ESSAYS ON RESOURCE FLEXIBILITIES IN SERVICES

Adelina Gnanlet Amal Samuel

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Kenan-Flagler Business School (Operations, Technology and Innovation Management).

Chapel Hill
2007

Approved by:

Dr. Wendell Gilland, Committee Chair
Dr. Barbara Mark, Committee Member
Dr. Ann Marucheck, Committee Member
Dr. Aleda Roth, Committee Member
Dr. Brian Tomlin, Committee Member

ABSTRACT

ADELINA GNANLET AMAL SAMUEL: ESSAYS ON RESOURCE FLEXIBILITIES IN SERVICES
(Under the direction of Dr. Wendell Gilland)

Service firms unable to store services, constantly search for strategies to minimize supply-demand imbalance in resources. Escalating costs, nursing shortages, and process inefficiencies have led the US healthcare system to operate at near-full capacity and maintain high utilization levels. Even though hospitals often function at full capacity, they are expected to treat all patients due to legal, social and moral obligations. Hospitals use a combination of resource flexibility strategies to reduce supply-demand imbalance due to demand variability. In this dissertation, we analyze the benefits of using flexibility strategies on critical resources such as labor (nursing staff), equipment (bed spaces), and patient assignment to contain demand variability in the short-term. Staffing flexibility, often referred to as cross-training, is widely used in many service firms to increase labor utilization, morale, and job enrichment. Although cross-training is cost-efficient, cross-trained nurses are less productive in new units. In the first essay, we analyze the cost savings vs. productivity changes when using staffing flexibility in hospitals. We find that productivity of cross-trained nursing staff has a significant effect on the optimal amount of cross-training. There is also an interaction effect between productivity of cross-trained nurses, cross-training cost, and demand variability. In the second essay, we add a second resource flexibility strategy referred to as demand upgrades. Patients who are typically treated in a simple unit can be upgraded to a complex unit when beds are unavailable in the simple unit. We compare the impact of these two resource flexibility strategies on a hospital's ability to meet stochastic patient demand. We also look at the impact of the timing of capacity and staffing decisions on system performance. In the third essay, we develop a resource flexibility strategy framework that utilizes internal

resources (labor, equipment, and patient assignment) to efficiently manage different levels of demand variability. The insights, issues and implications needed for managers to implement flexibility models are also presented in the third essay. Thus, this dissertation has analyzed the benefit, trade-offs, limitations, and outcomes in implementing critical resource flexibility strategies in hospitals to minimize supply-demand imbalance and address short-term demand variability.

ACKNOWLEDGMENTS

I would like to acknowledge everyone who assisted me during my doctoral work. I would especially like to thank my advisor, Wendell Gilland, for his generous time and commitment. Throughout my doctoral program he encouraged me to develop independent thinking and research skills. He continually stimulated my analytical thinking and greatly assisted me with academic writing.

I am also very grateful for having an exceptional doctoral committee and wish to thank Aleda Roth, Ann Marucheck, Barbara Mark, and Brian Tomlin for their timely guidance and support. I wish to thank all nursing personnel, directors and staff at UNC Hospital who shared their thoughts and opinions with me to accurately develop my models.

I am extremely thankful for my PhD colleagues, especially Muge Yayla-Kullu, Yimin Wang, Sriram Narayanan and Shankar Viswanathan, who consistently provided emotional support and advice throughout my four years of stay at Chapel Hill. I am exceedingly grateful and thankful to Lord God for my family: my wonderful parents Mr. and Mrs. John Samuel, my brother Benedict Samuel, my grandparents Mr. (Late) and Mrs. A.C. Solomon, without whose continued encouragement, support, nudging, and prayers, I could have never completed my doctoral work. Finally, I am deeply indebted to our Lord and Savior Jesus Christ for showering his abundant grace and mercies on me throughout my professional education.

TABLE OF CONTENTS

LIST OF TABLES x

LIST OF FIGURES xi

1 INTRODUCTION 1

1.1 Introduction and Motivation . . . 2

1.2 Essay 1 - Impact of Productivity on Optimal Cross-Training Decisions . . 4

1.3 Essay 2 - Sequential and Simultaneous Decision Making under Dual Resource Flexibilities . . . 6

1.4 Essay 3 - Resource Flexibility Framework and Managerial Implications . 7

2 IMPACT OF PRODUCTIVITY ON OPTIMAL CROSS-TRAINING DECISIONS 10

2.1 Introduction . . . 11

2.2 Literature Review . . . 14

2.3 Problem Formulation . . . 17

2.3.1 Notation . . . 19

2.3.2 Model Formulation . . . 21

2.4 Solving Two-Stage Stochastic Programming . . . 23

2.4.1 Second stage - Allocation 24

2.4.2 First stage - Optimal cross-training 30

2.5 Numerical Analysis 34

2.6 Conclusion 39

3 SEQUENTIAL AND SIMULTANEOUS DECISION MAKING UNDER DUAL RESOURCE FLEXIBILITIES 45

3.1 Introduction and Motivation 46

3.2 Literature Review 49

3.2.1 Manufacturing flexibility 50

3.2.2 Human resource flexibility 52

3.3 Problem Definition 54

3.3.1 Configurations 55

3.3.2 Timeline for decision making 58

3.4 Model Formulation and Analysis 59

3.4.1 Configuration 1 : No flexibility 63

3.4.2 Configuration 2 : Demand Upgrades 67

3.4.3 Configuration 3 : Staffing Flexibility 74

3.4.4 Configuration 4: Demand Upgrades and Staffing Flexibility 84

3.5 Numerical Experiments 96

3.5.1 Parameter Values 96

3.5.2 Results 97

3.6 Conclusions . . . 105

4 RESOURCE FLEXIBILITY FRAMEWORK AND MANAGERIAL IMPLICATIONS 107

4.1 Introduction . . . 108

4.2 Demand variability . . . 121

4.3 Degree of capacity flexibility . . . 123

4.4 Degree of staffing flexibility . . . 128

4.4.1 Float nurses . . . 129

4.4.2 Flex team or per diem nurses . . . 129

4.4.3 Travelers . . . 130

4.4.4 Contract nurses . . . 131

4.4.5 Implications . . . 131

4.5 Estimation of relevant costs . . . 133

4.5.1 Staffing costs . . . 134

4.5.2 Capacity costs . . . 137

4.6 Timing of staffing and capacity decisions . . . 137

4.7 Performance/productivity measurement . . . 139

4.7.1 Literature review . . . 139

4.7.2 Implications . . . 142

4.8 Managerial Insights . . . 143

4.9 Conclusion . . . 146

5 CONCLUSION 147

APPENDIX A - Configuration 2 **152**

A.1 Sequential Decision Making 152

A.2 Simultaneous Decision Making 153

APPENDIX B - Configuration 3 **155**

B.1 Sequential Decision Making - Case 1 155

B.2 Sequential Decision Making - Case 2 155

B.3 Simultaneous Decision Making 156

APPENDIX C - Configuration 4 **158**

C.1 Sequential Decision Making - Case 1 158

C.2 Sequential Decision Making - Case 2 159

C.3 Simultaneous Decision Making 160

REFERENCES **162**

LIST OF TABLES

2.1 Alternative cases used in numerical analysis 35

2.2 Parameter levels for numerical analysis 35

3.1 Values for Cost Parameters 96

3.2 Values for Demand Parameters 97

3.3 Sequential vs Simultaneous Decision Making - Average Percentage Cost Improvement 98

3.4 Sequential Decision Making Case 1 - Average Percentage Cost Improvement 99

3.5 Sequential Decision Making Case 2 - Average Percentage Cost Improvement 99

3.6 Simultaneous Decision Making - Average Percentage Cost Improvement . 99

3.7 Effect of Cost Parameters on the Decision Variables in Configuration 3 under Simultaneous Decision Making 103

4.1 Lower acuity units - Clinical grouping for medicine floor 127

4.2 Stepdown units - Clinical grouping intermediate care 127

4.3 ICU units - Clinical grouping for critical care 128

4.4 Pay rate for flex team 130

LIST OF FIGURES

2.1 Time Line for the Model 17

2.2 Network Structure 19

2.3 Regions for Second Stage Solution 24

2.4 Effect of training cost on optimal amount of cross-training with CV=0.2 41

2.5 Effect of training cost on optimal amount of cross-training with CV=0.5 41

2.6 Effect of training cost on optimal amount of cross-training with CV=0.7 41

2.7 Effect of demand variation in unit 1 and 2 on optimal amount cross-training at low training cost 42

2.8 Effect of demand variation in unit 2 on optimal amount cross-training at low training cost 42

2.9 Effect of demand variation in unit 1 and 2 on optimal amount cross-training at high training cost 42

2.10 Effect of demand variation in unit 2 on optimal amount cross-training at high training cost 43

2.11 Effect of contract cost on optimal amount of cross-training 43

2.12 Effect of training cost on optimal amount of cross-training when demand matches supply 43

2.13 Effect of demand variation on optimal amount of cross-training when demand matches supply 44

3.1 Four Configurations 55

3.2 Timeline for Configurations 59

3.3 Timeline for Configurations 1 and 2 63

3.4 Timeline for Configurations 3 and 4 74

3.5 Configuration 2 Simultaneous Decision Making - Effect of increase in penalty cost in unit 1 103

3.6 Configuration 4 Simultaneous Decision Making - Effect of increase in penalty cost in unit 2 104

4.1 Resource Flexibility Strategy Framework for Demand Variability 112

4.2 Demand variability about mean demand in short-term 114

4.3 Demand variability about mean demand with trend in long-term 114

4.4 Measurement Model for Productivity 144

CHAPTER 1

INTRODUCTION

1.1 Introduction and Motivation

US hospitals operate at near-full capacity most of the time due to escalating operating costs, shortage of nursing staff and inefficiencies in the process. Even though hospitals often function at full capacity, they are expected to treat all patients despite demand uncertainty. Demand variability in hospitals is more significant than demand variability in most manufacturing firms or other service firms for two main reasons. Hospitals running at near-full capacity have to accommodate patients seeking care due to legal, social and moral obligations. Additionally, health care services cannot be inventoried, so hospitals use a combination of resource flexibility strategies to reduce supply-demand imbalance. For these two reasons, it is essential for hospitals to understand the conditions under which various types of resource flexibilities are best used. In this dissertation, we analyze the benefits of using flexibility strategies on critical resources such as labor (nursing staff), equipment (bed spaces), and patient assignment, to address demand variability in the short-term. We also develop a resource flexibility strategy framework that utilizes internal resources (labor, equipment, and patient assignment) both in the long-term and short-term to manage different levels of demand variability.

Due to their inability to hold inventory, service institutions face an especially difficult challenge in matching supply and demand. High attrition rates in call centers as well as shortage of staff in hotels, restaurants and hospitals, increase the difficulty of effectively meeting customer demand. In order to alleviate the staffing shortage and effectively meet customer requirements, these service organizations began cross-training employees to work in more than one job role. The primary objective for cross-training is increased efficiency, although the following benefits are also obtained (Maggard and Globerson, 1986; Altimier and Sanders, 1999; Dela Cruz, 2003; Bergman, 1994).

- Lower absenteeism
- Increased job enrichment and job satisfaction

- Scheduling flexibility for the manager
- More marketability for employees
- Flexible organizational structure

Cross-training programs have led to substantial cost savings. "With three hospitals and 10 perinatal nursing units, nurse leaders developed a cross-training program that lead to increased morale, decreased turnover, and saved close to $1 million in just a year (Altimier and Sanders, 1999)." In a Michigan based chain of 23 stir-fry restaurants, "turnover has dropped to 30% from 150%. Productivity has increased, resulting in the elimination of two positions during non-peak period (Bertagnoli, 2004)."

In this dissertation, we examine the benefits and trades-offs in using staffing flexibility (cross-training), and also analyze the interaction between capacity flexibility and other operational decisions in a hospital setting. While it is cost effective to implement cross-training programs, the service managers should be aware of the outcomes, limitations and trade-offs involved in employing staffing flexibility as a course of action to meet uncertain demand. Hospitals increasingly use nurses from overstaffed units to meet the needs of understaffed units. In-depth interviews with practitioners at a south-east academic medical center, as well as health care and operations management literature, point out that cross-trained nurses are not as effective in the new unit as in their home (assigned) unit. Due to different patient populations and varying protocols in the new unit, cross-trained nurses are frequently not as productive as in their home unit. This dissertation addresses the following three broad research questions: What is the impact of varying productivity of cross-trained staff on system performance? How does staffing flexibility coordinate with other resource flexibility strategy and operational decisions? Under what circumstances should different types of resource flexibility strategies be used?

In the first essay of my dissertation, we analytically obtain the optimal amount of cross-training while accounting for productivity of cross-trained staff. Hospital managers co-ordinate two critical resources, (nurses and bed spaces) to meet patient needs. If one or both of these resources are not available, the patient cannot be treated. When a bed space is not available, the patient may be upgraded from a lower unit to a more sophisticated unit. When a dedicated nurse is not available, flexible nurses and/or travel nurses are used. The second essay of my dissertation ascertains the benefit of using two types of flexibility (staffing flexibility and capacity flexibility) while managing key resources (nursing staff and bed spaces). It compares the consequences of determining resource levels sequentially (staggered over two time periods) and simultaneously (in a single time period). The first two essays analytically evaluate the benefits and trade-offs in using two types of resource flexibility to meet short-term demand variability. In the third essay, we develop a resource flexibility strategy framework that categorizes flexibility strategies for three critical resources (labor, equipment, and patient assignment) into short-term and long-term strategies used to meet demand variability. The third essay also presents implications, insights and issues that managers need to consider when they implement resource flexibility models.

In the following section, a brief overview of each of these three essays is presented.

1.2 Essay 1 - Impact of Productivity on Optimal Cross-Training Decisions

One of the most important hospital resources, and one resource that has been among the most difficult to supply in recent years, is nursing staff. To cost effectively meet uncertain demand, hospitals must either hire more contract nurses from external agencies, incurring higher contract cost, or train more flexible nurses, incurring higher training cost. There is a trade-off between hiring contract nurses and training flexible nurses. In this paper, we address the following questions: What is the appropriate amount of cross-training (flexibility) required so that the service level is maintained and costs are minimized? How does flexibility change when there is demand variability in the units? What is the impact of costs on the amount of flexibility? What is the effect of productivity of flexible nurses on the amount of cross-training?

We optimize the amount of cross-training for nurses in two hospital units. At the beginning of the planning horizon, hospital administrators decide on the number of nurses who should be cross-trained. In the second stage, the nurses are allocated dynamically based on the stochastic patient requirements (demand) in each unit. The model is formulated as two-stage stochastic programming with recourse in the second stage. We derive a closed form expression for the optimal amount of cross-training in two units when demand follows a general continuous distribution. Our model considers cross-training costs, contract cost and productivity of flexible nurses. Campbell (1999) focuses on optimizing the allocation of cross-trained workers considering the worker capabilities. In our paper, as an extension of Campbell (1999), we solve an integrated model of planning and scheduling (allocation) of nurse staffing over a planning horizon.

When cost of cross-training is high, an increase in productivity leads to an increase in the amount of cross-training. When the cost of cross-training is relatively low, however, there is a productivity level beyond which further increases in productivity reduce the amount of cross-training. Above this threshold, the productivity of cross-trained nurses allows the benefits of flexibility to be obtained with fewer cross-trained nurses.

For low cross-training productivity, more demand variability leads to less cross-training. For high cross-training productivity, the effect of demand variability on the amount of cross-training depends on cross-training cost. When cross-training cost is high, more variability continues to cause less cross-training. For low cross-training cost, however, an increase in demand variability leads to more cross-training. This essay, titled "Impact of Productivity on Cross-training Decisions," comprises Chapter 2.

1.3 Essay 2 - Sequential and Simultaneous Decision Making under Dual Resource Flexibilities

Hospitals need two critical resources, nursing staff and beds, at the right time and right quantity to ensure the highest quality of care to patients. Hospital managers try to efficiently utilize these resources by employing two types of resource flexibility - staffing flexibility and demand upgrades - so that the maximum number of patients can be treated. In this paper, we compare four flexibility configurations, ranging from no flexibility of critical resources to full flexibility of both resources (staffing flexibility and demand upgrades), to meet stochastic demand at minimum cost. Under demand upgrades, patients from a lower unit can be accommodated in a higher/more sophisticated unit when beds are unavailable in the lower unit. Under staffing flexibility, flexible nurses cross-trained in more than one skill are used in addition to dedicated and travel nurses. The capacity (number of beds) and staffing (number of dedicated and flexible nurses) decisions can be made over a single time period (simultaneous decision making) or over multiple time periods (sequential decision making). In this paper, we address the following questions: For each flexibility configuration under sequential and simultaneous decision making, what is the optimal resource level required to meet stochastic demand at minimum cost? Is one type of flexibility (e.g. demand upgrades) better than the other type of flexibility (e.g. staffing flexibility)? What is the effect of timing of decisions on type of flexibility and system performance?

We use two-stage stochastic programming to find optimal capacity and staffing levels for each flexibility configuration under sequential and simultaneous decision making in two non-homogenous hospital units that face continuous, general demand distribution. We find that the benefit obtained in using staffing flexibility on average is higher than the benefit of using demand upgrades. The two types of flexibilities complement each other and have a positive interaction effect between them. The benefits of cross-training can be largely realized even if capacity and staffing levels have been determined prior to the establishment of a cross-training initiative. The timing of decisions is important, but timing of decisions and type of flexibility has an independent effect on system performance. This essay, titled "Simultaneous and Sequential Decision Making under Dual Resource Flexibilities," is presented as Chapter 3.

1.4 Essay 3 - Resource Flexibility Framework and Managerial Implications

In essay 1 and essay 2, we analyze the benefits of implementing two types of resource flexibility (staffing flexibility and demand upgrades) frequently used by hospitals as short-term strategies to gain operational flexibility and manage demand variability. In the third essay, we position those two types of resource flexibility strategies using Jack and Powers (2004) volume flexible strategy framework and develop a resource flexibility strategy framework. We discuss some of the propositions in Jack and Powers (2004) to see if it is consistent with our resource flexibility framework. We also address the practical issues that hospital administrators face when implementing the resource flexibility models in essay 1 and essay 2.

Demand variability causes supply-demand imbalance in the short-term. The supply-demand imbalance created by random fluctuations can be reduced by using two approaches: increasing the availability of supply to treat excess patients and/or reducing patient demand using demand management strategies. Although there are other supply-side resources that hospitals can utilize, we focus on two important resources required to meet patient demand: nursing staff and bedspaces, commonly categorized as labor and equipment resources. When demand exceeds existing capacity, operational (short-term) or strategic (long-term) flexibility of staffing and/or equipment are used to absorb demand variability. On the demand-side, demand management strategies such as staggering patient arrivals, transfer of patients, or increase in prices (such as in elective surgeries) are used to reduce supply-demand imbalance.

The latter part of the chapter elaborates on the implementation issues surrounding three operational flexibility strategies discussed in the resource flexibility strategy framework: labor flexibility, equipment flexibility, and demand management strategies. Six important implementation issues pertaining to the analytical models in essay 1 and 2 are explained in this essay using a public, academic medical center. This bridges the gap between the analytical models and their implementation in the health care industry. The key aspects in implementing our models are:

- estimating demand variability for each unit
- understanding the degree of capacity flexibility required to effectively use bedspace

- using staffing flexibility strategies to meet patient needs
- estimating relevant costs incurred in utilizing capacity and staffing flexibilities
- understanding the impact of timing of capacity and staffing decisions
- evaluating productivity of flexible staff

The processes and issues described in each of the sub-sections is compiled from in-depth interviews with practitioners and an extensive literature review. The hospital we have analyzed is a public, academic medical center operated by a south-east state government. Approximately 61,200 patients visited the emergency room (ER) and there were 741,980 clinic visits at the hospitals' 724 bed facility during 2006.

The details of this essay, titled "Resource Flexibility Framework and Managerial Implications," are given in Chapter 4.

The conclusions of this dissertation and an identification of related future work are discussed in Chapter 5.

CHAPTER 2

IMPACT OF PRODUCTIVITY ON OPTIMAL CROSS-TRAINING DECISIONS

2.1 Introduction

Due to their inability to hold inventory, service institutions face an especially difficult challenge in matching supply and demand. In the health care sector, hospital administrators can provide the highest level of service only when the resources required by the patients, both equipment and staff, are supplied at the right time and at the right quantity. One of the most important hospital resources, and one resource that has been among the most difficult to supply in recent years is nursing staff. Traditionally, hospitals handled shortfalls in nursing staff by hiring contract nurses for one or two shifts until the demand stabilizes. Hiring contract nurses from external agencies is expensive compared to regular nurses. In some units, contract nurses are paid twice as much as the regular nurses and in general contract nurses are always paid more than regular nurses. The current nursing shortage exacerbates this problem by increasing the demand for contract nurses. To provide the necessary service level for patients, all required nursing hours must be provided irrespective of associated cost.

In recent years, hospital administrators are under tremendous pressure to cut costs and have resorted to cross-training programs (Lyons, 1992, Siferd and Benton, 1992) for nurses, enabling them to float between units in the same specialization. These units have varying acuity level, but are similar enough to cross train nurses. Cross-training (floating / flexibility) of nurses helps to meet heavy demand in one unit by using nursing hours from another unit where the demand is lean. Many hospitals have reaped financial benefits from successfully implementing cross-training programs.

To cost effectively meet uncertain demand, hospitals must either hire more contract nurses from external agencies, incurring higher contract cost, or train more flexible nurses, incurring higher training cost. There is a trade off between hiring contract nurses and training flexible nurses. This observation leads to the question: What is the appropriate amount of cross-training (flexibility) required so that the service level is maintained and costs are minimized? Flexible nurses are not as productive as the regular nurses in the floated unit because of infrequent use of the skills required in the floated unit. The aspect of productivity plays a vital role when analyzing worker flexibility in both manufacturing and service settings. Yield of flexible machines is analogous to the productivity of flexible workers, but papers concerned with product-machine flexibility assume 100% yield in the allocated machines. In this paper, we explicitly model the productivity of flexible nurses in the floated unit and determine the effect of productivity of flexible nurses on flexibility. The other research questions addressed in this paper include: How does flexibility change when there is demand variability in the units? What is the impact of costs on the amount of flexibility?

In this paper, we optimize the amount of cross-training in multiple units. We formulate a two-stage stochastic programming model with recourse in the second stage to solve this problem. The first stage decision is the amount of cross-training in each unit. When demand is realized in the second stage, nurses are allocated to meet demand such that the resulting costs are minimized. We consider general, continuous demand distributions, and derive a closed form expression for the optimal amount of cross-training.

The analysis shows that at a given level of productivity for flexible nurses, the optimal amount of cross-training in unit i decreases with an increase in cross-training cost in unit

i, and increases with an increase in shortage cost in unit j as expected.

When cost of cross-training is high, increase in productivity leads to increase in amount of cross-training. When the cost of cross-training is relatively low, however, there is a productivity level beyond which further increases in productivity reduces the amount of cross-training. Above this threshold, the productivity of cross-trained nurses allow the benefits of flexibility to be obtained with fewer cross-trained nurses.

For low cross-training productivity, more demand variability leads to less cross-training. For high cross-training productivity, the effect of demand variability on the amount of cross-training depends on cross-training cost. When cross-training cost is high more variability continues to cause less cross-training. For low cross-training cost, however, increases demand variability leads to more cross-training (with high cross-training productivity).

This paper has integrated the planning and scheduling (allocation) phases of nurse staffing associated with utilizing flexible nurses in hospital. So far, literature has focussed independently on nurses planning problem and scheduling (allocation) of flexible nurses problem. In our model, we try to integrate both planning and scheduling phases of nurse staffing problem across a planning horizon.

This paper is organized as follows. §2.2 gives the literature review and discusses how our model differs from the existing literature. §2.3 introduces and formulates our model of cross-training and scheduling decisions. §2.4 analyzes the model as a two-stage stochastic program with recourse and §2.5 shows numerical analysis and highlights its implications for management. §2.6 concludes the paper and discusses possible extensions.

2.2 Literature Review

This section consists of literature related to implementation of cross-training in hospitals, as well as more general papers on multi-skill sharing and serial cross-training. We also review the literature on manufacturing flexibility, and discuss how it differs from our model. Cross-training of nurses in hospitals leads to both financial and non-financial benefits. Wheaton (1996) and Lyons (1992) list additional benefits such as increased job-satisfaction, decreased job stress, and increased marketability of nurses. Li and King (1999) develops a goal programming approach for optimizing the cross-trained staff for sub-divided tasks in health care.

Quite a few papers have been published in work force cross-training. Agnihothri et al. (2003) balances the trade off between customer delay cost and premium for flexibility and models a queueing system to determine the mix of dedicated and cross-trained servers for two job types using simulation, and extend this work to three job types in their 2004 paper (Agnihothri and Mishra, 2004). In our paper, we determine the optimal amount of cross-training in two units when demand follows general distribution. Agnihothri et al. (2003) use discrete values for mix of dedicated and cross-trained servers to study the effect of parameters on the mix ratio. Brusco and Johns (1998) present an ILP to evaluate cross-training configurations for a multi-skilled work force. Brusco et al. (1998) minimizes the total number of labor for two skill class considering productivity. They do not optimize on the level of flexibility but assume that all labors are totally flexible. McClain et al. (2000) show that work-in-process inventory has a significant effect on the productivity of workers when there is work sharing in a serial system. These papers in work force cross-training has not considered the impact of productivity, cross-training costs on optimal amount of cross-training particularly when demand follows general distributions.

Cross-training in manufacturing has widely been used to balance the work load in an assembly system in order to maximize throughput. A recent paper by Hopp et al. (2004) analyzes two different cross-training structures, skill chaining and cherry-picking, for a serial production system. They find that when capacity is fairly imbalanced but variability is low, cherry picking approach can be used. Jordan et al. (2004) evaluate the performance of three cross-training configurations in parallel systems using queueing theory and simulation. They conclude that complete chaining gives the maximum benefit and is also robust. Pinker and Shumsky (2000) analyzes a system with specialist and flexible servers when there is a trade off between the efficiency of the specialist and the quality of the flexible servers. Iravani et al. (2005) finds a structural flexibility index that chooses the best pattern among all the alternative patterns of flexibility in parallel systems without having to evaluate all the patterns. Vairaktarakis and Winch (1999) develops heuristics for scheduling work orders through assembly systems so that the cross-training costs are minimized when multi-skilled workers are used. The above papers deals in finding the best configuration or flexibility structure (level of cross-training) in different scenarios and does not deal in optimizing the amount of cross-training.

Van Mieghem (1998), studies the effect of cost and price differentials on the flexibility of two-product, three-plant system. He maximizes the total revenue by optimizing the capacities for two dedicated-capacity plants and one flexible-capacity plant. The differences between Van Mieghem's model and the model in this paper are given below:

1. Unlike the single flexible resource in Van Mieghem's model, nurses on both units become flexible with cross-training

2. In our paper, the impact of productivity of floated nurses on the amount of cross-training is analyzed.

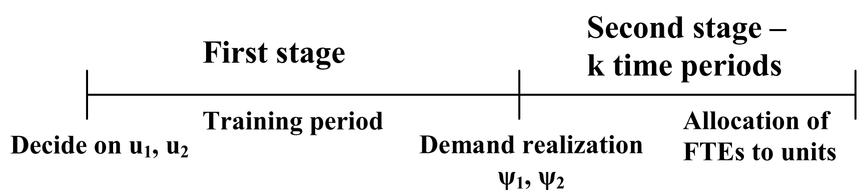
3. Flexible resources are associated with a home unit, and only floated if they are not needed at their home unit.

4. All demand must be met. An additional, more costly resource (contract nurses) is available to overcome capacity deficiencies.

Campbell (1999) develops a model for allocating cross-trained workers to a multi-department service environment. He determines the benefit of cross-training using a simulation study. His model maximizes utility (weighted sum of squared requirement that is satisfied) of meeting the requirements, considering the capabilities of the workers subject to their allocation in different departments and assuming different levels of cross-training (training breadth). Jordan and Graves (1995) expands the model to M products and N plants and analyzes the level of flexibility required to reap significant benefits. They conclude that small additional flexibility is sufficient to get results of total flexibility. Graves and Tomlin (2003), has extended the idea of Jordan and Graves (1995) to multi-stage supply chains.

In this paper, we analyze the effects of cross-training cost, contract cost, productivity and demand variability on the optimal amount of cross-training for two units in a health

FIGURE 2.1: Time Line for the Model



care setting. This analysis can be applied to other service settings, such as hotels, restaurants and call centers.

2.3 Problem Formulation

This section gives the notation used in the model and formulates the model. We model two hospital units, each of which has a pool of nursing resources and a (stochastic) demand for those resources. We measure supply and demand as the number of full-time-equivalent (FTE) nurses each day.

The sequence of events is as follows :

Stage 1 - A proportion of each unit's nurses are cross-trained, enabling them to float to the other unit as needed.

Stage 2 - Demand is realized and nurses are allocated to units. Excess demand is met by hiring contract nurses.

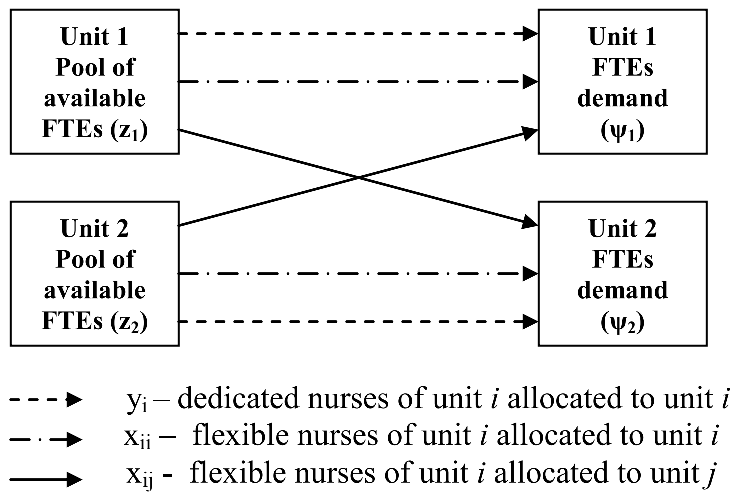
The time line for our model is shown in figure 2.1. There are two nurse pools consisting of dedicated and flexible nurses to meet demand across both units. The dedicated nurses in unit 1 and unit 2 can serve their home unit with 100% productivity. Flexible nurses can also serve their home units at 100% productivity but are generally less productive when required to float. Abernathy et al. (1973) defines a person's efficiency as a fraction of FTE. Campbell (1999) uses a capability measure between 0 and 1 to measure the capability of workers based on their training. In our model, e_{ij} is the measure of productivity for flexible nurses who are allocated from unit i to unit j. This measure varies between 0 and 1.

Allocation strategy :

Dedicated nurses who are not cross-trained can be allocated only to their home unit. Consistent with standard hospital practice, we assume that flexible nurses are first allocated to their home unit and only when the demand in their home unit is met can excess nurses be floated to the other unit. If the demand is still not met, then contract nurses are called in at higher cost. Demand for nursing hours is always met either by regular nurses or by contract nurses and shortage is never allowed to occur. The structure of the availability of nurses and the demand satisfied through allocations is shown in figure 2.2.

The objective is to determine the optimal amount of cross-training for each unit in order to minimize the total cost, which is the sum of contract cost and training cost, subject to meeting all the nursing hour requirements. The amount of cross-training is operationalized as the proportion of nurses who are cross-trained in each unit. The total number of nurses available is known a priori for both units. Demand in each unit is a function of the number of patients in each unit and their acuity. Conversion of the number of patients in a unit into required nursing hours based on their acuity is not considered

FIGURE 2.2: Network Structure



here. We assume that required nursing hours (demand) is stochastic and exogenous to the model. This problem is formulated using two stage stochastic programming with recourse in the second stage. Allocation of nurses in the second stage is the recourse for meeting the stochastic demand at the end of first stage.

2.3.1 Notation

N_i : Nurses required for unit i, which follows a general distribution

z_i : Nurses available in unit i

u_i : Amount of cross-training in unit i

y_i : Nurses in home unit i who are dedicated and are allocated to home unit i

x_{ii} : Nurses in home unit i who are flexible (cross-trained) and are allocated to home unit i

x_{ij} : Nurses in home unit i who are flexible (cross-trained) and are floated to unit j

e_{ij} : Productivity of nurses who are floated from unit i to unit j

s_i : Contract wages for hiring contract nurses to meet the demand in unit i

t_i : Training cost for a cross-trained nurses in unit i

k : length of second stage

The demand N_i in each unit is measured in terms of FTEs. The demand is stochastic and N_1 follows a general, continuous distribution with cdf $F(\cdot)$ and N_2 follows a general, continuous distribution with cdf $G(\cdot)$.

The total nurses available for allocation (z_i) in unit i is given a priori and cannot be varied in the model. Our model does not account for hiring, layoffs and absenteeism of regular nurses. It does not explicitly consider on-the-job regular training cost, but this cost can be added to the base wages for regular nurses and the model analysis will still hold. The base wages for regular nurses is a sunk cost for the hospital and so is not represented in the model. Of the total z_i, only $u_i \cdot z_i$ are cross-trained in each unit; u_i is the proportion of unit i nurses who are cross-trained.

Flexible nurses who are floated are not as productive as the nurses dedicated to that unit. The productivity parameter e_{ij} is used to capture this effect. Productivity is measured as the ratio of time taken by a dedicated nurse to do a task to the time taken by flexible nurse floated to that unit to do the same task. The productivity parameter varies from 0 to 1. 0 implies that cross-trained (flexible) nurses cannot perform any duties in the floated unit and 1 implies that they are as productive as the home unit (dedicated) nurses. If x_{ij} nurses are allocated from unit i to unit j, then the available nurse resource in unit j will be $e_{ij} \cdot x_{ij}$ nurses. Productivity of dedicated nurses (y_{ii}) and flexible nurses allocated to home unit (x_{ii}) are assumed to be 1.

When the demand of nursing hours is not met by flexible and dedicated nurses, contract nurses are called in to meet the demand at a higher cost. This contract cost is represented by s_i in the model. The productivity of contract nurses are assumed to be 1 since they are multi-skilled and experienced in a variety of tasks. If their productivity is less than 1, the model analysis will remain the same provided the contract cost s_i, is inflated by the productivity of the contract nurse. Fixed cross-training costs per nurse, per time period, are represented by $\frac{t_i}{k}$ where k is the length of the second stage planning period. These training costs are incurred at the beginning of the planning horizon, before demands are realized.

2.3.2 Model Formulation

The model is formulated using two stage stochastic programming with recourse Birge and Louveaux (1997). In the first stage, the training costs for both the units are incurred and then in the second stage demand is realized and nurses are allocated, with excess demand covered by contract nurses.

Stage 1

$$Min_{\ u1,\ u2} \quad t_1 \cdot u_1 \cdot z_1 + t_2 \cdot u_2 \cdot z_2 + k \cdot E_{N_1,N_2} Q(u_1, u_2, N_1, N_2) \tag{2.1}$$

subject to

$$0 \leq u_1, u_2 \leq 1 \tag{2.2}$$

The objective function (2.1) is the sum of training cost and expected contract cost determined from stage 2, subject to the constraint (2.2) that the amount of cross-training varies between 0 and 1. 0 represents no cross-training and 1 represents complete cross-training (all nurses in that unit are cross-trained).

Stage 2 $\quad\mathbf{Q(u_1, u_2, N_1, N_2)}$:

$$Min_{y_1,y_2,x_{11},x_{22},x_{12},x_{21}} \quad Q(u_1, u_2, N_1, N_2) \;\; =$$

$$\begin{aligned}&(N_1 - x_1 - e_{21} \cdot x_{21})s_1 + (N_2 - x_2 - e_{12} \cdot x_{12})s_2 - \\ &max(0, e_{12} \cdot s_1 - s_2 + c)x_1 - max(0, e_{21} \cdot s_2 - s_1 + c)x_2\end{aligned} \tag{2.3}$$

subject to

$$x_1 + e_{21} \cdot x_{21} \leq N_1 \tag{2.4}$$

$$x_2 + e_{12} \cdot x_{12} \leq N_2 \tag{2.5}$$

$$x_{12} \leq u_1 \cdot z_1 \tag{2.6}$$

$$x_{21} \leq u_2 \cdot z_2 \tag{2.7}$$

$$x_1 + x_{12} \leq z_1 \tag{2.8}$$

$$x_2 + x_{21} \leq z_2 \tag{2.9}$$

$$y_1, y_2, x_{11}, x_{22}, x_{12}, x_{21} \geq 0 \tag{2.10}$$

Aside conditions for parameters,

$$c \geq 0, 0 \leq e_{12}, e_{21} \leq 1, s_1, s_2 \geq 0 \tag{2.11}$$

In the above formulation, $x_1 = y_1 + x_{11}$ and $x_2 = y_2 + x_2$. In stage 2, the objective function (2.3) determines the contract cost for both units. Given a demand value N_i, we allocate y_i dedicated nurses to home unit i. Among the cross-trained nurses, we allocate x_{ii} to their home unit and altogether x_i nurses serve the home unit and the rest x_{ij} float to unit j.

Constraint (2.4) and (2.5) does not allocate more than the demand. Constraint (2.6) to (2.9) are the constraints for allocation of nurses to home unit and float unit. $(u_i \cdot z_i)$ are the number of nurses who are cross-trained. Constraint(2.10) is the non-negativity constraint.

2.4 Solving Two-Stage Stochastic Programming

In this section, we determine the solutions to the second stage problem based on the allocation policy and use it to solve the first stage optimization problem.

FIGURE 2.3: Regions for Second Stage Solution

$N_2 > z_2$	**7**	**8**	**9**
$(1\text{-}u_2)\, z_2 < N_2 \le z_2$	**4**	**5**	**6**
$N_2 \le (1\text{-}u_2)\, z_2$	**1**	**2**	**3**
	$N_1 \le (1\text{-}u_1)\, z_1$	$(1\text{-}u_1)\, z_1 < N_1 \le z_1$	$N_1 > z_1$

2.4.1 Second stage - Allocation

The solution is characterized based on whether demand can be Cattani et al. (2003):

1. satisfied with the non cross-trained nurse, $N_i \le (1 - u_i)z_i$
2. satisfied with the home unit nurses, $N_i \le z_i$
3. satisfied only with floated or contract nurses, $N_i \ge z_i$

The regions are explained below and are also represented in figure 2.3.

Region 1 ($N_1 \le (1 - u_1)z_1$, $N_2 \le (1 - u_2)z_2$)

$y_1^* = N_1$, $y_2^* = N_2$, $x_{ii}^* = 0$, $x_{ij}^* = 0$ for i, j = 1 and 2.

In this region, demand in both units is satisfied with the nurses in home unit who are not cross-trained.

Region 2 ($(1 - u_1)z_1 \le N_1 \le z_1$, $N_2 \le (1 - u_2)z_2$)

$y_1^* = (1 - u_1)z_1$, $x_{11}^* = N_1 - (1 - u_1)z_1$, $x_{21}^* = 0$

$y_2^* = N_2$, $x_{22}^* = 0$, $x_{12}^* = 0$

In this region, unit 1's demand is satisfied using home unit nurses not cross-trained and some of the home unit nurses who are cross-trained while unit 2 demand is satisfied using home unit nurses who are not cross-trained.

Region 3 ($N_1 \geq z_1$, $N_2 \leq (1 - u_2)z_2$)

$y_1^* = (1 - u_1)z_1$, $x_{11}^* = u_1 z_1$, $x_{21}^* = min(u_2 \cdot z_2, (N_1 - z_1)/e_{21})$

$y_2^* = N_2$, $x_{22}^* = 0$, $x_{12}^* = 0$

In this region, demand in unit 1 may not be satisfied using home unit nurses both cross-trained and not cross-trained and floated nurses from unit 2. Hence, the excess demand is met by contract nurses. In unit 2, demand is met using nurses in home unit who are not cross-trained.

Region 4 ($N_1 \leq (1 - u_1)z_1$, $(1 - u_2)z_2 \leq N_2 \leq z_2$)

$y_1^* = N_1$, $x_{11}^* = 0$, $x_{21}^* = 0$

$y_2^* = (1 - u_2)z_2$, $x_{22}^* = N_2 - (1 - u_2)z_2$, $x_{12}^* = 0$

In this region, demand in unit 1 is satisfied using home unit 1 nurses who are not cross-trained and in unit 2 the demand is met using home unit 2 nurses both cross-trained and not cross-trained.

Region 5 ($(1 - u_1) \cdot z_1 \leq N_1 \leq z_1$, $(1 - u_2)z_2 \leq N_2 \leq z_2$)

$y_1^* = (1 - u_1)z_1$, $x_{11}^* = N_1 - (1 - u_1)z_1$, $x_{21}^* = 0$

$y_2^* = (1 - u_2)z_2$, $x_{22}^* = N_2 - (1 - u_2)z_2$, $x_{12}^* = 0$

In this region, demand in both units are met using cross-trained nurses and not cross-trained in their respective home units.

Region 6 ($N_1 \geq z_1$, $(1-u_2)z_2 \leq N_2 \leq z_2$)

$y_1^* = (1-u_1)z_1$, $x_{11}^* = u_1 z_1$, $x_{21}^* = min(z_2 - N_2, (N_1 - z_1)/e_{21})$

$y_2^* = (1-u_2)z_2$, $x_{22}^* = N_2 - (1-u_2)z_2$, $x_{12}^* = 0$

In this region, demand in unit 1 may not be satisfied with nurses from home unit 1 both cross-trained and not cross-trained and from nurses floated from unit 2. The excess demand is met using contract nurses. Demand in unit 2 is met with nurses from home unit 2 who are cross-trained and nurses who are not cross-trained. The nurses are allowed to float only after they satisfy their home unit demand first.

Region 7 ($N_1 \leq (1-u_1)z_1$, $N_2 \geq z_2$)

$y_1^* = N_1$, $x_{11}^* = 0$, $x_{21}^* = 0$

$y_2^* = (1-u_2)z_2$, $x_{22}^* = u_2 z_2$, $x_{12}^* = min(u_1 \cdot z_1, (N_2 - z_2)/e_{12})$

In this region, unit 2 demand may be satisfied from nurses in its unit and also nurses from unit 1 who are cross-trained. Excess demand in unit 2 is met using contract nurses. Demand in unit 1 is satisfied using nurses in the home unit who are not cross-trained.

Region 8 ($(1-u_1)z_1 \leq N_1 \leq z_1$, $N_2 \geq z_2$)

$y_1^* = (1-u_1)z_1$, $x_{11}^* = N_1 - (1-u_1)z_1$, $x_{21}^* = 0$

$y_2^* = (1-u_2)z_2$, $x_{22}^* = u_2 z_2$, $x_{12}^* = min(z_1 - N_1, (N_2 - z_2)/e_{12})$

In this region, demand in unit 2 may be satisfied with nurses from home unit 2 who are both cross-trained and not cross-trained and also from unit 1 who are cross-trained. In unit 1, the demand is first satisfied using nurses in home unit who are not cross-trained and then nurses who are cross-trained. Only after demand in unit 1 is satisfied, the nurses who are cross-trained in unit 1 are floated to unit 2.

Region 9 ($N_1 \geq z_1$, $N_2 \geq z_2$)

$y_1^* = (1 - u_1)z_1$, $x_{11}^* = u_1 z_1$, $x_{21}^* = 0$

$y_2^* = (1 - u_2)z_2$, $x_{22}^* = u_2 z_2$, $x_{12}^* = 0$

In this region, nurses are not sufficient to meet the demand, so contract nurses have to be hired in both units.

Expectation of the solution in second stage :

The next step is to determine the expected value for the second stage problem. The solution from the nine regions are put into the objective function and the expectation over N_1 and N_2 are taken. Since the contract nurses are required only in region 3, 6, 7, 8 and 9, we see that the terms in the objective function (2.3) reduces and the resulting expectation terms are shown below.

$E_{N_1,N_2}Q(u_1, u_2, N_1, N_2)$:

Region 3

$$\int_0^{(1-u_2)z_2} \int_{z_1}^{\infty} \quad [N_1 - (1-u_1)z_1 - u_1 \cdot z_1 - e_{21} \cdot min(u_2 \cdot z_2, \frac{N_1 - z_1}{e_{21}})]s_1 \quad dF(N_1)dG(N_2) \tag{2.12}$$

Region 6

$$\int_{(1-u_2)z_2}^{z_2} \int_{z_1}^{\infty} \quad [N_1 - (1-u_1)z_1 - u_1 \cdot z_1 - e_{21} \cdot min(z_2 - N_2, \frac{N_1 - z_1}{e_{21}})]s_1 \quad dF(N_1)dG(N_2) \tag{2.13}$$

Region 7

$$\int_{z_2}^{\infty}\int_{0}^{(1-u_1)z_1} [N_2-(1-u_2)z_2-u_2\cdot z_2-e_{12}\cdot min(u_1\cdot z_1,\frac{N_2-z_2}{e_{12}})]s_2 \quad dF(N_1)dG(N_2) \tag{2.14}$$

Region 8

$$\int_{z_2}^{\infty}\int_{(1-u_1)z_1}^{z_1} [N_2-(1-u_2)z_2-u_2\cdot z_2-e_{12}\cdot min(z_1-N_1,\frac{N_2-z_2}{e_{12}})]s_2 \quad dF(N_1)dG(N_2) \tag{2.15}$$

Region 9

$$\int_{z_2}^{\infty}\int_{z_1}^{\infty}[N_1-(1-u_1)z_1-u_1\cdot z_1]s_1 dF(N_1)dG(N_2)+\int_{z_2}^{\infty}\int_{z_1}^{\infty}[N_2-(1-u_2)z_2-u_2\cdot z_2]s_2 dF(N_1)dG(N_2) \tag{2.16}$$

Splitting integrals in each of the regions 3, 6, 7 and 8 with a minimization operator we obtain following solutions:

Region 3

$min(u_2\cdot z_2,\frac{N_1-z_1}{e21})$: If $u_2\cdot z_2>\frac{N_1-z_1}{e_{21}}$ then $z_1+e_{21}\cdot u_2\cdot z_2>N_1$.

So, equation 2.12 becomes

$$\begin{aligned}\int_{0}^{(1-u_2)z_2}\int_{z_1}^{z_1+e_{21}\cdot u_2\cdot z_2} [N_1-z_1-e_{21}(\frac{N_1-z_1}{e_{21}})]s_1 \quad dF(N_1)dG(N_2)+\\ \int_{0}^{(1-u_2)z_2}\int_{z_1+e_{21}\cdot u_2\cdot z_2}^{\infty} [N_1-z_1-e_{21}\cdot u_2\cdot z_2]s_1 \quad dF(N_1)dG(N_2)\end{aligned} \tag{2.17}$$

Region 6

$min(z_2 - N_2, \frac{N_1 - z_1}{e_{21}})$: If $z_2 - N_2 < \frac{N_1 - z_1}{e_{21}}$ then $e_{21}(z_2 - N_2) + z_1 > N_1$.

So, equation 2.13 becomes

$$\int_{(1-u_2)z_2}^{z_2} \int_{z_1}^{e_{21}(z_2-N_2)} [N_1 - z_1 - e_{21}\frac{N_1 - z_1}{e_{21}}]s_1 \quad dF(N_1)dG(N_2) + \int_{(1-u_2)z_2}^{z_2} \int_{e_{21}(z_2-N_2)}^{\infty} [N_1 - z_1 - e_{21}(z_2 - N_2)]s_1 \quad dF(N_1)dG(N_2) \tag{2.18}$$

Region 7

$min(u_1 \cdot z_1, \frac{N_2 - z_2}{e_{12}})$: If $u_1 \cdot z_1 < \frac{N_2 - z_2}{e_{12}}$ then $z_2 + e_{12} \cdot u_1 \cdot z_1 > N_2$.

So, equation 2.14 becomes

$$\int_{0}^{(1-u_1)z_1} \int_{z_2}^{z_2+e_{12}\cdot u_1\cdot z_1} [N_2 - z_2 - e_{12}\frac{N_2 - z_2}{e_{12}}]s_2 \quad dG(N_2)dF(N_1) + \int_{0}^{(1-u_1)z_1} \int_{z_2+e_{12}\cdot u_1\cdot z_1}^{\infty} [N_2 - z_2 - e_{12} \cdot u_1 \cdot z_1]s_2 \quad dG(N_2)dF(N_1) \tag{2.19}$$

Region 8

$min(z_1 - N_1, \frac{N_2 - z_2}{e_{12}})$: If $z_1 - N_1 < \frac{N_2 - z_2}{e_{12}}$ then $e_{12}(z_1 - N_1) + z_2 > N_2$.

So, equation 2.15 becomes

$$\int_{(1-u_1)z_1}^{z_1} \int_{z_2}^{e_{12}(z_1-N_1)} [N_2 - z_2 - e_{12}\frac{N_2 - z_2}{e_{12}}]s_2 \quad dG(N_2)dF(N_1) + \int_{(1-u_1)z_1}^{z_1} \int_{e_{12}(z_1-N_1)}^{\infty} [N_2 - z_2 - e_{12}(z_1 - N_1)]s_2 \quad dG(N_2)dF(N_1) \tag{2.20}$$

Therefore,

$E_{N_1,N_2}Q(u_1,u_2,N_1,N_2) =$

$$\begin{aligned}
&\int_0^{(1-u_2)z_2}\int_{(z_1+e_{21}\cdot u_2\cdot z_2)}^{\infty}[N_1 - z_1 - e_{21}\cdot u_2\cdot z_2]s_1 \quad dF(N_1)dG(N_2)\\
&+\int_{(1-u_2)z_2}^{z_2}\int_{(z_1+e_{21}(z_2-N_2))}^{\infty}[N_1 - z_1 - e_{21}(z_2-N_2)]s_1 \quad dF(N_1)dG(N_2)\\
&+\int_0^{(1-u_1)z_1}\int_{(z_2+e_{12}\cdot u_1\cdot z_1)}^{\infty}[N_2 - z_2 - e_{12}\cdot u_2\cdot z_1]s_2 \quad dG(N_2)dF(N_1)\\
&+\int_{(1-u_1)z_1}^{z_1}\int_{(z_2+e_{12}(z_1-N_1))}^{\infty}[N_2 - z_2 - e_{12}(z_1-N_1)]s_2 \quad dG(N_2)dF(N_1)\\
&+\int_{z_2}^{\infty}\int_{z_1}^{\infty}[N_1 - z_1]s_1 \quad dF(N_1)dG(N_2)\\
&+\int_{z_2}^{\infty}\int_{z_1}^{\infty}[N_2 - z_2]s_2 \quad dF(N_1)dG(N_2)
\end{aligned} \tag{2.21}$$

2.4.2 First stage - Optimal cross-training

The expectation terms from the second stage is put into the first stage objective function (2.1).

$$Min_{\ u1,\ u2} \quad t_1\cdot u_1\cdot z_1 + t_2\cdot u_2\cdot z_2 + k\cdot E_{N_1,N_2}Q(u_1,u_2,N_1,N_2) \tag{2.22}$$

subject to

$$u_1 \geq 0 \tag{2.23}$$

$$u_2 \geq 0 \tag{2.24}$$

$$u_1 \leq 1 \tag{2.25}$$

$$u_2 \leq 1 \tag{2.26}$$

To determine if the above constrained minimization problem is convex, we define the Lagrangian equation and determine the Hessian matrix.

The Lagrangian function for the first stage is given by

$$L(u_1, u_2, \mu) = t_1 \cdot u_1 \cdot z_1 + t_2 \cdot u_2 \cdot z_2 + k \cdot E_{N_1,N_2} Q(u_1, u_2, N_1, N_2) - \mu_1 \cdot u_1 - \mu_2 \cdot u_2 + \mu_3 \cdot (u_1 - 1) + \mu_4 \cdot (u_2 - 1) \tag{2.27}$$

$$L'_{u_1,u_1}(u_1, u_2, \mu) = e_{12} \cdot k \cdot s_2 \cdot z_1^2 \cdot (\{1 - G[e_{12} \cdot u_1 \cdot z_1 + z_2]\} \cdot F'[(1-u_1)z_1] + e_{12} \cdot F[(1-u_1)z_1] \cdot G'[e_{12} \cdot u_1 \cdot z_1 + z_2]) \tag{2.28}$$

$$L'_{u_2,u_2}(u_1, u_2, \mu) = e_{21} \cdot k \cdot s_1 \cdot z_2^2 \cdot (\{1 - F[e_{21} \cdot u_2 \cdot z_2 + z_1]\} \cdot G'[(1-u_2)z_2] + e_{21} \cdot G[(1-u_2)z_2] \cdot F'[e_{21} \cdot u_2 \cdot z_1 + z_2]) \tag{2.29}$$

$$L'_{u_1,u_2}(u_1, u_2, \mu) = 0 \tag{2.30}$$

$$L'_{u_2,u_1}(u_1, u_2, \mu) = 0 \tag{2.31}$$

The Hessian matrix with second order conditions with respect to u_1 and u_2 for the above formulation shows that it is positive definite Blume and Simon (1994) and so the objective function (2.22) is convex in u_1 and u_2.

Since we proved that the objective function is convex, we can now take the first order conditions to determine optimal values, u_1^* and u_2^*.

The first order condition for first stage Lagrange function (2.27) is given below :

foc wrt u_1:

$$\mu_3 - \mu_1 + t_1 \cdot z_1 - k \cdot e_{12} \cdot s_2 \cdot F[(1-u_1)z_1] \cdot \{1 - G[e_{12} \cdot u_1 \cdot z_1 + z_2]\} = 0 \tag{2.32}$$

foc wrt u_2:

$$\mu_4 - \mu_2 + t_2 \cdot z_2 - k \cdot e_{21} \cdot s_1 \cdot G[(1-u_2)z_2] \cdot \{1 - F[e_{21} \cdot u_2 \cdot z_2 + z_1]\} = 0 \tag{2.33}$$

$$\mu_1 \cdot u_1 = 0 \tag{2.34}$$

$$\mu_2 \cdot u_2 = 0 \tag{2.35}$$

$$\mu_3 \cdot (1 - u_1) = 0 \tag{2.36}$$

$$\mu_4 \cdot (1 - u_2) = 0 \tag{2.37}$$

Solving for the lagrange multipliers we get the following cases:

Case 1 : ($\mu_1 > 0$ and $\mu_2 > 0$)

From the constraints (2.34) and (2.35), we see that $u_1^* = 0$ and $u_2^* = 0$, consequently from constraint (2.36) and (2.37) we get $\mu_3 = 0$ and $\mu_4 = 0$. So constraint (2.34) and (2.35) is binding.

Case 2 : ($\mu_1 > 0$ and $\mu_4 > 0$)

Implies $u_1^* = 0$ and consequently $\mu_3 = 0$. Since $\mu_4 > 0$, $u_2^* = 1$ from constraint (2.37) and $\mu_2 = 0$. Substituting $u_2^* = 1$ in constraint (2.33), we get $t_2 \cdot z_2 = -\mu_4$. This is not possible since all the values are positive and so this case is infeasible.

Case 3 : ($\mu_3 > 0$ and $\mu_2 > 0$)

Implies $u_2^* = 0$ and consequently $\mu_4 = 0$. But $\mu_1 = 0$ and so $u_1^* = 1$. Substituting in constraint (2.32) we see that $t_1 \cdot z_1 = -\mu_3$. So this case is infeasible.

Case 4 : ($\mu_1 = 0$ and $\mu_3 = 0$, $\mu_2 = 0$ and $\mu_4 = 0$)

The solution is $u_1^* = [0, 1)$ and $u_2^* = [0, 1)$ and satisfies the equations (2.32) and (2.33). The following two equations give the closed form expression for the optimum amount of cross-training for unit 1 and unit 2 when the demand follows a general, continuous distribution.

$$\mathbf{F}[(\mathbf{1} - \mathbf{u_1})\mathbf{z_1}] \;\cdot\; \{\mathbf{1} - \mathbf{G}[\mathbf{e_{12}} \cdot \mathbf{u_1} \cdot \mathbf{z_1} + \mathbf{z_2}]\} = \frac{\mathbf{t_1}}{\mathbf{k} \cdot \mathbf{e_{12}} \cdot \mathbf{s_2}} \tag{2.38}$$

$$\mathbf{G}[(\mathbf{1} - \mathbf{u_2})\mathbf{z_2}] \;\cdot\; \{\mathbf{1} - \mathbf{F}[\mathbf{e_{21}} \cdot \mathbf{u_2} \cdot \mathbf{z_2} + \mathbf{z_1}]\} = \frac{\mathbf{t_2}}{\mathbf{k} \cdot \mathbf{e_{21}} \cdot \mathbf{s_1}} \tag{2.39}$$

Looking at equations (2.38) and (2.39), we can infer some of the implications of parameters such as contract cost, training cost and productivity. We see that when training cost per period ($\frac{t_i}{k}$) increases, the amount of cross-training (u_i) decreases, as expected. An increase in the contract cost (s_j) results in an increase in the amount of cross-training in unit i (u_i). These relationships and some more results are analyzed using a numerical example in the following section.

Other observations are :

1. Amount of cross-training in unit i (u_i) does not depend on the contract cost of unit i (s_i).

2. Amount of cross-training in unit i (u_i) does not depend on the training cost per period of unit j ($\frac{t_j}{k}$).

3. Amount of cross-training in unit i (u_i) does not depend on the productivity of nurses floated from unit j to unit i (e_{ji}).

2.5 Numerical Analysis

Assuming N_1 and N_2 to be uniformly distributed between $[a_1, b_1]$ and $[a_2, b_2]$ we get the following equations.

For u_1^*:

$$\frac{(1-u_1)z_1 - a_1}{b_1 - a_1} \cdot \{1 - \frac{e_{12} \cdot u_1 \cdot z_1 + z_2}{b_2 - a_2}\} = \frac{t_1}{k \cdot e_{12} \cdot s_2} \tag{2.40}$$

Similarly, for u_2^*:

$$\frac{(1-u_2)z_2 - a_2}{b_2 - a_2} \cdot \{1 - \frac{e_{21} \cdot u_2 \cdot z_2 + z_1}{b_1 - a_1}\} = \frac{t_2}{k \cdot e_{21} \cdot s_1} \tag{2.41}$$

Solving (2.40) we get a quadratic equation (2.42) in u_1^*

$$e_{12} \cdot z_1^2 \cdot u_1^2 - \{(b_2 - z_2)z_1 - e_{12} \cdot z_1(z_1 - a_1)\} \cdot u_1 + \{(z_1 - a_1) \cdot (b_2 - z_2) - \frac{t_1}{k \cdot e_{12} \cdot s_2} \cdot (b_1 - a_1) \cdot (b_2 - a_2)\} = 0 \tag{2.42}$$

Solving (2.41) we get a quadratic equation (2.43) in u_2^*

$$e_{21}\cdot z_2^2\cdot u_2^2-\{(b_1-z_1)z_2-e_{21}\cdot z_2(z_2-a_2)\}\cdot u_2+\{(z_2-a_2)\cdot(b_1-z_1)-\frac{t_2}{k\cdot e_{21}\cdot s_1}\cdot(b_1-a_1)\cdot(b_2-a_2)\}=0 \tag{2.43}$$

Let $\overline{N_1}$ and $\overline{N_2}$ denote the mean of the uniform distribution for the random variables N_1 and N_2. In other words, $\overline{N_1}=\frac{a_1+b_1}{2}$ and $\overline{N_2}=\frac{a_2+b_2}{2}$.

The numerical analysis includes five cases as shown in table 1 depending on whether nurses available in unit i (z_i) is more or less than the mean demand in that unit ($\overline{N_i}$). Mean demand ($\overline{N_i}$) is taken as 40 for the numerical analysis. The values for z_i for different cases is given in table 1.

TABLE 2.1: Alternative cases used in numerical analysis

Cases	$\mathbf{z_1}$	$\mathbf{z_2}$
$z_1>\overline{N_1},\ z_2<\overline{N_2}$	50	30
$z_1=\overline{N_1},\ z_2=\overline{N_2}$	40	40
$z_1>\overline{N_1},\ z_2>\overline{N_2}$	50	50
$z_1<\overline{N_1},\ z_2>\overline{N_2}$	30	50
$z_1<\overline{N_1},\ z_2<\overline{N_2}$	30	30

TABLE 2.2: Parameter levels for numerical analysis

Parameters	**Level 1**	**Level 2**	**Level 3**
e_{12}	0.05 - 0.95 in steps of 0.05		
e_{21}	0.05 - 0.95 in steps of 0.05		
t_1/k	\$0 - \$75 in steps of 5		
t_2/k	\$0 - \$75 in steps of 5		
s_1	\$40	\$60	\$80
s_2	\$40	\$60	\$80
CV_1	0.2	0.5	0.7
CV_2	0.2	0.5	0.7

Table 2 indicates the parameters and their values used in the numerical analysis. Productivity parameters varies from 0.05 to 0.95 in steps of 0.05. The training cost per period varies from \$0 to \$75 or less in steps of \$5. Analysis is done for three levels of contract cost \$40, \$60 and \$80. Coefficient of variation (CV) is defined as the ratio of standard deviation to mean. Since mean is represented by $\overline{N_i}$, we get $CV_i = \frac{\sigma}{\overline{N_i}}$. If coefficient of variation is CV_i for unit i and $\overline{N_i}$ follows uniform distribution, then $[a_i, b_i] = [\overline{N_i} \cdot (1 - (CV_i/2)), \overline{N_i} \cdot (1 + (CV_i/2))]$.

Case 1 : $\mathbf{z_1 > \overline{N_1}}$ and $\mathbf{z_2 < \overline{N_2}}$

In this case, nurses available in unit 1 (z_1) is greater than the mean demand ($\overline{N_1}$) in unit 1. Nurses available in unit 2 (z_2) is less than the mean demand ($\overline{N_2}$) in unit 2. This indicates that unit 1 has surplus nurses most of the time and unit 2 has insufficient nurses most of the time. Assuming CV = 0.2 for demand in unit 1 and unit 2, we plot the amount of cross-training (u_1) versus productivity (e_{12}) for varying training cost per period t_1/k for unit 1 in figure: 2.4. Similar graphs are drawn for coefficient of variation 0.5 and 0.7 for demand as shown in the figure: 2.5 and 2.6.

0.2, 0.5 and 0.7 cov represents low, medium and high levels of coefficient of variation in demand.

For fixed demand variation in unit 1 and unit 2, we can conclude the following from figures: 2.4, 2.5, 2.6.

1. In general, higher productivity of cross-trained nurses fosters more cross-training by effectively reducing the cost of covering demand with cross-trained nurses. When

the cost of cross-training is relatively low, however, there is a productivity level beyond which further increases in productivity reduce the level of cross-training. Above this threshold, the productivity of cross-trained nurses allow the benefits of flexibility to be obtained with fewer cross-trained nurses.

2. When training cost is high, the optimal amount of cross-training is zero for lower values of productivity, but increases with increasing values of productivity.

3. In this case, there is a shortage of nurses in unit 2 and so there is no cross-training in unit 2.

When demand variation in unit 1 and unit 2 increases simultaneously or increases in unit 2, we can conclude the following from figures 2.7 and 2.9 for low training cost and 2.8 and 2.10 for high training cost.

1. For low cross-training cost per period:

   - at lower productivity level, optimal amount of cross-training decreases with increase in demand variation

   - at higher productivity level, optimal amount of cross-training increases with increase in demand variation

2. For high cross-training cost per period and at a productivity level increase in demand variation causes decrease in optimal amount of cross-training.

In all of the above analysis, the contract cost per period is assumed to be $60 per period. In figure 2.11, the contract cost is varied to determine the effect of the contract cost on optimal amount of cross-training. Increase in contract cost in unit j causes increase in cross-training in unit i.

Case 2 : $\mathbf{z_1 = \overline{N_1}}$ and $\mathbf{z_2 = \overline{N_2}}$

In this case, nurses available in both units equals the mean demand in their units. There is no cross-training when cross-training cost per period is high. Figure 2.12 shows that at lower values of training cost, the optimal amount of cross-training increases at higher values of productivity and is zero for lower values of productivity. There is no threshold value for productivity as seen in case 1. The optimal amount of cross-training increases with in increase in demand variability as seen in figure 2.13.

Case 3 : $\mathbf{z_1 > \overline{N_1}}$ and $\mathbf{z_2 > \overline{N_2}}$

In this case, nurses available in unit 1 (z_1) is more than the mean demand ($\overline{N_1}$) in unit 1. nurses available in unit 2 (z_2) is also greater than the mean demand ($\overline{N_2}$) in unit 2.

From figure 2.13, we infer that when training cost per period for unit i is low, the optimal amount of cross-training for unit i decreases with increase in coefficient of variation for unit i. If demand variation is high in the home unit then less nurses are cross-trained.

Case 4 : $\mathbf{z_1 < \overline{N_1}}$ and $\mathbf{z_2 < \overline{N_2}}$

In this case, nurses available in both units is less than the mean demand in their units. Since available nurses are less than the mean demand in both units there is little or no cross-training at all.

2.6 Conclusion

The model and analysis in this paper show that productivity of flexible nurses has a significant effect on the optimal amount of cross-training. We have derived a closed form expression that determines the optimal amount of cross-training to minimize the sum of cross-training cost and expected shortage cost. The analysis shows that at a given level of productivity for flexible nurses, the optimal amount of cross-training in unit i decreases with an increase in cross-training cost in unit i, and increases with an increase in shortage cost in unit j as expected.

When cost of cross-training is high, increase in productivity leads to increase in amount of cross-training. When the cost of cross-training is relatively low, however, there is a productivity level beyond which further increases in productivity reduces the amount of cross-training. Above this threshold, the productivity of cross-trained nurses allow the benefits of flexibility to be obtained with fewer cross-trained nurses.

For low cross-training productivity, more demand variability leads to less cross-training. For high cross-training productivity, the effect of demand variability on the amount of cross-training depends on cross-training cost. When cross-training cost is high more variability continues to cause less cross-training. For low cross-training cost, however, increases demand variability leads to more cross-training (with high cross-training productivity).

This paper has integrated the planning (amount of cross-training) and scheduling (allocation) phases of nurse staffing while utilizing flexible nurses in hospital. To date, the literature has focussed independently on nurses planning problem and allocation of flexible nurses problem. In our model, we try to integrate both planning and scheduling phases of nurse staffing across a planning horizon. The paper has provided insights on the interaction effects of productivity of flexible nurses, cross-training costs and contract costs on the optimum amount of cross-training.

Extending the model to more than two units introduces the possibility of different patterns of cross-training, so it becomes necessary to determine the pattern that best fits the situation. Another possible extension is to have varying costs for flexible nurses when they are floated from one unit to another. We have considered a model where an external resource (the contract nurse) is hired, to attain 100% service level. It is possible to consider a pool of highly flexible resource (similar to contract nurse) who are permanent employees, but are paid higher than the regular nurses and lower than the contract nurses. It would be interesting to determine the optimal number of such highly flexible nurses that should be hired at a given wage rate.

FIGURE 2.4: Effect of training cost on optimal amount of cross-training with CV=0.2

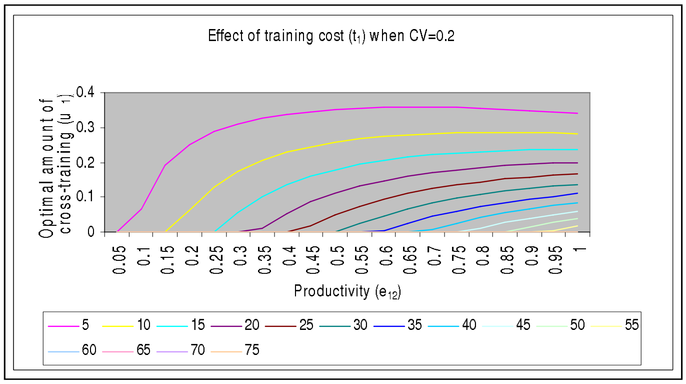


FIGURE 2.5: Effect of training cost on optimal amount of cross-training with CV=0.5

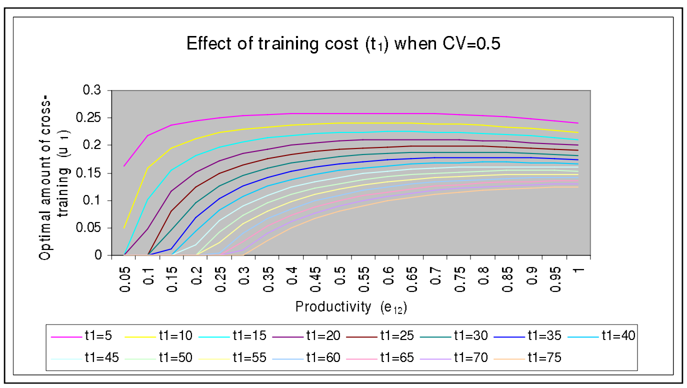


FIGURE 2.6: Effect of training cost on optimal amount of cross-training with CV=0.7

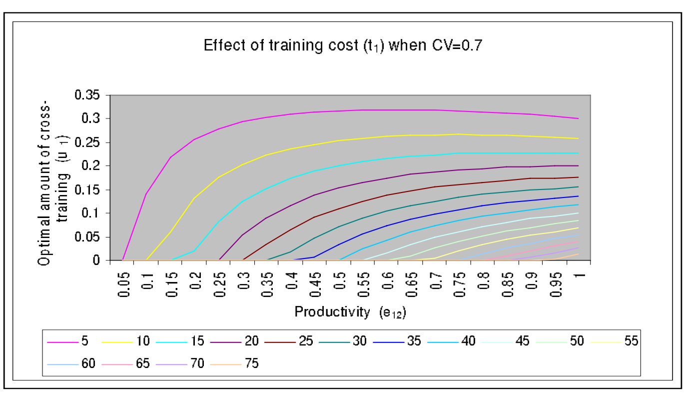


FIGURE 2.7: Effect of demand variation in unit 1 and 2 on optimal amount cross-training at low training cost

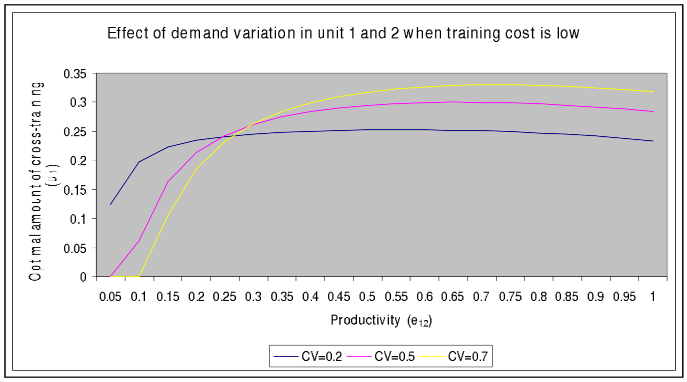


FIGURE 2.8: Effect of demand variation in unit 2 on optimal amount cross-training at low training cost

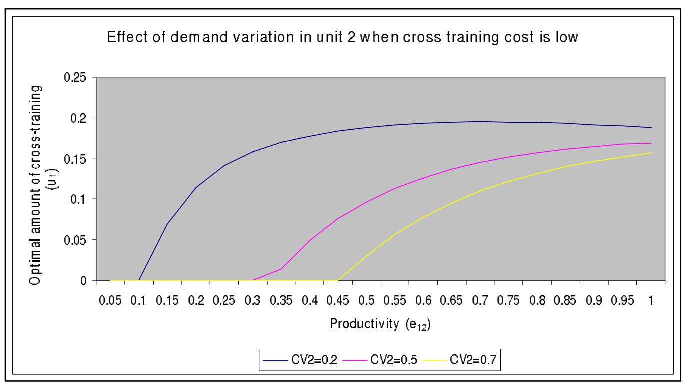


FIGURE 2.9: Effect of demand variation in unit 1 and 2 on optimal amount cross-training at high training cost

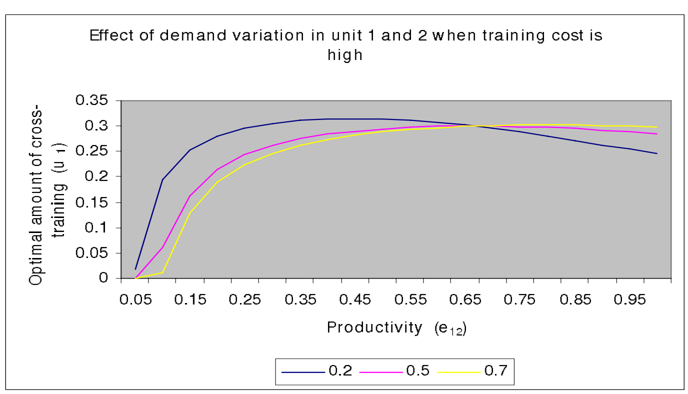


FIGURE 2.10: Effect of demand variation in unit 2 on optimal amount cross-training at high training cost

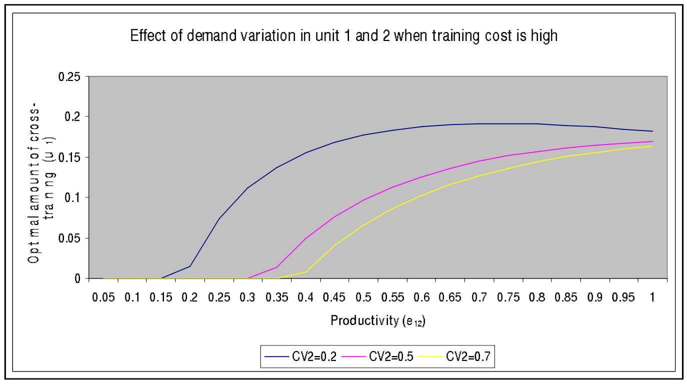


FIGURE 2.11: Effect of contract cost on optimal amount of cross-training

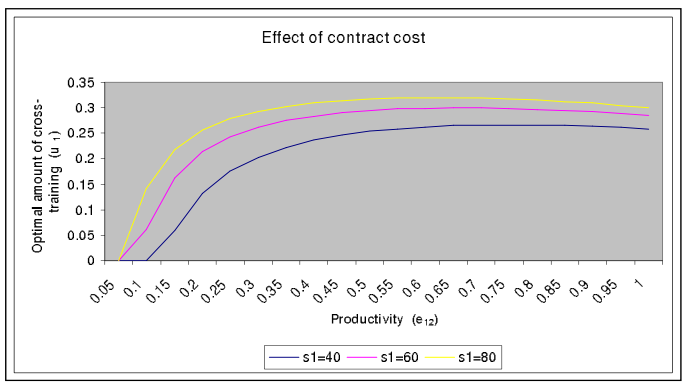


FIGURE 2.12: Effect of training cost on optimal amount of cross-training when demand matches supply

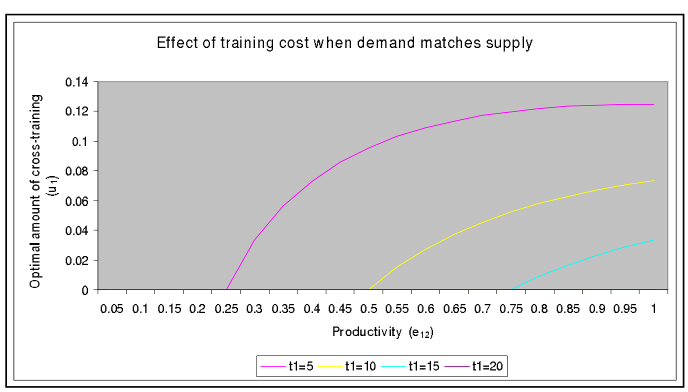
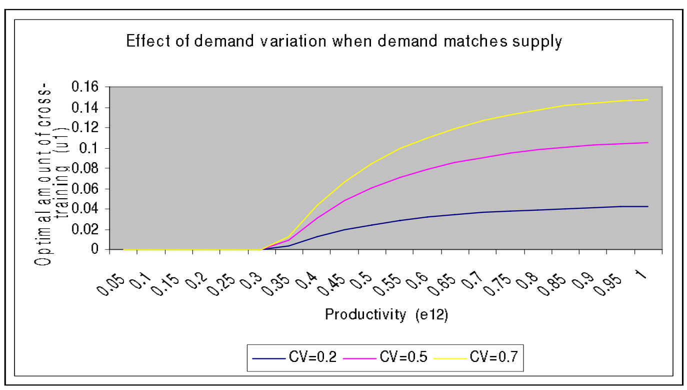


FIGURE 2.13: Effect of demand variation on optimal amount of cross-training when demand matches supply



CHAPTER 3

SEQUENTIAL AND SIMULTANEOUS DECISION MAKING UNDER DUAL RESOURCE FLEXIBILITIES

3.1 Introduction and Motivation

The primary objective of community hospitals is to provide the highest quality of care and reduce adverse outcomes, such as high mortality rate and longer length of stay for their patients. Nursing literature has studied extensively the positive effect of nurse staffing on quality of patient care (Aiken et al., 2002; Needleman et al., 2002; Hassmiller and Cozine, 2006). McCue et al. (2003) show that an increase in nurse staffing has positive impact on quality of patient care and financial performance of hospital. To provide high quality of care, hospital administrators should make sure that all resources, including staff and equipment, are available at the right time and right place to treat patients. The two most important resources are nursing staff and beds for patients. Hospitals have tried to co-ordinate these two resources so that maximum number of patients can be treated. This leads to the question of how to manage key resources so that large number of patients are treated at minimum cost?

In practice, when there are more patients than expected, hospital administrators efficiently utilize available resources in two ways. In the first scenario, when bed space is available but nursing staff (dedicated nurse) is unavailable, the patient is admitted and either flexible nurses or contract nurses are used. Typically, shortages of regular nurses are supplemented by hiring contract (travel) nurses for a shift or two until demand stabilizes. Contract nurses are more expensive than regular nurses. The American Hospital Directory indicates that contract labor as a percentage of total operating cost increased steadily from 1.4% to 3.8% over a five-year period (Shoemaker and Howell, 2005) .

In recent years, hospital administrators have been under tremendous pressure to cut costs and have resorted to cross-training programs (Lyons, 1992; Siferd and Benton, 1992) for nurses, enabling them to float between units in the same specialization. These units have varying acuity levels, but are similar enough to cross-train nurses. Cross-training (floating / flexibility) of nurses helps to meet heavy demand in one unit by using nursing hours from another unit where the demand is lean. Inman et al. (2005) show that cross-training not only helps to meet the variable demand, but can also reduce staffing costs for the hospital as well as improve morale and job satisfaction for the nurses. Many hospitals have reaped financial benefits from successfully implementing cross-training programs (Altimier and Sanders, 1999; Snyder and Nethersole-Chong, 1999). This way of floating/cross-training nurses is called "staffing flexibility or human resource flexibility" in operations literature.

In the second scenario, when bed space is not available, patients are admitted to a higher/more sophisticated unit where bed space is available. The more sophisticated unit where patient is upgraded usually has necessary equipment to more than meet the needs of patient moved in from the simple unit, and therefore more costly to operate. For example, assume the more sophisticated unit to be intensive care unit in the prenatal-floor while the other unit is the medical/surgical unit in the same pre-natal floor. A newborn who requires treatment from med/surg unit can easily be treated at PICU, since both units have sufficient resources to provide pre-natal care. This type of upgrades is called "demand upgrade or downward substitution or capacity flexibility" in operations literature.

In the case where bed space is unavailable in any unit, the patient is directed to nearby hospital.

Hospitals widely use both staffing flexibility and demand upgrades as a means to meet variable demand. In this paper, we examine the following questions: How much of each resource - nursing staff and bed space - is required under different types of flexibility (e.g.) staffing flexibility, demand upgrades and both? How does type of flexibility affect optimal capacity and staffing decisions? How much cross-training should be conducted under each type of flexibility?

Hospitals plan for staffing and bed space within different time windows. Usually bed space, hereafter called the capacity of a unit, is determined when the unit is constructed. Staffing decisions, on the other hand, are typically made once per year. It is quite difficult to anticipate future needs of the hospital, including type of flexibility to be used, so hospitals often make myopic decisions regarding both capacity and staffing. As an example, a academic medical center in south-east state very recently decided to hire a pool of nurses who can be floated in multiple units. Though staffing and capacity decisions were made earlier, when units were constructed, flexibility of staff is considered only now. So, this leads to the question: Is timing of capacity and staffing decisions important? In this paper, we study the extent to which timing of staffing and capacity decisions affect system performance. We consider two types of decision making, simultaneous decision making where capacity and staffing decisions are made at the same point of time and sequential decision making where capacity decisions precede staffing decisions. How much impact does timing of operational decisions have on hospital performance? The model was motivated from in-depth interviews with practitioners and extensive literature review. The hospital we analyzed is a public, academic medical center operated by a south-east state government. This academic medical center includes a main hospital, children's hospital, neurosciences hospital, and women's hospital. Construction of the new cancer hospital as a part of the academic medical center is now under way, with completion tentatively expected in 2009. Approximately 61,200 patients visited the emergency room (ER) and there were 741,980 clinic visits at the hospitals' 724 bed facility during 2006. Here on, this hospital will be referred as south-east academic, medical center.

Since we model two types of flexibility that are widely used in hospitals, it is interesting to check if both flexibilities are required or one of them is sufficient. Is there an interaction effect between the two types of flexibilities? Does one type of flexibility complement or supplement the other? Does demand variation affect the benefit of flexibility? What is the effect of cost parameters on staffing and capacity decisions?

This paper is motivated by a hospital setting and so the paper will be geared towards the application of the model to hospitals. The model can also be applied to call-centers, hotels and other service applications where two input resources are required to meet uncertain demand.

3.2 Literature Review

Flexibility literature in operations management is huge and has been widely researched since the 90's. Flexibility literature under operations management can be broadly divided into literature focussing on manufacturing flexibility and literature focussing on human resource flexibility. Our research derives elements from manufacturing flexibility (downward substitution) and human resource flexibility (cross-training in parallel systems). The following section, explains the literature related to manufacturing flexibility and human resource flexibility and points out the differences between our research and relevant papers in both the categories.

3.2.1 Manufacturing flexibility

Van Mieghem (1998) studies the effect of cost and price differentials on flexibility of two-product, three-plant system. He maximizes total revenue by optimizing capacities in two dedicated-capacity plants and one flexible-capacity plant. While capacities are unconstrained in Van Mieghem's paper, the equivalent resource in our paper, nursing staff is constrained by available bed spaces. The demand faced by nursing staff is not the true patient demand but demand constrained by available bed space capacity. Fine and Freund (1990) uses two stage stochastic programming to determine optimal investment capacities for 'n' products having 'n' dedicated capacity and one flexible capacity under discrete demand distribution. They prove that there exists an optimal solution when acquisition and production costs are linear. Harrison and Van Mieghem (1999) uses multi-dimensional news vendor model to determine optimal investment strategy for multiple resource that can dynamically readjust investment levels based on uncertain demand over successive time periods. They do not consider flexibility of those resources as we have done in our paper. Netessine et al. (2002) consider a firm with multiple services. The demand for lower level services can be satisfied with resources for higher level services. For single level upgrades and correlated demand they obtain sufficient and necessary conditions for optimal solution. They assume one resource only that is totally flexible. Similarly, Rao et al. (2004) use a combination of optimization techniques to determine optimal production decisions in a multi-product inventory planning problem with downward substitution and setup cost. In our paper, we coordinate for two resources to satisfy uncertain demand and both resources have different levels of flexibility.

Iravani et al. (2005) obtain a structural flexibility index that quantifies the ability of flexible systems to respond to variability. They evaluate this index through simulation to determine the performance of serial and parallel networks. Jordan and Graves (1995) consider 'M' products in 'N' plants and analyze the level of flexibility required to reap significant benefits. They conclude that small additional flexibility is sufficient to get results of total flexibility. Graves and Tomlin (2003) has extended the idea of Jordan and Graves (1995) to multi-stage supply chains.

Chod and Rudi (2005) and Bish and Wang (2004) consider a two-product, price setting firm which makes investment decisions before demand is realized and then allocates resources and they determine prices after the demands are realized. Chod and Rudi optimize for flexible resource given dedicated resources while Bish and Wang optimize for resource portfolio of two dedicated and one flexible resource. Bish et al. (2005) show that system performance depends heavily on allocation mechanism when flexible capacity is used to hedge against demand uncertainty for make-to-order products. Goyal and

Netessine (2005) obtain conditions when technologies using volume and product flexibility are chosen to mitigate uncertain demand. Chod et al. (2006) determine the impact of three types of flexibilities viz., product mix, time and volume on the capacity output and pricing decisions for a two-product firm.

3.2.2 Human resource flexibility

Cross-training in manufacturing has widely been used to balance work load in an assembly system in order to maximize throughput. A recent paper by Hopp et al. (2004) analyze two different cross-training structures, skill chaining and cherry-picking, for a serial production system. They find that when capacity is fairly imbalanced but variability is low, cherry picking approach can be used. Inman et al. (2004) show that chaining is an effective cross-training strategy to mitigate losses due to absenteeism of workers in assembly line. Vairaktarakis and Winch (1999) develops heuristics for scheduling work orders through assembly systems so that cross-training costs are minimized when multi-skilled workers are used. The above papers finds the best configuration or flexibility structure (level of cross-training) in different scenarios and does not deal in optimizing for amount of cross-training.

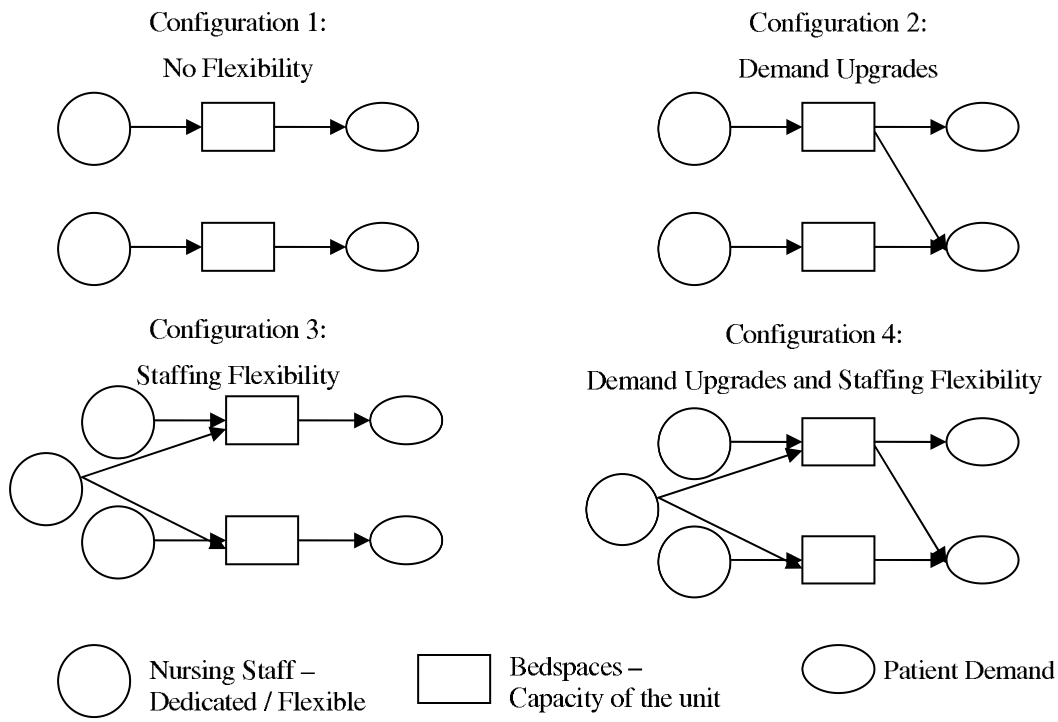
Campbell (1999) develops a model for allocating cross-trained workers to a multi-department service environment. He determines the benefit of cross-training using a simulation study. His model maximizes utility (weighted sum of squared requirement that is satisfied) of meeting the requirements, considering the capabilities of workers subject to their allocation in different departments and assuming different levels of cross-training (training breadth). Pinker and Shumsky (2000) analyzes a system with specialist and flexible servers when there is a trade off between efficiency of specialist and quality of flexible servers. Jordan et al. (2004) evaluate performance of three cross-training configurations in parallel systems using queueing theory and simulation. They conclude that complete chaining gives maximum benefit and is also robust. Inman et al. (2005) simulate and show economic advantages of cross-training nurses and also develop simple spread sheets formulas to help in staffing.

Agnihothri et al. (2003) balances the trade off between customer delay cost and premium for flexibility and models a queueing system to determine mix of dedicated and cross-trained servers for two job types using simulation, and extend this work to three job types in their 2004 paper (Agnihothri and Mishra, 2004). Chakravarthy and Agnihothri (2005) use discrete values for mix of dedicated and cross-trained servers to study the effect of parameters on choice of server mix. Brusco and Johns (1998) present an ILP to evaluate cross-training configurations for a multi-skilled work force. Brusco et al. (1998) minimizes the total number of labor for two-skill class considering productivity. They do not optimize on level of flexibility but assume that all labors are totally flexible. McClain et al. (2000) show that work-in-process inventory has a significant effect on productivity of workers when there is work sharing in a serial system. Tekin et al. (2004) assume that all skilled workers for a call-center are pooled and using queuing theory study the effect of first-come-first-served and non-preemptive priority service disciplines on system parameters. The above papers determine the benefit of cross-training using either queuing or simulation and does not account for capacity constrains on staffing and possibility of demand upgrades.

3.3 Problem Definition

This section explains the structure of the model and is divided into two sub-sections. The first sub-section elaborates on the four types of flexibility configurations used. The second sub-section explains the timeline for decision making that are used to make capacity and staffing decisions for each flexibility configuration. We consider two non-homogeneous hospital units, one being a complex unit with high patient acuity and the other being a simple unit with lower patient acuity. Patients in the simple unit, in some configurations, can be upgraded to complex unit but not vice-versa. For notation purposes, we assume unit 1 to be a complex unit and unit 2 to be a simple unit. For example, assume unit 1 to be the intensive care unit in the pre-natal floor while unit 2 is the medical/surgical unit in the same pre-natal floor. Both the units are under the same clinical grouping (pre-natal) so if there is not enough capacity to admit a newborn in the med/surg unit, the new born is upgraded to ICU. These two units can be staffed with three types of nursing staff, depending on availability. The in-house regular nurses are assigned to their home unit (hereafter referred to as dedicated nurses) while the in-house flexible nurses (hereafter referred to as flexible nurses) are assigned to either of the two units based on demand. The dedicated nurses and flexible nurses are called total staff in the notation and formulation. If patient demand is still not met, then travel/contract nurses from an outside agency (hereafter referred to as contract nurses) are hired (at higher cost than either unit). The two non-homogenous units have different capacities, measured in

FIGURE 3.1: Four Configurations



terms of number of bed spaces. We assume that one patient needs only one bed space for his/her treatment and that one nurse treats one patient.

3.3.1 Configurations

This section describes the model using two types of flexibility under four configurations. Figure 3.1 shows the network representation of the four configurations. Nursing staff (number of dedicated and flexible nurses) and capacity (number of bed spaces) for each unit are the decision variables in our model. **Configuration 1 : Base case No flexibility**

When both bed space and nursing staff are available, the patient is admitted and treatment proceeds. In this configuration nursing staff consists of only dedicated nurses; no nurses have been cross-trained. If bed space is available when a patient arrives but nursing staff is not available, contract nurses are hired at a cost s_i for unit i. If a bed space is not available, the patient is directed to another hospital and the system incurs a penalty cost of p_i for unit i.

Configuration 2 : Demand Upgrades

When both capacity and nursing staff are available, the patient for unit i is admitted to unit i. Here again, there are no flexible nurses. Nursing staff consists of dedicated nurses and contract nurses.

Unlike configuration 1, in this configuration, when capacity is not available in unit 2, patient is admitted to unit 1 provided unused capacity is available in unit 1. Such upgrades to unit 1 are allowed only when demand in unit 1 is first met. The dedicated nurses in unit 1 are not trained to handle patients from unit 2 and so contract nurses are hired to treat patients upgraded to unit 1. Since unit 1 is a complex/sophiticated unit, the equipment is sufficient to treat patients from unit 2 but the staff are not trained in that skill. Therefore, contract nurses are hired at a higher cost to treat the upgraded patients in unit 1. Patients for unit 2 are turned away at a cost of p_2 if capacity is not available in unit 2 and if upgrade to unit 1 is not possible. If enough capacity in unit 1 is not available, patients for unit 1 are turned away at a cost of p_1.

If capacity is available in unit i but dedicated staff is not available in unit i then contact nurses are hired for unit i.

Configuration 3 : Staffing Flexibility

In this configuration we use three types of nursing staff (dedicated, flexible and contract nurses). Dedicated nurses are trained to work only in their home unit. Cost of wages for them is h_i for unit i. Flexible nurses are cross-trained to work in both unit 1 and unit 2. These cross-trained nurses, also called float nurses, can work in either unit 1 or unit 2. Cost of wages for flexible nurses is t. We assume flexible nurses are equally productive in both unit 1 and unit 2. The third type of nursing staff are the contract nurses who are hired at an additional cost of s_i for unit i.

Unlike configuration 2, in this configuration we do not have demand upgrades. If both capacity and nursing staff are available, patient is admitted for treatment at the appropriate unit. If capacity is not available, patients are turned away, incurring a penalty cost of p_i.

In each unit, at first dedicated nurses are assigned. If demand exceeds the number of dedicated nurses, flexible nurses are used. Flexible nurses are first assigned to unit 1 because unit 1 is the complex unit and hiring contract nurses for unit 1 is more expensive. Any remaining flexible nurses are assigned to unit 2. If dedicated and flexible nurses are still not able to meet demand, contract nurses are hired as needed.

Configuration 4 : Demand Upgrades and Staffing Flexibility

This configuration is highly flexible. Both types of flexibility, demand upgrades and staffing flexibility, are used in this model. If capacity is not available in both units, patients are turned away incurring a penalty cost of p_i. When capacity is not enough in unit 2 patients are upgraded to unit 1, provided there are beds not being used by unit 1 patients.

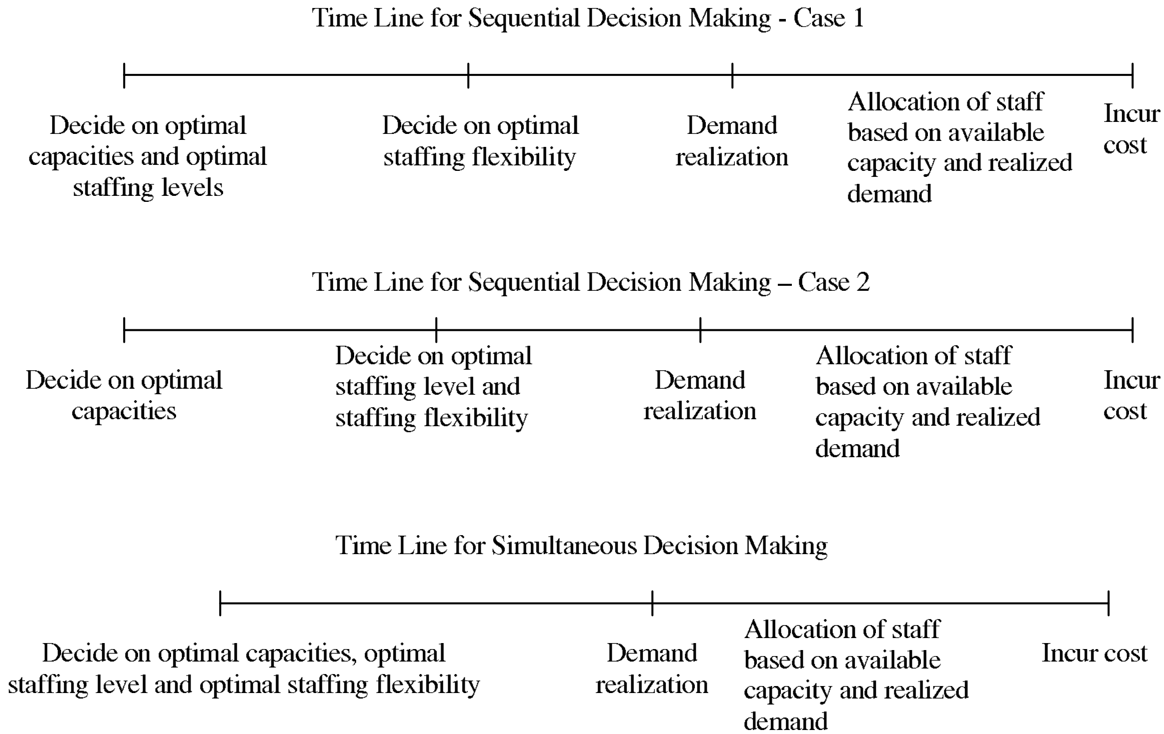
Here again three types of nursing staff are used as in configuration 3. Dedicated nurses are assigned first to meet the demand in unit i, followed by assigning flexible nurses to unit 1. Any excess flexible nurses are assigned to unit 2, as needed. Finally, contract nurses are hired to meet the remaining demand in each unit.

3.3.2 Timeline for decision making

This section motivates and explains the framework for the timing of decision making. In all cases, the actual allocation of nursing staff and beds are made after demand is realized. The timing issues we discuss here relate to decisions regarding capacity (number of beds) and nurse staffing levels (both dedicated and flexible). As indicated in Figure 2, we consider two types of sequential decision making and simultaneous decision making The south-east academic, medical center, for example, had decided on capacity and total staffing levels initially when the units were created. Now, in order to reduce labor costs, they are planning to implement cross-training programs. The cross-training program chooses some dedicated nurses and trains them to float to other units. This motivated us to study the impact of decision timing on system performance. We model sequential decision making that closely represents the process followed by the south-east academic, medical center. Ideally, both capacity and staffing decisions has to be made at the same point of time. So, in our paper we also model simultaneous decision making.

We consider two cases of sequential decision making. In sequential decision making case 1, capacity and total staffing decisions are made initially. Given capacity and total staff, we later determine number of flexible nurses to train in order to minimize total expected cost. This type of decision making is applicable only in configuration 3 and

FIGURE 3.2: Timeline for Configurations



configuration 4, where staffing flexibility is modeled.

In sequential decision making case 2, the capacity decisions are made originally and all staffing decisions (including dedicated and flexible staff) are made later. This type of decision making is applicable in all four configurations.

With simultaneous decision making, both capacity and staffing (both dedicated and flexible staff) decisions are made at the same time.

3.4 Model Formulation and Analysis

This section defines the notation, discusses assumptions, formulates and analyzes models under each flexibility configuration for sequential and simultaneous decision making.

The two hospital units, each of which has a pool of nursing resources and fixed capacity (bed spaces), faces a stochastic demand. In practice, supply and demand for each unit is measured as the number of full-time-equivalent (FTE) nurses required each day. The full-time-equivalents can also be represented as nursing hour requirements for each unit. In our model, all allocations and demand are measured on a continuous scale, indicating that the unit of measurement is nursing hours. Capacity is also approximated to continuous scale for simplicity. Stochastic demand is represented as the number of nursing hours required. Utilizing one bed space is equivalent to serving one patient or one full-time equivalent nursing hours.

Demand for each unit i is stochastic and follows a general, continuous distribution with the cumulative distribution function Φ_i. The realization of demand is represented as d_i for unit i. The following list summarizes the notations including cost parameters and decision variables used in our models.

Cost parameters

s_i : contract wages for hiring contract nurses in unit i for an hour

h_i : wages for a dedicated nurse for unit i per hour

t : wages for a cross-trained nurse per hour

f_i : operating cost per bed space in unit i per hour

p_i : penalty for losing patients per hour in unit i

Staffing decisions

z_i : number of total staff in unit i

n_i : number of dedicated nurses in unit i

e_i : number of flexible (cross-trained) nurses in unit i

e : number of flexible (cross-trained) nurses available for both units

Staffing allocations

x_i : number of dedicated nurses allocated to unit i

y_i : number of flexible (cross-trained)nurses allocated to i

Capacity decisions

k_i : capacity (number of bed spaces) available in unit i

Capacity allocations

a_i : number of bed spaces allocated to treat patients in unit i

a_u : number of bed spaces in unit *1* allocated to treat patients in unit *2*

Demand

d_i : number of patients to be treated in unit i per hour - realization of stochastic demand

Φ_i : stochastic demand for unit i

Under all configurations, total staff in unit i is the sum of dedicated nurses and flexible nurses $z_1 + z_2 = n_1 + n_2 + e$. By design configuration 1 and 2 do not use any flexible nurses and so $z_i = n_i$. To compare sequential case 1 and case 2 decision making, we take $\sum_i e_i = e$.

The following list enumerates the assumptions that have to hold between cost parameters so that trivial solutions are eliminated from the model.

1. $h_2 < h_1 < t < s_2 < s_1$
2. $s_2 < h_1 + h_2$
3. $f_2 < f_1 < p_2 < p_1$
4. $p_i > s_i + f_i, \forall\, i = 1, 2$

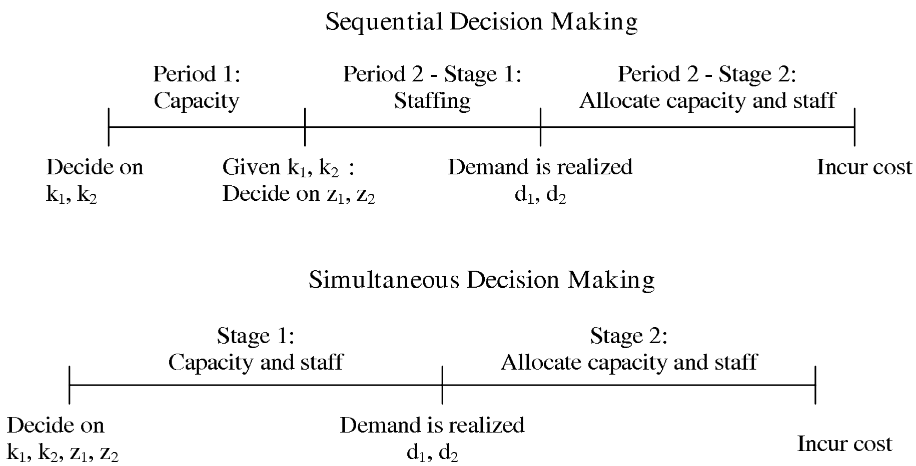
5. $p_2 > s_1 + f_1$

The first assumption prioritizes allocation of nursing staff. Contract nurses are used only when dedicated and flexible nurses are insufficient to meet demand. The second assumption is required to avert the possibility of hiring dedicated nurses from both unit 1 and unit 2 instead of using a contract nurse. The third assumption prioritizes allocations for capacity. Since unit 1 is complex, the per unit capacity and staffing costs are expensive compared to the equivalent cost in unit 2. The fourth assumption prevents the model from losing all demand. The fifth assumption prevents the scenario where it is better off to lose patients in unit 2 than upgrade them to unit 1.

All models are formulated as two-stage stochastic programming, with second stage being the actual assignment of patients and nurses to floors, after demand has been realized. Demand in both units follow general, continuous distribution. The second stage decisions for all configurations are convex in their objective functions, so decisions are determined using first order conditions. The expected value of the second stage objective function is then substituted into the first stage objective function. After presenting the formulation, we prove convexity of first stage objective function, and determine first order conditions. Solving first order conditions, we get optimal values for decisions variables of interest.

Under sequential decision making cases we consider two periods while in simultaneous decision making we consider only one time period. Sequential case 1 decision making does not apply to configuration 1 and 2 because of lack of staffing flexibility. So, sequential case 2 is termed as just sequential decision making under configuration 1 and 2.

FIGURE 3.3: Timeline for Configurations 1 and 2



3.4.1 Configuration 1 : No flexibility

Configuration 1 is the base case configuration without any type of flexibility. The formulation and analysis for sequential and simultaneous decision making is given in this section. Given the lack of flexibility in configuration 1, the second stage allocations are straightforward in all cases.

Sequential Decision Making

Sequential decision making in configuration 1 has two periods. Capacity (k_i) is decided in period 1 according to news vendor quantity. We minimize sum of fixed capacity cost and expected penalty cost when demand follows a general distribution. After capacity is decided, staffing decisions are made in period 2 using news vendor approach. Sum of nurse wages and expected contract nurse cost is minimized to determine optimal staffing decisions (z_i). Sequential decision making for configuration 1 can be seen in Figure 3.3.

The following formulation is $\forall\ i = 1, 2$.

Period 1 :

$$\begin{aligned} \text{Min}_{k_i} \quad & f_i \cdot k_i + E_{\Phi_i} p_i \cdot (d_i - min(d_i, k_i)) \\ \text{subject to} & \\ k_i & \geq 0, \forall \ i = 1, 2 \end{aligned}$$

Period 2 determines staffing decisions given capacity decisions from period 1.

Period 2 Stage 1:

$$\begin{aligned} \text{Min}_{z_i} \quad & h_i \cdot z_i + E_{\Phi_i} \Omega(z_i, d_i) \\ \text{subject to} & \\ z_i & \geq 0, \forall \ i = 1, 2 \end{aligned}$$

Period 2 Stage 2 : $\Omega(z_i, d_i)$:

$$\begin{aligned} \text{Min}_{x_i, a_i} \quad & s_i \cdot (a_i - x_i) + p_i \cdot (d_i - a_i) \\ \text{subject to} & \\ a_i & \leq min(k_i, d_i) \\ x_i & \leq z_i \\ a_i - x_i & \geq 0 \\ x_i, a_i & \geq 0, \forall \ i = 1, 2 \end{aligned}$$

In stage 2 of period 2, allocated capacity (a_i) cannot be more than available capacity or demand. Also, allocated dedicated nurses (x_i) cannot be more than available dedicated nurses (z_i) or allocated capacity (a_i).

Optimal bed capacity is determined in period 1, by minimizing (3.1), and their staffing levels are set in period 2 by minimizing (3.2) (subject to $z_i \leq k_i$).

$$\Lambda_{11}(k_i) = f_i \cdot k_i + p_i \int_{k_i}^{\infty} (d_i - k_i) d\Phi_i \tag{3.1}$$

$$\Lambda_{12}(z_i) = h_i \cdot z_i + s_i \int_{z_i}^{k_i} (d_i - z_i) d\Phi_i + s_i \int_{k_i}^{\infty} (k_i - z_i) d\Phi_i \tag{3.2}$$

First order conditions lead to closed form expressions:

$$\Phi_i(k_i^*) = \frac{p_i - f_i}{p_i} \tag{3.3}$$

$$\Phi_i(z_i^*) = \frac{s_i - h_i}{s_i} \tag{3.4}$$

Since the formulation closely represents newsvendor problem, proof for convexity is not shown here. Optimal capacity and optimal staffing are the newsvendor quantities as shown in equations (3.3) and (3.4). Optimal staffing (z_i^*) is less than optimal capacity obtained from equations (3.3) and (3.4) only when wages for nurses h_i is more than operating cost f_i. If operating cost is more than wages then, we set $z_i^* = k_i^*$ since objective function (3.1) is not feasible when z_i^* is greater than k_i^*.

Simultaneous Decision Making

Simultaneous decision making has only one period. In stage 1, both capacity (k_i) and total staff (z_i) are decided. In stage 2, allocations are made after demand is realized. At the end of the period cost is incurred. Simultaneous decision making for configuration 1 can be seen in Figure 3.3.

Stage 1 :

$$\begin{aligned}
\text{Min}_{k_i,z_i} \qquad & f_i \cdot k_i + h_i \cdot z_i + E_{\Phi_i}\Omega(k_i, z_i, d_i) \\
\text{subject to} \qquad & \\
k_i, z_i \ \geq \ & 0, \forall \ i = 1,2
\end{aligned}$$

Stage 2 : $\Omega(k_i, z_i, d_i)$:

$$\begin{aligned}
\text{Min}_{x_i,a_i} \qquad & s_i \cdot (a_i - x_i) + p_i \cdot (d_i - a_i) \\
\text{subject to} \qquad & \\
a_i \ \leq \ & min(k_i, d_i) \\
x_i \ \leq \ & z_i \\
a_i - x_i \ \geq \ & 0 \\
x_i, a_i \ \geq \ & 0, \forall \ i = 1,2
\end{aligned}$$

Second stage constraints restricts allocated capacity not to exceed demand or available capacity and restricts allocated staff not to exceed available staff or allocated capacity.

The first stage objective (3.5) is a function of capacity and staffing.

$$\Lambda_{21}(k_i, z_i) = f_i \cdot k_i + h_i \cdot z_i + s_i \int_{z_i}^{k_i} (d_i - z_i) d\Phi_i + s_i \int_{k_i}^{\infty} (k_i - z_i) d\Phi_i + p_i \int_{k_i}^{\infty} (d_i - k_i) d\Phi_i \quad (3.5)$$

The optimal levels of capacity (k_i^*) and staffing (z_i^*) are found by solving equations (3.6) and (3.7).

$$\Phi_i(k_i^*) = \frac{p_i - f_i - s_i}{p_i - s_i} \quad (3.6)$$

$$\Phi_i(z_i^*) = \frac{s_i - h_i}{s_i} \quad (3.7)$$

Optimal staffing and optimal capacity are still determined by a news vendor-type relationship, but optimal capacity now depends also on cost of contract nurses. If by using equations (3.6) and (3.7), optimal staffing is more than optimal capacity we set $z_i^* = k_i^*$. It is easily shown that result that simultaneous k_i^* is less than sequential k_i^*.

3.4.2 Configuration 2 : Demand Upgrades

Configuration 2 allows for one type of flexibility, demand upgrades. The formulations and analysis for sequential and simultaneous decision making are given below.

Sequential Decision Making

Sequential decision making in configuration 2 has two periods. In the first period, capacity decisions are made assuming possibility for demand upgrades in stage 2. Capacity is decided by minimizing sum of fixed capacity cost and expected penalty cost under stochastic demand allowing for upgrades. In the next period, given capacity from period 1, total staff (z_i) is determined minimizing sum of total wage and expected contract nurse cost. Sequential decision making can be seen in Figure 3.3.

Period 1 :

$$\text{Min}_{k_i} \qquad \sum_i (f_i \cdot k_i) + E_{\Phi_i} p_1(d_1 - min(d_1, k_1)) + p_2(d_2 - min(d_2, k_2 + (k_1 - d_1)^+))$$

subject to

$$k_i \;\geq\; 0, \forall \; i = 1, 2$$

Period 2 Stage 1 :

$$\text{Min}_{z_i} \qquad \sum_i (h_i \cdot z_i) + E_{\Phi_i} \Omega(z_i, d_i)$$

subject to

$$z_i \;\geq\; 0, \forall \, i = 1, 2$$

Period 2 Stage 2 : $\Omega(z_i, d_i)$:

$$\begin{aligned}
\text{Min}_{x_i,a_i} \quad & s_1 \cdot (a_1 + a_u - x_1) + s_2 \cdot (a_2 - x_2) + \\
& p_1 \cdot (d_1 - a_1) + p_2 \cdot (d_2 - a_2 - a_u)
\end{aligned}$$

subject to

$$\begin{aligned}
a_i &\leq min(k_i, d_i), \forall\, i = 1, 2 \\
x_i &\leq z_i, \forall\, i = 1, 2 \\
a_i - x_i &\geq 0, \forall\, i = 1, 2 \\
a_1 + a_u &\leq k_1 \\
a_2 + a_u &\leq d_2 \\
x_i, a_i &\geq 0, \forall\, i = 1, 2, u
\end{aligned}$$

Objective function in stage 2 of period 2 minimizes sum of cost of hiring contract nurses and penalty cost of not meeting the demand. Constraints indicate upper and lower limits for capacity and staffing allocations.

The first period objective (3.8) is a function of capacity alone.

$$\begin{aligned}
\Gamma_{11}(k_i) &= \sum_i (f_i \cdot k_i) + p_1 \int_{k_1}^{\infty} \int_0^{k_2} (d_1 - k_1)\, d\Phi_2\, d\Phi_1 \\
&\quad + p_2 \int_0^{k_1} \int_{k_1+k_2-d_1}^{\infty} (d_1 + d_2 - k_1 - k_2)\, d\Phi_2\, d\Phi_1 \\
&\quad + \int_{k_1}^{\infty} \int_{k_2}^{\infty} (p_1(d_1 - k_1) + p_2(d_2 - k_2))\, d\Phi_2\, d\Phi_1
\end{aligned} \tag{3.8}$$

Analysis of second order condition proves that first period objective function (3.8) is convex (see appendix A.1) in (k_i). Solving the following first order conditions gives optimal capacity (k_i^*) for period 1.

$$\begin{aligned}
\frac{\partial \Gamma_{11}}{\partial k_1} &= (f_1 - p_1) + (p_1 - p_2)\Phi_1(k_1) + p_2 \int_0^{k_1} \Phi_2'(k_1 + k_2 - d_1) d\Phi_1 \\
\frac{\partial \Gamma_{11}}{\partial k_2} &= (f_2 - p_2) + (1 - \Phi_1(k_1))\Phi_2(k_2) + p_2 \int_0^{k_1} \Phi_2'(k_1 + k_2 - d_1) d\Phi_1
\end{aligned}$$

In the second period, equation (3.9) is the first stage objective which is a function of total staff.

$$\begin{aligned}
\Gamma_{12}(z_i) &= \sum_i (h_i \cdot z_i) + \Gamma_{12A} + \Gamma_{12B} + \Gamma_{12C} + \Gamma_{12D} \qquad (3.9) \\
\Gamma_{12A} &= s_1 \int_{z_1}^{k_1} \int_0^{z_2} (d_1 - z_1)\, d\Phi_2\, d\Phi_1 + s_2 \int_0^{z_1} \int_{z_2}^{k_2} (d_2 - z_2)\, d\Phi_2\, d\Phi_1 \\
&\quad + \int_{z_1}^{k_1} \int_{z_2}^{k_2} (s_1(d_1 - z_1) + s_2(d_2 - z_2))\, d\Phi_2\, d\Phi_1 \\
\Gamma_{12B} &= s_1 \int_{k_1}^{\infty} \int_0^{k_2} (k_1 - z_1)\, d\Phi_2\, d\Phi_1 + s_2 \int_{k_1}^{\infty} \int_{z_2}^{k_2} (d_2 - z_2)\, d\Phi_2\, d\Phi_1 \\
\Gamma_{12C} &= s_1 \int_0^{z_1} \int_{k_2}^{k_1+k_2-d_1} (d_2 - k_2)\, d\Phi_2\, d\Phi_1 + s_1 \int_{z_1}^{k_1} \int_{k_2}^{k_1+k_2-d_1} (d_1 + d_2 - k_2 - z_1)\, d\Phi_2\, d\Phi_1 \\
&\quad + s_1 \int_0^{k_1} \int_{k_1+k_2-d_1}^{\infty} (k_1 - z_1)\, d\Phi_2\, d\Phi_1 + s_2 \int_0^{k_1} \int_{k_2}^{\infty} (k_2 - z_2)\, d\Phi_2\, d\Phi_1 \\
\Gamma_{12D} &= \int_{k_1}^{\infty} \int_{k_2}^{\infty} (s_1(k_1 - z_1) + s_2(k_2 - z_2))\, d\Phi_2\, d\Phi_1
\end{aligned}$$

Again, the second order conditions show that objective function (3.9) is convex (see appendix A.1) in (z_i). Equating the following first order conditions to zero and solving them simultaneously we determine optimal staffing (z_i^*).

$$\begin{aligned}
\frac{\partial \Gamma_{12}}{\partial z_1} &= (h_1 - s_1) + s_1 \int_0^{z_1} \Phi_2(k_1 + k_2 - d_1) d\Phi_1 \\
\frac{\partial \Gamma_{12}}{\partial z_2} &= (h_2 - s_2) + s_2 \Phi_2(z_2) \\
\text{So, } \Phi_2(z_2^*) &= \frac{s_2 - h_2}{s_2}
\end{aligned}$$

Simultaneous Decision Making

In simultaneous decision making, capacity (k_i) and staffing decisions (z_i) are made in stage 1. In stage 2 when demand is realized, staffing and capacity allocations are made assuming that demand upgrades are possible. At the end of the period fixed capacity cost, wage, expected penalty and contract nurse costs are incurred. Simultaneous decision making can be seen in Figure 3.3.

Stage 1 :

$$\begin{aligned}
\text{Min}_{k_i, z_i} \qquad & \sum_i (f_i \cdot k_i + h_i \cdot z_i) + E_{\Phi_i} \Omega(k_i, z_i, d_i) \\
\text{subject to} \qquad & \\
k_i, z_i \geq{} & 0, \forall\, i = 1, 2
\end{aligned}$$

Stage 2 : $\Omega(k_i, z_i, d_i)$:

$$\begin{aligned}
\text{Min}_{x_i,a_i} \quad & s_1 \cdot (a_1 + a_u - x_1) + s_2 \cdot (a_2 - x_2) + \\
& p_1 \cdot (d_1 - a_1) + p_2 \cdot (d_2 - a_2 - a_u)
\end{aligned}$$

subject to

$$\begin{aligned}
a_i &\leq min(k_i, d_i), \forall\, i = 1, 2 \\
x_i &\leq z_i, \forall\, i = 1, 2 \\
a_i - x_i &\geq 0, \forall\, i = 1, 2 \\
a_1 + a_u &\leq k_1 \\
a_2 + a_u &\leq d_2 \\
x_i, a_i &\geq 0, \forall\, i = 1, 2, u
\end{aligned}$$

All the constraints explained in sequential decision making are applicable in simultaneous case as well.

Beds are allocated to patients in their respective units. Any excess patients in unit 2 are assigned to unit 1 if beds are available in unit 1. The optimal capacity allocation variables are a_1^*, a_2^* and a_u^*, with a_u^* representing the beds in unit 1 that are used for patients of unit 2. Nurses are allocated to beds in their units that have appropriate patients; any nursing shortfall is met by using contract nurses.

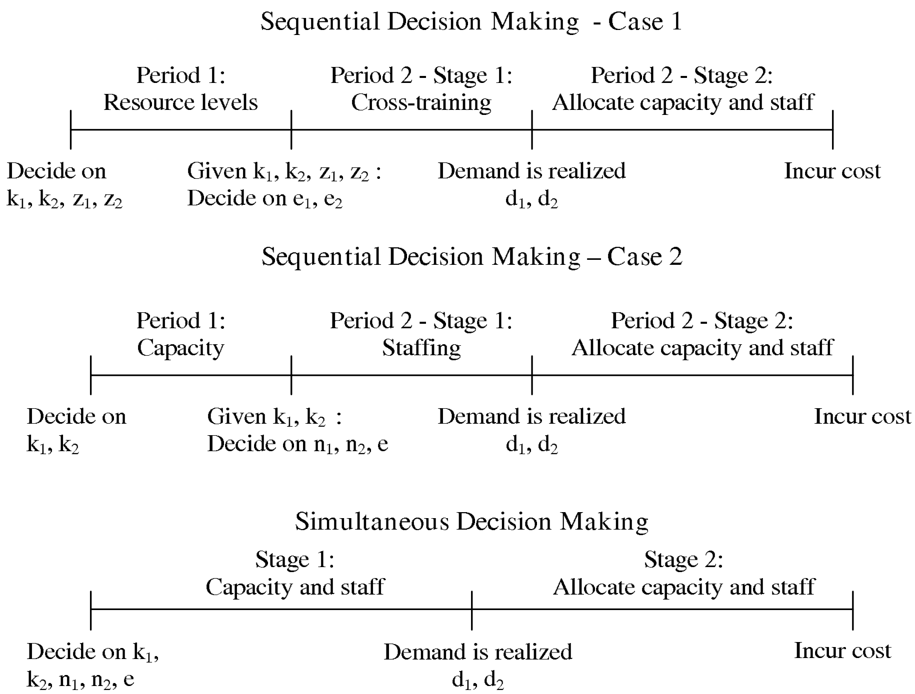
First stage objective function (3.10) is the expected value of capacity and staffing allocations substituted in second stage objective function.

$$
\begin{aligned}
\Gamma_{21}(k_i, z_i) &= \sum_i (f_i \cdot k_i + h_i \cdot z_i) + \Gamma_{21A} + \Gamma_{21B} + \Gamma_{21C} + \Gamma_{21D} + \Gamma_{21E} \qquad (3.10) \\
\Gamma_{21A} &= s_1 \int_{z_1}^{k_1} \int_0^{z_2} (d_1 - z_1)\, d\Phi_2\, d\Phi_1 + s_2 \int_0^{z_1} \int_{z_2}^{k_2} (d_2 - z_2)\, d\Phi_2\, d\Phi_1 \\
&\quad + \int_{z_1}^{k_1} \int_{z_2}^{k_2} (s_1(d_1 - z_1) + s_2(d_2 - z_2))\, d\Phi_2\, d\Phi_1 \\
\Gamma_{21B} &= s_1 \int_{k_1}^{\infty} \int_0^{k_2} (k_1 - z_1)\, d\Phi_2\, d\Phi_1 + s_2 \int_{k_1}^{\infty} \int_{z_2}^{k_2} (d_2 - z_2)\, d\Phi_2\, d\Phi_1 \\
\Gamma_{21C} &= s_1 \int_0^{z_1} \int_{k_2}^{k_1+k_2-d_1} (d_2 - k_2)\, d\Phi_2\, d\Phi_1 + s_1 \int_{z_1}^{k_1} \int_{k_2}^{k_1+k_2-d_1} (d_1 + d_2 - k_2 - z_1)\, d\Phi_2\, d\Phi_1 \\
&\quad + s_1 \int_0^{k_1} \int_{k_1+k_2-d_1}^{\infty} (k_1 - z_1)\, d\Phi_2\, d\Phi_1 + s_2 \int_0^{k_1} \int_{k_2}^{\infty} (k_2 - z_2)\, d\Phi_2\, d\Phi_1 \\
\Gamma_{21D} &= \int_{k_1}^{\infty} \int_{k_2}^{\infty} (s_1(k_1 - z_1) + s_2(k_2 - z_2))\, d\Phi_2\, d\Phi_1 \\
\Gamma_{21E} &= p_1 \int_{k_1}^{\infty} \int_0^{k_2} (d_1 - k_1)\, d\Phi_2\, d\Phi_1 + p_2 \int_0^{k_1} \int_{k_1+k_2-d_1}^{\infty} (d_1 + d_2 - k_1 - k_2)\, d\Phi_2\, d\Phi_1 \\
&\quad + \int_{k_1}^{\infty} \int_{k_2}^{\infty} (p_1(d_1 - k_1) + p_2(d_2 - k_2))\, d\Phi_2\, d\Phi_1
\end{aligned}
$$

Analysis of second order conditions (see appendix A.2) show that first order conditions can only be used to determine optimal values of staffing (z_i^*), as follows:

$$
\begin{aligned}
\frac{\partial \Gamma_{21}}{\partial z_1} &= (h_1 - s_1) + s_1 \int_0^{z_1} \Phi_2(k_1 + k_2 - d_1)\, d\Phi_1 \\
\frac{\partial \Gamma_{21}}{\partial z_2} &= (h_2 - s_2) + s_2 \Phi_2(z_2) \\
\text{So, } \Phi_2(z_2^*) &= \frac{s_2 - h_2}{s_2}
\end{aligned}
$$

FIGURE 3.4: Timeline for Configurations 3 and 4



3.4.3 Configuration 3 : Staffing Flexibility

Configuration 3 allows for staffing flexibility but no demand upgrades. The formulations and analysis for sequential and simultaneous decision making are presented in this section.

Sequential Decision Making - Case 1

Sequential decision making case 1, has two periods. In period 1, capacity (k_i) and staffing decisions (z_i) (only dedicated nurses) are made without the knowledge that flexibility will be allowed in period 2. The results of this period 1 optimization are the same as for simultaneous decision making in configuration 1. In period 2, given the capacity and total staff available, the optimal number of flexible nurses is determined. Sequential decision making case 1 for configuration 3 is shown in Figure 3.4.

Period 1 :

$$\begin{aligned}
\text{Min}_{k_i,z_i} \quad & \sum_i (f_i \cdot k_i + h_i \cdot z_i) + E_{\Phi_i}\Omega(k_i, z_i, d_i) \\
\text{subject to} \quad & \\
\text{where } \ \Omega(k_i, z_i, d_i) \ = \ & \sum_i (s_i \cdot (min(k_i, d_i) - min(z_i, k_i, d_i)) \\
& + p_i \cdot (d_i - min(k_i, d_i))
\end{aligned}$$

In period 2, optimal number of flexible nurses (e_i) is determined under the assumption that preliminary staffing decisions have already been made. Stage 2 optimization assigns dedicated and flexible nurses to each floor based on patient demand.

Period 2 Stage 1 :

$$\begin{aligned}
\text{Min}_{e_i} \quad & \sum_i ((t - h_i) \cdot e_i) + E_{\Phi_i}\Omega(e_i, d_i) \\
\text{subject to} \quad & \\
e_i \ \leq \ & z_i, \forall\, i = 1, 2 \\
e_i \ \geq \ & 0, \forall\, i = 1, 2
\end{aligned}$$

Period 2 Stage 2 : $\Omega(e_i, d_i)$:

$$\begin{aligned}
\text{Min}_{x_i,y_i,a_i} \quad & \sum_i s_i \cdot (a_i - x_i - y_i) + p_i \cdot (d_i - a_i) \\
\text{subject to} \quad & \\
a_i &\leq min(k_i, d_i), \forall\, i = 1, 2 \\
x_i &\leq z_i, \forall\, i = 1, 2 \\
y_j &\leq e_i, \forall\, i, j = 1, 2\; and\; i \neq j \\
a_i - x_i - y_i &\geq 0, \forall\, i = 1, 2 \\
x_i, y_i, a_i &\geq 0, \forall\, i = 1, 2
\end{aligned}$$

In the first period of configuration 3, we optimize for k_i and z_i. This formulation is similar to optimizing k_i and z_i simultaneously in configuration 1, so we use k_i and z_i from configuration 1 - simultaneous decision making. In the second period, we determine optimal flexible resource e_i, given k_i and z_i from period 1. (3.11) gives the first stage objective as a function of the flexible nurses e_i.

$$
\begin{aligned}
\Delta_{11}(e_i) &= \sum_i (t_i \cdot e_i) + \Delta_{11A} + \Delta_{11B} + \Delta_{11C} + \Delta_{11D} && (3.11) \\
\Delta_{11A} &= s_1 \int_{z_1+e_2}^{k_1} \int_0^{k_2} (d_1 - z_1 - e_2)\, d\Phi_2\, d\Phi_1 + s_2 \int_{z_1+e_2}^{k_1} \int_{z_2-e_2}^{k_2} (d_2 - z_2 + e_2)\, d\Phi_2\, d\Phi_1 \\
&\quad + s_2 \int_{z_1}^{z_1+e_2} \int_{z_1+z_2-e_2}^{k_2} (d_1 + d_2 - z_1 - z_2)\, d\Phi_2\, d\Phi_1 \\
&\quad + s_2 \int_0^{z_1-e_1} \int_{z_2+e_1}^{k_2} (d_2 - z_2 - e_1)\, d\Phi_2\, d\Phi_1 \\
&\quad + s_2 \int_{z_1-e_1}^{z_1} \int_{z_1+z_2-d_1}^{k_2} (d_1 + d_2 - z_1 - z_2)\, d\Phi_2\, d\Phi_1 \\
\Delta_{11B} &= s_1 \int_{k_1}^{\infty} \int_0^{k_2} (k_1 - z_1 - e_2)\, d\Phi_2\, d\Phi_1 + s_2 \int_{k_1}^{\infty} \int_{z_2-e_2}^{k_2} (d_2 - z_2 + e_2)\, d\Phi_2\, d\Phi_1 \\
\Delta_{11C} &= s_2 \int_0^{z_1-e_1} \int_{k_2}^{\infty} (k_2 - z_2 - e_1)\, d\Phi_2\, d\Phi_1 + s_2 \int_{z_1-e_1}^{z_1+e_2} \int_{k_2}^{\infty} (k_2 - z_2 - z_1 + d_1)\, d\Phi_2\, d\Phi_1 \\
&\quad + \int_{z_1+e_2}^{k_1} \int_{k_2}^{\infty} (s_1(d_1 - z_1 - e_2) + s_2(k_2 - z_2 + e_2))\, d\Phi_2\, d\Phi_1 \\
\Delta_{11D} &= \int_{k_1}^{\infty} \int_{k_2}^{\infty} (s_1(k_1 - z_1 - e_2) + s_2(k_2 - z_2 + e_2))\, d\Phi_2\, d\Phi_1
\end{aligned}
$$

Analysis of second order conditions (see appendix B.1) shows that first order conditions can be used to determine optimal values of flexible nurses (e_i^*).

$$
\begin{aligned}
\frac{\partial \Delta_{11}}{\partial e_1} &= t_1 + s_2 \phi_1(z_1 - e_1)(1 + \Phi_2(z_2 + e_1)) \\
\frac{\partial \Delta 11}{\partial e_2} &= s_2 + t_2 - s_1 + (s_1 - s_2)\Phi_1(z_1 + e_2) + s_2 \Phi_2(z_2 - e_2)(1 - \Phi_1(z_1 + e_2))
\end{aligned}
$$

Sequential Decision Making - Case 2

Sequential decision making case 2 has two periods. In the first period, capacity (k_i) is determined. In the second period, given the capacity from first period, optimal dedicated (n_i) and flexible nurses (e) are determined. Sequential decision making case 2 for configuration 3 is shown in Figure 3.4.

Period 1 :

$$\begin{aligned} \text{Min}_{k_i} \quad & \sum_i (f_i \cdot k_i) + E_{\Phi_i} p_i \cdot (d_i - min(k_i, d_i)) \\ \text{subject to} & \\ k_i \ \geq \ & 0, \forall\, i = 1, 2 \end{aligned}$$

Period 2 Stage 1 :

$$\begin{aligned} \text{Min}_{n_i, e} \quad & \sum_i (h_i \cdot n_i) + t \cdot e + E_{\Phi_i} \Omega(n_i, e, d_i) \\ \text{subject to} & \\ n_i, e \ \geq \ & 0, \forall\, i = 1, 2 \end{aligned}$$

Period 2 Stage 2 : $\Omega(n_i, e, d_i)$:

$$
\begin{aligned}
\text{Min}_{x_i,y_i,a_i} \quad & \sum_i s_i \cdot (a_i - x_i - y_i) + p_i \cdot (d_i - a_i) \\
\text{subject to} \quad & \\
a_i &\leq min(k_i, d_i), \forall\, i = 1, 2 \\
x_i &\leq n_i, \forall\, i = 1, 2 \\
a_i - x_i - y_i &\geq 0, \forall\, i = 1, 2 \\
\sum_i y_i &\leq e, \forall\, i = 1, 2 \\
x_i, y_i, a_i &\geq 0, \forall\, i = 1, 2
\end{aligned}
$$

In period 1, we only determine capacity for each unit by minimizing the objective function (3.12). First order conditions (3.13) prove that these capacities follow a simple newsvendor relationship.

$$
\begin{aligned}
\Delta_{21}(k_i) &= \sum_i (f_i \cdot k_i) + \Delta_{21A} \qquad (3.12) \\
\Delta_{21A} &= p_1 \int_0^{k_2} \int_{k_1}^{\infty} (d_1 - k_1)\, d\Phi_2\, d\Phi_1 + p_2 \int_{k_2}^{\infty} \int_0^{k_1} (d_2 - k_2)\, d\Phi_2\, d\Phi_1 \\
&\quad + \int_{k_2}^{\infty} \int_{k_1}^{\infty} (p_1(d_1 - k_1) + p_2(d_2 - k_2))\, d\Phi_2\, d\Phi_1
\end{aligned}
$$

Leads to closed form expression:

$$
\Phi_i(k_i^*) = \frac{p_i - f_i}{p_i} \qquad (3.13)
$$

The following equation (3.14), shows period 2 objective as a function of dedicated and flexible nurses, given capacity determined in period 1.

$$
\begin{aligned}
\Delta_{22}(n_i, e) &= \sum_i (h_i \cdot n_i) + t \cdot e + \Delta_{22A} + \Delta_{22B} + \Delta_{22C} + \Delta_{22D} \qquad (3.14)\\
\Delta_{22A} &= s_1 \int_{n_1+e}^{k_1} \int_0^{n_2} (d_1 - n_1 - e)\, d\Phi_2\, d\Phi_1 + s_2 \int_0^{n_1} \int_{n_2+e}^{k_2} (d_2 - n_2 - e)\, d\Phi_2\, d\Phi_1\\
&\quad + s_2 \int_{n_1}^{n_1+e} \int_{n_1+n_2+e-d_1}^{n_2+e} (d_1 + d_2 - n_1 - n_2 - e)\, d\Phi_2\, d\Phi_1\\
&\quad + s_2 \int_{n_1}^{n_1+e} \int_{n_2+e}^{k_2} (d_1 + d_2 - n_1 - n_2 - e)\, d\Phi_2\, d\Phi_1\\
&\quad + s_2 \int_{n_1+e}^{k_1} \int_{n_2}^{n_2+e} (s_1(d_1 - n_1 - e) + s_2(d_2 - n_2))\, d\Phi_2\, d\Phi_1\\
&\quad + s_2 \int_{n_1+e}^{k_1} \int_{n_2+e}^{k_2} (s_1(d_1 - n_1 - e) + s_2(d_2 - n_2))\, d\Phi_2\, d\Phi_1\\
\Delta_{22B} &= s_1 \int_{k_1}^{\infty} \int_0^{k_2} (k_1 - n_1 - e)\, d\Phi_2\, d\Phi_1 + s_2 \int_{k_1}^{\infty} \int_{n_2}^{k_2} (d_2 - n_2)\, d\Phi_2\, d\Phi_1\\
\Delta_{22C} &= s_1 \int_{n_1+e}^{k_1} \int_{k_2}^{\infty} (d_1 - n_1 - e)\, d\Phi_2\, d\Phi_1 + s_2 \int_{n_1+e}^{k_1} \int_{k_2}^{\infty} (k_2 - n_2)\, d\Phi_2\, d\Phi_1\\
&\quad + s_2 \int_0^{n_1} \int_{k_2}^{\infty} (k_2 - n_2 - e)\, d\Phi_2\, d\Phi_1 + s_2 \int_{n_1}^{n_1+e} \int_{k_2}^{\infty} (k_2 - n_2 - n_1 - e + d_1)\, d\Phi_2\, d\Phi_1\\
\Delta_{22D} &= \int_{k_1}^{\infty} \int_{k_2}^{\infty} (s_1(k_1 - n_1 - e) + s_2(k_2 - n_2))\, d\Phi_2\, d\Phi_1
\end{aligned}
$$

Second order conditions prove that objective function (3.14) is convex in decision variables (see appendix B.2). First order conditions are sufficient and necessary to find optimum dedicated nurses (n_i^*) and (e^*), given capacity (k_i) determined in period 1.

$$\begin{aligned}
\frac{\partial \Delta_{22}}{\partial n_1} &= h_1 - s_1 + s_2\Phi_1(n_1) + (s_1 - s_2)\Phi_1(n_1+e) + s_2\int_{n_1}^{n_1+e}\Phi_2(n_1+n_2+e-d_1)d\Phi_1'(d_1) \\
\frac{\partial \Delta_{22}}{\partial n_2} &= h_2 - s_2 + s_2(1-\Phi_1(n_1+e))\Phi_2(n_2) + s_2\Phi_1(n_1)\Phi_2(n_2+e) \\
&\quad + s_2\int_{n_1}^{n_1+e}\Phi_2(n_1+n_2+e-d_1)d\Phi_1'(d_1) \\
\frac{\partial \Delta_{22}}{\partial e} &= t - s_1 + s_2\Phi_1(n_1)\Phi_2(n_2+e) + (s_1 - s_2)\Phi_1(n_1+e) \\
&\quad + s_2\int_{n_1}^{n_1+e}\Phi_2(n_1+n_2+e-d_1)d\Phi_1'(d_1)
\end{aligned}$$

Simultaneous Decision Making

In simultaneous decision making both capacity and staffing decisions are made in the same period. Capacity (k_i), dedicated nurses (n_i) and flexible nurses (e) are decided in stage 1 while capacity and staffing allocations are made in stage 2. It should be noted that total staff ($z_1 + z_2$) for this configuration is the sum of dedicated ($n_1 + n_2$) and flexible nurses (e). Simultaneous decision making for configuration 3 is shown in Figure 3.4.

Stage 1 :

$$\text{Min}_{k_i,n_i,e} \quad \sum_i (f_i \cdot k_i + h_i \cdot n_i) + t \cdot e + E_{\Phi_i}\Omega(k_i, n_i, e, d_i)$$

subject to

$$k_i, n_i, e \geq 0, \forall\, i = 1, 2$$

Stage 2 : $\Omega(k_i, n_i, e, d_i)$:

$$\begin{aligned}
\text{Min}_{x_i,y_i,a_i} \quad & \sum_i (s_i \cdot (a_i - x_i - y_i) + p_i \cdot (d_i - a_i)) \\
\text{subject to} \quad & \\
a_i &\leq min(k_i, d_i), \forall\, i = 1, 2 \\
x_i &\leq n_i, \forall\, i = 1, 2 \\
\sum_i y_i &\leq e, \forall\, i = 1, 2 \\
a_i - x_i - y_i &\geq 0, \forall\, i = 1, 2 \\
x_i, y_i, a_i &\geq 0, \forall\, i = 1, 2
\end{aligned}$$

For simultaneous decision making, the optimal stage 2 allocations can be summarized as follows:

- Demand for each floor is accommodated until all beds on that floor are filled
- Dedicated nurses are used, as needed, to care for patients on their floor
- Flexible nurses are first used on floor 1 to cover the shortfall between patients and dedicated staff
- Any remaining flexible nurses are used on floor 2 to cover the shortfall between patients and dedicated staff.

The first stage objective function (3.15) is obtained by substituting second stage allocations in it. Second order conditions prove that objective function (3.15) is convex in its decision variables k_i, n_i and e (see appendix B.3).

$$
\begin{aligned}
\Delta_{31}(k_i, n_i, e) &= \sum_i (f_i \cdot k_i + h_i \cdot n_i) + t \cdot e + \Delta_{31A} + \Delta_{31B} + \Delta_{31C} + \Delta_{31D} + \Delta_{31E} \qquad (3.15) \\
\Delta_{31A} &= s_1 \int_{n_1+e}^{k_1} \int_0^{n_2} (d_1 - n_1 - e)\, d\Phi_2\, d\Phi_1 + s_2 \int_0^{n_1} \int_{n_2+e}^{k_2} (d_2 - n_2 - e)\, d\Phi_2\, d\Phi_1 \\
&\quad + s_2 \int_{n_1}^{n_1+e} \int_{n_1+n_2+e-d_1}^{n_2+e} (d_1 + d_2 - n_1 - n_2 - e)\, d\Phi_2\, d\Phi_1 \\
&\quad + s_2 \int_{n_1}^{n_1+e} \int_{n_2+e}^{k_2} (d_1 + d_2 - n_1 - n_2 - e)\, d\Phi_2\, d\Phi_1 \\
&\quad + s_2 \int_{n_1+e}^{k_1} \int_{n_2}^{n_2+e} (s_1(d_1 - n_1 - e) + s_2(d_2 - n_2))\, d\Phi_2\, d\Phi_1 \\
&\quad + s_2 \int_{n_1+e}^{k_1} \int_{n_2+e}^{k_2} (s_1(d_1 - n_1 - e) + s_2(d_2 - n_2))\, d\Phi_2\, d\Phi_1 \\
\Delta_{31B} &= s_1 \int_{k_1}^{\infty} \int_0^{k_2} (k_1 - n_1 - e)\, d\Phi_2\, d\Phi_1 + s_2 \int_{k_1}^{\infty} \int_{n_2}^{k_2} (d_2 - n_2)\, d\Phi_2\, d\Phi_1 \\
\Delta_{31C} &= s_1 \int_{n_1+e}^{k_1} \int_{k_2}^{\infty} (d_1 - n_1 - e)\, d\Phi_2\, d\Phi_1 + s_2 \int_{n_1+e}^{k_1} \int_{k_2}^{\infty} (k_2 - n_2)\, d\Phi_2\, d\Phi_1 \\
&\quad + s_2 \int_0^{n_1} \int_{k_2}^{\infty} (k_2 - n_2 - e)\, d\Phi_2\, d\Phi_1 \\
&\quad + s_2 \int_{n_1}^{n_1+e} \int_{k_2}^{\infty} (k_2 - n_2 - n_1 - e + d_1)\, d\Phi_2\, d\Phi_1 \\
\Delta_{31D} &= \int_{k_1}^{\infty} \int_{k_2}^{\infty} (s_1(k_1 - n_1 - e) + s_2(k_2 - n_2))\, d\Phi_2\, d\Phi_1 \\
\Delta_{31E} &= p_1 \int_{k_1}^{\infty} \int_0^{k_2} (d_1 - k_1)\, d\Phi_2\, d\Phi_1 + p_2 \int_0^{k_1} \int_{k_1+k_2-d_1}^{\infty} (d_1 + d_2 - k_1 - k_2)\, d\Phi_2\, d\Phi_1 \\
&\quad + \int_{k_1}^{\infty} \int_{k_2}^{\infty} (p_1(d_1 - k_1) + p_2(d_2 - k_2))\, d\Phi_2\, d\Phi_1
\end{aligned}
$$

Following first order conditions are sufficient and necessary to find k_i^*, n_i^* and e^*.

$$
\begin{aligned}
\frac{\partial \Delta_{31}}{\partial k_1} &= f_1 - p_1 + s_1 + (p_1 - s_1)\Phi_1(k_1) \\
\frac{\partial \Delta_{31}}{\partial k_2} &= f_2 - p_2 + s_2 + (p_2 - s_2)\Phi_2(k_2) \\
\frac{\partial \Delta_{31}}{\partial n_1} &= h_1 - s_1 + s_2\Phi_1(n_1) + (s_1 - s_2)\Phi_1(n_1 + e) + s_2 \int_{n_1}^{n_1+e} \Phi_2(n_1 + n_2 + e - d_1) d\Phi_1'(d_1) \\
\frac{\partial \Delta_{31}}{\partial n_2} &= h_2 - s_2 + s_2(1 - \Phi_1(n_1 + e))\Phi_2(n_2) + s_2\Phi_1(n_1)\Phi_2(n_2 + e) \\
&\quad + s_2 \int_{n_1}^{n_1+e} \Phi_2(n_1 + n_2 + e - d_1) d\Phi_1'(d_1) \\
\frac{\partial \Delta_{31}}{\partial e} &= t - s_1 + s_2\Phi_1(n_1)\Phi_2(n_2 + e) + (s_1 - s_2)\Phi_1(n_1 + e) \\
&\quad + s_2 \int_{n_1}^{n_1+e} \Phi_2(n_1 + n_2 + e - d_1) d\Phi_1'(d_1)
\end{aligned}
$$

Solving for k_i, using first order conditions we get equations (3.16) and (3.17), indicating that under simultaneous decision making, capacity decisions (k_i) are independent of staffing decisions (n_i) and e.

$$\Phi_1(k_1^*) = \frac{p_1 - f_1 - s_1}{p_1 - s_1} \tag{3.16}$$

$$\Phi_2(k_2^*) = \frac{p_2 - f_2 - s_2}{p_2 - s_2} \tag{3.17}$$

3.4.4 Configuration 4: Demand Upgrades and Staffing Flexibility

Configuration 4 has the highest level of flexibility among our models. It uses both demand upgrades and staffing flexibility. The timeline for decision making for configuration 4 is exactly the same as in configuration 3 except that capacity decisions include the possibility for demand upgrades. The three types of decision making for configuration 4 is shown in Figure 3.4.

Sequential Decision Making - Case 1

In sequential decision making case 1, there are two periods. In the first period both total staff and optimal capacity is determined, assuming demand upgrades. In the second period, given capacity and total staff from period 1, optimal number of flexible nurses required to meet demand at minimum cost is determined.

Period 1 :

$$
\begin{aligned}
\mathrm{Min}_{k_i,z_i} \quad & \sum_i (f_i \cdot k_i + h_i \cdot z_i) + E_{\Phi_i}\Omega(k_i, z_i, d_i) \\
\text{subject to} \quad & \\
k_i, z_i \;\; & \geq \;\; 0, \forall\, i = 1, 2 \\
\text{where } \; \Omega(k_i, z_i, d_i) \;\; & = \;\; s_1 \cdot (a_1 + a_u - x_1) + s_2 \cdot (a_2 - x_2) + \\
& \quad\;\; p_1 \cdot (d_1 - a_1) + p_2 \cdot (d_2 - a_2 - a_u) \\
\text{subject to} \quad & \\
a_i \;\; & \leq \;\; min(k_i, d_i), \forall\, i = 1, 2 \\
x_i \;\; & \leq \;\; z_i, \forall\, i = 1, 2 \\
a_i - x_i \;\; & \geq \;\; 0, \forall\, i = 1, 2 \\
a_1 + a_u \;\; & \leq \;\; k_1 \\
a_2 + a_u \;\; & \leq \;\; d_2 \\
x_i, a_i \;\; & \geq \;\; 0, \forall\, i = 1, 2, u
\end{aligned}
$$

The function $\Omega(\cdot)$ minimizes the cost of hiring contract nurses and penalty cost of not meeting demand, assuming demand upgrades. Staffing flexibility is determined in period 2.

Period 2 Stage 1 :

$$
\begin{aligned}
\text{Min}_{e_i} \qquad & \sum_i (t_i \cdot e_i) + E_{\Phi_i} \Omega(e_i, d_i) \\
\text{subject to} \qquad & \\
e_i &\leq z_i, \forall\, i = 1, 2 \\
e_i &\geq 0, \forall\, i = 1, 2
\end{aligned}
$$

Period 2 Stage 2 : $\Omega(e_i, d_i)$:

$$\begin{aligned}
\text{Min}_{x_i,y_i,a_i} \quad & s_1 \cdot (a_1 + a_u - x_1 - y_1) + s_2 \cdot (a_2 - x_2 - y_2) + \\
& p_1 \cdot (d_1 - a_1) + p_2 \cdot (d_2 - a_2 - a_u) \\
\text{subject to} \quad & \\
a_i &\leq min(k_i, d_i), \forall i = 1, 2 \\
x_i &\leq z_i, \forall\ i = 1, 2 \\
a_1 - x_1 &\geq 0 \\
a_1 + a_u - x_1 - y_1 &\geq 0 \\
a_2 - x_2 - y_2 &\geq 0 \\
a_1 + a_u &\leq k_1 \\
a_2 + a_u &\leq d_2 \\
y_i &\leq e_i, \forall\ i = 1, 2 \\
x_i, y_i, a_i &\geq 0, \forall\ i = 1, 2, u
\end{aligned}$$

In period 2, flexible nurses can cater to the needs of both unit 1 and unit 2. In unit 1, dedicated staff caters to the needs of patients from unit 1 and not patients from unit 2. Flexible and contract nurses in unit 1 are assigned to meet the needs of both unit 1 patients and upgraded patients.

Optimal capacity (k_i^*) and staffing (z_i^*) for configuration 4 is the same optimal values got from simultaneous decision making in configuration 2.

In period 2, given the values for k_i and z_i, we determine optimal number of flexible nurses (e_i). The objective as a function of flexible nurses is given in equation (3.18).

$$
\begin{aligned}
\Theta_{11}(e_i) &= \sum_i (t_i \cdot e_i) + \Theta_{11A} + \Theta_{11B} + \Theta_{11C} + \Theta_{11D} \qquad (3.18)\\
\Theta_{11A} &= s_1 \int_{z_1+e_2}^{k_1} \int_{0}^{k_2} (d_1 - z_1 - e_2)\, d\Phi_2\, d\Phi_1 + s_2 \int_{z_1+e_2}^{k_1} \int_{z_2-e_2}^{k_2} (d_2 - z_2 + e_2)\, d\Phi_2\, d\Phi_1\\
&\quad s_2 \int_{z_1}^{z_1+e_2} \int_{z_1+z_2-d_1}^{k_2} (d_1 + d_2 - z_2 - z_1)\, d\Phi_2\, d\Phi_1\\
&\quad + s_2 \int_{0}^{z_1-e_1} \int_{z_2+e_1}^{k_2} (d_2 - z_2 - e_1)\, d\Phi_2\, d\Phi_1\\
&\quad + s_1 \int_{z_1-e_1}^{z_1} \int_{z_1+z_2-d_1}^{k_2} (d_2 + d_1 - z_2 - z_1)\, d\Phi_2\, d\Phi_1\\
\Theta_{11B} &= s_1 \int_{k_1}^{\infty} \int_{0}^{k_2} (k_1 - z_1 - e_2)\, d\Phi_2\, d\Phi_1 + s_2 \int_{k_1}^{\infty} \int_{z_2-e_2}^{k_2} (d_2 - z_2 + e_2)\, d\Phi_2\, d\Phi_1\\
\Theta_{11C} &= s_2 \int_{0}^{z_1-e_1} \int_{k_2}^{k_2+e_1+e_2} (d_2 - z_2 - e_1)\, d\Phi_2\, d\Phi_1\\
&\quad + \int_{0}^{z_1-e_1} \int_{k_2+e_1+e_2}^{k_1+k_2-d_1} (s_1(d_2 - k_2 - e_1 - e_2) + s_2(k_2 - z_2 + e_2))\, d\Phi_2\, d\Phi_1\\
&\quad + s_2 \int_{z_1-e_1}^{z_1+e_2} \int_{k_2}^{z_1+e_2+k_2-d_1} (d_1 + d_2 - z_1 - z_2)\, d\Phi_2\, d\Phi_1\\
&\quad + \int_{z_1-e_1}^{e_1+e_2} \int_{z_1+e_2+k_2-d_1}^{k_1+k_2-d_1} (s_1(d_1 + d_2 - k_2 - z_1 - e_2) + s_2(k_2 - z_2 + e_2))\, d\Phi_2\, d\Phi_1\\
&\quad + \int_{z_1+e_2}^{k_1} \int_{k_2}^{k_1+k_2-d_1} (s_1(d_1 + d_2 - k_2 - z_1 - e_2) + s_2(k_2 - z_2 + e_2))\, d\Phi_2\, d\Phi_1\\
&\quad + \int_{0}^{k_1} \int_{k_1+k_2-d_1}^{\infty} (s_1(k_1 - z_1 - e_2) + s_2(k_2 - z_2 + e_2))\, d\Phi_2\, d\Phi_1\\
\Theta_{11D} &= \int_{k_1}^{\infty} \int_{k_2}^{\infty} (s_1(k_1 - z_1 - e_2) + s_2(k_2 - z_2 + e_2))\, d\Phi_2\, d\Phi_1
\end{aligned}
$$

The objective function is convex (see Appendix C.1), so first order conditions are sufficient and necessary to obtain optimal flexible nurses (e_i^*).

$$\begin{aligned}
\frac{\partial \Theta_{11}}{\partial e_1} &= t_1 + (s_1 - s_1)\Phi_1(z_1 - e_1)\Phi_2(e_1 + e_2 + k_2) - s_2\Phi_1(z_1 - e_1)\Phi_2(z_2 + e_1) \\
&\quad -s_1 \int_0^{z_1 - e_1} \Phi_2(k_1 + k_2 - d_1) d\Phi_1 \\
\frac{\partial \Theta_{11}}{\partial e_2} &= s_2 + t_2 - s_1 + (s_1 - s_2)\Phi_1(z_1 - e_1)\Phi_2(e_1 + e_2 + k_2) + s_2\Phi_2(z_2 - e_2)(\Phi_1(z_1 + e_2) - 1) \\
&\quad +(s_1 - s_2) \int_{z_1 - e_1}^{z_1 + e_2} \Phi_2(k_2 + e_2 + z_1 - d_1) d\Phi_1
\end{aligned}$$

Sequential Decision Making - Case 2

Sequential decision making case 2 has two periods. In the first period, capacity is determined assuming demand upgrades in period 2, while in the second period optimal dedicated and flexible staff is determined.

Period 1 :

$$\begin{aligned}
\text{Min}_{k_i} \qquad & \sum_i (f_i \cdot k_i) + E_{\Phi_i}(p_1(d_1 - min(k_1, d_1)) + \\
& p_2 \cdot (d_2 - min(d_2, k_2 + max(0, k_1 - d_1))))
\end{aligned}$$

subject to

$$k_i \geq 0, \forall i = 1, 2$$

Period 2 Stage 1 :

$$
\begin{aligned}
\text{Min}_{n_i,e} \quad & \sum_i (h_i \cdot n_i) + t \cdot e + E_{\Phi_i}\Omega(n_i, e, d_i) \\
\text{subject to} \quad & \\
n_i + e & \leq k_i, \forall\, i = 1, 2 \\
n_i, e & \geq 0, \forall\, i = 1, 2
\end{aligned}
$$

Period 2 Stage 2 : $\Omega(n_i, e, d_i)$:

$$
\begin{aligned}
\text{Min}_{x_i,y_i,a_i} \quad & s_1 \cdot (a_1 + a_u - x_1 - y_1) + s_2 \cdot (a_2 - x_2 - y_2) \\
& p_1 \cdot (d_1 - a_1) + p_2 \cdot (d_2 - a_2 - a_u) \\
\text{subject to} \quad & \\
a_i & \leq min(k_i, d_i), \forall\, i = 1, 2 \\
x_i & \leq n_i, \forall\, i = 1, 2 \\
a_1 + a_u - x_1 - y_1 & \geq 0 \\
a_2 - x_2 - y_2 & \geq 0 \\
a_1 + a_u & \leq k_1 \\
a_2 + a_u & \leq d_2 \\
\sum_i y_i & \leq e, \forall\, i = 1, 2 \\
x_i, y_i, a_i & \geq 0, \forall\, i = 1, 2, u
\end{aligned}
$$

Optimal capacity (k_i^*) under period 1 is the same as optimal capacity in configuration 2 under sequential decision making - case 2 period 1.

In the second period, given k_i from period 1, we determine optimal number of dedicated staff, n_i and flexible staff, e. The objective as a function of n_i and e is given in equation (3.19).

$$
\begin{aligned}
\Theta_{21}(n_i, e) &= \sum_i (h_i \cdot n_i) + t \cdot e + \Theta_{21A} + \Theta_{21B} + \Theta_{21C} + \Theta_{21D} \qquad (3.19) \\
\Theta_{21A} &= s_1 \int_{n_1+e}^{k_1} \int_0^{n_2} (d_1 - n_1 - e)\, d\Phi_2\, d\Phi_1 + s_2 \int_0^{n_1} \int_{n_2+e}^{k_2} (d_2 - n_2 - e)\, d\Phi_2\, d\Phi_1 \\
&\quad + s_2 \int_{n_1}^{n_1+e} \int_{n_1+n_2+e-d_1}^{n_2+e} (d_1 + d_2 - n_1 - n_2 - e)\, d\Phi_2\, d\Phi_1 \\
&\quad + s_2 \int_{n_1}^{n_1+e} \int_{n_2+e}^{k_2} (d_1 + d_2 - n_1 - n_2 - e)\, d\Phi_2\, d\Phi_1 \\
&\quad + \int_{n_1+e}^{k_1} \int_{n_2}^{n_2+e} (s_1(d_1 - n_1 - e) + s_2(d_2 - n_2))\, d\Phi_2\, d\Phi_1 \\
&\quad + \int_{n_1+e}^{k_1} \int_{n_2+e}^{k_2} (s_1(d_1 - n_1 - e) + s_2(d_2 - n_2))\, d\Phi_2\, d\Phi_1 \\
\Theta_{21B} &= s_1 \int_{k_1}^{\infty} \int_0^{k_2} (k_1 - n_1 - e)\, d\Phi_2\, d\Phi_1 + s_2 \int_{k_1}^{\infty} \int_{n_2}^{k_2} (d_2 - n_2)\, d\Phi_2\, d\Phi_1 \\
\Theta_{21C} &= s_2 \int_0^{n_1} \int_{k_2}^{k_2+e} (d_2 - n_2 - e)\, d\Phi_2\, d\Phi_1 \\
&\quad + \int_0^{n_1} \int_{k_2+e}^{k_1+k_2-d_1} (s_1(d_2 - k_2 - e) + s_2(k_2 - n_2))\, d\Phi_2\, d\Phi_1 \\
&\quad + s_2 \int_{n_1}^{n_1+e} \int_{k_2}^{n_1+e+k_2-d_1} (d_1 + d_2 - n_1 - n_2 - e)\, d\Phi_2\, d\Phi_1 \\
&\quad + \int_{n_1}^{n_1+e} \int_{n_1+e+k_2-d_1}^{b_2} (s_1(d_1 + d_2 - k_2 - n_1 - e) + s_2(k_2 - n_2))\, d\Phi_2\, d\Phi_1 \\
&\quad + \int_{n_1+e}^{k_1} \int_{k_2}^{k_1+k_2-d_1} (s_1(d_1 + d_2 - k_2 - n_1 - e) + s_2(k_2 - n_2))\, d\Phi_2\, d\Phi_1 \\
&\quad + \int_0^{k_1} \int_{k_1+k_2-d_1}^{\infty} (s_1(k_1 - n_1 - e) + s_2(k_2 - n_2))\, d\Phi_2\, d\Phi_1 \\
\Theta_{21D} &= \int_{k_1}^{\infty} \int_{k_2}^{\infty} (s_1(k_1 - n_1 - e) + s_2(k_2 - n_2))\, d\Phi_2\, d\Phi_1
\end{aligned}
$$

Convexity of the objective function (see appendix C.2) makes first order conditions sufficient and necessary to obtain optimal dedicated nurses (n_i^*) and flexible nurses (e^*).

$$\begin{aligned}
\frac{\partial \Theta_{21}}{\partial n_1} &= h_1 - s_1 + s_1 \int_0^{z_1} \Phi_2(k_1 + k_2 - d_1) d\Phi_1 \\
&\quad + s_2 \int_{z_1}^{z_1+e} \Phi_2(z_1 + z_2 + e - d_1) d\Phi_1 + (s_1 - s_2) \int_{z_1}^{z_1+e} \Phi_2(k_1 + z_1 + e - d_1) d\Phi_1 \\
\frac{\partial \Theta_{21}}{\partial n_2} &= h_2 - s_2 + s_2 \Phi_2(z_2) + s_2 \Phi_1(z_1) \Phi_2(z_2 + e) + s_2 \Phi_1(z_1 + e) \Phi_2(z_2) \\
&\quad + s_2 \int_{z_1}^{z_1+e} \Phi_2(z_1 + z_2 + e - d_1) d\Phi_1 \\
\frac{\partial \Theta_{21}}{\partial e} &= t - s_1 + s_1 \Phi_1(k_1) + (s_1 - s_2) \Phi_1(z_1) \Phi_2(k_2 + e) + s_2 \Phi_1(z_1) \Phi_2(z_2 + e) \\
&\quad + (s_1 - s_2) \int_{z_1}^{z_1+e} \Phi_2(k_2 + z_1 + e - d_1) d\Phi_1 + s_2 \int_{z_1}^{z_1+e} \Phi_2(z_1 + z_2 + e - d_1) d\Phi_1
\end{aligned}$$

Simultaneous Decision Making

Under simultaneous decision making, all capacity and staffing decisions are made simultaneously assuming demand upgrades and staffing flexibility.

Stage 1 :

$$\begin{aligned}
\text{Min}_{k_i, n_i, e} \quad & \sum_i (f_i \cdot k_i + h_i \cdot n_i) + t \cdot e + E_{\Phi_i} \Omega(k_i, n_i, e, d_i) \\
\text{subject to} \quad & \\
k_i, n_i, e &\geq 0, \forall\, i = 1, 2
\end{aligned}$$

Stage 2 : $\Omega(k_i, n_i, e, d_i)$:

$$\begin{aligned}
\text{Min}_{x_i,y_i,a_i} \quad & s_1 \cdot (a_1 + a_u - x_1 - y_1) + s_2 \cdot (a_2 - x_2 - y_2) \\
& +p_1 \cdot (d_1 - a_1) + p_2 \cdot (d_2 - a_2 - a_u) \\
\text{subject to} \quad & \\
a_i \leq\ & min(k_i, d_i), \forall\, i = 1, 2 \\
x_i \leq\ & n_i, \forall\, i = 1, 2 \\
\sum_i y_i \leq\ & e, \forall\, i = 1, 2 \\
a_1 + a_u - x_1 - y_1 \geq\ & 0 \\
a_2 - x_2 - y_2 \geq\ & 0 \\
a_1 + a_u \leq\ & k_1 \\
a_2 + a_u \leq\ & d_2 \\
x_i, y_i, a_i \geq\ & 0, \forall\, i = 1, 2, u
\end{aligned}$$

For simultaneous decision making, stage 2 allocations can be summarized as follows:

- Patients are admitted to their desired floor as long as beds are available; any excess beds on floor 1 are used to accommodate overflow patients from floor 2 as needed
- Dedicated nurses are used to care for traditional patients on their floor
- Flexible nurses are used on floor 1 to care for any upgraded patients from floor 2 and/or traditional patients who were not assigned a dedicated nurse
- Any remaining flexible nurses are used on floor 2 to care for patients not assigned

a dedicated floor 2 nurse.

First stage objective function is obtained by substituting second stage allocations in first stage objective. Stage 1 decision is to choose capacity and staffing levels to minimize the first stage objective function in equation (3.20).

$$
\begin{aligned}
\Theta_{31}(k_i, n_i, e) &= \sum_i (f_i \cdot k_i + h_i \cdot n_i) + t \cdot e + \Theta_{31A} + \Theta_{31B} + \Theta_{31C} + \Theta_{31D} + \Theta_{31E} \qquad (3.20) \\
\Theta_{31A} &= s_1 \int_{n_1+e}^{k_1} \int_0^{n_2} (d_1 - n_1 - e)\, d\Phi_2\, d\Phi_1 + \\
&\quad s_2 \int_0^{n_1} \int_{n_2+e}^{k_2} (d_2 - n_2 - e)\, d\Phi_2\, d\Phi_1 + \\
&\quad s_2 \int_{n_1}^{n_1+e} \int_{n_1+n_2+e-d_1}^{n_2+e} (d_1 + d_2 - n_1 - n_2 - e)\, d\Phi_2\, d\Phi_1 + \\
&\quad s_2 \int_{n_1}^{n_1+e} \int_{n_2+e}^{k_2} (d_1 + d_2 - n_1 - n_2 - e)\, d\Phi_2\, d\Phi_1 + \\
&\quad \int_{n_1+e}^{k_1} \int_{n_2}^{n_2+e} (s_1(d_1 - n_1 - e) + s_2(d_2 - n_2))\, d\Phi_2\, d\Phi_1 + \\
&\quad \int_{n_1+e}^{k_1} \int_{n_2+e}^{k_2} (s_1(d_1 - n_1 - e) + s_2(d_2 - n_2))\, d\Phi_2\, d\Phi_1 + \\
\Theta_{31B} &= s_1 \int_{k_1}^{\infty} \int_0^{k_2} (k_1 - n_1 - e)\, d\Phi_2\, d\Phi_1 \\
&\quad s_2 \int_{k_1}^{\infty} \int_{n_2}^{k_2} (d_2 - n_2)\, d\Phi_2\, d\Phi_1 + \\
\Theta_{31C} &= s_2 \int_0^{n_1} \int_{k_2}^{k_2+e} (d_2 - n_2 - e)\, d\Phi_2\, d\Phi_1 + \\
&\quad \int_0^{n_1} \int_{k_2+e}^{k_1+k_2-d_1} (s_1(d_2 - k_2 - e) + s_2(k_2 - n_2))\, d\Phi_2\, d\Phi_1 + \\
&\quad s_2 \int_{n_1}^{n_1+e} \int_{k_2}^{n_1+e+k_2-d_1} (d_1 + d_2 - n_1 - n_2 - e)\, d\Phi_2\, d\Phi_1 + \\
&\quad \int_{n_1}^{n_1+e} \int_{n_1+e+k_2-d_1}^{k_1+k_2-d_1} (s_1(d_1 + d_2 - k_2 - n_1 - e) + s_2(k_2 - n_2))\, d\Phi_2\, d\Phi_1 + \\
&\quad \int_{n_1+e}^{k_1} \int_{k_2}^{k_1+k_2-d_1} (s_1(d_1 + d_2 - k_2 - n_1 - e) + s_2(k_2 - n_2))\, d\Phi_2\, d\Phi_1 + \\
&\quad \int_0^{k_1} \int_{k_1+k_2-d_1}^{\infty} (s_1(k_1 - n_1 - e) + s_2(k_2 - n_2))\, d\Phi_2\, d\Phi_1 \\
\Theta_{31D} &= \int_{k_1}^{\infty} \int_{k_2}^{\infty} (s_1(k_1 - n_1 - e) + s_2(k_2 - n_2))\, d\Phi_2\, d\Phi_1 \\
\Theta_{31E} &= p_1 \int_{k_1}^{\infty} \int_0^{k_2} (d_1 - k_1) d\Phi_2 d\Phi_1 + p_2 \int_0^{k_1} \int_{k_1+k_2-d_1}^{\infty} (d_1 + d_2 - k_1 - k_2) d\Phi_2 d\Phi_1 + \\
&\quad \int_{k_1}^{\infty} \int_{k_2}^{\infty} (p_1(d_1 - k_1) + p_2(d_2 - k_2)) d\Phi_2 d\Phi_1
\end{aligned}
$$

Analysis of second order conditions (see appendix C.3) reveals that we can only conclude analytically that the staffing levels (n_1, n_2, e) are unique for a given set of capacity decisions (k_1, k_2).

3.5 Numerical Experiments

In this section, we select parameter values and discuss the results obtained through numerical experiments.

3.5.1 Parameter Values

To fully understand the effect of flexibility, decision timing, and parameter values on resource decisions and performance of a multi-floor hospital unit, we have developed a full-factorial scenario analysis for optimal policies. We consider two levels for each cost

TABLE 3.1: Values for Cost Parameters

Cost Parameter Levels	$\mathbf{s_1}$ **($/hr)**	$\mathbf{s_2}$ **($/hr)**	$\mathbf{f_1}$ **($/hr)**	$\mathbf{f_2}$ **($/hr)**	$\mathbf{p_1}$ **($/hr)**	$\mathbf{p_2}$ **($/hr)**
Low	30	25	9	5	51	45
High	35	28	12	7	60	50

parameter as shown in Table 3.1. Wages for dedicated nurses are taken as $22/hr for unit 1 and $20/hr for unit 2. Most cost values used in the numerical analysis are chosen to represent closely the values used in hospitals. The average pay for regular nurses in The south-east academic, medical center varies between $20/hr and $25/hr. The cost of contract nurses typically varies from $28 to $32/hr.

The demand is assumed to follow a uniform distribution in both unit 1 and 2. As shown in Table 3.2, three scenarios are used for demand distribution in the numerical analysis. In each case, the lower bound of the uniform distribution is set to zero. (A lower bound that is greater than zero would add a fixed cost to each configuration in our model.) Scenario 1 models two units facing same demand, whereas scenario 2 assumes higher demand in unit 2 and scenario 3 captures higher demand in unit 1.

TABLE 3.2: Values for Demand Parameters

Demand Parameters	**Scenario 1**	**Scenario 2**	**Scenario 3**
Unit 1	[0,20]	[0,20]	[0,25]
Unit 2	[0,20]	[0,25]	[0,20]

3.5.2 Results

This section presents the summary of results and explains the insights derived from the numerical analysis.

Optimal Capacity and Staffing Levels

First order conditions are used to find optimal capacity and staffing levels for all sequential cases as well as the simultaneous cases of configurations 1 and 3. Under simultaneous decision making in configurations 2 and 4, our numerical analysis suggests that the objective function is convex in both capacity and staffing decisions, but analytically we could only prove that given a capacity level, the objective function is convex in staffing decisions. A search of costs associated with each capacity level generated optimal decisions for simultaneous decision making in configurations 2 and 4.

Simultaneous vs Sequential Decision Making

This section compares simultaneous and sequential decision making for the four configurations. The percentage cost improvement for the timing alternatives under each configuration is shown in Table 3.3. Not surprisingly, increasing the integration of

TABLE 3.3: Sequential vs Simultaneous Decision Making - Average Percentage Cost Improvement

	Configurations			
Percentage Improvement	**1**	**2**	**3**	**4**
Simultaneous vs Sequential case 1	-	-	0.32%	0.39%
Simultaneous vs Sequential case 2	2.88%	2.87%	2.86%	2.75%
Sequential case 1 vs Sequential case 2	-	-	2.53%	2.37%

capacity and staffing decisions leads to better performance. Simultaneous decision making yields the lowest costs, while separating capacity decisions from all staffing decisions (sequential case 2) leads to the highest costs. In sequential case 1, where only nurse cross-training decisions are made after initial capacity and staffing decisions, performance is very close to the full simultaneous case. This implies that the *benefits of cross-training can be largely realized even if capacity and staffing levels have been determined prior to the establishment of a cross-training initiative.* The percentage cost improvement in integrating capacity and staffing decisions on average is 2.8%, irrespective of the level of flexibility, which indicates that *there is no interaction effect between type of flexibility and time of decision making.*

No Flexibility vs Staffing Flexibility vs Demand Upgrades

This section compares the percentage performance (cost) improvement obtained by shifting from one configuration to another configuration and explains the results and insights.

The percentage cost improvement under sequential decision making case 1 and 2 are shown in Table 3.4 and Table 3.5 respectively. Table 3.6 shows the benefit of added flexibility in the simultaneous decision making scenarios.

TABLE 3.4: Sequential Decision Making Case 1 - Average Percentage Cost Improvement

From / To	Configuration 4
Configuration 3	1.56%

TABLE 3.5: Sequential Decision Making Case 2 - Average Percentage Cost Improvement

	Configurations			
From / To	1	2	3	4
Configuration 1	-	1.39%	1.94%	3.78%
Configuration 2		-	0.55%	2.37%
Configuration 3			-	1.94%
Configuration 4				-

TABLE 3.6: Simultaneous Decision Making - Average Percentage Cost Improvement

	Configurations			
From / To	1	2	3	4
Configuration 1	-	1.38%	1.83%	3.42%
Configuration 2		-	0.45%	2.06%
Configuration 3			-	1.61%
Configuration 4				-

Under simultaneous decision making, the average percentage cost improvement for configuration 3 (with demand upgrades) and configuration 2 (staffing flexibility) from configuration 1 (base case) is 1.83% and 1.38%. The percentage cost improvement in using staffing flexibility is on average 0.45% higher than using demand upgrades. *Thus, the benefit of using staffing flexibility on average is higher than using demand upgrades.* The intuition behind this is as follows: once capacity is allocated to meet demand, staffing needs are met by any means even if it leads to using more contract nurses and incurring higher staffing costs. Employing staffing flexibility helps to minimize some of this staffing cost.

But a system that has both flexibilities (configuration 4) yields a higher benefit than the sum of the benefits from using staffing flexibility and demand upgrades separately. Under sequential decision making (case 2), the additive benefit of using staffing flexibility and demand upgrade is 3.33% while benefit of using staffing flexibility and demand upgrade in a single system (configuration 4) is 3.78%. Similarly, under simultaneous decision making, the additive benefit is 3.21% while the benefit of using both staffing flexibility and demand upgrade in a single system (configuration 4) is 3.42%. Staffing flexibility and demand upgrades have slight positive interaction. *Thus, the two types of flexibilities complement each other as evidenced by a slight positive interaction effect.*

Comparing staffing and capacity decisions among the four configurations, the following observations can be made.

1. Optimal capacity in unit 1 under simultaneous decision making is less than optimal capacity in unit 1 under sequential decision making, irrespective of type of

flexibility.

2. Optimal capacity in unit 2 is lower in demand upgrade configuration than in no flexibility or staffing flexibility configuration. The possibility of upgrading patients from unit 2 to unit 1 lowers capacity in unit 2.

3. Optimal capacity in unit 1 in staffing flexibility configuration is lower than in no flexibility or demand upgrade configuration.

4. Total available staff $(n_1 + n_2 + e)$ in staffing flexibility configuration is greater than total available staff $(z_1 + z_2)$ in no flexibility or demand upgrade configuration.

Optimal staffing decisions for no flexibility and staffing flexibility (configurations 1 and 3) are same and independent of capacity decisions under sequential and simultaneous decision making. This indicates that *making early capacity decision is the main cause of sub-optimality in sequential decision making.*

Implications of Demand

The benefit of staffing flexibility is highest when demand in unit 1 is more than demand in unit 2. Higher demand in unit 1 leads to higher utilization of flexible nurses since flexible nurses have priority allocation to unit 1. When demand in unit 2 is more than demand in unit 1, the benefit of staffing flexibility is least, but still significant.

Benefit of demand upgrades is highest when demand in unit 2 is more than demand in unit 1. When there is higher demand in unit 2 than in unit 1, upgrading patients from unit 2 to unit 1 is helpful. When demand in unit 1 is more than demand in unit 2, the benefit of upgrade is least, but significant.

Effect of Cost Parameters on Capacity and Staffing Decisions

In this section we analyze the trends exhibited by capacity and staffing variables when cost parameters vary under four configurations in simultaneous and sequential decision making.

The benefit of using staffing flexibility is higher when cost of contract nurses in unit 1 (s_1) is high irrespective of values taken by other cost parameters. This is because unit 1, being a complex unit, gets priority over unit 2 to utilize flexible nurses. Under *demand upgrades, maximum/ higher benefit of demand upgrades is obtained when contract nurse cost in unit 2 (s_2) is high* irrespective of values taken by other cost parameters.

Under simultaneous decision making in demand upgrades, as penalty cost in unit 1 increases optimal capacity and dedicated staff increases in unit 1 as seen in Figure 3.5. After a threshold value of penalty cost, capacity in unit 1 keeps increasing while dedicated staff in unit 1 decreases. This phenomenon is observed because at high penalty cost p_1, capacity in unit 1 is quite high and capacity in unit 2 is relatively low leading to greater possibility of upgrades to unit 1. Higher upgrades implies more use of contract nurses who can treat upgraded patients in unit 1. *The benefit of using contract nurses in unit 1, for unit 1 patients and upgraded patients is higher than the benefit of increasing (permanent) dedicated staff in unit 1 for patients in unit 1 and hiring contract nurses for upgraded patients.* Therefore, dedicated staff in unit 1 decreases at high penalty cost.

Under staffing flexibility configuration, the effect of cost parameters on capacity and staffing decisions under simultaneous decision making is summarized in Table 3.7. The table indicates that, when capacity-related cost parameter changes, change in unit 1

FIGURE 3.5: Configuration 2 Simultaneous Decision Making - Effect of increase in penalty cost in unit 1

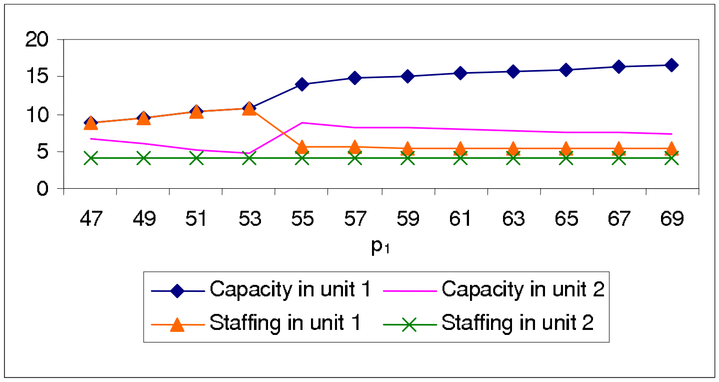


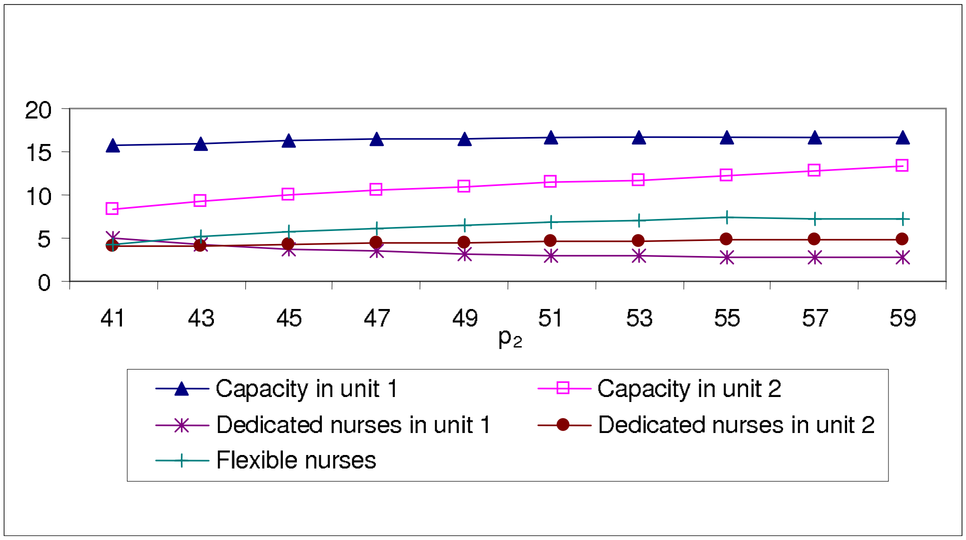
TABLE 3.7: Effect of Cost Parameters on the Decision Variables in Configuration 3 under Simultaneous Decision Making

Parameter Increase	$\mathbf{k_1^*}$	$\mathbf{k_2^*}$	$\mathbf{n_1^*}$	$\mathbf{n_2^*}$	$\mathbf{e^*}$
$\mathbf{p_1}$	$\uparrow$	$\leftrightarrow$	$\uparrow$	$\downarrow$	$\uparrow$
$\mathbf{p_2}$	$\leftrightarrow$	$\uparrow$	$\leftrightarrow$	$\leftrightarrow$	$\leftrightarrow$
$\mathbf{f_1}$	$\downarrow$	$\leftrightarrow$	$\downarrow$	$\uparrow$	$\downarrow$
$\mathbf{f_2}$	$\leftrightarrow$	$\downarrow$	$\leftrightarrow$	$\leftrightarrow$	$\leftrightarrow$
$\mathbf{s_1}$	$\downarrow$	$\leftrightarrow$	$\uparrow$	$\downarrow$	$\uparrow$
$\mathbf{s_2}$	$\downarrow$	$\downarrow$	$\uparrow$	$\uparrow$	$\downarrow$

capacity shows similar trend changes in flexibility. If capacity in unit 1 increases (when p_1 increases), flexibility increases, if capacity in unit 1 decreases (when f_1 increases), flexibility decreases, if capacity in unit 1 remains constant (when p_2 and f_2 increases) flexibility remains constant. *Flexibility (e^*) is more sensitive to directional change in capacity of unit 1 (k_1^*) than capacity of unit 2 because of priority assignment of flexible nurses to unit 1.*

Under demand upgrades (configuration 2) and staffing flexibility (configuration 3),

FIGURE 3.6: Configuration 4 Simultaneous Decision Making - Effect of increase in penalty cost in unit 2



when penalty cost in unit 2 increases, optimal capacity in unit 2 increases while optimal capacity in unit 1 either decreases or remains constant. Unlike configuration 2 and 3, under configuration 4 when p_2 increases optimal capacity in unit 1 increases under certain cost scenarios. When p_1 and f_2 is high, f_1 is low, either s_1, s_2 or both are high as shown in Figure 3.6, increase in p_2 increases capacity in unit 1. The intuition is that *high value of cost parameters that favors increase in capcaity in unit 1 such as* p_1*,* f_2*,* s_1 *and* s_2 *has greater impact on* k_1^**, forcing the system to increase capacity in unit 1 and overpowering the effect of increase in* p_2 *on* k_1^*. Highest number of flexible nurses are utilized when s_2 is low, s_1 is high and f_1 is low irrespective of demand pattern. High contract nurses cost in unit 1 prevents hiring more contract nurses while higher operating cost in unit 1 allows greater patient upgrades. The next highest number of flexible nurses is seen when s_2 is high, s_1 is high and f_1 is low.

The effect of all other cost parameters on the four configuration under simultaneous and sequential decision making are as expected.

3.6 Conclusions

In this paper, we consider two types of flexibility used to coordinate two critical resources (nursing staff and beds) and satisfy stochastic demand at minimum cost. We analyze four flexibility configurations (no flexibility, staffing flexibility, demand upgrades and both) under simultaneous decision making and sequential decision making. We prove convexity of the objective cost function for all models and determine optimal capacity and staffing decisions under different types of flexibility and timing of decisions.

On average, the benefit of using staffing flexibility (configuration 3) is higher than the benefit of using demand upgrades (configuration 2). Shifting from configuration 1 (No flexibility) to configuration 4 (Demand upgrades and staffing flexibility) creates significant average percentage cost improvement in the system. We also find that the percentage cost improvement is relatively consistent across parameter values.

The two types of flexibility, demand upgrades and staffing flexibility, have a positive interaction effect between them. The benefit of using demand upgrades and staffing flexibility together as a single system is slightly higher than the sum of using demand upgrades and staffing flexibility separately.

We find that simultaneously determining capacity, staffing and flexibility levels offers only a small improvement over a system where flexibility decisions are made later. The benefits of cross-training can be largely realized even if capacity and staffing levels have been determined prior to the establishment of a cross-training initiative.

The percentage cost improvement obtained by using simultaneous decision making over sequential decision making is approximately the same across different flexibility scenarios. The timing of decisions and type of flexibility have limited interaction; their effect on system performance is largely independent.

We find that the benefit of staffing flexibility is highest when demand in unit 1 is more than demand in unit 2, while the benefit of demand upgrades is highest when demand in unit 2 is more than demand in unit 1. The benefit of staffing flexibility is highest when the cost of hiring contract nurses in unit 1 is high, while the benefit of demand upgrades is highest when the cost of hiring contract nurses in unit 2 is high.

In this paper, we explore benefits and trade-offs in employing different types of flexibility while coordinating two key resources, staff and capacity. The consequences of making capacity and staffing decisions at different time periods is examined. The results in this paper will help managers not only understand advantages of staffing flexibility, but also identify its negative effects, and their inter-relationship among other system decisions. It will also enable administrators to recognize the importance of including staffing flexibility in the planning phases rather than using it as an ad-hoc mechanism to meet daily demand.

CHAPTER 4

RESOURCE FLEXIBILITY FRAMEWORK AND MANAGERIAL IMPLICATIONS

4.1 Introduction

In Chapter 2 and 3, we analyzed the benefits of implementing two types of resource flexibilities (staffing flexibility and demand upgrades) frequently used by hospitals as short-term strategies to gain operational flexibility and manage demand variability. In this chapter, we position those two types of resource flexibility strategies using Jack and Powers (2004) (JP) volume flexible strategy framework and develop a resource flexibility strategy (RFS) framework. We discuss some of the propositions in JP to see if it is agreeable with our RFS framework. We also address the practical issues that hospital administrators face when implementing the resource flexibility models in Chapters 2 and 3.

US hospitals operate at near-full capacity most of the time due to escalating operating costs, shortage of nursing staff and inefficiencies in the process. Even though hospitals often function at full capacity, they are expected to treat all patients despite demand uncertainty. Demand variability in hospitals is more significant than demand variability in most manufacturing firms or other service firms for two main reasons. Hospitals running at near-full capacity have to accommodate patients seeking care due to legal, social and moral obligations. Additionally, health care services cannot be inventoried, so hospitals use a combination of resource flexibility strategies to reduce supply-demand imbalance. For these two reasons, it is essential for hospitals to understand the conditions under which various types of resource flexibilities are best used.

Hospitals typically determine the capacity level of resources using "turnaway probability," the probability that they will turn away new patients for lack of capacity or discharge patients early (Baker et al., 2004). When patient arrivals are random, meaning arrival fluctuations about a level demand or a trend, capacity is not always sufficient to meet patient demand. Patient demand, measured as patient census, exhibits variability. It is high, earlier in the day and at beginning of the week while census (demand) is low during nights and weekends causing a supply-demand imbalance (Baker et al., 2004). Planning for appropriate resource flexibility strategy reduces supply-demand imbalance in both the short-term and long-term caused by demand variability.

Jack and Powers (2004) (JP) develops two frameworks that categorize volume flexible strategies in health services: one based on level of demand variability and range of flexibility, and another based on time and source of flexibility. "Volume flexibility is concerned primarily with organizations's ability to efficiently manage its output level in response to fluctuations in demand for its current products and services (Jack and Powers, 2004, p. 232)". Depending on range of flexibility and level of demand variability, one of four volume flexible strategies is likely to be most appropriate. The four strategies are (shielding strategy (for high demand variability and low range of flexibility), absorbing strategy (for low demand variability and low range of flexibility), containing strategy (for low demand variability and low range of flexibility), mitigating strategy (high demand variability and high range of flexibility). The framework also indicates that health care services use three shielding strategies (pricing and rationing, demand management models, and manage care control strategies) to shield against demand variability, two absorbing strategies (time buffers and slack capacity) to absorb fluctuations in demand, three containing strategies (workforce flexibility, efficiency measures and informational technologies) to contain demand variability, three mitigating strategies (restructuring, risk pooling, and outsourcing and strategic alliance) to mitigate demand variability.

Jack and Powers (2004) also conclude that when hospitals plan to use volume flexible strategies, the characteristics of different service lines indicate which strategy is most appropriate. For example, a shielding strategy provides high performance for primary care, elective plastic surgery, geriatric services, and health education due to the high demand variability and low flexibility typically experienced in those lines.

The volume flexible strategies and thereby the tactics can also be categorized using dimensions of time (short-term vs long-term) and source or capability (internal vs external). Workforce flexibility, time buffers, and demand management models are also identified as short-term, internal capability strategies. Pricing and rationing, and managed care controls are short-term, external capability strategies. Slack capacity buffer, informational technologies, efficiency measures, risk pooling, and restructuring are long-term, internal capability strategies. Outsourcing and strategic alliances are long-term, external capability strategies.

Demand variability causes supply-demand imbalance in the short-term. The supply-demand imbalance created by random fluctuations can be reduced by using two approaches: increasing the availability of supply to treat excess patients and/or reducing patient demand using demand management strategies. Although there are other supply-side resources that hospitals can utilize, we focus on two important resources required to meet patient demand: nursing staff and bedspaces, commonly categorized as labor and equipment resources.

When demand exceeds existing capacity, operational (short-term) or strategic (long-term) flexibility of staffing and/or equipment are used to absorb demand variability. On the demand-side, demand management strategies such as staggering patient arrivals, transfer of patients or increase in prices (such as in elective surgeries) are used to reduce supply-demand imbalance. Figure 4.1 displays a resource flexibility framework for accommodating demand variability. The supply-side strategies utilizing labor and equipment resources are listed in cells 1, 2, 4 and 5 in Figure 4.1, and demand management strategies using patients alone (and not external organizations like HMO, PPO etc) are listed in cells 3 and 6 in Figure 4.1.

In this essay, we develop a framework similar to the time vs source framework given by JP, but in our resource flexibility strategies framework we discuss only internal sources and not external sources as in JP. We elaborate on the flexibility strategies that can be applied to internal resources (labor, equipment, and patient) in both the long-term and short-term to contain different levels of demand variability. JP framework utilizes multiple resources to implement volume flexible strategies. In addition to to mapping three critical internal resources with volume flexible strategies from JP framework, we describe in detail the different flexibility strategies for the three resources and the interaction among their flexibility strategies. As we proceed to explain our resource flexibility framework, we also discuss the suitability of some of JP's propositions in our framework.

FIGURE 4.1: Resource Flexibility Strategy Framework for Demand Variability

Time Interval

	Supply-side resources		Demand-side strategies
	Labor	**Equipment**	**Patient**
Long Term RFS (Strategic flexibility)	*Increase capacity:* –Hiring & training –Foreign trained nurses **4b** *Increase utilization:* –Retaining (by salary, by benefits) –Increase productivity using technology **4a**	*Increase capacity:* –Expansion –Increase licensed beds **5b** *Increase utilization:* –Reallocation of beds from one unit to another –Improve process efficiencies **5a**	–Rationing –Waitlist for chemotherapy and organ donor in organ transplant –Pricing for elective surgery **6**
	Labor	**Equipment**	**Patient**
Short term RFS (Operational flexibility)	–Cross-training –Contract nurses –Agency nurses –Travel nurses **1**	–Upgrades –Redeployment –Increase holding area –Improve triage care –Monitor bed occupancy –Unstaffed beds **2**	–Transfer patients on hold –ER patients in triage and holding area –Early discharges –Diversions of critical care patients –Elective surgery rescheduled **3**

Low dispersion **High dispersion** ***Demand Variability***

RF S – Resource Flexibility Strategies

Operational flexibility: As indicated in Figure 4.2, over a short time period such as days, demand can be highly variable or less variable depending on the unit's characteristics and other factors. Short-term resource flexibility strategies, also referred to as operational flexibilities, are appropriate to absorb demand variability around a stable mean while long-term resource flexibility strategies, which are expensive and time consuming are used to absorb demand variability that causes shift in the mean level of demand. When demand is greater than capacity, managers should have the ability to change resource availability within a short period of time in order to meet the needs of patients. "Operational Flexibility is the ability to change day to day, or within a day as a matter of course (Upton, 1994, p.79)" The lower tier cells 1, 2 and 3 in the framework (Figure 4.1) list the strategies for supply-side and demand-side resources that will provide operational flexibility in the system. These strategies are efficient to meet only short-term demand variability. Using the short-term resource flexibilities such as part-time or contract employees, demand upgrades, or redeployment to meet long-term demand changes is not cost effective for service firms. Jack and Powers (2004) refer to some of these resource flexibility strategies as workforce flexibility, and demand management models. They do not consider short-term equipment flexibility strategies such as redeployment and demand upgrades.

Strategic Flexibility: Over a longer time interval such as months or years, the mean demand can shift or gradually change with a trend, but still have high or low demand variability as depicted in Figure 4.3. Growth, mergers, or organizational restructuring are some of the reasons for demand to change. When mean demand shifts,

FIGURE 4.2: Demand variability about mean demand in short-term

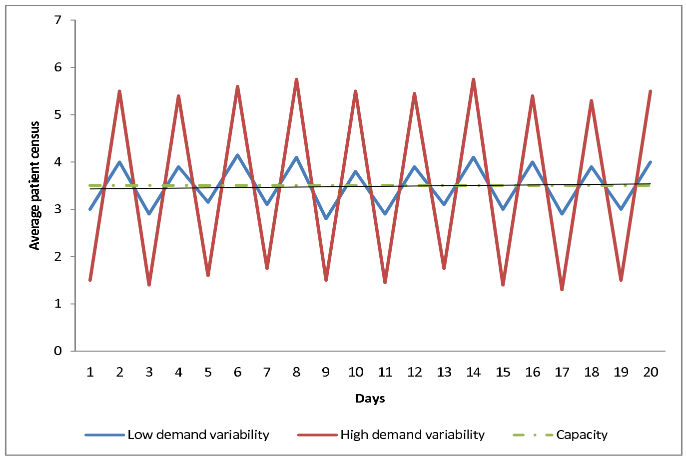
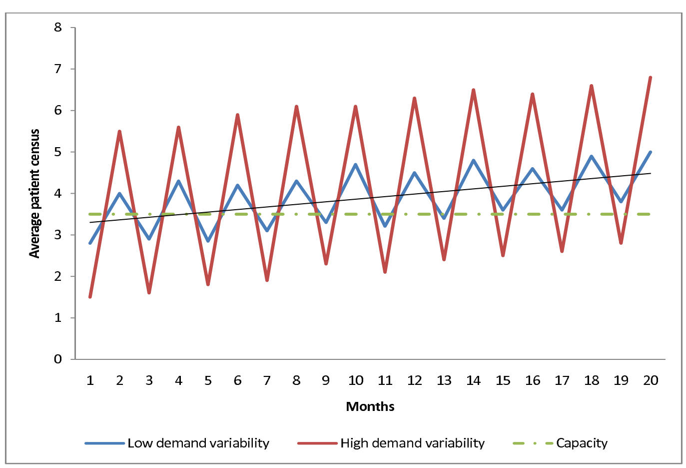


FIGURE 4.3: Demand variability about mean demand with trend in long-term



it is effective to make a one time change in resource capacities (*increase capacity* by buying new capacity, outsourcing demand) or a permanent *increase in utilization*, as opposed to absorbing excess demand with operational flexibility strategies. Although at first glance the strategy of adding capacity or increasing utilization does not seem to involve flexibility, based on Upton's strategic flexibility definition we can safely consider the above strategies as long-term resource flexibility strategies. "Strategic flexibility is the ability to make one-way, long-term changes which, in general, involve significant change, commitment or capital and which occur infrequently, say every few years or so (Upton, 1994,p.79)." Before deciding to increase capacity, it is essential to maximize the use of existing capacity, or in other words, increase utilization by eliminating inefficiencies, increasing productivity through technology, minimizing unoccupied/unstaffed beds, etc. Jack and Powers (2004) refer to these resource flexibility strategies as slack capacity buffers, and efficiency measures.

Proposition 1: Hospital units utilizing a combination of long-term and short-term resource flexibility strategies (when unit characteristics allow utilization of long-term and short-term RFS) are efficient in reducing supply-demand imbalance due to demand variability.

Proposition 2: Hospitals utilizing only long-term (strategic) supply side strategies (cells 4 and 5) but not short-term (operational) supply-side strategies (cells 1 and 2) are not cost-efficient in reducing supply-demand imbalance.

Although proposition 11 and 12 in JP indicate that volume flexible strategies can be categorized based on time and source dimensions, they do not explain the interaction between long-term and short-term strategies.

Short-term resource flexibility strategies: Demand variability may be high or low based on the type of unit. Units that have relatively less demand variability should use labor flexibility strategies (cell 1 in Figure 4.1). As demand variability increases, hospital administrators should start using labor and equipment flexibility strategies (cells 1 and 2). When demand variability is very high, then flexibility strategies of all three resources (cells 1, 2 and 3) are put into practice. The reasoning behind this progressive use of labor flexibility (cell 1 alone), labor and equipment flexibility (cells 1 and 2), labor, equipment and patient flexibility (cell 1, 2 and 3) as demand variability increases is as follows: Labor capacity is typically less than or equal to equipment capacity. So, when demand variability is low, there is a greater probability of having insufficient nurses than having insufficient beds. Labor flexibility is used when nurses are insufficient but beds are available. As demand variability increases, both labor and beds are insufficient and so both labor and equipment flexibilities are used. At demand variability, the probability of not having enough nurses and beds are very high and so we resort to patient management strategies.

At low demand variability, hospitals use flexible nurses (flexibility in terms of skill) to absorb demand (assuming all other resources are available). Some of the labor (nursing staff) flexibility strategies are cross-trained nurses, contract nurses, agency nurses and travel nurses (May et al., 2006, Li and King, 1999, Campbell, 1999, Lyons, 1992, Altimier and Sanders, 1999). Refer to later sections of this chapter for implementation and managerial issues in employing staffing flexibility when demand is uncertain.

When demand variability is slightly higher, hospitals redeploy patients to different units where bed spaces are available (Netessine et al., 2002 and Rao et al., 2004). They also increase the occupancy in holding area and employ contract staff for unstaffed beds to maintain high bed occupancy rates in all units. Therefore, as demand variability increases, flexibility strategies from cell 1 and cell 2 are used. Refer to later sections of this chapter for implementation and managerial issues in employing labor and equipment flexibility strategies to reduce supply-demand imbalance.

When demand variability is high, in addition to supply-side (labor and equipment) strategies (cells 1 and 2 in Figure 4.1), hospitals use demand-side strategies (cell 3), also referred to as demand management models in JP. They hold transfer patients from other clinics, prioritize patients at triage in ER, allow triaged patients to wait in holding area, organize early discharges, divert critical care patients on their way to the hospital in ambulance, and/or reschedule or cancel elective surgery (Dara et al., 2005) in order to smooth patient demand.

Proposition 3: It is cost-effective for hospital units with low short-term demand variability to use labor flexibility (when permissible by unit policies) as an operational flexibility strategy.

This proposition agrees with proposition 3 in JP where workforce flexibility can be used for an environment with low demand variability and high range of flexibility.

Proposition 4: Units with higher short-term demand variability should use both labor and equipment flexibility.

Proposition 5: At high short-term demand variability, hospital units should resort to demand-side strategies only after effectively utilizing supply-side flexibility strategies.

JP indicates that labor flexibility is a containing strategy and used in low demand variability and high range of flexibility environments. We argue that labor flexibility is a recourse for both low and high demand variability, in addition to equipment and demand management strategies in high demand variability environments.

JP also suggest that different volume flexible strategies are best suited for different service lines. As long as unit policies and unit characteristics permit operational flexibilities of labor, equipment and demand management, we propose that all service lines utilize them to reduce supply-demand imbalance and not categorize deployment of volume flexible strategies based on service lines. For example, JP suggests absorbing strategies (time buffers and slack capacity) for intensive care, although labor flexibility (containing strategy) in certain intensive care units is feasible (e.g. The academic medical center uses labor flexibility for its medical and surgical ICUs) and cost-effective to implement in the presence of demand variability.

Long-term resource flexibility strategies: Under long-term strategies there are two ways of adjusting supply-side resources to increase the number of patients a hospital can serve. The first option is to increase the *utilization of resources*; the second option is to increase the *physical capacity of resources*. JP refers to the earlier strategy as efficiency measures and latter strategy as slack capacity buffers.

Similar to short-term resource flexibility strategy (RFS), long-term strategies can focus on the supply-side (labor and resources) and the demand side. Long-term RFS are used to absorb a shift or trend in demand over longer time interval. The labor capacity (nursing staff) can be increased by hiring and training nurses or increasing the proportion of foreign-trained nurses (since the local pool of trained nurses is often already drained, not on 4.1), hiring already cross-trained nurses (May et al., 2006). Labor utilization is increased by increasing retention rates for nurses (increased through salary, flexible schedule, increased benefits, etc.,) or improving labor productivity by implementing more efficient processes and technology (May et al., 2006). Equipment capacity is increased by expansion of units, or addition of licensed beds. The effective capacity of equipment is increased by reducing process inefficiencies, or reallocation of unutilized beds to other units.

Proposition 6: For long-term demand changes, it is cost effective to employ strategies that increase utilization (cells 4a and 5a in Figure 4.1) before employing strategies that increase capacity (cells 4b and 5b in Figure 4.1) when using supply-side long-term resource flexibility strategies.

The demand-side RFS include commonly used strategies such as rationing and pricing. Patients who need organ transplant or chemotherapy are waitlisted until critical resources such as a donor organ or chemotherapy equipment is available. Once the critical resource, an organ donor in this instance is available, then the treatment process starts. Patients in rehabilitation and psychiatry are served as outpatients until a bed is available to admit them. Specialty hospitals increase the price of elective surgery such as cosmetic surgery to smooth the demand. These types of demand-side RFS can be implemented for certain types of service line where patient treatments can be delayed within reasonable limits.

For service lines like rehabilitation, psychiatry etc. where demand variability is relatively low, it is efficient for hospital administrators to use long-term resource flexibility strategies (upper tier in the framework in Figure 4.1) instead of using short-term resource flexibility strategies (lower tier in the framework). Due to the nature of the patients (long-term stay), these units experience high occupancy rate. When demand exceeds capacity, they start employing demand-side long-term resource flexibility strategies (cell 6 like rationing, waitlist and pricing) and then use supply-side long-term resource flexibility strategies (RFS) (such as expansion and hiring nurses).

In summary,we develop resource flexibility strategies framework where we present in detail various flexibility strategies utilized for three internal resources (labor, equipment and patient). Based on our research, interviews and literature review, we suggest propositions for flexibility strategies in the resource flexibility strategies (RFS) framework. These propositions should be tested in future for level of cost vs benefit and determine situations under which certain resource flexibility strategies are more/less cost-effective.

The rest of the chapter elaborates on the implementation issues surrounding three operational flexibility strategies discussed in the RFS framework: labor flexibility, equipment flexibility, and demand management strategies. The chapter is divided into six sections which explain the implementation issues pertaining to the analytical models in chapters 2 and 3. This bridges the gap between the analytical models and their implementation in the industry. The key aspects in implementing our models are:

- estimating demand variability for each unit
- understanding the degree of capacity flexibility to effectively use bedspace

- using staffing flexibility strategies to meet patient needs
- estimating relevant costs incurred in utilizing capacity and staffing flexibilities
- understanding the impact of timing of capacity and staffing decisions
- evaluating productivity of flexible staff

The processes and issues described in each of the sub-sections is compiled from in-depth interviews with practitioners and extensive literature review. The hospital we have analyzed is a public, academic medical center operated by a south-east state government. This academic medical center includes a main hospital, children's hospital, neurosciences hospital, and women's hospital. Construction of the new cancer hospital as a part of the academic medical center is now under way, with completion tentatively expected in 2009. Approximately 61,200 patients visited the emergency room (ER) and there were 741,980 clinic visits at the hospitals' 724 bed facility during 2006.

4.2 Demand variability

Staffing needs for each shift are estimated as the number of *FTE (full-time-equivalent)* RNs (registered nurses), LPNs/LVNs (licensed practical nurses or licensed vocational nurses) and NAs (nursing assistants) required. Individual patient care requires a mix of nursing care from RNs, LPNs and NAs. While RNs assess, diagnose, plan, implement and evaluate the appropriate plan of care for patients, LPNs help RNs with the plan of care, and NAs perform the routine tasks and chores that aid patient care. Therefore, a patient needs a varying proportion of nursing hours from RNs, LPNs and NAs. Consequently, the number of FTE RNs, LPNs and NAs needed varies for every shift.

The RN, LPN and NA staffing for each unit is calculated using a combination of factors such as patient acuity, patient census, skill mix for that unit, expected admissions, expected discharges and transfers. *Patient acuity* refers to the extent to which the patient is sick. There are 6 patient acuity categories ranging from Type 1 to Type 6. Type 6 patients are the most sick, and need one RN throughout their stay. For example, units with patient acuity 3 or 4 will typically be staffed with 50% RNs, 30% NAs and 20% LPNs. *Patient census* refers to the number of beds occupied or the number of patients in the unit at a given point of time. Of the total nursing care hours required for a patient, *skill mix* determines the relative proportion of hours served by RNs, LPNs and NAs. Using past history, the skill mix for a given unit and a given patient acuity is predetermined. In south-east academic medical center, the staffing needs for the next 24 hours are determined using the midnight census, acuity, and expected discharges, transfers and admissions. Using this information, a staffing software system calculates the FTE requirement for each of the three skill categories required during the next 24 hours. For example, if the output indicates that 7.2 RNs and 2.4 NAs are required in the critical care unit, then this becomes the demand for that shift. Any callouts (absenteeism of nurses without sufficient notice) or sudden changes in acuity are monitored 3 hours ahead of the shift start and once during the day shift , with any demand changes met by using one of the staffing flexibility strategies (see §4).

Implications The demand in our models is expressed in terms of full-time equivalent nurses required within a particular skill category. Our model should be applied to one skill category (e.g., RNs) at a time. The south-east academic medical center float both RNs and NAs, and also use per diem RNs and NAs and so our staffing model is applicable to RN staffing and NA staffing. The percentage of LPNs at the south-east academic medical center is sufficiently small that they are not cross-trained or hired as contract nurses. For hospitals with LPNs who float and LPN contract nurses, our staffing model is applicable to LPNs as well. This hospital uses FTEs per shift as a base unit to determine daily staffing needs. Past staffing data can be used to fit a demand distribution and then used in our model. For example, the budgeted staffing for the next 24 hrs in PFD (pulmonary and infectious disease unit) could be 7.2 FTEs for RN, 5.4 FTEs for NA and 2.6 FTEs for LPNs based on 95% occupancy rate in this 16 bed unit. *Managers should be careful to use budgeted staffing data and not the actual staffing because the unit may be under or over staffed and only budgeted staffing reflects the actual daily demand.* Also, other staffing systems might use nursing hours instead of FTEs and so careful understanding of the measurement units for staffing is required. As nursing managers move closer towards demand realization, they possess additional information about the demand distributions and the extent of demand uncertainty. Our model assumes that demand is independently and identically distributed throughout the planning horizon.

4.3 Degree of capacity flexibility

Hospitals use various strategies on the demand side to smooth the patient admission process into each unit. Many of the demand management strategies are listed in the resource flexibility framework (see Figure 4.1). Patients arrive through four different channels. Based on the type of patient arrival and their triage report, patients are allocated beds/capacity in the unit. The following list enumerates the four different channels of patient arrivals and also explains the relative importance of a patient gaining a bed in the recommended unit based on the channel through which they arrive.

1. Emergency Room (ER) patients: These patients are admitted to the emergency triage and categorized into five levels based on the level of service required. The level of service indicates the extent to which patients are sick. For the year 2006-2007, the south-east academic medical center had 61,200 ER visits. About one-third of these patients were not critical and were treated without admitting them to the units. If enough beds/capacity is not available in the recommended unit for high triage patients (patients requiring immediate medical attention), patients wait in the ER holdout area until beds become available. The emergency department at the south-east academic medical center has a holdout area of 6-8 beds. Patients not very critical (lower triage) are seated in the lounge and treated like outpatients without being admitted. This reduces blocking of ER capacity by non-critical patients. ER patients get the highest priority in bed allocation when a bed becomes available.

2. Clinical visits: Most patients usually contact their general practitioner when they get sick. If these patients are very sick, the doctors refer them to the nearest hospital. Patients who arrive at the hospital after a doctor's reference are called clinic visits. For the year 2006-2007 there were 741,980 clinic visits at the south-

east academic medical center. Clinic visits get second priority in bed allocation, after the ER patients.

3. Transfers: The third category of patients are transfers from another hospital. These patients are admitted in a non-specialty hospital but are transferred to a speciality hospital like the south-east academic medical center for further treatments. Some examples are chemotherapy, organ transplant, etc. Since transfer patients are already undergoing treatment in another hospital, their priority in bed allocation is third, after clinic visits.

4. Direct admissions: The direct admission patients are those who have planned treatments like chemotherapy, elective surgery, etc. These patients have the lowest priority in bed allocation.

As indicated above, patients are allocated beds under the following channel priority system: ER patients, clinic visits, transfers, and direct admissions. Patients who reach the hospital cannot be turned away or diverted to another hospital because of unavailable capacity. It is a legal, social and moral obligation of the hospital to treat any patient who reaches the hospital. However, critical patients such as trauma care and burn patients are sometimes diverted to another hospital as reported by a nurse in the south-east academic medical center.

Triage patients are allowed to wait either in the ER or holdout area, depending on the severity of their illness, until a bed becomes available. When a bed becomes available, patients are not admitted on FCFS (first come, first served) basis but based on their triage level. The allocation between ER patients, transfers, clinic visits and direct admissions are monitored once every shift by the house supervisor who also monitors the staffing for all units. The occupancy rate for each unit is also monitored twice every 24 hours and reported to the house supervisor. The average occupancy for medicine, surgery, women's and children service lines are around 95% at the south-east academic medical center. This hospital is running at almost full capacity in most of its service lines.

When beds in the desired unit are not available, patients are assigned to another unit similar to the one to which they are supposed to be admitted (i.e., units within the same clinical grouping). The following three tables show three sets of clinical groupings with increasing acuity levels where patients from one unit are upgraded / admitted to another unit. Table 4.1 indicates a sample of lower acuity units within the medicine floor. When beds in one unit (e.g., nephrology) are unavailable patients are allocated to another unit for this unit (e.g., pulmonary and infectious disease). For higher acuity units, such as the stepdown units and ICU units, patients are admitted between stepdown unit (Table 4.2) and between ICU units (Table 4.3) even if the speciality areas are different. For example, MICU patients are admitted in CCU or CTICU even though MICU is under medicine service line and CCU and CTICU are under heart center. Demand upgrades in the south-east academic medical center are allowed between some or all of the units within these clinical groupings.

Implications

For demand upgrade models (capacity flexibility), we considered two extremes of satisfying patient demand. First, when beds are not available, patients are upgraded to similar units within the same clinical grouping as seen in the lower acuity units at the

TABLE 4.1: Lower acuity units - Clinical grouping for medicine floor

Units under medicine floor clinical grouping
MSS - med/surg short stay
NFM - nephrology & family medicine
PID - pulmonary & infectious disease
GM - gerontology/medicine
RHB - rehabilitation

TABLE 4.2: Stepdown units - Clinical grouping intermediate care

Units under intermediate care clinical grouping
MPCU - medicine progressive care unit
ISCU - intermediate surgical care unit
CTSU - cardio thoracic surgical unit
ICCU - intermediate coronary care unit
BCIU - Burns center intermediate care

south-east academic medical center. Second, when beds are not available in both units, patients are diverted to another hospital. Depicting these two scenarios, we assume independent demand in our model. However, when beds are unavailable even after considering upgrades, hospitals place patients in the holdout area or ER for up to 24 hours. Patients in the holdout area or ER waiting for beds to become available cause additional cost to be incurred by the hospital. These patients usually are admitted usually within 24 hours. Demand between successive days is slightly correlated due to the fact that patients in the holdout or ER are eventually admitted. We do not consider this scenario in our model.

Since our model uses flexible nurses and demand upgrades between similar units in the same clinical grouping, our model has a better real-life applicability to lower acuity

TABLE 4.3: ICU units - Clinical grouping for critical care

Units under critical care clinical grouping
MICU - medicine intensive care unit
CCU - coronary care unit
CTICU - cardio thoracic intensive care unit
SICU - surgery intensive care unit
NSICU - neurosurgery intensive care
BICU - burn intensive care
ED - emergency department

units than ICU or step down units. When applying our model to ICU or stepdown units, managers should be careful to check if nurses are flexible enough to float between these units. UNC Hospital uses flexible nurses between all units within each clinical grouping shown in Table 4.1, 4.2, and 4.3 and demand upgrades between most units within these clinical groupings.

4.4 Degree of staffing flexibility

In Chapters 2 and 3, we used staffing flexibility to meet customer demands. Float nurses cross-trained to work in more than one unit and contract nurses (who are more expensive then staff nurses) are used as short-term staffing flexibility strategies to meet excess demand.

Hospitals use four types of flexibility strategies for nurse staffing. In order of increasing degree of flexibility, they are regular staff who float (called float nurses), flex team or per diem nurses, travelers or travel nurses, and contract nurses. A brief description of each type of flexibility strategy and its associated cost to the hospital is given below.

4.4.1 Float nurses

Float nurses are regular staff (full-time employees) assigned to a home unit. Staffing needs arising from last minute changes to either supply or demand, such as absenteeism, callouts in the current shift, and sudden changes in patient acuity, are met using float nurses from another unit in the same clinical grouping. The cost of using these float nurses (their base wages) vary between \$20 to \$24 per hour not including benefits. Float nurses float to all the units within the clinical grouping in lower acuity units, step down units and ICUs as shown in Tables 4.1, 4.2 and 4.3. Certain ICUs such as neonatal critical care (NCCU), pediatric intensive care (PICU) belong to one clinical grouping and hence only regular staff are used.

4.4.2 Flex team or per diem nurses

Flex team is a newly created system at many hospitals whereby nurses are hired or trained to work in several units across one or more clinical groupings. Flex team members are qualified to work in more units than float nurses, and they have a higher degree of flexibility. The south-east academic medical center created a flex team consisting of 15 RNs in July 2006. The flex team consists of both RNs (registered nurses) and NAs (nursing assistants), although flexible NAs are referred to as per diem nurses within the hospital.

This team provides RN coverage for units whose staffing needs are known in advance (i.e. open positions, maternity leave, etc.) as well as those that occur on short notice (i.e. callouts, changes in patient acuity). The flex team has three options depending on the number of clinical groupings in which they are hired to work. Under option I, they work in one clinical grouping (similar to float nurses), but can work no less than 24 hours per month. Under option II, flex team members can work in two clinical groupings but should work no less than 48 hours per month. Under option III, flex team nurses are allowed to work in three clinical groupings but should work no less than 48 hours. Critical care and intensive care units permit flex team nurses under option I only. The pay rate for flex team nurses under each option and type of shift is given in Table 4.4.

TABLE 4.4: Pay rate for flex team

Shift	**Option I**	**Option II**	**Option III**
Days	$28	$35	$37
Evening/nights	$33	$36	$38
Weekend days	$35	$38	$40
Weekend-E/N	$38	$40	$42

4.4.3 Travelers

Travel nurses are hired for a fixed period of time from an external agency and are not hospital employees. Travel nurses move from one hospital to another for a pre-specified period of time. Ususally travel nurses are hired for 13 weeks to a particular unit and are not allowed to work in any other unit unless otherwise specified in the contract. The hiring unit for travel nurses cannot float them to any other unit even if they are overstaffed and a sister unit (units under the same clinical grouping) is understaffed. The contract of travel nurses cannot be terminated unless they are proven to be too risky to handle patients. Travel nurses typically cost more than flex team nurses and their agency charges anywhere between \$45 and \$58 per hour for a travel nurse. This includes benefits, such as housing and travel allowance, given by the agency. Travel nurses are also called agency nurses.

4.4.4 Contract nurses

The last category of nurses are the contract nurses, who are the most expensive. They are hired from an agency to meet short term needs arising due to callouts and patient acuity changes. Their usual callout window is 3 hours prior to the start of the shift. These nurses are hired for a shift or two and typically have prior experience (from another hospital) in the unit they are hired to work. The cost for contract nurses is between \$55 and \$60 per hour.

4.4.5 Implications

In our model, we consider two extreme types of flexible nurses used to meet immediate or short-term staffing needs in hospitals. The first type of staffing flexibility strategy we use is the regular staff who float to sister units under the same clinical grouping or service line. The other type of staffing flexibility that we use in our model is the contract nurses who are hired at short notice for a shift or two. Our model determines the optimal number of float nurses to train, considering the fact that short term variability in staffing needs is met by float nurses and contract nurses within a specific clinical grouping. While implementing our model, managers should understand that the total staff pool is constant over a period of time. We do not account for long term absenteeism, such as maternity leave or unfilled positions. The change in nurse supply arising from long-term absenteeism is usually filled by travel nurses or flex team.

Health care literature has widely discussed the impact of nurse staffing on patient outcomes and costs (Robertson et al., 1997, Behner et al., 1990, McCue et al., 2003). Hospital administrators should not use staffing flexibility strategies, such as float nurse or cross-trained nurse, to reduce labor costs by decreasing the proportion of skill mix or decreasing the staffing intensity (number of nurses relative to the amount of care provided) for different units. Using multi-skilled nurses to replace RNs or reduce staffing requirements will only increase the overall costs for the hospital (Robertson et al., 1997). Although this result seems contrary to our findings - that use of multi-skilled nurses are cost-effective - the basis of cost comparisons are different. In Robertson et al. (1997), they use multi-skilled nurses but assume their nurses provide lower quality of care to patients, resulting in higher patient outcome costs for the hospital. In our model, we assume that use of cross-trained nurses does not lower the quality of care for patients. The use of cross-trained nurses is beneficial only when patients get adequate care from float nurses.

The recent legislation maintaining mandatory nurse-to-patient ratios across California hospitals has sent a wave of concern through all hospitals, especially as the nursing shortage becomes more severe. Maintaining nurse-to-patient ratios may well provide a higher level of patient care, but it is also important to look at the associated labor costs. The trade-off is between a decrease in post-operative or patient outcome costs due to higher staffing levels and higher labor costs due to an increase in nurse-to-patient ratios.

The legislation of mandatory nurse-to-patient ratios will increase labor cost for California by almost 1 billion dollars by 2008 (DHS, 2004). Wright et al. (2006) formulates an integer-programming model to determine the impact of mandatory nurse-to-patient ratios on nurse wages, and nurse scheduling. They find that nurse wage costs can be highly nonlinear with respect to changes in mandatory nurse-to-patient ratio, but a desirable scheduling policy can be implemented for nurses under mandatory nurse-to-patient ratio with a small impact on total wage cost. In our model in Chapter 3, we use 1:1 nurse-to-patient ratio. Higher nurse-to-patient ratios will be accounted for in the nursing hours requirement for each unit for models in Chapter 2. For example, a nurse-to-patient ratio of 1:2 (one nurse responsible for two patients), implies demand from two patients translates into one FTE nurse.

4.5 Estimation of relevant costs

The models in Chapter 2 and Chapter 3 consider relevant costs for the two resources, nurse staffing and bedspaces. Staffing costs include wages and training costs for the regular staff and different types of flexible nurses. Bed capacity costs include investment and operating costs per bed for each unit. Different methods for estimating staffing and capacity costs at the south-east academic medical center are discussed in the following sub-sections.

4.5.1 Staffing costs

The staffing costs vary depending on the level of flexibility. For a manufacturing plant, when components are outsourced because of insufficient plant capacity, the cost of outsourcing is usually less than the cost of producing in-house because of economies of scale for the contractor. Unlike a manufacturing plant, in hospitals when nurse staffing needs are outsourced, hiring contract nurses is typically more expensive than using in-house flexible staff.

The nursing wages for regular staff vary from state to state within the US, with California being the highest payer (even after considering the cost of living) while North Carolina is one of the lowest paying states. There is a significant cost difference for the hospital in employing a regular staff and flexible staff. The wages also differ depending on the type of day and shift. The weekday shift, weekday evening and night shifts, and weekend shifts have increasing hourly rates for RNs. The hourly wages, excluding benefits, for RNs at the south-east academic medical center vary from \$20 to \$24. In our model, we have considered the average hourly rate and have not segregated the regular staff into weekday/weekend or day/evening/night shifts. The numerical analysis considers both lower percentile wage rate and higher percentile wage rate.

Staff nurses in the south-east academic medical center are not paid additional wages when they float to another unit. They receive the same wage rate as a regular staff, but in certain units float nurses incur cross-training costs. The float nurses have assigned preceptors for two shifts at the unit to which they may float in the future. The cross-training costs include the cost of preceptors' time in the float unit and also the competency tests conducted by the HR department. The costs are budgeted into the units' cost center for utilizing preceptors and the HR cost center for the initial training. If patient acuity is similar between two units in a service line / clinical grouping, float nurses do not receive cross-training through preceptors.

Although there are significant cost benefits in using float nurses, Nicholls et al. (1996) indicate that most float nurses have negative feelings and apprehensions of floating because of new work environment, non-familiar protocols, and different skill set requirements. Some of the common perceptions of float nurses are disorientation, anxiety and uneasiness, and unfriendly receiving unit nurses. On the other hand, a few nurses look forward to new challenges, interactions with different types of patients, and the ability to expand their skill set. The lack of monetary incentives in floating could also contribute to the reluctance of regular staff to float, forcing hospital administrators to rely more heavily on contract nurses.

RNs who are members of the newly created flex teams at the south-east academic medical center are paid hourly rates higher than that of traditional staff nurses. There are three different options for flex teams. The flex team under option I have similar responsibilities as float nurses but receive differential wages. The wage rates for flex team members in Option II and Option III are higher and are shown in table 4.4.

The travel nurses and contract nurses are employed by a third party agency and contracted by the hospital. The hospital incurs a very high hourly rate ranging from \$55 - \$60. This includes the commission for the agency and the travel and housing allowances. Though the travel nurses and contract nurses incur the same hourly wage rate, travel nurses tend to be more expensive since nurses are hired for a 13 week period and must be paid-even for shifts when they are not needed.

In addition to the hourly wage rate, the travel nurses and flex team members incur a cross-training cost. The components of cross-training cost include costs for conducting a two-week hospital orientation to teach them the policies and procedures, administering an exam to test competencies, and two shifts of preceptors in the unit the nurses are planning to work.

Implications

Our model has considered the wages for dedicated staff and flexible staff (also referred as float nurses). While using this model, managers should understand that all full-time staff, irrespective of whether they float or not, are paid their base wages. Nurses who volunteer to float for over-time are not included in our model since their salary will exceed their base wage. McHugh (1997) simulated two staffing policies. One is unrestricted unit floating (UUF) where nurses are allowed to float to units across different clusters, and cluster unit floating (CUF) where nurses are allowed to float only within a cluster of similar units. She found that the wage cost differences between CUF and UUF were negligible, while the UUF policy had well-staffed units and the CUF policy had understaffed units.

The cross-training costs incurred as a result of orientation for flexible staff, conducting competency exams, and precepting in multiple units are included in our models. Hospitals may account for these cross-training costs to different cost centers, but it is important to include each cost component to capture the total cost of cross-training.

4.5.2 Capacity costs

There are two types of costs incurred by a hospital to build and maintain a unit. The initial investment cost, which includes the construction of the building and buying of equipment. This is a fixed cost and is approximately 1 million dollars per bed. Hospital beds can either be bought or leased. Most of the beds are owned by the south-east academic medical center but some are leased. The cost of owning a non-speciality bed is around $45,000 and the cost for specialty beds goes up to $95,000 (based on interview). The second type of cost is the daily operating cost for maintaining the bed. The daily operating charges include any specialty equipment that is leased such as pumps, cylinders and cost of supplies such as tubes, hoses etc. If the beds are leased, daily operating costs include the leasing charges as well.

Implications

In our model, we have considered both the initial investment cost (fixed cost) and daily operating cost for a bed. The initial construction cost for the building and equipment should be amortized to a unit time period. The operating cost should include all costs related to maintaining and using a bed such as leasing charges if leased, and cost of all supplies and leased equipments. In our model, operating costs are incurred whether or not the bed is used, depicting the real hospital system.

4.6 Timing of staffing and capacity decisions

This section discusses the timing of staffing and capacity decisions made by hospitals. Currently, the south-east academic medical center is running at a very high occupancy rate of 95%. As the occupancy rate increases, the south-east academic medical center submits a Certificate of Need to the State Hospital Review and Planning Council. The Certificate of Need indicates the total cost and benefit analysis of increasing the number of beds, either through expansion of the existing unit or the opening of a new unit. UNC Hospital has approval to increase their licensed beds from 724 to 800 beds for the year 2007. The new Cancer Hospital under construction will receive 50 beds and the rest of the beds will be allocated to other departments based on need. This process of approval is specific to a state hospital, but any private hospital will have to get approval for expansion from their planning and budget committees.

Implications

The staffing and capacity expansion decisions are made by the hospital planning and financing department. While calculating the budget for capacity expansion, the planning and financing department includes staffing needs and staffing budget. Although the hospital received approval for increasing regular staff before expansion, the nursing applicant pool is not large enough to fill all required regular staff positions. The staffing budget did not include float nurses or per diem nurses in their initial planning for staffing needs. Therefore, the south-east academic medical center used the sequential decision making - case 2 model (see Chapter 3) for timing of capacity and staffing decisions. Planning for total staffing and capacity was done initially, with decisions on the number of nurses to float determined at a later stage. When hospitals plan for physical capacity expansion, our research suggests they will benefit from accounting simultaneously for the use of flexible staff in addition to the hiring and use of regular staff.

4.7 Performance/productivity measurement

Qualitative research and discussion with practitioners indicate that cross-trained nurses do not perform as well as regular nurses because of new work environment, non-familiar protocols in float unit, and different skill set requirements (Nicholls et al., 1996). Manufacturing literature in worker cross-training corroborates the above phenomenon and shows that there are negative effects when workers are trained to do multiple tasks (Schultz et al., 2003; Stratman et al., 2004). Performance of cross-trained workers is lower than the performance of specialist workers, due to interruptions in work combined with learning and forgetting effects while shifting from one task to another (McCreery et al., 2004).

4.7.1 Literature review

Productivity measurement among manufacturing workers and knowledge workers: Productivity in the manufacturing literature is defined as the ratio of output to input, typically measured in terms of cost or quantity. Productivity is also defined as "output per employee-hour, quality considered (Sutermeister, 1976)." Though the above definitions of productivity are widely used to measure labor productivity in manufacturing, they cannot be easily applied to knowledge workers in service sectors because part of the output is intangible. Drucker (1999) states that "the challenge today is not to increase manual worker's productivity but to measure and increase knowledge worker's productivity ." Researchers have attempted to quantify the productivity of knowledge workers and have failed to find a measure that is both suitable and general to measure productivity of individual and teams applicable to any service industry (Ray and Sahu, 1989; Schroeder et al., 1985; Ramirez and Nembhard, 2004). Schroeder et al. (1985) discuss the relationship between productivity and performance: "In a strict sense, productivity refers to a relationship between outputs and inputs, and as such tends to emphasize efficiency. In a broader sense, productivity refers to all performance on the job, including effectiveness and efficiency." Drucker (1973) has characterized the difference between efficiency and effectiveness as "efficiency is doing things right, while effectiveness is doing the right things".

Productivity measurement of nurses: Productivity of nurses in the nursing literature is measured in an aggregate level. The aggregate productivity measure for a DRG/unit/hospital is commonly quantified in two ways, first as a ratio of output to inputs and second as patient outcomes. These two aggregate productivity measures widely seen in the nursing literature are not applicable to measure performance of float nurse as elaborated here:

1. Ratio of output to inputs: Productivity of units/DRG is measured as ratio of outputs to inputs; inputs and outputs are measured in terms of monetary value (paid dollars), patient hours worked, hours per patient day (hppd), and /or FTE per admission, FTE per occupied bed (Spitzer, 1986; Curtin, 1995; Eastaugh, 2002; Longest, 1977; Moody, 2004; Edwardson, 1985; Holcomb et al., 2002; Jordan, 1994; Strasen, 1987). The performance of float nurses should also include an intangible element of patient care. Patient care is the ability of nurses to make patients comfortable over and above completing tasks. Hence, measuring productivity as a

ratio of outputs to inputs is not a suitable measure of productivity for a float nurse or for that matter, a regular nurse.

2. Patient outcomes: Patient outcome in a unit is also used to measure productivity of nurses. Outcome indicators are defined as the states or levels of well-being, which result from care processes. Some of the variables that are used to measure patient outcomes are infection rates, length of stay, mortality rate, medication errors (Weech-Maldonado et al., 2005; Aiken et al., 2002; Hall et al., 2004; Czaplinski and Donna, 1998; Blegen et al., 1998). Patient outcome is the consequence of the performance of a group of nurses in a unit, and is the result of actions and decisions made by nurses, physicians and the health care team throughout the stay of the patient and cannot be surely attributed to the performance of a float nurse.

Productivity measurement in the organizational behavior literature: In the organizational behavior literature, job performance of employees consists of in-role and extra-role job performance. The in-role job performance determines the amount of work completed and the time taken to get the work done. The extra-role job performance identifies the extent to which the employee is proactive and initiates tasks for the betterment of the department. The scales used to measure in-role and extra-role job performance are referred to as the Organizational Citizenship Behavior Scales (Janssen and Van Yperen, 2004; Williams and Anderson, 1991).

McNeese-Smith in many of her papers (McNeese-Smith, 1996; McNeese-Smith, 2001; McNeese-Smith, 1997; McNeese-Smith, 1999; Loke, 2001) defines productivity as the "contribution toward an end result in relation to resources consumed" and develops a

15 dimension scale which is the self-assessment of the employee's contribution to unit productivity.

Performance of nurses cannot be measured solely on time taken to complete tasks or counting the number of errors because of variability, uncertainty and instability in patient conditions. Variability occurs because patient conditions differ from one patient to another. Secondly, there is no certainty on the outcome for a patient even when treated in a particular way. Patients respond differently to the same treatments. Thirdly, patient condition is unpredictable at different points of time causing instability in patients' condition.

4.7.2 Implications

The operations and health care literature described above prescribes ways to measure individual and aggregate productivity of service workers. Customizing the measures to effectively measure the contribution or productivity of cross-trained nurses from one unit to another, will help hospital managers to implement our model.

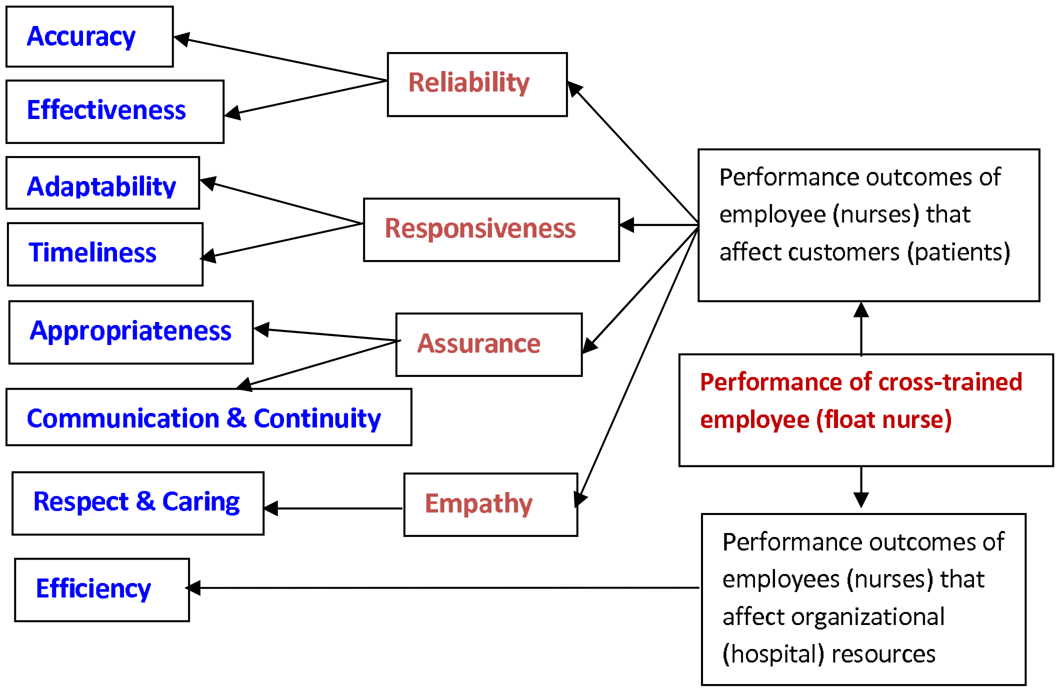
The south-east academic medical center uses performance appraisal forms that are completed at the end of the shift for contract nurses and float nurses. Using this information, and feedback provided by the charge nurse or nursing manager from the host unit, performance values can be aggregated to determine the level of productivity of nurses from one unit to another. Our models indicate that productivity of float nurses does impact the optimal level of cross-training, but numerical analysis shows that slight changes in productivity do not impact the solution drastically. In other words, the optimal solution is robust to slight productivity changes, but it is very important to identify whether floating causes significant decrease in productivity.

Performance of cross-trained staff affects both the customers and the organization (see Figure 4.4). The work done by a float nurse (cross-trained employee) will affect the perception of quality of service in the minds of patients (customer) and will also affect the hospital (organizational) resources. Figure 4.4 shows how performance of cross-trained staff affects both the customer and the organization. Parasuraman et al. (1988) and Parasuraman et al. (1985) develop and validate 22 items to measure service quality that load onto five factors (Tangibles and four intangibles: Reliability, Responsiveness, Assurance and Empathy) called SERVQUAL measures. JCAHO (JCAHO, 1996) developed nine theoretical dimensions (efficacy, efficiency, respect & caring, safety, availability, timeliness, appropriateness, continuity, and effectiveness) used to measure patient satisfaction towards their hospital. Modifying the JCAHO scales to measure service quality of float nurses will be one other way of measuring productivity of cross-trained nurses. Further study is required to test if the framework in Figure 4.4 is valid and reliable to measure the performance of cross-trained nurses.

4.8 Managerial Insights

In Chapter 2, we derived the optimal number of flexible nurses for two units. After identifying the relevant costs (e.g., wages, cross-training costs), demand distribution, total staff for each unit, and worker productivity, these parameter values are used in the closed form expression given in Chapter 2 to determine the optimal number of float

FIGURE 4.4: Measurement Model for Productivity



nurses in each unit. When the level of demand variability is similar across the two units under consideration, the degree of cross-training is driven by cross-training costs and employee productivity. If cross-training costs are low (because the two units are similar in terms of medication distribution, protocols and patient population) and the productivity of float nurses (based on past performance appraisals) is high, the *number of nurses to be cross-trained should be high.* If the cost of cross-training is high (because the units are dissimilar) and expected productivity is not very high, the *number of nurses to be cross-trained should be low.* These insights are as expected for a given scenario.

The more interesting insight is obtained when demand variability across the two units is different. Consider a unit with high demand variability (called unit B) and another unit under the same clinical grouping with average demand variability (called unit A). If the anticipated productivity of nurses floating from unit A to unit B is low, *the number of float nurses in unit A should be low.* If the anticipated productivity of nurses floating from unit A to unit B is high and cross-training cost is low, *the number of float nurses in unit A should be high.* If the expected productivity of float nurses from unit A to unit B is high but cross-training cost is high, *the number of float nurse in unit A should be low.*

The impact of productivity on staffing decisions when hospitals use multiple staffing flexibility strategies (such as regular staff, float nurses, flex team, travelers and contract nurses) for multiple units within a clinical grouping is not studied in this dissertation and will be interesting to study in the future. Additionally, future research should include more units within a clinical grouping than the two units considered in this dissertation. Chaining of cross-trained nurses is more likely for staffing flexibility in more than two units. Higher productivity in a unit might lead to fewer cross-trained nurses and chaining might result in an interaction effect as well.

In Chapter 3, we modeled two types of resource flexibility: staffing flexibility and capacity flexibility. We considered two types of staffing flexibility (float nurses and contract nurses) and two types of capacity flexibility (upgrades and diverted patients). When both staffing and capacity flexibility are available, our research suggests that on *average staffing flexibility will have a greater impact on the hospital's financial performance.*

When hospitals expand or build new units, it is essential to make all *capacity, regular staffing and flexible staffing decisions simultaneously to attain minimum cost in the presence of stochastic patient demand.* In situations like the south-east academic medical center where float staff are used as an emergency resource and not as a policy to meet short term staffing needs, planning for staffing flexibility (float nurse) in initial decision making is not feasible. In that case, capacity and total staff decision making should be made simultaneously to reap maximum benefit (Sequential decision making - Case 2).

4.9 Conclusion

Chapters 2 and 3 use analytical models to represent the issues of staffing and capacity flexibility for two units in a hospital. This chapter positions those two resource flexibilities in a broader framework and compares it with Jack and Powers (2004) framework. This chapter also describes hospital processes, discusses ways to determine parameter estimates from existing systems, and explains the implications of our models to hospital managers who implement them. Each of the six sub-sections details the implementation issues in our model, presents the process of estimating parameters from existing systems, and also discusses the limitations and assumptions of our models. Based on the results that we obtained in Chapters 2 and 3, we derive insights for managers when they utilize staffing and capacity flexibility strategies to absorb short-term demand variability.

CHAPTER 5

CONCLUSION

The benefits realized through employing a cross-training program in service and manufacturing firms are well known. In this dissertation, we weigh the benefits of cross-training against the negative effects when cross-training (staffing flexibility) is used as a mechanism to meet uncertain demand. Cross-trained staff are not as productive in their non-primary department as they are in their primary department. In a series of three essays, we address the following research questions: What is the impact of varying productivity of a cross-trained staff on system performance in a service firm? How does staffing flexibility coordinate with other operational flexibility and operational decisions? Under what circumstances should different types of resource flexibility strategies be used?

In the first essay, we determine an optimal cross-training policy accounting for the lower productivity of cross-trained staff in the non-primary unit. We derive a closed form expression for the optimal amount of cross-training in two units when demand follows a general, continuous distribution. When the cost of cross-training is high, an increase in productivity leads to an increase in the amount of cross-training. When the cost of cross-training is relatively low, however, there is a productivity level beyond which further increases in productivity reduce the amount of cross-training. Above this threshold, the productivity of cross-trained nurses allows the benefits of flexibility to be obtained with fewer cross-trained nurses. For low cross-training productivity, more demand variability leads to less cross-training. For high cross-training productivity, the effect of demand variability on the amount of cross-training depends on cross-training cost. When the cross-training cost is high, more variability continues to cause less cross-training. For low cross-training cost, however, an increase in demand variability leads to more cross-training.

In addition to the staffing flexibility (cross-training), we also consider demand upgrades as a type of flexibility. In order to satisfy patient demand we need both nurses and bed spaces at the right time and right quantity. When facing insufficient resources (nurses and bedspaces), hospitals use cross-trained nurses or upgrade patients to a higher or more sophisticated unit. We analyze the implication of using no flexibility, only staffing flexibility, only demand upgrades, and flexibility for both resources. Also, we determine the optimal decisions (number of nurses and number of bedspaces) when choices are made over a single time period (simultaneous decision making) and over staggered time periods (sequential decision making). In the second essay, we answer the following questions: For each flexibility configuration, under sequential and simultaneous decision making, what is the optimal resource level required to meet stochastic demand at minimum cost? Is one type of flexibility (e.g. demand upgrades) better than the other type of flexibility (e.g. staffing flexibility)? What is the effect of timing of decisions on type of flexibility and system performance? We use two-stage stochastic programming to find optimal capacity and staffing levels for each flexibility configuration under sequential and simultaneous decision making, in two non-homogenous hospital units that face continuous, general demand distribution. We find that the benefit obtained in using staffing flexibility on average is higher than the benefit of using demand upgrades. The two types of flexibilities complement each other and have a positive interaction effect between them. The benefits of cross-training can be largely realized even if capacity and staffing levels have been determined prior to the establishment of a cross-training initiative. The timing of decisions is important, but timing of decisions and type of flexibility has an independent effect on system performance.

In essay 1 and 2, we analyzed the benefits of implementing two types of resource flexibilities (staffing flexibility and demand upgrades) frequently used by hospitals as short-term strategies to gain operational flexibility and manage demand variability. In the third essay, we position those two types of resource flexibility strategies using Jack and Powers (2004) volume flexible strategy framework and develop a resource flexibility strategy (RFS) framework. We discuss some of the propositions in Jack and Powers (2004) to see if they are consistent with our resource flexibility framework. Various types of resource flexibility strategies needed to meet short-term variability and long-term demand uncertainty are discussed in this essay. We also address the practical issues that hospital administrators face when implementing resource flexibility models from essay 1 and 2. The essay describes hospital processes, discusses ways to estimate parameters from existing systems, and explains the implications of our models to hospital managers who want to implement them. The key implementation issues that we present, with the aid of a case study from a south-east, academic medical center are: estimating demand variability for each unit, understanding the degree of capacity flexibility to effectively use bedspace, using staffing flexibility strategies to meet patient needs, estimating relevant costs incurred in utilizing capacity and staffing flexibilities, understanding the impact of timing of capacity and staffing decisions, and evaluating productivity of flexible staff.

Thus, this dissertation has analyzed the benefit, trade-offs, limitations, and outcomes in implementing critical resource flexibility strategies in hospitals to minimize supply-demand imbalance and manage short-term demand variability. As an extension to essay 1, the results from the numerical study can be tested on different demand distributions. Further research is needed to generalize the results in essay 1 to more than two units. Conducting numerical analysis for additional demand distributions in essay 2 might lead to unforeseen results. Demand data can be collected from a hospital to evaluate the effectiveness of the model. Essay 3 positions the two types of flexibilities we used in the models in essay 1 and essay 2. Further research is required to test the interactions between different types of resource flexibility strategies that are used to manage demand variability in the short-term and demand uncertainty in the long-term.

APPENDIX A - Configuration 2

A.1 Sequential Decision Making

Period 1

Hessian for equation (3.8) is

$$\begin{bmatrix} \kappa + \alpha & \kappa \\ \kappa & \kappa + \beta \end{bmatrix}$$

$$\begin{aligned} \text{where,}\ \kappa &= p_2 \int_0^{k_1} \Phi_2'(k_1 + k_2 - d_1)\, d\Phi_1 > 0 \\ \alpha &= (p_1 - p_2)\Phi_1'(k_1) + p_2 \Phi_2(k_2)\Phi_1'(k_1) > 0 \\ \beta &= p_2(1 - \Phi_1(k_1))\Phi'(k_2) > 0 \end{aligned}$$

This shows that the hessian is strictly diagonally dominant and so $\Gamma_{11}(k_i)$ (3.8) is convex in k_i. Therefore, we can use first order conditions to determine optimal capacities.

Period 2

Hessian for equation (3.9) is

$$\begin{bmatrix} s_1\, \Phi_2(k_1 + k_2 - z_1)\Phi_1(z_1) & 0 \\ 0 & s_2\, \Phi_2'(z_2)(\Phi_(k_1) - \Phi_(z_1) + 1) \end{bmatrix}$$

This matrix is strictly diagonally dominant and so the $\Gamma_{12}(z_i)$ is convex in z_i (3.9).

A.2 Simultaneous Decision Making

Hessian for the first stage unconstrained objective function in equation (3.10) is

$$\begin{bmatrix} \alpha & 0 & \kappa_1 & \kappa_1 \\ 0 & \beta & 0 & 0 \\ \kappa_1 & 0 & \kappa_2 + \eta + \varphi_1 & \kappa_2 + \eta \\ \kappa_1 & 0 & \kappa_2 + \eta & \kappa_2 + \eta + \varphi_2 \end{bmatrix}$$

$$\begin{aligned} \text{where, } \alpha &= s_1\Phi_2'(k_1 + k_2 - z_1)\Phi_1(z_1) > 0 \\ \beta &= s_2\Phi_2'(z_2) > 0 \\ \kappa_1 &= s_1 \int_0^{z_1} \Phi_2'(k_1 + k_2 - d_1)\, d\Phi_1 > 0 \\ \kappa_2 &= (p_2 - s_1) \int_0^{k_1} \Phi_2'(k_1 + k_2 - d_1)\, d\Phi_1 > 0 \\ \eta &= s_1 \int_0^{z_1} (z_1 - d_1)\Phi_2\text{''}(k_1 + k_2 - d_1)\, d\Phi_1 > 0 \\ \varphi_1 &= (p_1 - p_2)\Phi_1'(k_1) > 0 \\ \varphi_2 &= ((p_2 - s_2) + (s_1 - p_2)\Phi_1(k_1))\Phi_2'(k_2) > 0 \end{aligned}$$

The first minor is α and second minor is $\alpha \cdot \beta$. The entire hessian is not positive definite but the first and second minor is positive and so we can conclude that z_i is convex in its objective function $\Gamma_{21}(k_i, z_i)$ (3.10) given k_i. Therefore, first order conditions can only be used to determine optimal values of staffing *(z_i^*)*.

APPENDIX B - Configuration 3

B.1 Sequential Decision Making - Case 1

The following matrix shows the hessian for unconstrained objective function in equation (3.11).

$$\begin{bmatrix} s_2\Phi_1'(z_1 - e_1) & 0 \\ 0 & (s_1 - s_2)\Phi_1'(z_1 + e_2) + s_2\Phi_2'(z_2 - e_2) \end{bmatrix}$$

The hessian is strictly diagonally dominant and so first order conditions are necessary and sufficient to find optimal values of e_i.

B.2 Sequential Decision Making - Case 2

Hessian for objective function (3.14) is

$$\begin{bmatrix} \alpha + \beta + \delta & \alpha & \alpha + \delta \\ \alpha & \alpha + \gamma + \epsilon & \alpha + \epsilon \\ \alpha + \delta & \alpha + \epsilon & \alpha + \delta + \epsilon \end{bmatrix}$$

$$
\begin{aligned}
\text{where, } \alpha &= s_2 \int_{n_1}^{n_1+e} \Phi_2'(n_1+n_2+e-d_1)\, d\Phi_1 > 0 \\
\beta &= s_2\Phi_1'(n_1)(1-\Phi_2(n_2+e)) > 0 \\
\gamma &= s_2\Phi_2'(n_2)(1-\Phi_1(n_1+e)) > 0 \\
\delta &= s_2\Phi_2(n_2)\Phi_1'(n_1+e) + s_1 - s_2)\Phi_1'(n_1+e) > 0 \\
\epsilon &= s_2\Phi_1(n_1)\Phi_2'(n_2+e) > 0 \\
\text{First principal minor} &= \alpha+\beta+\delta > 0 \\
\text{Second principal minor} &= \alpha\beta+\alpha\delta+\alpha\gamma+\beta\gamma+\delta\gamma+\alpha\epsilon+\beta\epsilon+\delta\epsilon > 0 \\
\text{Determinant} &= \gamma\delta\epsilon+\alpha(\gamma+\delta)(\beta+\epsilon)+\beta\delta\epsilon+\beta\gamma(\delta+\epsilon) > 0
\end{aligned}
$$

The hessian is positive definite and so first order conditions are used to determine optimal staffing *(n_i^*)* and *(e^*)*.

B.3 Simultaneous Decision Making

The following matrix shows the hessian and proof of convexity for the objective function in equation (3.15).

$$
\begin{bmatrix}
\alpha+\beta+\delta & \alpha & \alpha+\delta & 0 & 0 \\
\alpha & \alpha+\gamma+\epsilon & \alpha+\epsilon & 0 & 0 \\
\alpha+\delta & \alpha+\epsilon & \alpha+\delta+\epsilon & 0 & 0 \\
0 & 0 & 0 & \eta & 0 \\
0 & 0 & 0 & 0 & \zeta
\end{bmatrix}
$$

$$
\begin{aligned}
\text{where, } \alpha &= s_2 \int_{n_1}^{n_1+e} \Phi_2'(n_1+n_2+e-d_1)\, d\Phi_1 > 0 \\
\beta &= s_2 \Phi_1'(n_1)(1-\Phi_2(n_2+e)) > 0 \\
\gamma &= s_2 \Phi_2'(n_2)(1-\Phi_1(n_1+e)) > 0 \\
\delta &= s_2 \Phi_2(n_2)\Phi_1'(n_1+e) + (s_1-s_2)\Phi_1'(n_1+e) > 0 \\
\epsilon &= s_2 \Phi_1(n_1)\Phi_2'(n_2+e) > 0 \\
\eta &= (p_1-s_1)\Phi_1'(k_1) > 0 \\
\zeta &= (p_2-s_2)\Phi_2'(k_2) > 0 \\
\text{First principal minor} &= \alpha+\beta+\delta > 0 \\
\text{Second principal minor} &= \alpha\beta+\alpha\delta+\alpha\gamma+\beta\gamma+\delta\gamma+\alpha\epsilon+\beta\epsilon+\delta\epsilon > 0 \\
\text{Third principal minor} &= \gamma\delta\epsilon+\alpha(\gamma+\delta)(\beta+\epsilon)+\beta\delta\epsilon+\beta\gamma(\delta+\epsilon) > 0 \\
\text{Fourth principal minor} &= (\beta\gamma\delta+\gamma\delta\epsilon+\beta\epsilon(\gamma+\delta)+\alpha(\gamma+\delta)(\beta+\epsilon))\eta > 0 \\
\text{Determinant} &= (\beta\gamma\delta+\gamma\delta\epsilon+\beta\epsilon(\gamma+\delta)+\alpha(\gamma+\delta)(\beta+\epsilon))\eta\zeta > 0
\end{aligned}
$$

Since the hessian is positive definite, first order conditions are sufficient and necessary to find optimal capacity (k_i^*) and staffing (n_i^*) and (e^*).

APPENDIX C - Configuration 4

C.1 Sequential Decision Making - Case 1

Hessian for equation (3.18):

$$\begin{bmatrix} \alpha+\beta & \alpha \\ \alpha & \alpha+\gamma \end{bmatrix}$$

$$\begin{aligned}
\text{where, } \alpha &= (s_1-s_2)\Phi_1(z_1-e_1)\Phi_2'(e_1+e_2+k_2) > 0 \\
\beta &= s_1\Phi_1(z_1+e)(\Phi_2(k_1+k_2+e_1-z_1)-\Phi_2(e_1+e_2+k_2) \\
&\quad +s_2\Phi_1'(z_1+e_1)\Phi_2(e_1+e_2+k_2)+\Phi_1(z_1+e_1)\Phi_2(z_2+e_1) > 0 \\
\gamma &= (s_1-s_2)\Phi_2(k_2)\Phi_1'(z_1+e_2)+s_2(1-\Phi_1(z_1+e_2))\Phi_2'(z_2-e_2) \\
&\quad +(s_1-s_2)\int_{z_1-e_1}^{z_1+e_2}\Phi_2'(k_2+z_1+e_2-d_1)d\Phi_1 > 0 \\
\text{First principal minor} &= \alpha+\beta > 0 \\
\text{Determinant} &= \alpha\beta+\alpha\gamma+\beta\gamma > 0
\end{aligned}$$

Since the hessian of equation (3.18) is positive definite by strict diagonal dominance, first order conditions are sufficient and necessary to obtain optimal flexible nurses (e_i^*).

C.2 Sequential Decision Making - Case 2

Hessian for equation (3.19):

$$\begin{bmatrix} \alpha+\beta+\gamma+\epsilon & \alpha & \alpha+\beta+\gamma \\ \alpha & \alpha+\delta+\zeta & \alpha+\delta \\ \alpha+\beta+\gamma & \alpha+\delta & \alpha+\beta+\gamma+\delta+\eta \end{bmatrix}$$

$$\begin{aligned}
\text{where, } \alpha &= s_2 \int_{n_1}^{n_1+e} \Phi_2'(n_1+n_2+e-d_1)d\Phi_1 > 0 \\
\beta &= (s_1-s_2)\int_{n_1}^{n_1+e} \Phi_2(k_2+n_1+e-d_1)\Phi_1 > 0 \\
\gamma &= \Phi_1'(n_1+e)((s_1-s_2)\Phi_2(k_2)+s_2\Phi_2(n_2) > 0 \\
\delta &= s_2\Phi_1(n_1)\Phi_2'(n_2+e) > 0 \\
\epsilon &= s_1\Phi_1'(n_1)(\Phi_2(k_1+k_2-n-1)-\Phi_2(k_2+e)) \\
&\quad +s_2\Phi_1'(n_2)(\Phi_2(k_2+e)-\Phi(n_2+e)) > 0 \\
\zeta &= s_2\Phi_2'(n_2)(1-\Phi_1(n_1+e) > 0 \\
\eta &= (s_1-s_2)\Phi_1(n_1)\Phi_2'(k_2+e) > 0 \\
\text{First principal minor} &= \alpha+\beta+\gamma+\epsilon > 0 \\
\text{Second principal minor} &= \alpha(\beta+\gamma+\epsilon)+(\delta+\zeta)(\alpha+\beta+\gamma+\epsilon) > 0 \\
\text{Determinant} &= \gamma(\delta(\epsilon+\zeta+\eta)+\zeta(\epsilon+\eta)) \\
&\quad +\alpha(\zeta(\delta+\epsilon)+\eta(\delta+\epsilon+\zeta)+\beta(\delta+\epsilon+\eta)+\gamma(\delta+\epsilon+\eta)) \\
&\quad +\beta(\zeta(\epsilon+\eta)+\delta(\epsilon+\zeta+\eta)) > 0
\end{aligned}$$

Since hessian of equation (3.19) is positive definite, first order conditions are sufficient and necessary to obtain optimal n_i and e.

C.3 Simultaneous Decision Making

Hessian of objective function in equation (3.20)

$$\begin{bmatrix} \alpha+\beta+\gamma+\epsilon & \alpha & \alpha+\beta+\gamma & \kappa_1 & \kappa_1+\beta \\ \alpha & \alpha+\delta+\zeta & \alpha+\delta & 0 & 0 \\ \alpha+\beta+\gamma & \alpha+\delta & \alpha+\beta+\gamma+\delta+\eta & 0 & \eta+\beta \\ \kappa_1 & 0 & 0 & \kappa_2+\phi_1 & \kappa_2 \\ \kappa_1+\beta & 0 & \eta+\beta & \kappa_2 & \kappa_2+\beta+\eta+\phi_2 \end{bmatrix}$$

$$
\begin{aligned}
\alpha &= s_2 \int_{n_1}^{n_1+e} \Phi_2'(n_1+n_2+e-d_1)d\Phi_1 > 0 \\
\beta &= (s_1-s_2) \int_{n_1}^{n_1+e} \Phi_2(k_2+n_1+e-d_1)\Phi_1 > 0 \\
\gamma &= \Phi_1'(n_1+e)((s_1-s_2)\Phi_2(k_2)+s_2\Phi_2(n_2) > 0 \\
\delta &= s_2\Phi_1(n_1)\Phi_2'(n_2+e) > 0 \\
\epsilon &= s_1\Phi_1'(n_1)(\Phi_2(k_1+k_2-n-1)-\Phi_2(k_2+e)) + s_2\Phi_1'(n_2)(\Phi_2(k_2+e)-\Phi(n_2+e)) > 0 \\
\zeta &= s_2\Phi_2'(n_2)(1-\Phi_1(n_1+e) > 0 \\
\eta &= (s_1-s_2)\Phi_1(n_1)\Phi_2'(k_2+e) > 0 \\
\kappa_1 &= s_1 \int_0^{n_1} \Phi_2'(k_1+k_2-d_1)d\Phi_1 > 0 \\
\kappa_2 &= (p_2-s_1) \int_0^{k_1} \Phi_2'(k_1+k_2-d_1)d\Phi_1 > 0 \\
\Phi_1 &= (p_1-p_2)\Phi_1'(k_1)+(p_2-s_1)\Phi_2(k_2)\Phi_1(k_1) > 0 \\
\Phi_1 &= (s_1-s_2)\Phi_2'(k_2)(\Phi_1(k_1)-\Phi_1(n_1+e))+(p_2-s_2)\Phi_2'(k_2)(1-\Phi_1(k_1) > 0
\end{aligned}
$$

REFERENCES

Abernathy, J., Baloff, N., Hershey, J. C. and Wandel, S. (1973). A three-stage manpower planning and scheduling model - a service sector example. *Operations Research* **21**, 693–711.

Agnihothri, S. R. and Mishra, A. K. (2004). Cross-training decisions in field services with three job types and server-job mismatch. *Decision Sciences* **35**, 239–257.

Agnihothri, S. R., Mishra, A. K. and Simmons, D. (2003). Workforce cross-training decisions in field service systems with two job types. *Journal of the Operationsl Reserach Society* **54**, 410–418.

Aiken, L. H., Clarke, S. P., Sloane, D. M., Sochalski, J. and Silber, J. H. (2002). Hospital nurse staffing and patient mortality, nurse burnout, and job dissatisfaction. *Journal of the American Medical Association* **288**, 1987–1993.

Altimier, L. B. and Sanders, J. M. (1999). Cross-training in 3D. *Nursing Management* **30**, 59–62.

Baker, L., Phibbs, C., Guarino, C., Supina, D. and Reynolds, J. (2004). Within-year variation in hospital utilization and its implications for hospital costs. *Journal of Health Economic* **23**, 191–211.

Behner, K., Fogg, L., Fournier, L., Frankenbach, J. and Robertson, S. (1990). Nursing resource management: Analyzing the relationship between costs and quality in staffing decisions. *Health care management review* **15**, 63–71.

Bergman, R. (1994). The key is cross-training. *Hospitals and Health Networks* **68**.

Bertagnoli, L. (2004). Cross-training staff. *Restaurants and Institutions* .

Birge, J. R. and Louveaux, F. (1997). *Introduction to Stochastic Programming*. Springer.

Bish, E. K., Muriel, A. and Biller, S. (2005). Managing flexible capacity in a make-to-order environment. *Management Science* **51**, 167–180.

Bish, E. K. and Wang, Q. (2004). Optimal investment strategies for flexible resources, considering pricing and correlated demands. *Operations Research* **52**, 954–964.

Blegen, M. A., Goode, C. J. and Reed, L. (1998). Nurse staffing and patient outcomes. *Nursing Research* **47**, 43–50.

Blume, C. P. and Simon, L. (1994). *Mathematics for Economists*. W. W. Norton and Company.

Brusco, M. J. and Johns, T. R. (1998). Staffing a multi-skilled workforce with varying levels of productivity: An analysis of cross-training policies. *Decision Sciences* **29**, 499–515.

Brusco, M. J., Johns, T. R. and Reed, J. H. (1998). Cross-utilization of a two-skilled workforce. *International Jouranl of Operations and Production Management* **18**, 555–564.

Campbell, G. M. (1999). Cross-utilization of workers whose capabilities differ. *Management Science* **45**, 722–732.

Cattani, K., Ferrer, G. and Gilland, W. (2003). Simultaneous production of market-specific and global products: A two-stage stochastic program with additional demand after recourse. *Naval Research Logistics* **50**, 438–461.

Chakravarthy, S. R. and Agnihothri, S. R. (2005). Optimal workforce mix in service systems with two types of customers. *Production and Operations Management* **14**, 218–231.

Chod, J. and Rudi, N. (2005). Resource flexibility with responsive pricing. *Operations Research* **53**, 532–548.

Chod, J., Rudi, N. and Van Mieghem, J. A. (2006). Mix, time and volume flexibility: Valuation and corporate diversification. Working paper.

Curtin, L. L. (1995). Nursing productivity from data to definition. *Nursing Management* **26**, 25–36.

Czaplinski, C. and Donna, D. (1998). The effect of staff nursing on length of stay and mortality. *Medical Care* **36**, 1626–1638.

Dara, S. I., Ashton, R. W. and Farmer, C. (2005). Engendering enthusisam for sustainable disaster critical care response:why this is of consequence to critical care professionals. *Critical care* **9**, 125–127.

Dela Cruz, T. (2003). Hotel use many strategies to maintain brand standards. *Hotel and Motel Management* .

DHS (2004). *Finding of emergency R-01-04E*. California Department of Health Service.

Drucker, P. (1973). *Management: Tasks, Responsibilities and Practices*. Harper and Row.

Drucker, P. F. (1999). Knowledge-worker productivity: The biggest challenge. *California Management Review* **41**, 79–94.

Eastaugh, S. R. (2002). Hospital nurse productivity. *Journal of Health Care Finance* **29**, 14–22.

Edwardson, S. (1985). Measuring nursing productivity. *Nursing Economics* **3**.

Fine, C. H. and Freund, R. M. (1990). Optimal investments in product-flexible manufacturing capacity. *Management Science* **36**, 449–466.

Goyal, M. and Netessine, S. (2005). Capacity investment and the interplay between volume flexibility and product flexibility. Working paper.

Graves, S. C. and Tomlin, B. T. (2003). Process flexibility of supply chains. *Management Science* **49**, 907–919.

Hall, L. M., Doran, D. and Pink, G. H. (2004). Nurse staffing models, nursing hours, and patient safety outcomes. *Journal of Nursing Administration* **34**, 41–45.

Harrison, M. and Van Mieghem, J. A. (1999). Multi-resource investment strategies: Operational hedging under demand uncertainty. *European Journal of Operational Research* **113**, 17–29.

Hassmiller, S. B. and Cozine, M. (2006). Addressing the nurse shortage to improve the quality of patient care. *Health Affairs* **25**, 268–274.

Holcomb, B. R., Hoffart, N. and Fox, M. H. (2002). Defining and measuring nursing productivity: A concept analysis and pilot study. *Journal of Advanced Nursing* **38**, 378–386.

Hopp, W. J., Tekin, E. and Van Oyen, M. P. (2004). Benefits of skill chaining in serial produciton lines with cross-trained workers. *Management Science* **50**, 83–98.

Inman, R. R., Blumenfeld, D. E. and Ko, A. (2005). Cross-training hospital nurses to reduce staffing costs. *Health Care Management Review* **30**, 116–125.

Inman, R. R., Jordan, W. C. and Blumenfeld, D. E. (2004). Chained cross-training of assembly line workers. *International Journal of Production Research* **42**, 1899–1910.

Iravani, S. M., Van Oyen, M. P. and Sims, K. T. (2005). Structural flexibility: A new perspective on the design of manufactuirng and service operations. *Management Science* **51**, 151–166.

Jack, E. P. and Powers, T. L. (2004). Volume flexible strategies in health services: A research framework. *Production and Operations Management* **13**, 230–244.

Janssen, O. and Van Yperen, N. W. (2004). Employees' goal orientations, the quality of leader-member exchange and the outcomes of job performance and job satisfaction. *Academy of Management Journal* **47**, 368–384.

JCAHO (1996). *Accreditation Manual for Hospitals*. Joint Commission on Accreditation of Healthcare Organizations.

Jordan, S. D. (1994). Nursing productivity in rural hospitals. *Nursing Management* **25**, 58–62.

Jordan, W. C. and Graves, S. C. (1995). Principles on the benefits of manufacturing process flexibility. *Management Science* **41**, 577–594.

Jordan, W. C., Inman, R. R. and Blumenfeld, D. E. (2004). Chained cross-training of workers for robust performance. *IIE Transactions* **36**, 953–967.

Li, L. L. X. and King, B. E. (1999). A health care staff decision model considering the effects of cross-training. *Health Care Management Science* **2**, 53–61.

Loke, J. C. F. (2001). Leadership behaviors: Effects of job satisfaction, productivity and organizational commitment. *Journal of Nursing Management* **9**, 191–204.

Longest, B. B. (1977). Productivity in the provision of hospital services: A challenge to the management community. *The Academy of Management Review* **2**, 475–483.

Lyons, R. (1992). Cross-training: A richer staff for leaner budgets. *Nursing Management* **23**, 43–44.

Maggard, M. J. and Globerson, S. (1986). Employee cross-training. *Training and Development Journal* pages 10–12.

May, J., Bazzoli, G. and Gerland, A. (2006). Hopitals' response to nurse staffing shortages. *Health Affairs* **25**, w316–w323.

McClain, J. O., Schultz, K. L. and Thomas, L. J. (2000). Management of worksharing systems. *Manufacturing and Service Operations Management* **2**, 32–48.

McCreery, J. K., Krajewski, L. J., Leong, G. K. and Ward, P. T. (2004). Perfromance implications of assembly work teams. *Journal of Operations Management* **22**, 387–412.

McCue, M., Mark, B. A. and Harless, D. W. (2003). Nurse staffing, quality and financial performance. *Journal of Health Care Finance* **29**, 54–76.

McHugh, M. (1997). Cost-effectiveness of clustered unit vs unclustered nurse floating. *Nursing Economics* **15**, 294–300.

McNeese-Smith, D. (1996). Increasing employee productivity, job satisfaction, and organizational commitment. *Hospital and Health Services Administration* **41**, 160–175.

McNeese-Smith, D. (1997). The influence of manager behavior on nurses' job satisfaction, productivity , and commitment. *Journal of Nursing Administration* **27**, 47–55.

McNeese-Smith, D. (1999). The relationship between managerial motivation, leadership, nurse outcomes and patient satisfaction. *Journal of Organizational Behavior* **20**, 243–259.

McNeese-Smith, D. (2001). Staff nurse views of their productivity and nonproductivity. *Health Care Management Review* **26**, 7–19.

Moody, R. C. (2004). Nurse productivity measures for the 21st century. *Health Care Management Review* **29**, 98–106.

Needleman, J., Buerhaus, P., Mattke, A., Stewart, M. and Zelevinsky, K. (2002). Nurse-staffing levels and the quality of care in hospitals. *New England Journal of Medicine* **346**, 1715–1722.

Netessine, S., Dobson, G. and Shumsky, R. A. (2002). Flexible service capacity: Optimal investment and the impact of demand correlation. *Operations Research* **50**, 375–388.

Nicholls, D. J., Duplaga, E. A. and Meyer, L. M. (1996). Nurses' attitude about floating. *Nursing Management* **27**.

Parasuraman, A., Zeithaml, V. A. and Berry, L. L. (1985). A conceptual model of service quality and its implications for future research. *Journal of Marketing* pages 41–50.

Parasuraman, A., Zeithaml, V. A. and Berry, L. L. (1988). Servqual: A multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing* **64**, 12–40.

Pinker, E. J. and Shumsky, R. A. (2000). The efficiency-quality trade-off of cross-trained workers. *Manufacturing and Service Operations Management* **2**, 49–67.

Ramirez, Y. W. and Nembhard, D. A. (2004). Measuring knowledge worker productivity: A taxonomy. *Journal of Intellectual Capital* **5**, 602–628.

Rao, U. S., Swaminathan, J. M. and Zhang, J. (2004). Multi-product inventory planning with downward substitution, stochastic demand and setup costs. *IIE Transactions* **36**, 59–71.

Ray, P. K. and Sahu, S. (1989). The measurement and evaluation of white-collar productivity. *International Journal of Operations and Production Management* **9**, 28–46.

Robertson, R., Dowd, S. and Hassan, M. (1997). Skill-specific staffing intensity and the cost of hospital care. *Health care management review* **22**, 61–71.

Schroeder, R. G., Anderson, J. C. and Scudder, G. D. (1985). Measurement of white collar productivity. *International Journal of Operations and Production Management* **5**, 25–54.

Schultz, K. L., McClain, J. O. and Thomas, L. J. (2003). Overcoming the dark side of worker flexibility. *Journal of Operations Management* **21**, 81–92.

Shoemaker, P. and Howell, D. (2005). Trends in the use of contract labor among hospitals. Technical report, American Hospital Directory.

Siferd, S. P. and Benton, W. (1992). Workforce staffing and scheduling: Hospital nursing and specific models. *European Journal of Operational Research* **60**, 233–246.

Snyder, J. and Nethersole-Chong, D. (1999). Is cross-training medical/surgical RN's to ICU the answer? *Nursing Management* **30**, 58–60.

Spitzer, R. (1986). *Nursing Productivity: The Hospital's Key to Survival and Profit*. S-N Publications, Inc, first edition.

Strasen, L. (1987). Standard costing/productivity model for nursing. *Nursing Economics* **5**, 158–161.

Stratman, J. K., Roth, A. V. and Gilland, W. G. (2004). The deployment of temporary production workers in assembly operations: A case study of the hidden costs of learning and forgetting. *Journal of Operations Management* **21**, 689–707.

Sutermeister, R. A. (1976). *People and Productivity*. McGraw-Hill Book Company, third edition.

Tekin, E., Hopp, W. J. and Van Oyen, M. P. (2004). Pooling strategies for call center agent cross-training. Working paper.

Upton, D. (1994). The management of manufacturing flexibility. *California Management Review* **36**, 72–89.

Vairaktarakis, G. and Winch, J. K. (1999). Worker cross-training in paced assembly lines. *Manufacturing and Service Operations Management* **1**, 112–131.

Van Mieghem, J. A. (1998). Investment strategies for flexible resources. *Management Science* **44**, 1071–1078.

Weech-Maldonado, R., Meret-Hanke, L., Neff, M. C. and Mor, V. (2005). Nurse staffing patterns and quality of care in nursing homes. *Health Care Management Review* **29**, 107–116.

Wheaton, M. (1996). Cross training: Meeting staffing needs in ICU. *Health Care Management Review* **27**.

Williams, L. J. and Anderson, S. E. (1991). Job satisfaction and organizational commitment as predictors of organizational citizenship and in-role behaviors. *Journal of Management* **17**, 601–617.

Wright, D., Bretthauer, K. and Cote, M. (2006). Reexamining the nurse scheduling problem: Staffing ratios and nursing shortages. *Decision Sciences* **37**, 39–70.