically assign inbound and outbound calls. The objective is to minimize the staffing cost, by deciding the shift setting and workload allocation. The inbound calls service level and staffing utilization are taken into consideration in the constraints. Numerical experiments based on actual operational data are included. Results show that the model is effective to optimize the shift planning and reduce the call centers' cost.

The future research includes several aspects. The first thing is to improve the model inputs. The parameters we use are from current operating data of the sample hospital call center and it fits the sample hospital's requirements. We need consider how to make the policies robust when the parameters vary in a wider range. The second task is to continue model refinement and validation. In this way, we can decrease the gap between model and real situation. Finally, the call centers we consider in this research is small or medium size. In the future research, we will study the way to extend this model to large scale systems.

REFERENCES

Adan, I. and Weiss, G., (2012). Exact fcfs matching rates for two infinite multitype sequences. Operations Research, 60(2) 475-489.

Aksin, Z., Armony, M. and Mehrotra, V., (2007). The modern call center: A multi‐disciplinary perspective on operations management research. Production and Operations Management, 16(6) 665-688.

Aktekin, T. and Soyer, R., (2011). Call center arrival modeling: A bayesian state‐space approach. Naval Research Logistics (NRL), 58(1) 28-42.

Andradóttir, S., Ayhan, H. and Down, D. G., (2013). Design principles for flexible systems. Production and Operations Management, 22(5) 1144-1156.

Armony, M. and Maglaras, C., (2004). Contact centers with a call-back option and real-time delay information. Operations Research, 52(4) 527-545.

Ásgeirsson, E. I., (2014). Bridging the gap between self schedules and feasible schedules in staff scheduling. Annals of Operations Research, 218(1) 51-69.

Baron, O. and Milner, J., (2009). Staffing to maximize profit for call centers with alternate service-level agreements. Operations Research, 57(3) 685-700.

Batta, R., Berman, O. and Wang, Q., (2007). Balancing staffing and switching costs in a service center with flexible servers. European journal of operational research, 177(2) 924-938.

Bostrom, J., Caldwell, J., Mcguire, K. and Everson, D., (1996). Telephone follow-up after discharge from the hospital: Does it make a difference? Applied Nursing Research, 9(2) 47-52.

Bowles, K. H., Holland, D. E. and Horowitz, D. A., (2009). A comparison of in-person home care, home care with telephone contact and home care with telemonitoring for disease management. Journal of Telemedicine and Telecare, 15(7) 344-350.

Cordone, R., Piselli, A., Ravizza, P. and Righini, G., (2011). Optimization of multi-skill call centers contracts and work-shifts. Service Science, 3(1) 67-81.

Darkins, A., Ryan, P., Kobb, R., Foster, L., Edmonson, E., Wakefield, B. and Lancaster, A. E., (2008). Care coordination/home telehealth: The systematic implementation of health informatics, home telehealth, and disease management to support the care of veteran patients with chronic conditions. Telemedicine and e-Health, 14(10) 1118-1126.

Deslauriers, A., L'ecuyer, P., Pichitlamken, J., Ingolfsson, A. and Avramidis, A. N., (2007). Markov chain models of a telephone call center with call blending. Computers & operations research, 34(6) 1616-1645.

Dietz, D. C., (2011). Practical scheduling for call center operations. Omega, 39(5) 550-557.

Erlang, A. K., (1909). The theory of probabilities and telephone conversations. Nyt Tidsskrift for Matematik B, 20(6) 87-98.

Gans, N., Koole, G. and Mandelbaum, A., (2003). Telephone call centers: Tutorial, review, and research prospects. Manufacturing & Service Operations Management, 5(2) 79-141.

Gariti, P., Alterman, A. I., Holub-Beyer, E., Volpicelli, J. R., Prentice, N. and O'brien, C. P., (1995). Effects of an appointment reminder call on patient show rates. Journal of substance abuse treatment, 12(3) 207-212.

Giordano, A., Scalvini, S., Zanelli, E., Corrà, U., Gl, L., Ricci, V., Baiardi, P. and Glisenti, F., (2009). Multicenter randomised trial on home-based telemanagement to prevent hospital readmission of patients with chronic heart failure. International journal of cardiology, 131(2) 192-199.

Goldberg, Y., Ritov, Y. A. and Mandelbaum, A., (2014). Predicting the continuation of a function with applications to call center data. Journal of Statistical Planning and Inference, 147(2014) 53-65.

Grabowski, D. C. and O’malley, A. J., (2014). Use of telemedicine can reduce hospitalizations of nursing home residents and generate savings for medicare. Health Affairs, 33(2) 244-250.

Green, L. V., Kolesar, P. J. and Soares, J., (2001). Improving the sipp approach for staffing service systems that have cyclic demands. Operations Research, 49(4) 549-564.

Haynes, J. M. and Sweeney, E. L., (2006). The effect of telephone appointment-reminder calls on outpatient absenteeism in a pulmonary function laboratory. Respiratory care, 51(1) 36-39.

Ibrahim, R. and L'ecuyer, P., (2013). Forecasting call center arrivals: Fixed-effects, mixed-effects, and bivariate models. Manufacturing & Service Operations Management, 15(1) 72-85.

Janssen, A., Van Leeuwaarden, J. S. and Zwart, B., (2011). Refining square-root safety staffing by expanding erlang c. Operations Research, 59(6) 1512-1522.

Jouini, O., Koole, G. and Roubos, A., (2013). Performance indicators for call centers with impatient customers. Iie Transactions, 45(3) 341-354.

Keblis, M. and Chen, M., (2006). Improving customer service operations at amazon. Com. Interfaces, 36(5) 433-445.

Kim, J. W. and Ha, S. H., (2010). Consecutive staffing solution using simulation in the contact center. Industrial Management & Data Systems, 110(5) 718-730.

Kirsch, S. D., Wilson, L. S., Harkins, M., Albin, D. and Del Beccaro, M. A., (2014). Feasibility of using a pediatric call center as part of a quality improvement effort to prevent hospital readmission. Journal of pediatric nursing, 30(2015) 333-337.

Kocaga, Y. L., Armony, M. and Ward, A. R., (2014). Staffing call centers with uncertain arrival rates and co - sourcing. Production and Operations Management, 0(0) xxxx-xxxx.

Koole, G. and Mandelbaum, A., (2002). Queueing models of call centers: An introduction. Annals of Operations Research, 113(1-4) 41-59.

Legros, B., Jouini, O. and Dallery, Y., (2015). A flexible architecture for call centers with skill-based routing. International Journal of Production Economics, 159(2015) 192-207.

Liao, S., Koole, G., Van Delft, C. and Jouini, O., (2012). Staffing a call center with uncertain non-stationary arrival rate and flexibility. OR spectrum, 34(3) 691-721.

Mandelbaum, A. and Zeltyn, S., (2009). Staffing many-server queues with impatient customers: Constraint satisfaction in call centers. Operations Research, 57(5) 1189-1205.

Mehrotra, V. and Fama, J. (2003). Call center simulations: Call center simulation modeling: Methods, challenges, and opportunities. Winter Simulation Conference.

Mehrotra, V., Ozlük, O. and Saltzman, R., (2010). Intelligent procedures for intra‐day updating of call center agent schedules. Production and Operations Management, 19(3) 353-367.

Nah, J. E. and Kim, S., (2013). Workforce planning and deployment for a hospital reservation call center with abandonment cost and multiple tasks. Computers & Industrial Engineering, 65(2) 297-309.

Pang, G. and Perry, O., (2014). A logarithmic safety staffing rule for contact centers with call blending. Management Science, 61(1) 73-91.

Perry, O. and Whitt, W., (2011). A fluid approximation for service systems responding to unexpected overloads. Operations research, 59(5) 1159-1170.

Pichitlamken, J., Deslauriers, A., L'ecuyer, P. and Avramidis, A. N. (2003). Modelling and simulation of a telephone call center. IEEE. 2: 1805-1812

Ren, Z. J. and Zhou, Y.-P., (2008). Call center outsourcing: Coordinating staffing level and service quality. Management Science, 54(2) 369-383.

Robbins, T. R. and Harrison, T. P., (2010). A stochastic programming model for scheduling call centers with global service level agreements. European Journal of Operational Research, 207(3) 1608-1619.

Roubos, A., Bhulai, S. and Koole, G. (2011). Flexible staffing for call centers with nonstationary arrival rates. Submitted.

Roubos, A., Koole, G. and Stolletz, R., (2012). Service-level variability of inbound call centers. Manufacturing & Service Operations Management, 14(3) 402-413.

Roubos, D. and Bhulai, S., (2012). Approximate dynamic programming techniques for skill-based routing in call centers. Probability in the Engineering and Informational Sciences, 26(4) 581-591.

Roubos, D. and Bhulai, S., (2010). Approximate dynamic programming techniques for the control of time-varying queuing systems applied to call centers with abandonments and retrials. Probability in the Engineering and Informational Sciences, 24(1) 27-45.

Samuelson, D. A., (1999). Predictive dialing for outbound telephone call centers. Interfaces, 29(5) 66-81.

Schrieck, J. C., Akşin, Z. and Chevalier, P., (2014). Peakedness ‐ based staffing for call center outsourcing. Production and Operations Management, 23(3) 504-524.

Stubbings, T., Miller, C., Humphries, T. L., Nelson, K. M. and Helling, D. K., (2005). Telepharmacy in a health maintenance organization. Am J Health Syst Pharm, 62(4) 406-410.

Tarakci, H., Ozdemir, Z. and Sharafali, M., (2009). On the staffing policy and technology investment in a specialty hospital offering telemedicine. Decision Support Systems, 46(2) 468-480.

Taylor, J. W., (2012). Density forecasting of intraday call center arrivals using models based on exponential smoothing. Management Science, 58(3) 534-549.

Van Zyl, M. A., Brown, L. L. and Pahl, K., (2015). Using a call center to encourage linkage to care following mobile hiv counseling and testing. AIDS care, ahead-of-print

Yao, C. and Cassandras, C. G., (2011). Perturbation analysis and optimization of multiclass multiobjective stochastic flow models. Discrete Event Dynamic Systems, 21(2) 219-256.

Zhang, B., Van Leeuwaarden, J. S. and Zwart, B., (2012). Staffing call centers with impatient customers: Refinements to many-server asymptotics. Operations Research, 60(2) 461-474.

Zhao, Y., Yang, K. and Yu, S., (2014). Efficient staffing for hospital call centers with service blending. Proc. of the IIE Conference, Montreal, Canada.

ABSTRACT

MODELING AND OPTIMIZATION OF NON-PROFIT HOSPITAL CALL CENTERS WITH SERVICE BLENDING

by

YANLI ZHAO

May 2015

Advisor: Dr. Kai Yang

Major: Industrial Engineering

Degree: Doctor of Philosophy

This dissertation focuses on the operations problems in non-profit hospital call centers with inbound and outbound calls service blending.

First, the routing policy for inbound and outbound calls is considered. The objective is to improve the system utilization under constraints of service quality and operators' quantity. A collection of practical staffing assignment methods, separating and mixing staffing policy are evaluated. Erlang C queuing model is used to decide the minimum number of operators required by inbound calls. Theoretical analysis and numerical experiments illustrate that through dynamically assigning the inbound and outbound calls to operators under optimal threshold policy, mixing staffing policy is efficient to balance the system utilization and service quality. Numerical experiments based on real-life data demonstrate how this method can be applied in practice.

Second, we study the staffing shift planning problem based on the inbound and outbound calls routing policies. A mathematical programming model is developed, based on a hospital call center with one kind of inbound calls and multiple kinds of outbound calls. The objective is to minimize the staffing numbers, by deciding the shift setting and workload allocation. The inbound calls service level and staffing utilization are taken into consideration in the constraints. Numerical experiments based on actual operational data are included. Results show that the model is effective to optimize the shift planning and hence reduce the call centers' cost.

Third, we model the staffing shift planning problem for a hospital call center with two kinds of service lines. Each kind of service is delivered through both inbound calls and outbound calls. The inbound calls can be transferred between these two service lines. A mathematical programming model is developed. The objective is to minimize the staffing cost, by deciding the shift setting and workload allocation. The inbound calls service level and staffing utilization are taken into consideration in the constraints. Numerical experiments are carried out based on actual operational data. Results show that the model is effective to reduce the call centers' labor cost.

AUTOBIOGRAPHICAL STATEMENT

Yanli Zhao was born in Jiaozuo, Henan province of China. She received her Bachelor of Science and Master of Science degree in Manufacturing Engineering in 2004 and 2007, respectively, from Beijing University of Aeronautics and Astronautics, China. Then she did one year study in Mines ParisTech as a Post-Master. She joined the PH.D. program in the Department of Industrial & Systems Engineering, Wayne State University in 2009. Her research areas are Operations Research, Data Analytics and Statistics.