RISK-AVERSE APPOINTMENT SCHEDULING FOR HEALTHCARE INSTITUTIONS
(SAĞLIK KURUMLARI İÇİN RİSKE DUYARLI RANDEVU PLANLAMASI)

by

NAZMİ ŞENER, B.S.

Thesis

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE

in

INDUSTRIAL ENGINEERING

in the

INSTITUTE OF SCIENCE AND ENGINEERING

of

GALATASARAY UNIVERSITY

January 2014

RISK-AVERSE APPOINTMENT SCHEDULING FOR HEALTHCARE INSTITUTIONS

(SAĞLIK KURUMLARI İÇİN RİSKE DUYARLI RANDEVU PLANLAMASI)

by

NAZMİ ŞENER, B.S.

Thesis

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE

Date of Submission : January 3, 2014

Date of Defense Examination : January 15, 2014

Supervisor	:	Assoc. Prof. Dr. Orhan FEYZİOĞLU
Committee Members	:	Assist. Prof. Dr. S. Şebnem AHISKA
		Assist. Prof. Dr. Hasan H. TURAN

ACKNOWLEDGMENTS

Firstly, I am grateful to my advisor Professor Feyzioğlu for his invaluable support and encouragements. Without his recommendations, this thesis would never be appeared. During my study, he has also endured my unnecessary problems. I am also grateful for this.

I want also thank my colleague and my spiritual brother, Dr. Hasan H. Turan, for our relaxing time and academic talks. I thank to my cousin Dr. Ömer Özçağlayan for supporting us with the data.

Last but not least, I want to thank my family, my father Taner, my mother and my first teacher Ümmühan, my sister Nazlı and my brother Enes. They are my source of happiness. Even when they don't understand what I do, they encourage me.

I want to dedicate this thesis to my beloved grandfathers Selahattin Tanrıtanır and Nazmi Şener.

Nazmi ŞENER
January 3rd, 2014

TABLE OF CONTENTS

ACKNOWLEDGEMENTS ii
TABLE OF CONTENTS iii
LIST OF FIGURES v
LIST OF TABLES vi
ABSTRACT vii
RÉSUMÉ viii
ÖZET ix

1 INTRODUCTION 1
2 LITERATURE REVIEW 3
2.1 Arrival Characteristics of Patients 5
2.1.1 No-show 5
2.1.2 Walk ins 6
2.2 Solution Methods for ASP 7
2.2.1 Queuing Studies 7
2.2.2 Simulation Studies 8
2.2.3 Markov Decision Process Studies 9
2.2.4 Exact and Heuristic Studies 10
3 MODEL FORMULATION 13
3.1 Stochastic Programming 13
3.2 Measuring Risks 14
3.2.1 Value at Risk 16
3.2.2 Conditional Value at Risk 16
3.2.3 Coherent Risk Measure 17

3.3 Risk Averse Optimization 18
3.4 ASP Formulation 20
3.4.1 Risk Neutral Formulation 20
3.4.2 Risk Averse Formulation 23
3.4.3 ASP in the Presence of Walk-in Patients 24
4 SOLUTION METHODOLOGY 26
4.1 L-Shaped Method 26
4.2 Solution Methodology of RNASP 29
4.3 Solution Methodology of RAASP 30
5 COMPUTATIONAL STUDIES 33
5.1 Generation of the Problem Instances 33
5.2 Results for RNASP 34
5.3 Results for RAASP 38
5.4 RAWASP Results 42
6 CONCLUSION 45

REFERENCES 46
BIOGRAPHICAL SKETCH 50

LIST OF FIGURES

Figure 3.1 CVaR vs. VaR 17

Figure 3.2 Single Doctor Scheduling with Stochastic Examination Times 21

Figure 5.1 Optimal solution of RNASP ($\gamma = c^o/c^w, Z_i = \mathcal{N}(7, 0.07), d = 70$) . . 35

Figure 5.2 Optimal solution of RNASP for different C_v levels 36

Figure 5.3 Changes in RD1 and RD2 for C_v and γ 39

Figure 5.4 Optimal solution of RAASP for different C_v levels 40

Figure 5.5 Optimal solution of RAWASP 44

LIST OF TABLES

Table 3.1 Notation of the Risk Neutral ASP 21

Table 5.1 Optimal solution of RNASP ($\gamma = c^o/c^w, Z_i \sim \mathcal{N}(7, 0.07), d = 70$) . . . 34

Table 5.2 Optimal solution of RNASP ($Z_i \sim \mathcal{N}(7, 0.07), d = 70$) 37

Table 5.3 Comparison of Relative Differences 38

Table 5.4 Optimal solution of RAASP ($Z_i \sim \mathcal{N}(7, 0.07), d = 70$) 41

Table 5.5 RD1 values with respect to model parameters 42

Table 5.6 RD2 values with respect to model parameters 43

ABSTRACT

In line with the consistent rise in health expenses in the last decades, operations research based decision-making has gained an important place in the healthcare management. Appointment scheduling is a widely addressed research area for the healthcare management as its performance is very influential in enhancing on-site services offered and reduces operational costs. In this thesis, we develop two different two-stage risk-averse stochastic programming models to solve the appointment scheduling problem for diagnostic/treatment clinics while considering different sources of uncertainty. The no-show and waiting times of the patients and overtime working conditions of the doctors are included in the formulation of both models. The first model is total cost reduction oriented while the scope of the second model is raising revenue with the addition of walk-in patients. We characterize the random parameters by finite sets of scenarios and use conditional value-at-risk measure to control the possible large realizations of random outcomes. We obtain the optimum appointment times by the variants of the L-Shaped algorithm developed. We also conduct a computational study to illustrate the effectiveness of the proposed modelling approaches.

Keywords : Appointment Scheduling Problem, Stochastic Programming, Healthcare Management, Risk Aversion, Price Discrimination

RÉSUMÉ

Conformément à l'augmentation constante des dépenses de santé dans les dernières décennies, les décisions basées sur la recherche opérationnelle ont gagné une place importante dans la gestion des services de santé. La programmation des rendez-vous est un sujet de recherche très répandu pour la gestion de la santé, comme la performance des systèmes de rendez-vous est très influente pour l'amélioration des services offerts sur place et pour réduire les couts opérationnels. Dans cette thèse, nous avons développé deux différents modèles de programmation en deux étapes afin de résoudre le problème du programme de rendez-vous pour les cliniques de diagnostic/traitement en prenant compte les différentes sources d'incertitude et en limitant les risques. La non-présence et la durée d'attente des malades et les heures supplémentaires des docteurs sont inclues à la formulation des deux modèles. Le premier modèle vise la réduction du coût total tandis que le deuxième modèle est en mesure d'augmenter les revenues en considérant les malades qui arrivent à la dernière minute. Nous caractérisons les paramètres aléatoires par des ensembles dénombrables de scénarios et utilisons la mesure de valeur-à-risque conditionnelle afin de contrôler la possibilité des grandes pertes selon les résultats aléatoires. Nous avons obtenu les temps de rendez-vous optimaux en utilisant les variantes de l'algorithme L-Shaped développés. Nous avons aussi effectué une étude computationnelle afin de démontrer l'efficacité des approches de modélisation proposées.

Mots Clés : Problème de Rendez-vous, Programmation Stochastique, Gestion des Soins de Santé, Aversion au Risque, Discrimination de Prix

ÖZET

Son yıllardaki sağlık harcamalarındaki sürekli artışla birlikte, yöneylem araştırması temelli karar verme sağlık kurumları yönetiminde önemli bir yer kazanmaktadır. Randevu çizelgeleme, performansını önerilen hizmetlerin yerinde arttırılmasında ve işletme maliyetinin azaltımında fazlaca etkili olarak, sağlık kurumları yönetimi için çokça değinilen bir araştırma konusudur. Bu tezde farklı sebeplerden kaynaklanan belirsizlikleri dikkate alan, görüntüleme/tanı klinikleri için oluşturulan randevu çizelgeleme problemini çözebilmek adına iki farklı iki aşamalı riskten kaçınan rassal programlama modelleri geliştirdik. Hastanın gelmemesi ve hastanın beklemesinin maliyeti ve doktorun fazla mesai koşulları her iki modelin kurulumunda da dikkate alındı. İlk model geliri dikkate almadan toplam maliyet minimizasyonu temelli iken, ikinci model gelmeyen hastaları dikkate alarak karı arttırır. Sınırlı sayıdaki senaryo için rasgele parametreleri nitelendirdik ve rastgele çıktıların gerçeklenmesindeki olası büyük senaryoları kontrol edebilmek için koşullu risk değeri ölçütünü kullandık. L şekilli algoritmanın geliştirilmiş varyantları kullanarak en iyi randevu zamanlarını elde ettik. Ayrıca önerilen modelleme yöntemlerinin etkinliğini göstermek için bir sayısal çalışma yaptık.

Anahtar Kelimeler : Randevu Çizelgeleme Problemi, Rassal Programlama, Sağlık Yönetimi, Riskten Kaçınma, Ücret Ayrıştırma

1 INTRODUCTION

Appointment scheduling problem (ASP) is a well known problem for healthcare management. The management aims to fulfill patients demand with proper time arrangements. For the arrangement, the managers want to minimize waiting times of patients and idle time and overtime of doctors. There are differences between well-known service system queues and ASP queues. First, the capacity of the queue is limited in ASP, which means that only limited number of patients take an appointment in a session. Secondly, the session length is finite, which means that the length of queue is prearranged.

Within ASP context, healthcare managements face with several complications to find the optimum solution. These problems are no show of patients, walk in patients, interruption of services, uncertainty of examination times, etc. Stochastic examination times are studied in the literature (Mercer, 1960; Ho and Lau, 1992). Stochastic examination times occur due to the several environmental factors depending on the industry and the nature of service being processed. Examination times in outpatient procurement center are usually extremely uncertain. This uncertainty is related to the diagnosis type, patient's age, information acquired from patients, etc. Hereby, basic assumptions are not appropriate for this problem.

Another complicating factor is the uncertainty of the patient's arrival time. This problem has two parts. The first one is no-show, the absence of the patient during the scheduled time. This causes idle time for the doctor and deteriorate the optimal schedule. The second one is walk-in patients. Walk-in patients cause the interruption of the service and deterioration of the schedule.

Risk is an important issue for the decision makers. The general modelling approach for ASP is risk neutral. But changing environment conditions (for example, no show of patients, stochastic consultation times etc.) impose to use a risk averse modelling approach. In risk averse modelling, choosing the proper risk measure is crucial. Here we adopt conditional value-at risk as the risk measure and propose a model, which minimizes this risk measure.

This thesis is organized as follows. In Chapter 2, we provide a brief literature review about ASP. We also classify modelling methodologies under four titles : queuing models, simulation models, Markov decision process models, exact and heuristic models. In Chapter 3, we formulate the proposed models and describe risk measures. In this chapter, we propose three mathematical models : risk neutral ASP, risk averse ASP, risk averse walk-in ASP. In Chapter 4, the well-known two stage stochastic programming solution algorithm, namely L-Shaped algorithm, and its variant are introduced. In this chapter, properties of ASP are also described. In Chapter 5, insights gained from our models are presented. We also present a sensitivity analysis related the proposed models. In Chapter 6, we summarize the fundamental findings of the thesis and suggest future works.

2 LITERATURE REVIEW

In this chapter, we emphasize appointment scheduling literature from modelling issues and operations research point. We classify the literature in two ways : arrival characteristics of the patients and solution methodologies. "No show" and "walk in" concepts are examined to understand patient's arrival characteristics. Solution methodologies are grouped into four : Queuing studies, Simulation Studies, Markov decision processes studies and exact or/and heuristic studies.

For modern societies, healthcare expenditures have always an increasing trend. The average public health expenditures of OECD countries is expected to increase from 5.5% of GDP to 8% of GDP from 2010 to 2060 (Maisonneuve and Martins, 2013). This significant statistic clearly indicates that countries' healthcare budgets must be under control. Maisonneuve and Martins (2013) point out that outpatient expenditures constitute the second largest expenditure item for the healthcare system. In this context, different problems are investigated for increasing profits and keeping systems sustainable, such as appointment scheduling problem, staff scheduling problem, healthcare supply chain management, diagnosis improvement, decision support systems for consultations, etc.

Staff scheduling has significant effects on a healthcare facility's productivity. Naidu et al. (2000) focus on different methodologies such as linear optimization, expert systems and heuristic methods for discovering an optimal staff schedule that meets the requirement at the lowest cost. Cai and Li (2000) define a multi-criteria model with three objectives important in order to schedule homogeneous skilled staff. In their model, they aim to maximize excess staff and minimize the variation of excess staff.

Appointment scheduling problem (ASP) aims to fulfill patients expectations and increase profits (Gupta and Denton, 2008; Green and Savin, 2008). The aim of appointment scheduling problem is to assign the start time of each patients' examination. ASP objectives generally consist of improving the resource utilization and reducing waiting time of patients (Cayirli and Veral, 2003). In ASP, practitioners face with several complicated situations, such as uncertain consultation times, no show of patients, non-punctuality of patient, etc. For these types of problems, researchers generally construct a model to increase the performance of clinics by including expected patient waiting time, server idle time, and overtime (Erdogan and Denton, 2013).

There is only a few literature survey articles about ASP for the outpatient clinics. Cayirli and Veral (2003) present a review on the classification of outpatient appointment scheduling methodologies in the context of outpatient services. A problem which includes the complicating environmental factors such as the number of different services and physicians, uncertainty in service durations, punctuality of patients, no-shows, walk-ins etc is posed. They expose diverse performance criteria including total cost of waiting time of patients and overtime and idle time of service providers, diverse time based, congestion based, productivity based performance criteria in the previous appointment scheduling literature. An elaborated classification of methodologies which are used in the literature is provided.

Gupta and Denton (2008) classify characteristics of ASP in primary care, speciality clinics and surgery clinics. They contend that indirect patient waiting time, which is the waiting time between the appointment requested by the patient and the scheduled appointment date, might be considered in addition to patient waiting time on the day of the service. They also emphasize a number of open research areas for ASP including the effect of late cancellations and no-shows, patient preferences (such as doctor choice, slot choice, etc.), and resource allocation depending on different patient types.

Different appointment rules are examined in the literature. Appointment rules are classified into seven different classes. First rule appoints see patients to a single block. It means that patient appointments only include dates without specific appointment time. Second rule appoints patients individual appointment times with fixed intervals to each patients. In this rule, one patient is assigned to a slot. Main difference between second and third rules is the first slot capacity, which means that in third rule, there is more than one patients in first slot. Fourth rule explains that a predetermined number of the patients (more than one patient) is assigned to one slot with fixed time length.

Fifth rule is a combination of fourth rule with differentiated initial block. Sixth rule differentiates block sizes during all sessions with fixed appointment intervals. Last rule comprises individual appointments to each patient with varying appointment intervals (Cayirli and Veral, 2003). In this thesis, second rule is applied.

Consultation times depend on stage of the healthcare facilities. According to Gupta and Denton (2008), there are three types of consultation durations : constant, diagnosis dependent and random. Constant consultation durations generally occur in primary care, while diagnosis dependent consultation durations are frequently observed in primary care and speciality clinics. Random consultation durations occur in surgeries and hospital stays. Bailey (1952) reports that a little change in consultation durations cause huge effects on the appointment system's performance. Coefficient of variation ($C_v = \frac{\sigma}{\mu}$), is a measure of variation of consultation time. Greater coefficient of variance for consultation duration causes patients' waiting times and doctor's idle times to increase (Bailey, 1952; Vissers and Wijngaard, 1979; Ho and Lau, 1992; Denton and Gupta, 2003).

2.1 Arrival Characteristics of Patients

The arrival characteristics of patients affect appointment system's performance, which include no shows and walk-ins. No show is the absence of an appointed patient. No show increases idle time of the doctors. A walk-in patient does not have a previously scheduled appointment and/or is a urgent patient. Walk in patients deteriorate the schedule and cause interruption of the ongoing examination.

2.1.1 No-show

Both doctors and system analysts want to mitigate effects of no-show (Cayirli and Veral, 2003; Jonas, 1971; Schroeder, 1973). According to Ho and Lau (1992), among no show, service time variability and number of patients per session, no show has significant effects on a healthcare facility's performance. No show rates given in the literature range from 3% to 30% (Cayirli and Veral, 2003; Rust et al., 1995; Deyo and Inui, 1980). These rates reveal that source utilization is violated by no shows. Till 2003, nearly 30% of the articles take into account no show (Cayirli and Veral, 2003).

Fetter and Thompson (1966) conducted a research about impact of walk-ins response to no shows. They also indicate that walk-in patients increase the waiting time of patients and decrease idle time of the doctor. Because of this, walk in concept mitigates effect of the no show. Vissers and Wijngaard (1979) arranged the service times with mean and variance to retrieve no shows and walk ins. They show that the effect of no show and walk ins can be found by their influence on mean consultation time and coefficient of variation.

Kaandorp and Koole (2007) conducted their research with homogeneous no show probabilities. They assume that no show behaviour of the patients are independent from each other. They find that if no show probability becomes larger, the mean waiting time of the patients and the idle time of the doctor and tardiness of consultation become larger.

Green and Savin (2008) have explained the state-dependent no show rate with a function of queue length. Their model include rescheduling of no shows with a probability. Zeng et al. (2010) extended no show probabilities with heterogeneous. They compare the model with homogeneous probability and the model with heterogeneous probability. Turkcan et al. (2011) have used patient type dependent no show probabilities. They also indicate that the performance measures are sensitive to the no show prediction error.

2.1.2 Walk ins

Walk-in is generally ignored in literature especially in analytical papers (Cayirli and Veral, 2003). However, administrations of clinics take into account walk-ins in clinic schedules (Cayirli and Veral, 2003). Walter (1973) discovers that if percentage of walk in patients drops, efficiency of system increases (so there is reduction in doctor's idle time and patient's waiting time). Vissers and Wijngaard (1979) explain the effect of walk-ins on the mean and variance of consultation times. In Vissers and Wijngaard (1979)'s paper, the effect of walk ins can be found in a similar way by interpreting the fraction of walk-ins as the probability that patients need a revised consultation time that is equal to the sum of two consultation times of two different clinics.

In outpatient studies , emergency walk-in is a less addressed issue than overall walk in studies. Preemptive emergency walk ins interrupt consultation of the last patient. This interruption makes harder to model the system. Due to this, Fetter and Thompson (1966) prefer to model "non" preemptive emergency walk-ins in their simulation model (Cayirli and Veral, 2003). Fetter and Thompson (1966) assign probabilities to arrivals in order to solve the walk-in problem. They consider that different specialities have different walk in probabilities.

2.2 Solution Methods for ASP

Different solution techniques have been used to solve ASP. In literature, there are many articles involving ASP that are queueing theory, discrete event simulation models or exact or heuristic solution algorithms due to the difficulty of finding an analytical solution for ASP including more than two customers (Erdogan and Denton, 2013).

2.2.1 Queuing Studies

Queueing articles commonly include restrictive assumptions, such as equal intervals for appointments, independent consultation times, and infinite number of patients (Mercer, 1960, 1973; Jansson, 1966). Mercer (1960, 1973) introduces a queuing system and analyzes its performance of when patients arrive late or no-show situation occurs. In his study, the nonequilibrium distribution of the queue length is analyzed and also the results for the equilibrium distribution are discussed.

Jansson (1966) use simulation to observe optimal interarrival times of the patients. A queuing system with time dependent Markovian arrival rate and discrete examination duration distributions (M(t)/G/s queue) are developed with a finite number of patients at any time assumption (Brahimi and Worthington, 1991). The results of the model are used to indicate that there resemble to be remarkable scope for reducing patients' waiting times without increasing doctors' idle times.

Wang (1993) uses phase-type distributions to observe the transient solution of a Markovian server with a general arrival distribution (S(n)/M/1 queue) for establishing patients start times. Vanden Bosch and Dietz (2000) develop a queueing model with deterministic arrivals and no-shows. They group patients to different groups depending of their service durations. They develop an algorithm based on the special structure of their model to find the optimal solution and the sequence of patients efficiently.

A queuing model is improved to reduce negative influence of no shows with overbooking of appointment schedules (Zeng et al., 2010). Zeng et al. (2010) use patient waiting time costs and doctor's idle time costs as objective. They prove that if patients are homogeneous (all patients have same no show probability), objective is multimodular. However heterogeneous patients' model is not multimodular. They use local search to find local optimal solution to heterogeneous patients' model. In this queuing model, overbooking is advantageous for "open access" ASP which permits patients to request appointment on the same day of consultation.

In ASP, queuing models has a few deficiencies. In most of them, it is assumed that "system reaches steady state" and this may cause regular service interruptions. On the other hand, exponential examination durations are common in the queuing literature and as the coefficient of variance of exponential distribution always equals to one, this causes lack of reflecting to variability of the ASP.

2.2.2 Simulation Studies

Discrete event simulation helps to avoid some of the restrictive assumptions adopted for queueing systems. Vissers and Wijngaard (1979) model with the first simulation model to minimize patient waiting time and specialist's idle time. They use four decision variables : mean consultation time, coefficient of variation of consultation time, mean earliness and standard deviation of patient punctuality. Ho and Lau (1992) model the ASP with simulation model to interpret different scheduling rules for diverse scheduling environments characteristics such as the probability of no-shows, number of patients to schedule, and the service distribution. They use an easily implementable technique. For example, two or more patients are scheduled at the start of the session, other patients are scheduled in intervals which equal to the mean of the consultation time's distribution. They also observe the effect of changing interval time between patients. They use Pareto optimal set of scheduling rules with respect to the expected idle time and expected patient waiting time.

Lowery (1992)'s simulation model is focused on the patient flow in a hospital including intensive care units beds, operation rooms beds and post-anaesthesia care units beds. Jerbi and Kamoun (2011), develop a simulation optimization algorithm to solve a multi objective ASP with no-show and patient walk-ins. Simulation studies generally include complex queue structures. They infer that each management satisfaction function decreases linearly with an increase in the deviations. So computational studies based on simulation need more time to analysing queues. Lu et al. (2011) develop a simulation model of information flow in the scheduling process based on the analysis of the central scheduling process. They also conduct a "what if" analysis to identify potential process improvement strategies in the form of recommendations to the hospital management.

2.2.3 Markov Decision Process Studies

Markov decision process methodology is used in ASP with finite and infinite horizon approaches. Green et al. (2006) model a diagnosis facility with a computed tomography (CT) machine's patient demand as a finite horizon Markov decision process. They aim is to maximize total profit is the difference between the includes revenue of scanned patients and the waiting and penalty costs. Penalty costs are due to not being able to scan patients by the end of the work-day. The authors assume that outpatients have the probability of no show.

Patrick et al. (2008) model a outpatient schedule system which assigns patients to future slots. The authors aim to minimize total penalty cost being assumed by the administration. The total penalty cost includes the waiting (patients had to wait longer than a maximum recommended waiting-time) and rejection costs of the patients. They use an infinite horizon Markov decision process formulation and solve it with approximate dynamic programming approach.

Kolisch and Sickinger (2008) enlarge the model proposed by Green et al. (2006) in two dimensions. First dimension is adding an identical CT machine to the system. The other dimension is the number of outpatient appointments per time-slot that can exceed the number of available scanners. The authors also investigate different arrival probabilities and cost parameters. Their study proves that applying first in first out queuing discipline is simple and has better performance.

Gocgun et al. (2011) develop a Markov decision process model for multi category patients and CT machines. They categorize emergency patients as critical and non critical, and use finite horizon Markov decision process methodology with expected net revenue. Optimal solution is compared with five heuristics : first in first out heuristic, R-1 heuristic (one patient from a randomly chosen type is scanned), R-2 heuristic (one patient is randomly scanned from the patients that are waiting), OP-1 heuristic (outpatients have the highest priority, and non-critical emergency patients have higher priority than inpatients), OP-2 heuristic (outpatients have the highest priority, and inpatients have higher priority than non-critical emergency patients.). The average net revenue gap between the optimal policy and the five heuristic decision rules ranged from 5% to 12% for various scenarios.

2.2.4 Exact and Heuristic Studies

There are some analytical models in the literature, such due to Bailey (1952) which involve two patients system. In this study, appointments are given in regular intervals. Bailey (1952) discuss the effect of variations in the appointment interval, the number of patients attending the clinics, and the distribution of queue-size. The author observe the performance of the system is very sensitive to small changes in appointment intervals that small changes. Fries and Marathe (1981) model single server variable sized multiple block system with dynamic programming approach, and explain advantages and disadvantages of the system. Weiss (1990) gives a closed form solution to obtain the optimal estimated inter arrival time between two patients. Examination times are assumed stochastic in Weiss (1990)'s model. The total cost has two parts which are doctor's waiting and idle time costs. Vanden Bosch et al. (1999) derive upper and lower bounds for the optimal cost appointment schedule. For the construction of these bounds, they use submodularity property which is indeed multimodularity on subset of the equations.

Denton and Gupta (2003) introduce a two stage stochastic linear programming (SLP) model and utilize the problem framework for developing a methodology that gives bounds on the optimal solution. In their study, interarrival times are *dome shaped* which means shorter interarrival time at the beginning of the session and near the end of session. Robinson and Chen (2003) model ASP as a stochastic program and solve it with a fast heuristic for calculating robust interarrival times.

Kaandorp and Koole (2007), use local search heuristic to solve ASP with weighted average of the expected waiting times of the patients, idle time of the doctor and tardiness as the objective. Chern et al. (2008) study on health examination scheduling problem. The authors model ASP as binary integer programming model using an analogy to sequence dependent flow shop scheduling problem. Binary integer programming problems are hard to solve and consuming much time. Because of this, they use a heuristic algorithm to solving the problem.

Muthuraman and Lawley (2008) introduce a new myopic scheduling algorithm to maximize the profit. While they improve a scheduling algorithm and reproduce optimal stopping criteria, optimal schedule is not described. Turkcan et al. (2011), introduce multi objective SLP model with two service criteria and solve the problem using Pareto optimal set. They observe fairness/revenue tradeoff and use constraint based approach.

Erdogan and Denton (2013) develop multi stage stochastic model to analyze patient flow of the surgery clinics. To solve problem efficiently, they adapt an algorithm which is commonly used to solve multi stage stochastic programming models. The authors also extend Denton and Gupta (2003)'s model with no show and extract that no show structure deteriorates dome shape structure.

In this thesis we aim to show how to risk averse and rational healthcare managers decide their optimal choices under the risk. In particular, we attract how to risk aversion of healthcare managers affect their optimal decision in ASP. Our enthusiasm in this line of the research introduce when we observe a space on the impacts of risk aversion in the ASP literature. That is, ASP literature principally assumes that healthcare managers aim to minimize the expected value of costs, but they don't take into account variations of random outcomes.

3 MODEL FORMULATION

In this chapter, basic modelling and specialities are defined. First, stochastic programming and coherent risk measure concept is discussed. After that model of ASP and basic properties of model are studied. At the end of this chapter risk averse walk-in ASP model introduced.

3.1 Stochastic Programming

Stochastic programming models allow decision makers to observe different realizations and to model risk with different approaches. Stochastic Programming or Optimization concept is first introduced by Dantzig (1955). Over decades, different approaches are developed for solving stochastic programming models. There is a widely adapted to solve stochastic problems which is replacing random variables with their mean values. This expectation approach transform stochastic programming model to deterministic problem which called *expected value* or *mean value problem*. However this deterministic structure does not take into account variation so this model is not capable to reflect all stochastic programming model's features.

A major drawback of the stochastic programming modelling is that problem size expand when number of realizations of random variables increases. Because of this incline, different solution approaches are developed. As for example, Benders' Decomposition and Dantzig-Wolfe Decomposition approaches are widely used. In these decomposition methods, the model is divided to two parts : master problem and sub problems. Master problem is the deterministic part of the model. Sub problem is realizations of stochastic part of the problem. In Benders (2005)s' decomposition, a new constraint is added to master problem regards to solutions of sub problems to proceed algorithm. Dantzig and Wolfe (1960) decomposition progresses with adding a new decision variable (column) to the master problem in relation to the solutions of sub problems.

Two stage stochastic programming models has two decision stages. First stage decisions,

$\mathbf{x} \in \mathbb{R}^n$, are determined with complete results of random variables, $\xi \in \Xi$, where Ξ denotes all possible outcomes of the random variables. Second stage decisions, $\mathbf{y} \in \mathbb{R}^m$, are the results of each scenario outcome. General formulation of two stage stochastic programming is :

$$
\begin{aligned}
z = \min c^T\mathbf{x} + \mathbb{E}_\omega \left[Q\left(x, \omega\right)\right] \\
s.t. \quad A\mathbf{x} = \mathbf{b} \\
\mathbf{x} \geq 0
\end{aligned}
\tag{3.1}
$$

where

$$
Q\left(\mathbf{x}, \omega\right) = \min \left\{ \mathbf{q}(\omega)^T \mathbf{y}(\omega) \mid \mathbf{W}(\omega)\mathbf{y}(\omega) = \mathbf{h}(\omega) - \mathbf{T}(\omega)\mathbf{x}, \mathbf{y}(\omega) \geq 0 \right\} \tag{3.2}
$$

denotes the recourse function (Birge and Louveaux, 2011). First stage parameters $(\mathbf{A}, \mathbf{b})$ are deterministic. Second stage parameters $(\mathbf{W}, \mathbf{T}, \mathbf{q}, \mathbf{h})$ are stochastic which means that the parameters are changing with respect to realizations. $\mathbb{E}_\omega$ denotes the expectation of the outcomes of all realizations.

3.2 Measuring Risks

Risk has different definitions in different contexts. Our risk definition is uncertainty in realizations' outcomes. It means that the probability of each possible random realization is given before. There are different risk approaches to random outcomes of realizations : risk averse, risk neutral, risk seeking. Risk averse approach is the most conservative approach of all. Conservativeness of risk aversion is dependent to modeller. The main aim of risk averse in optimization is lowering the effects of undesirable results of realizations without regarding that the problem is minimization or maximization problem. Risk neutral approach is based only expected value of realizations in any case of risk of possible outcomes of realizations. (3.1)-(3.2) is an example for risk neutral stochastic programming model. Risk seeking approach is comparison of the best possible outcomes of realizations from risky performance alternative and fixed outcomes of realizations performance alternative. After the comparison, best performance alternative is chosen regardless the guaranteed outcomes. Risk averse and risk neutral approaches are widely used compared to risk seeking. Healthcare management must keep facilities persistent and economically stable, so risk seeking approach is not preferable.

Risk measures are reflected differently in optimization models, such as expected utility theory, stochastic dominance theory and mean-risk models :

- *Expected Utility Theory* : Decision makers has a non decreasing utility function, $u(\cdot)$, such that random outcome X over Y is preferred if and only if $\mathbb{E}[u(X)] \leq \mathbb{E}[u(Y)]$ and the expected value of the utility function is optimized instead of the expected outcome.

- *Stochastic Dominance Theory* : This method is the generalization of expected utility theory. The distribution of a random outcome X is chosen instead of a random outcome of Y concerning a stochastic dominance relation if and only if expected utility of X is preferable instead of expected utility of Y for all utility functions in definite class, named the generator of the relation.

- *Mean–Risk Analysis* : This method measure the problem in a luminous form of two criteria : the expected value of the outcome and a scalar measure of the variability of the outcome. Decision makers ses a particularized functional $r : \chi \to \mathbb{R}$, where χ is a definite space of quantitative function on a probability space (Ω, $\mathfrak{F}$, $\mathbb{P}$) to represent variability of the random outcomes (Choi, 2009).

Mean risk analysis has many advantages : the method permits to formulate the problem as a parametric optimization problem and simplifies the trade-off analysis between the expected value of the outcome and a scalar measure of variability of the outcome. Because of this, it is applied in this thesis. With a little modification, the objective becomes $\min_{x \in X} \mathbb{E}(x, \omega) + \lambda \rho f(x, \omega)$, where λ is the positive trade off coefficient representing the exchange rate of mean cost for risk (Noyan, 2012).

3.2.1 Value at Risk

Value at risk (VaR) is a commonly used risk measure, which has the aim to determine the maximum loss of the revenue for a given confidence level. VaR is developed for aggregating various source of the risk into a unique quantitative measure. The mathematical definition of VaR is given below (Noyan, 2012) :

Definition 3.1. *Let the $F_\chi(\cdot)$ represent the cumulative distribution function (cdf) of a random variable χ. In financial studies, the VaR at the confidence level α (α quantile)*

$$\inf\{\eta \in \mathbb{R} : F_\chi(\eta) \geq \alpha\} \tag{3.3}$$

is denoted by $VaR_\alpha(\chi)$, where $\alpha \in (0, 1]$.

3.2.2 Conditional Value at Risk

When VaR is used as a risk measure, the magnitude of the losses exceeding VaR is not taken into account . Conditional value at risk (CVaR) is commonly used to be an alternative to VaR. The common names of CVaR is tail VaR or excess loss. A graphical representation of VaR and CVaR is given at 3.1. This graphic explains

difference between CVaR and VaR. Mathematical definition of CVaR at the confidence level, α, is given below (Rockafellar and Uryasev, 1969) :

$$CVaR_\alpha = \frac{1}{1-\alpha} \int_{-\infty}^{VaR_\alpha} x f_X(x) dx \tag{3.4}$$

or equivalently

$$CVaR_\alpha(\chi) = \mathbb{E}[x | x \leq VaR_\alpha]. \tag{3.5}$$

From above formulation, CVaR is the conditional expectation exceeding VaR for the confidence level of α. For the minimization problems, VaR_α is the α quantile of distribution of the sub problems' objectives. Because of this, VaR_α generates an upper bound. The bound is exceeded only with a small probability $(1 - \alpha)$ (Noyan, 2012). Open form of $CVaR_\alpha$ is given in below definition.

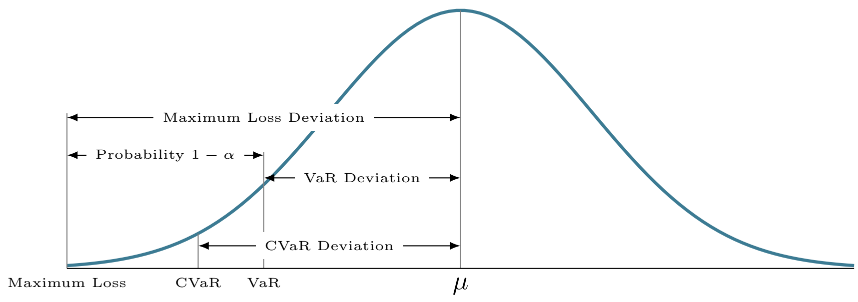


Figure 3.1: CVaR vs. VaR

Definition 3.2. *The CVaR of random variable χ at the confidence level α is equals to*

$$CVaR_{\alpha}(\chi) = \inf_{\eta \in \mathbb{R}} \{\eta + \frac{1}{1-\alpha}\mathbb{E}[\chi - \eta]^{+}\} \tag{3.6}$$

where $[z]^{+} = \max\{0, z\}, \;\; z \in \mathbb{R}$. Infinimum of (3.6) is acquired at a α quantile of χ.

3.2.3 Coherent Risk Measure

If a risk measure satisfies the desirable properties of risk measure, this risk measure is called as *coherent*. Artzner et al. (1999) propose and demonstrate that four properties is enough to describe risk measure as coherent. These properties are described below (Coherent risk measure ρ is functional and $\rho Z \to \mathbb{R}$) :

- Convexity : $\rho(tX + (1-t)Y) \leq t\rho(X) + (1-t)\rho(Y)$,
- Monotonicity : If $X, Y \in Z, X \succeq Y$ then $\rho(X) \geq \rho(Y)$,
- Translational invariance : If $t \in \mathbb{R},$ and $X \in Z,$ then $\rho(X + t) = \rho(X) + t,$
- Positive homogeneity : If $t \geq 0$ and $X \in Z$ then $\rho(tX) = t\rho(X)$.

Although VaR is widely used risk measure, it is not a coherent risk measure because of lack of convexity and subadditivity. VaR is coherent only when normal distribution is used. Optimizing VaR is difficult when it is calculated from realizations (Rockafellar and Uryasev, 1969). Moreover CVaR has better properties than VaR (Artzner et al., 1999). Pflug (2000) proves that CVaR is coherent risk measure and it has positive homogeneity, translational invariance, monotonicity, convexity specialities. In addition, CVaR is easily modelled with linear programming. Due to these reasons, we use CVaR as a risk measure.

3.3 Risk Averse Optimization

According to Noyan (2012), decision vector, $\mathbf{x}$, is efficient in the mean-risk sense if and only if for a predetermined level of expected outcome, the total cost function, $\mho(x, \omega)$, has the lowest possible level of CVaR, and for a predetermined level of CVaR which has the lowest possible level of expected outcome. The mean-risk efficient frontier by calculating the efficient solutions for different risk coefficients can be constructed. By revising the objective from above informations, new formulation of the model is given below :

$$\min_{x \in X} \{\mathbb{E}[f(x, \omega)] + \lambda CVaR_\alpha[f(x, \omega)]\}. \tag{3.7}$$

Proposition 3.1. *In a finite probability space, where $\Omega = \{\omega_1, \omega_2, ..., \omega_N\}$ with corresponding probabilities $p_1, p_2, ..., p_N$, mean-risk problem in (3.7) can be reformulated as the following problem (Noyan, 2012) :*

$$\min_{x \in X} \quad (1+\lambda)\mathbf{c}^T\mathbf{x} + \sum_{s=1}^{N} p_s[\mathbf{q}_s]^T\mathbf{y}_s + \lambda\left(\eta + \frac{1}{1-\alpha}\sum_{s=1}^{N} p_s\nu_s\right) \tag{3.8}$$

$$s.t. \quad W_s\mathbf{y}_s = \mathbf{h}_s - T_s\mathbf{x}, \qquad s = 1, ..., N,$$

$$A\mathbf{x} = \mathbf{b}, \tag{3.9}$$

$$\mathbf{y}_s \geq 0, \qquad s = 1, ..., N \tag{3.10}$$

$$\nu_s \geq (\mathbf{q}_s)^T\mathbf{y}_s - \eta, \qquad s = 1, ..., N, \tag{3.11}$$

$$\nu_s \geq 0 \qquad s = 1, ..., N, \tag{3.12}$$

$$\eta \in \mathbb{R} \tag{3.13}$$

Proof From CVaR's translational invariance speciality, Noyan (2012) prove that :

$$CVaR_\alpha(f(\mathbf{x},\omega)) = \mathbf{c}^T\mathbf{x} + CVaR_\alpha(Q(\mathbf{x},\xi(\omega)))$$

and

$$\mathbb{E}[f(\mathbf{x},\omega)] + \lambda CVaR_\alpha(f(\mathbf{x},\omega))$$

$$= \mathbf{c}^T\mathbf{x} + \mathbb{E}[Q(\mathbf{x},\xi(\omega))] + \lambda(\mathbf{c}^T\mathbf{x} + CVaR_\alpha(Q(\mathbf{x},\xi(\omega)))$$

$$= (1+\lambda)\mathbf{c}^T\mathbf{x} + \mathbb{E}[Q(\mathbf{x},\xi(\omega))] + \lambda Q(\mathbf{x},\xi(\omega))) \quad (3.14)$$

For finite number of realizations of random parameters $\xi(\omega)$, realizations of the recourse function equal to $Q(\mathbf{x},\xi_1) = q_1^T\mathbf{y}_1, ..., Q(\mathbf{x},\xi_N) = q_N^T\mathbf{y}_N$. From the special structure of the sub problem (4.4) and by the definition (3.2), the proposition proved. □

The model (3.8)-(3.13) is reformulated in master-sub problem format as below :

Master Problem :

$$\min\, (1+\lambda)\mathbf{c}^T\mathbf{x} + \theta_1 + \lambda\theta_2 \quad (3.15)$$

$$\text{s.t. } A\mathbf{x} = \mathbf{b}, \quad (3.16)$$

$$\theta_1 \geq \sum_{\ell=1}^{s} p^\ell(\mathbf{u}^{\ell t})^T(\mathbf{h}^\ell - T^\ell\mathbf{x}) \qquad t = 1,...,\tau, \quad (3.17)$$

$$\theta_2 \geq \eta^t + \frac{1}{1-\alpha}\sum_{\ell=1}^{s} p^\ell \nu^{\ell t} \qquad t = 1,...,\tau, \quad (3.18)$$

$$(T^\ell)^T\sigma^{g_\ell}\mathbf{x} - (h^\ell)^T\sigma^{g_\ell} \qquad \ell = 1,...,s, g_\ell = 1,...,G_\ell, \quad (3.19)$$

$$\nu^{\ell t} \geq (\mathbf{u}^{\ell t})^T(\mathbf{h}^\ell - T^\ell\mathbf{x}) - \eta^t, \quad \ell = 1,...,s, t = 1,...,\tau, \quad (3.20)$$

$$\nu^{\ell t} \geq 0, \eta^t \in \mathbb{R}, \qquad \ell = 1,...,s, t = 1,...,\tau. \quad (3.21)$$

Subproblems are same as (4.4)

where, η^t represent the α quantile of recourse cost and $\nu^{\ell t} = [(\mathbf{u}^{\ell t})^T(\mathbf{h}^\ell - T^\ell \mathbf{x}) - \eta^t]^+, \ t = 1, ..., \tau,$. In the risk averse model, equations (3.17) and (3.18) correspond to optimality cuts of the model. (3.17) is used in risk neutral model. (3.18) is introduced by Noyan (2012) for risk averse problems. (3.19) is feasibility cut of the model and it is also used in the risk neutral model.

3.4 ASP Formulation

In the ASP the objective is to create the best possible schedule for patients and doctors. The maximum number of patients that can be examined in a day is predetermined. Examination times differ from one patient to another. Patients can be punctual or not, or even not come to the clinic the whole appointment day (no-show). They can take their appointments at an earlier time or they can just apply within a day they want to be examined. A bad schedule make patients to wait, sometimes at very long times. Moreover, it may also cause doctors to work overtime, even be idled during the day. These events result in direct and indirect costs for the system. In this study, we assume that there is a single doctor. Denton and Gupta (2003)'s model has computational efficiency due to using two stage stochastic linear programming formulation. Erdogan and Denton (2013) revise the model by considering some propositions. We make use of both studies in our model formulation. We transform their base model formulated with risk-neutral modelling approach to a risk-averse model. After that we reformulate the model for analysing walk-in patients capacity. An instance of the problem is visualized in Figure 3.2 and the used notation is summarized in Table 3.1.

3.4.1 Risk Neutral Formulation

We aim to optimize the appointment schedule by arranging the arrival times of n patients. Doctor examination times are stochastic so examination time is assumed be a random variable. Objective function minimizes waiting time cost of patients and overtime cost of the doctors relating to session length. We assume that our patients

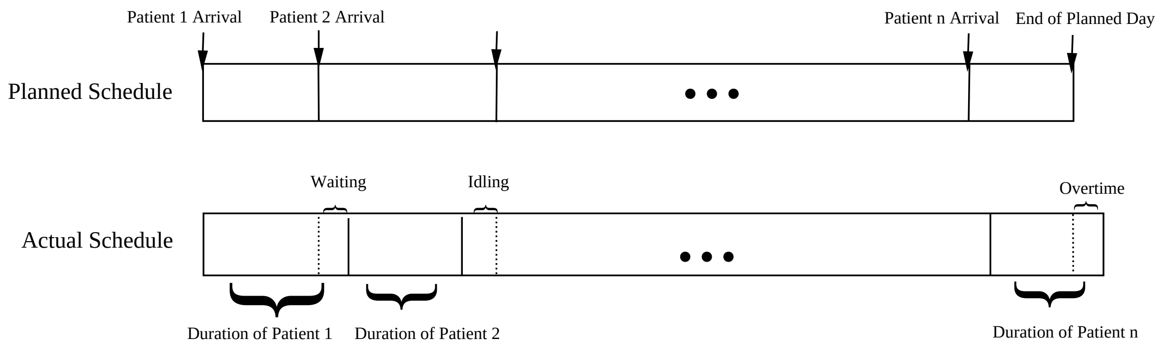


Figure 3.2: Single Doctor Scheduling with Stochastic Examination Times

Table 3.1: Notation of the Risk Neutral ASP

Parameters	
n	Number of appointed patients
ξ	Index for realizations
d	Length of session which we schedule
c_o	Coefficient of overtime cost
c_w	Coefficient of waiting time cost
c_s	Coefficient of idle time cost
$\mathbf{A}(\xi)$	Random indicating vector in presence of no show(0) or arrival(1)
$\mathbf{Z}(\xi)$	Vector of random examination duration for n patients
p_i	No show probability of patient i
Decision Variables	
$\mathbf{x}$	Examination time allowance vector for the first n patients
$\mathbf{w}(\xi)$	Patient waiting times vector
$\mathbf{s}(\xi)$	Doctor idle times vector
$o(\xi)$	Overtime with respect to session length

are homogeneous. It means that all patients have same no show probability. Lindley (1952) recursion is used to schedule single server scheduling. Denton and Gupta (2003) use this recursion for modelling single doctor ASP.

The scheduled examination time of patient i is calculated as summation of examination time allowances of patient 1 to $i-1$. In this way, patient 1 is assigned to time 0, patient 2 assigned to time x_1, patient 3 is assigned to time $x_1 + x_2$, and so forth. Due to unrestricted examination time of last patient, the dimension of the vector x is equals to $n-1$. The vector of examination time allowance, $\mathbf{x} \in \mathbb{R}^{n-1}$, denotes the first stage decision. Cost parameters, $c_o, c_w, c_s \in \mathbb{R}$, are in unit time. Patient waiting times vector, doctor idle times vector and overtime are second stage decisions with respect to realizations ($\xi \in \Xi$).

From Lindley (1952) recursion , patient waiting time, doctor idle time and overtime equations can be written as below :

$$w_i(\xi) = [w_{i-1}(\xi) + Z_{i-1}(\xi) - x_{i-1}]^+ \qquad i = 1, ..., n \qquad (3.22)$$

$$s_i(\xi) = [-w_{i-1}(\xi) - Z_{i-1}(\xi) + x_{i-1}]^+ \qquad i = 1, ..., n \qquad (3.23)$$

$$o(\xi) = [w_n(\xi) + Z_n(\xi) + \sum_{i=1}^{n-1} x_i - d]^+ \qquad (3.24)$$

In Chapter 2, We commented on the general objective structure of the ASP. Common point between Lindley recursion and the objective structure is that both of them include patient waiting time, doctor idle time and overtime. Doctor idle times can be excluded from the objective due to Denton and Gupta (2003)'s proposition. The proposition is :

Proposition 3.2. *At the time zero, expected doctor idle time is equal to expected overtime minus expected total examination time :*

$$\mathbb{E}[\sum_{i=1}^{n} s_i] = \mathbb{E}[o] - \mathbb{E}[\sum_{i=1}^{n} Z_i]$$

Proof Proved by Denton and Gupta (2003) □

Reflecting no-shows to the model is a complicating issue. Erdogan and Denton (2013), propose a data structure to reflect no-shows to the model. This structure multiplies two random variables, the no show indicator random variable and the examination time random variable($\widehat{Z}_i(\xi) = A_i(\xi)Z_i(\xi)$), where

$$A_i(\xi) = \begin{cases} 1, & \text{with } 1 - p_i, \\ 0, & \text{with } p_i. \end{cases}$$

From the proposition, the literature and the proposed structure is combined as :

$$\min c_w \sum_{i=1}^{n} \mathbb{E}[w_i] + c_o \mathbb{E}[o]. \qquad (3.25)$$

The last form of the model is given (Erdogan and Denton, 2013) :

$$
\begin{aligned}
& \min \mathbb{E}_\xi[\sum_{i=2}^{n} c^w w_i(\xi) + c^o o(\xi)] \\
& \text{s.t. } w_2(\xi) && \geq \widehat{Z}_1(\xi) - x_1, \forall \xi \\
& \quad - w_2(\xi) + w_3(\xi) && \geq \widehat{Z}_2(\xi) - x_2, \forall \xi \\
& \qquad \ddots \ddots && \vdots \\
& \qquad\qquad - w_{n-1}(\xi) + w_n(\xi) && \geq \widehat{Z}_{n-1}(\xi) - x_{n-1}, \forall \xi \qquad (3.26) \\
& \qquad\qquad\quad - w_n(\xi) + o(\xi) \geq \widehat{Z}_n(\xi) + \sum_{i=1}^{n-1} x_i - d, \forall \xi \\
& \quad \mathbf{x} \geq 0, \mathbf{w}(\xi), o(\xi) \geq 0, \forall \xi.
\end{aligned}
$$

Through the rest of the thesis, above model is referred as risk neutral no-show appointment scheduling model (RNASP).

3.4.2 Risk Averse Formulation

Aforesaid, we discussed on risk averse mathematical modelling structure and RNASP model. Risk averse model is built on these informations. Risk averse model is given below :

$$
\begin{aligned}
\min \sum_{\xi=1}^{s} p(\xi) \sum_{i=2}^{n} c^w w_i(\xi) + c^o o(\xi) + \lambda \left(\eta + \frac{1}{1-\alpha} \sum_{\xi=1}^{s} p(\xi)\nu(\xi) \right) & \\
\text{s.t. } w_2(\xi) \qquad\qquad\qquad & \geq \widehat{Z}_1(\xi) - x_1, \forall\xi \\
-w_2(\xi) + w_3(\xi) \qquad\qquad & \geq \widehat{Z}_2(\xi) - x_2, \forall\xi \\
\ddots \ddots \qquad\qquad & \qquad \vdots \\
-w_{n-1}(\xi) + w_n(\xi) & \geq \widehat{Z}_{n-1}(\xi) - x_{n-1}, \forall\xi \\
-w_n(\xi) + o(\xi) & \geq \widehat{Z}_n(\xi) + \sum_{i=1}^{n-1} x_i - d, \forall\xi \\
\nu(\xi) \geq (\mathbf{q}(\xi))^T \mathbf{y}(\xi) - \eta, \ \forall\xi, & \\
\mathbf{x} \geq 0, \mathbf{w}(\xi), o(\xi), \nu(\xi) \geq 0, \forall\xi. &
\end{aligned} \tag{3.27}
$$

In this model, risk aversion and no-show are modelled. Rest of the thesis, above model is called as risk averse no-show appointment scheduling model (RAASP).

3.4.3 ASP in the Presence of Walk-in Patients

Walk-in is important for the management because capacity allocation is should be made in relation to the management revenue targets. The patients are separated into two groups : patients who take their appointments before the examination day form type 1 patient group, whereas walk-in patients form type 2 patient group. The facility total capacity equals to n patients. While the allocated capacity for the type 1 patients is denoted as $k1$, type 2 patients' allocated capacity is denoted as $k2$. Time allowance vectors type 1 and type 2 patients are denoted as $\mathbf{x_1}$ and $\mathbf{x_2}$ respectively. Within a revenue management framework, the unit revenues obtained from type 1 and type 2 patients are differentiated as r_1 and r_2 respectively. As the same-day appointments may have severe consequence on the system, and the management we will later on compare the cases $r_1 = r_2$ and $r_1 \leq r_2$. Examination time vectors for the patient types are $\widehat{\mathbf{Z}}_\mathbf{1}, \widehat{\mathbf{Z}}_\mathbf{2}$

respectively. Binary variables y1 are introduced to decide on how many type 1 patients will be appointed given the predetermined maximum capacity of k1. Binary variables y2 are introduced with a similar purpose for type 2 patients. We will always have k1+k2=n. However, we may have $\mathbf{1}^T.\mathbf{y_1^*} \leq k1$ or $\mathbf{1}^T.\mathbf{y_2^*} \leq k2$ at optimality, in other words maximum capacities $k1$ and/or $k2$ can be found more than enough. Therefore, the model proposed below can be solved for different values of $k1$ (as $k2 = n - k1$) to identify the best profit generating capacity to allocate for type 1 patients (and thus also for type 2 patients). If we reformulate the problem, the model transforms to :

$$
\begin{aligned}
\max \sum_{\xi=1}^{s} p(\xi) & \left[r_1 \sum_{i=1}^{k_1} y_{1i} + r_2 \sum_{j=(k_1+1)}^{n-1} y_{2i} - \sum_{i=2}^{n} c^w w_i(\xi) - c^o o(\xi) \right] \\
& - \lambda \left(\eta + \frac{1}{1-\alpha} \sum_{\xi=1}^{s} p(\xi)\nu(\xi) \right) \\
\text{s.t. } & w_{i+1}(\xi) - w_i \geq \widehat{Z}_{1i}(\xi) - x_{1i}, \forall \xi, i = 1, ..., k1, \\
& w_{j+1}(\xi) - w_j(\xi) \geq \widehat{Z}_{2j}(\xi) - x_{2j}, \forall \xi, j = (k1+1), ..., (n-1) \\
& - w_n(\xi) + o(\xi) \geq \widehat{Z}_{1n}(\xi) + \widehat{Z}_{2n}(\xi) + \sum_{i=1}^{k1} x_{1i} + \sum_{j=k1+1}^{n-1} x_{2j} - d, \forall \xi \qquad (3.28) \\
& \nu(\xi) \geq (\mathbf{q}(\xi))^T \mathbf{y}(\xi) - \eta, \ \forall \xi, \\
& M y_{1i} \geq x_{1i}, \qquad i = 1, ..., k1, \\
& y_{1i} \leq x_{1i}, \qquad i = 1, ..., k1, \\
& M y_{2j} \geq x_{2j}, \qquad j = (k1+1), ..., (n-1) \\
& y_{2j} \leq x_{2j}, \qquad j = (k1+1), ..., (n-1) \\
& \mathbf{x_{1i}}, \mathbf{x_{2j}} \geq 0, \mathbf{w}(\xi), o(\xi), \nu(\xi) \geq 0, \mathbf{y_{1i}}, \mathbf{y_{2j}} \in \{0,1\} \forall \xi
\end{aligned}
$$

Hereafter, above model is referred as risk averse walk in ASP (RAWSP).

4 SOLUTION METHODOLOGY

In Chapter 3, two different mathematical models are introduced. In this chapter, structural properties and solution procedure of the RNASP and the RAASP models are explained.

4.1 L-Shaped Method

Because of tendency of increase of number of realizations and realization structure, decomposition technique is desirable technique for two stage stochastic programming model. L-Shaped Method is usually used for solving two stage stochastic programming models. L-Shaped method is introduced by Slyke and Wets (1969). This method is basically a special variant of Benders Decomposition for solving two stage stochastic programming models which is divided master-sub problem structure. By changing recourse function in main problem ($Q(.)$) with θ and using dual solutions of each realization, we are capable to generate *feasibility cuts* (4.2) and *optimality cuts* (4.3) for the master problem. L Shaped Algorithm is reformulated problem with Birge and Louveaux (2011)s' cuts :

Master Problem :

$$\min z = c\mathbf{x} + \theta \tag{4.1}$$

$$s.t. A\mathbf{x} = \mathbf{b}$$

$$D_\ell \mathbf{x} \geq d_\ell \qquad \ell = 1, .., r, \tag{4.2}$$

$$E_\ell + \theta \geq e_\ell \qquad \ell = 1, .., s, \tag{4.3}$$

$$x \geq 0, \theta \in \mathbb{R}$$

Subproblems (k=1,..,K) :

$$\min w = q_k^T \mathbf{y}$$

$$s.t. W\mathbf{y} = h_k - T_k \mathbf{x}^{\nu} \qquad (4.4)$$

$$\mathbf{y} \geq 0$$

A new recourse terminology is used to explain further speciality of the structure of the problem :

- **fixed recourse :** if recourse matrix, W, is deterministic, i.e., does not changing with realizations. Otherwise problem is called *stochastic recourse.*

- **simple recourse :** if recourse matrix, W, equals identity matrix with same dimensions.

- **complete recourse :** if second stage problem is feasible for any given first stage decision vector.

- **integer recourse :** if all of the second stage variables are integer.

- **mixed integer recourse :** if some of the second stage variables are integer.

- **linear recourse :** if all of the second stage variables are linear.

In complete recourse stochastic programming models, feasibility cuts are unnecessary.

L-Shaped Methods algorithm is given below (adapted from Birge and Louveaux (2011) :

Algorithm 1: L-Shaped Algorithm

Step 0 Set $r = s = \nu = 0$.

Step 1 Set $\nu = \nu + 1$. **Solve** the mathematical model (4.1)-(4.3).

Let the x^ν, θ^{nu} be an optimal solution. If no constraint in (4.3) is present, θ^ν is equal to $-\infty$ and it is not considered in the computation of x^ν.

Step 2

if $x \in K_2$ **then**
  Go to **Step 1**,
  else add least one cut (4.2) and **return Step 1**.;

end

Step 3

for $k \in K$ **do**
  Solve the the mathematical model (4.4).

end

Let the ϕ_k^ν the optimum value of dual variables associated with the model k of (4.4).

Define :

$$E_{s+1} = \sum_{k=1}^{K} p_k \phi_k^\nu T_k$$

and

$$e_{s+1} = \sum_{k=1}^{K} p_k \phi_k^\nu h_k.$$

Let $w^\nu = e_{s+1} - E_{s+1} x^\nu$.

if $\theta^\nu \leq w^\nu$ **then**
  $s = s + 1$ **add** the constraint set (4.1) and **return Step 1**
  else Stop, x^ν is an optimal solution ;

end

4.2 Solution Methodology of RNASP

Assumptions of problem is explained in this section. First assumption is whole patients are punctual. Second one is there is no walk in patients including emergency cases. Third assumption is there is no price discrimination in waiting time costs. Last assumption is that all patients have same no show probability. Because of aforesaid problems in 4.1, RNASP problem can be written in master-sub problem format. Master problem can be written :

$$\min\{Q(\mathbf{x})$$

$$\text{s.t.}\mathbf{x} \geq 0. \tag{4.5}$$

where $Q(\mathbf{x}) = \mathbb{E}_{\xi}[Q(x,\xi)]$ is the recourse function denoting the expectation of whole sub problems in whole scenarios.

The sub problem comprises separate realizations of scenarios $Q(x,\xi)$. Sub problem structure includes a minimization of objective which equals waiting time costs and overtime cost, constraints for computing waiting times and overtime. A sub problem of RNASP can be written as below :

$$Q(x,\xi) = \min c^w \sum_{i=2}^{n} A_i(\xi) w_i(\xi) + c^o o(\xi)$$

$$\begin{aligned}
\text{s.t.} w_2(\xi) & \geq \widehat{Z}_1(\xi) - x_1, \forall \xi \\
-w_2(\xi) + w_3(\xi) & \geq \widehat{Z}_2(\xi) - x_2, \forall \xi \\
\ddots \ddots & \quad \vdots \\
-w_{n-1}(\xi) + w_n(\xi) & \geq \widehat{Z}_{n-1}(\xi) - x_{n-1}, \forall \xi \\
-w_n(\xi) + o(\xi) & \geq \widehat{Z}_n(\xi) + \sum_{i=1}^{n-1} x_i - d, \forall \xi
\end{aligned} \tag{4.6}$$

$$\mathbf{x} \geq 0, \mathbf{w}(\xi), o(\xi) \geq 0, \forall \xi$$

From the definition of complete recourse, RNASP is a complete recourse type of two stage stochastic programming problem. Because for all values of overtime and waiting times, **x** vector is always feasible. So, we can use L-Shaped Algorithm's a variant for the complete recourse problems which is introduced by Erdogan (2010).

Algorithm 2: A variant of L-Shaped Algorithm

Step 0 $\nu = 1, k = 1$

Step 1 **Start** with a arbitrary solution **x**

Step 2 **while** *Current Bound* - $\theta \geq 0$ **do**

  $\nu \leftarrow \nu + 1$

  Solve Master problem (4.5)

  Solve Sub problem (4.6) for each realization

  Add Optimality cut to master problem

end

There is a structural property of the model related with no show and double booking which means that two or more patients arrives simultaneously. Erdogan (2010) proposed this structure using similarity of well known newsboy problem in two patients environment.

Proposition 4.1. *The optimal schedule for 2 patients and* $d = 0$*, with no show probability* p*, can allow to double booking patients if* $\dfrac{c^w}{c^w + c^o} \leq p$.

Proof This proposition is proved by Erdogan (2010). □

4.3 Solution Methodology of RAASP

Risk averse two stage stochastic programming model is solved via a specialized L-Shaped method which includes calculation of α quantile of the sub problems for calculating CVaR values. This algorithm is proposed by Noyan (2012). Master-sub problem structure is main structure of the algorithm. We explain the cuts and the structure in

Chapter 3. If we rewrite RAASP in this format :

Master Problem :

$$\begin{aligned}
&\min \theta_1 + \lambda\theta_2 \\
&\text{s.t.}\theta_1 \geq \sum_{\ell=1}^{s} p^\ell (\mathbf{u}^{\ell t})^T(\mathbf{h}^\ell - T^\ell \mathbf{x}) \qquad t = 1, ..., \tau, \\
&\theta_2 \geq \eta^t + \frac{1}{1-\alpha}\sum_{\ell=1}^{s} p^\ell \nu^{\ell t} \qquad t = 1, ..., \tau, \\
&\nu^{\ell t} \geq (\mathbf{u}^{\ell t})^T(\mathbf{h}^\ell - T^\ell \mathbf{x}) - \eta^t, \quad \ell = 1, ..., s, t = 1, ..., \tau, \\
&\nu^{\ell t} \geq 0, \eta^t \in \mathbb{R}, \qquad \ell = 1, ..., s, t = 1, ..., \tau.
\end{aligned} \tag{4.7}$$

Sub Problems :

$$\begin{aligned}
&\min c^w \sum_{i=2}^{n} A_i(\xi) w_i(\xi) + c^o o(\xi) \\
&\text{s.t.} w_2(\xi) \qquad \geq \widehat{Z}_1(\xi) - x_1, \forall\xi \\
&- w_2(\xi) + w_3(\xi) \qquad \geq \widehat{Z}_2(\xi) - x_2, \forall\xi \\
&\qquad \ddots \quad \ddots \qquad \vdots \\
&- w_{n-1}(\xi) + w_n(\xi) \geq \widehat{Z}_{n-1}(\xi) - x_{n-1}, \forall\xi \\
&- w_n(\xi) + o(\xi) \geq \widehat{Z}_n(\xi) + \sum_{i=1}^{n-1} x_i - d, \forall\xi \\
&\mathbf{x} \geq 0, \mathbf{w}(\xi), o(\xi) \geq 0, \forall\xi
\end{aligned} \tag{4.8}$$

For every solution of the sub problems, the master problem is always feasible. Because of this, the model is complete recourse stochastic programming model. In complete recourse stochastic programming model, feasibility cut can be omitted from the model and the algorithm. This algorithm is also proper to use for solving RAWSP model. The structure of algorithm is given below :

Algorithm 3: Algorithm for RAASP

Step 0 $\tau = 0, s = 1, ..., N$.

Step 1 Solve the master problem (4.7).

Let $(\overline{\mathbf{x}}, \overline{\theta_1}, \overline{\theta_2})$ be an optimal solution and $\overline{\theta} = \overline{\theta_1} + \lambda\overline{\theta_2}$. (When $\tau = 0$ ignore θ_1, θ_2 and the optimality cuts.)

Step 2 Solve all the sub problems (4.8).

Let $\tau = \tau + 1$ and $\mathbf{u}^{\ell\tau}$ denote the dual vector corresponding to the optimal solution of the sub problems.

Step 3 Observe that $(\mathbf{u}^{\ell\tau})^T(h^\ell - T^\ell\overline{\mathbf{x}}), \ell = 1, ..., s$, are the realizations of $Q(\overline{\mathbf{x}}, \xi)$.

Find the α quantile of $Q(\overline{\mathbf{x}}, \xi)$ denoted by $\overline{\eta}_\alpha$, and calculate $CVaR_\alpha(Q(\overline{\mathbf{x}}, \xi))$:

$$CVaR_\alpha(Q(\overline{\mathbf{x}}, \xi)) = \overline{\eta}_\alpha + \frac{1}{1-\alpha}\left(\sum_{\ell=1}^{s} p^\ell[Q(\overline{\mathbf{x}}, \xi^\ell) - \overline{\eta}_\alpha]^+\right).$$

Step 4 Calculate the mean-risk function value of the recourse cost at the current solution :

$$\theta^* = \sum_{\ell=1}^{s} p^\ell \mathbf{u}^{\ell\tau})^T(h^\ell - T^\ell\overline{\mathbf{x}}) + \lambda CVaR_\alpha(Q(\overline{\mathbf{x}}, \xi)).$$

Step 5 if $\overline{\theta} \geq \theta^*$ **then**

Start $\overline{\mathbf{x}}$ is the optimal solution vector of the master problem.

else Introduce the optimality cuts for $\mathbf{u}^{\ell\tau}$ and go to **Step 1**.;

end

---

5 COMPUTATIONAL STUDIES

5.1 Generation of the Problem Instances

For the purpose of testing the computational performance of the proposed algorithm, we consider different problem instances of different sizes. Because of the existence of two methodologies, we generate two groups of data sets. Standard deviation and mean of the examination times are determined via the previous data which has been collected for three months from a radiology clinic in a state hospital in Turkey. Selected process is the computed tomography (CT) machine consultation. Because CT machine is commonly used as a visualization machine in radiology clinics and the initial investment cost of CT machine is very high.

Data Set I

- The first data set has 30 problem instances and 10 patients.
- From the collected data, the consultation time distribution is normal with mean 7 minutes.
- Variation of data is changing week to week. So, we run the model with different standard deviation levels.
- No show probability of appointed patients is 20%.
- The session length is assumed to be equal to the sum of the mean consultation durations.
- Scenario probabilities are set to be equal.
- $\gamma = c^o/c^w$ rates are taken in different levels.

Data Set II

- The problem instances and the capacity of the system is same in Data Set I.
- Durations of the consultations are generated as in Data Set I.
- No show probability of appointed patients is also 20% and no show probability of walk in patients is 5%.
- Length of the session is same as in Data Set I.
- Scenario probabilities are set to be equal.

– $\gamma = c^o/c^w$ rates are taken in different levels.

Data set I is used while solving RNASP, RAASP and RAWASP. Data set II is only used while solving RAWASP. All proposed models are coded with GAMS 23.5 mathematical programming language running on CPLEX 12.2 solver. The numerical experiments are performed on a computer with 64 bit A 10 (4 physical cores and 4 virtual cores) AMD 4600M CPU with 2.3 GHz processor and 8 GB of memory. In our numerical experiments, we terminate the solver when 500000 iterations are completed.

5.2 Results for RNASP

We begin with a simple example that points out the optimal solutions' patterns with respect to changes in coefficient of variance, c_v ; and to changes in relative cost of waiting, c^w, and overtime, c^o. The capacity of clinic 10 patients in 70 minutes. The consultation durations are normally distributed with mean 7 minutes and standard deviation 0.07. Table 5.1 underlines the optimal interarrival times of patients with $c_v = 0.01$ and different levels of the *gamma*.

Table 5.1: Optimal solution of RNASP ($\gamma = c^o/c^w, Z_i \sim \mathcal{N}(7, 0.07), d = 70$)

Interarrival Times	$\gamma = 10$	$\gamma = 1$	$\gamma = 0.1$
x_1	6.9530	7.0078	7.0940
x_2	6.9870	7.0144	7.0760
x_3	6.9860	7.0344	7.0820
x_4	7.0000	7.0267	7.0900
x_5	6.9900	7.0222	7.0920
x_6	6.9800	7.0211	7.0910
x_7	6.9990	7.0211	7.0800
x_8	7.0000	7.0167	7.0900
x_9	7.0130	7.0044	7.0870

The optimal interarrival times for three different levels of γ are listed in Table 5.1 and are graphically illustrated in Figure 5.1. If Figure 5.1 is analyzed, the *dome shape* structure for interarrival times can be easily observed. Dome shape structure implies more frequent appointments at the start and end of the daily session, and less

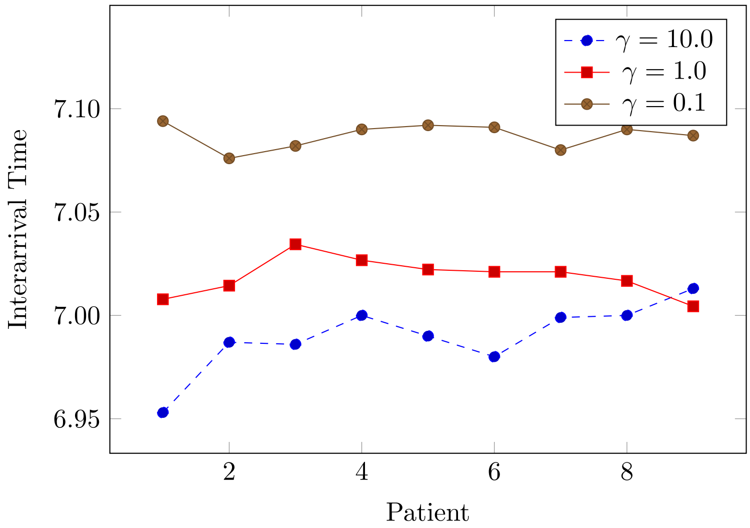
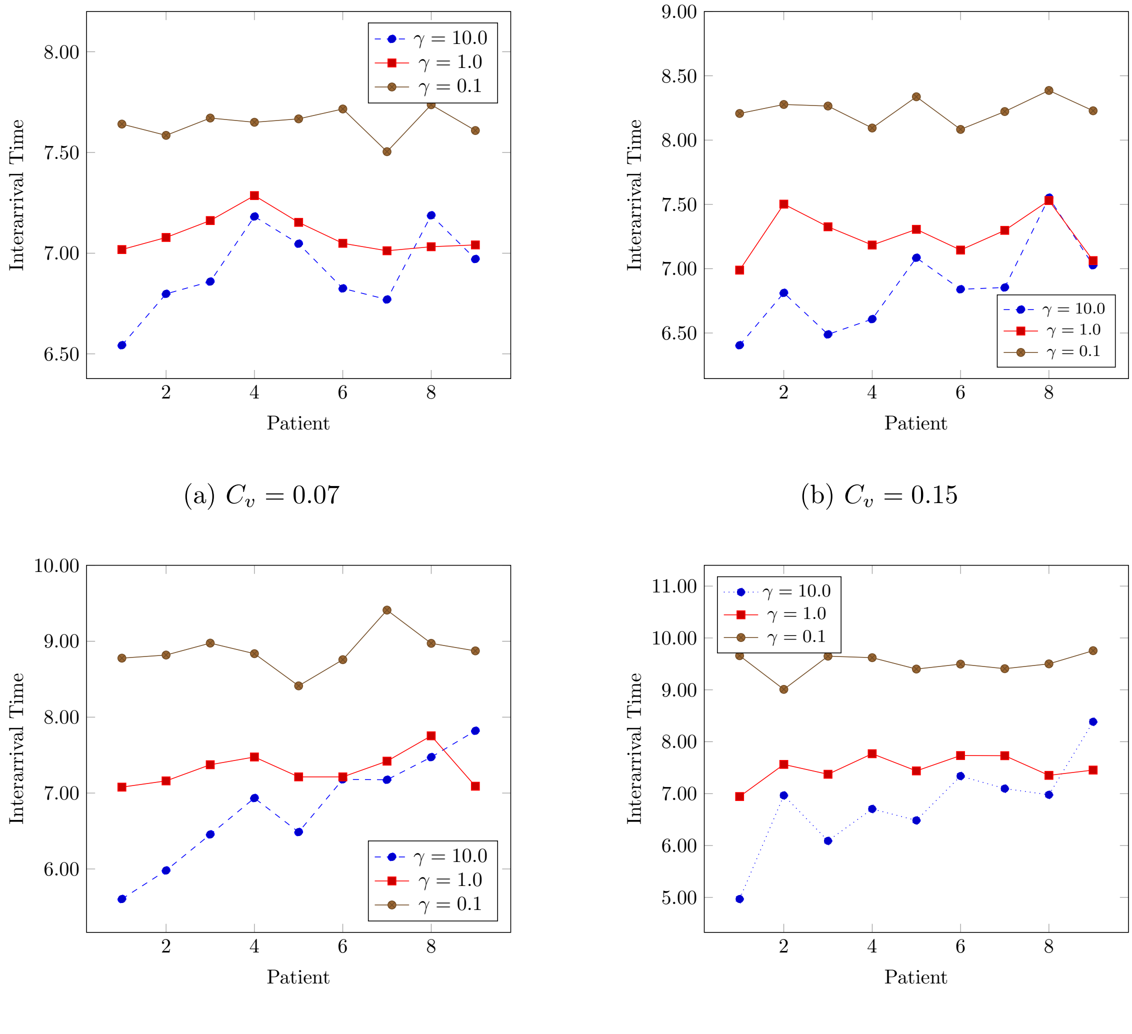


Figure 5.1: Optimal solution of RNASP ($\gamma = c^o/c^w, Z_i = \mathcal{N}(7, 0.07), d = 70$)

frequent appointments at the middle of the session. This shape appears at optimality mainly to prevent overtime and compensate random events during the session. This dome shaped structure has been discovered for the risk neutral problem in Denton and Gupta (2003). When *gamma* increases from 0.1 to 10, the consultation allowance times increase. When *gamma* equals to 0.1, the time allowances are become equal.

Coefficient of variation of consultation duration is another parameter in our model. Figure 5.2 shows interarrival times of the patients with consultation times for different level of coefficient of variation. When coefficient of variation increases (from Figure 5.2a to Figure 5.2d), dome shape structure is deformed significantly. When *gamma* increases, *dome shaped* structure preserve its shape. In Figures 5.2a- 5.2d, maximum interarrival times are approximately equal to 7.8, 8.5, 9, 10 minutes. In same way, minimum interarrival times are approximately equal to 6.6, 6.5, 5.9, 5 minutes.

In Table 5.2 , interarrival times of patients are given to explain changes in respect to coefficient of variation. To underline the changes, we fix *gamma*. When coefficient of variation increases, range between maximum interarrival times and minimum interarrival times are also increasing. Deformation level of the dome shape structure increases parallel with coefficient of variance shown in Table 5.2 detailed.



(a) $C_v = 0.07$

(b) $C_v = 0.15$

(c) $C_v = 0.23$

(d) $C_v = 0.30$

Figure 5.2: Optimal solution of RNASP for different C_v levels

Table 5.2: Optimal solution of RNASP ($Z_i \sim \mathcal{N}(7, 0.07), d = 70$)

$\gamma = 10$					
Interarrival Times	$C_v = 0.01$	$C_v = 0.08$	$C_v = 0.15$	$C_v = 0.23$	$C_v = 0.30$
x_1	6.9530	6.5430	6.4050	5.6040	4.9690
x_2	6.9870	6.7980	6.8120	5.9800	6.9660
x_3	6.9860	6.8590	6.4890	6.4550	6.0900
x_4	7.0000	7.1820	6.6080	6.9340	6.7050
x_5	6.9900	7.0470	7.0850	6.4860	6.4840
x_6	6.9800	6.8250	6.8400	7.1800	7.3390
x_7	6.9990	6.7700	6.8540	7.1750	7.0970
x_8	7.0000	7.1880	7.5520	7.4730	6.9780
x_9	7.0130	6.9710	7.0270	7.8200	8.3850
$\gamma = 1$					
Interarrival Times	$C_v = 0.01$	$C_v = 0.08$	$C_v = 0.15$	$C_v = 0.23$	$C_v = 0.30$
x_1	7.0078	7.0180	6.9890	7.0780	6.9440
x_2	7.0144	7.0780	7.5020	7.1600	7.5630
x_3	7.0344	7.1620	7.3260	7.3730	7.3730
x_4	7.0267	7.2860	7.1840	7.4750	7.7690
x_5	7.0222	7.1530	7.3060	7.2130	7.4380
x_6	7.0211	7.0490	7.1450	7.2130	7.7340
x_7	7.0211	7.0120	7.2980	7.4200	7.7310
x_8	7.0167	7.0320	7.5300	7.7540	7.3520
x_9	7.0044	7.0410	7.0630	7.0900	7.4540
$\gamma = 0.1$					
Interarrival Times	$C_v = 0.01$	$C_v = 0.08$	$C_v = 0.15$	$C_v = 0.23$	$C_v = 0.30$
x_1	7.0940	7.6410	8.2070	8.7770	9.6580
x_2	7.0760	7.5850	8.2770	8.8180	9.0090
x_3	7.0820	7.6710	8.2650	8.9750	9.6490
x_4	7.0900	7.6500	8.0940	8.8360	9.6200
x_5	7.0920	7.6670	8.3370	8.4120	9.4010
x_6	7.0910	7.7160	8.0830	8.7570	9.4960
x_7	7.0800	7.5040	8.2220	9.4100	9.4080
x_8	7.0900	7.7370	8.3860	8.9720	9.5010

5.3 Results for RAASP

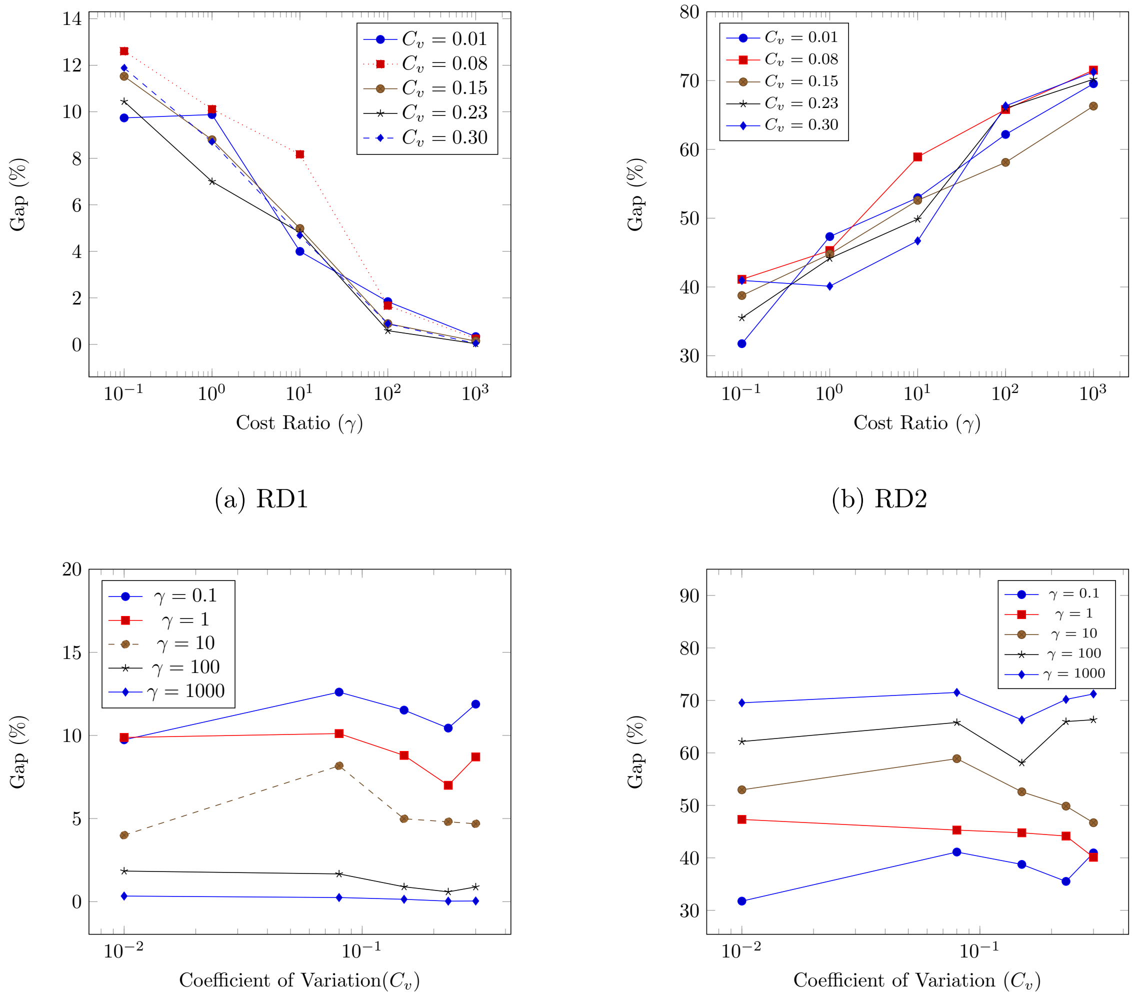
Let us denote the optimum objective value of RAASP in 3.27 as E1 when $\lambda = 0$ and C1 when $\lambda = 1$. Moreover, let CVaR$_\alpha$ value calculated based on the optimum solution of RAASP with $\lambda = 0$ be C2, and the expectation value calculated based on the optimum solution of RAASP with $\lambda = 1$ be E2. Then, the validity of the model can be checked based on the ratios (or relative differences) $RD1 = (E2 - E1)/E1$ and $RD2 = (C1 - C2)/C1$. When $RD1$ is low and $RD2$ is high, the proposed CVaR model is meaningful in the sense that it can avoid large losses. We investigate these ratios in respect to coefficient of variation of consultation times and the cost ratios. In Table 5.3, it is clear that risk averse modelling is significant for ASP because RD1 is significantly lower than RD2 at all γ values of 0.1, 1, 10, 100, 1000 and C_v values 0.01, 0.07, 0.15, 0.23, 0.30.

Table 5.3: Comparison of Relative Differences

Relative Difference in Expected Value (%)					
$\gamma \setminus C_v$	0.01	0.08	0.15	0.23	0.3
0.1	9.7369	12.6069	11.5238	10.4424	11.8826
1	9.8799	10.1123	8.7982	6.9997	8.7078
10	3.9985	8.1721	4.9844	4.8067	4.6839
100	1.8394	1.6687	0.8936	0.5911	0.8855
1000	0.3366	0.2477	0.1417	0.0317	0.0400
Relative Difference in $CVaR_\alpha$ (%)					
$\gamma \setminus C_v$	0.01	0.08	0.15	0.23	0.3
0.1	-31.7596	-41.1134	-38.7552	-35.5188	-40.9514
1	-47.3254	-45.2966	-44.7792	-44.1585	-40.1125
10	-52.9709	-58.9018	-52.5902	-49.8603	-46.6975
100	-62.1686	-65.7963	-58.1081	-65.9728	-66.3216
1000	-69.5435	-71.5195	-66.2891	-70.1914	-71.2349

Figure 5.3 illustrates RD1 and RD2 in respect to changes in C_v and *gamma*. When *gamma* increases(illustrated in Figures 5.3a and 5.3b), increment of RD1 declines.

In sharp contrast to this, decrement of RD2 rises. Although, both facts improves our solution. Figures 5.3c and 5.3d illustrate C_v effect on RD1 and RD2. When C_v increases, there is no significant effect on both relative differences.



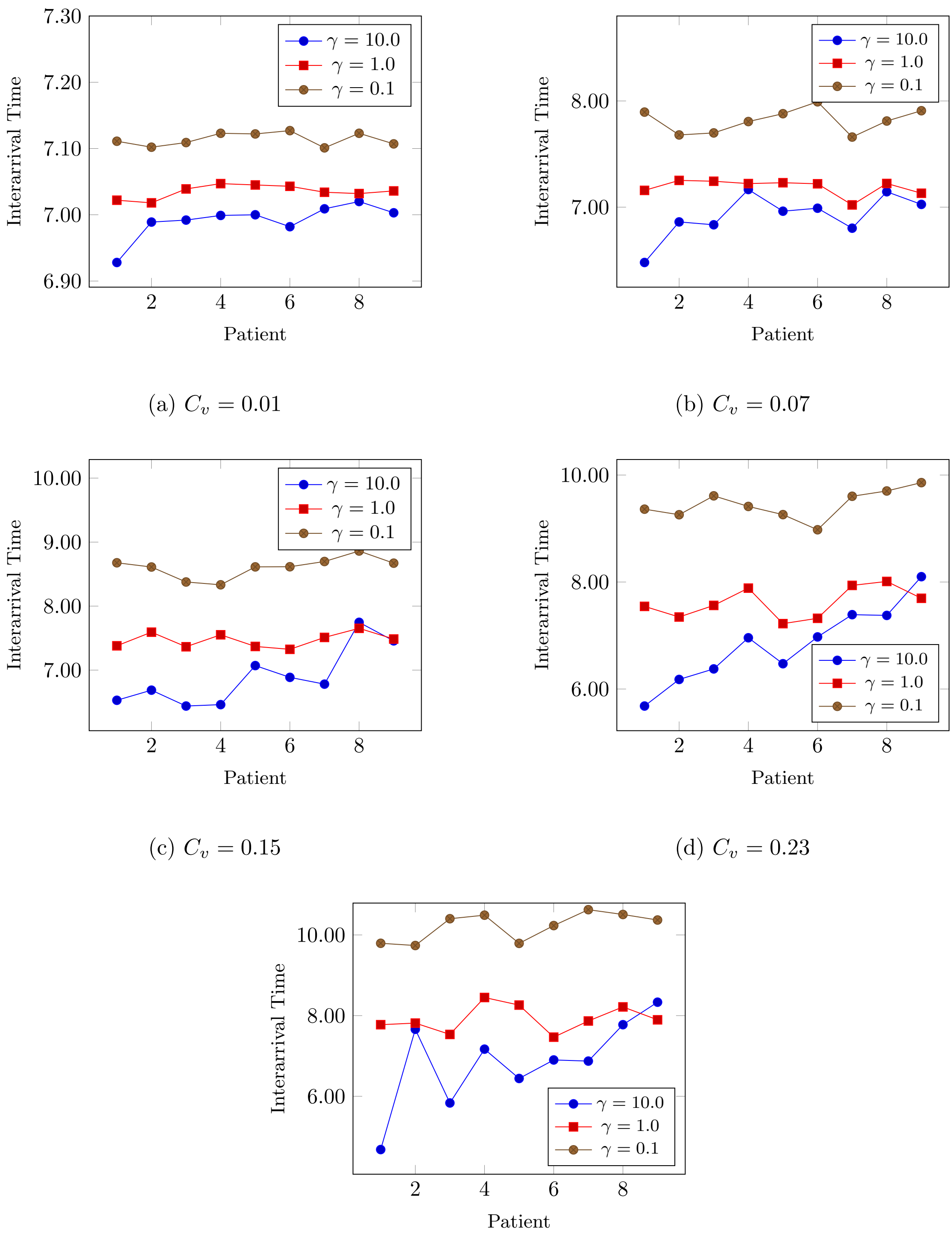
(a) RD1 (b) RD2

(c) RD1 (d) RD2

Figure 5.3: Changes in RD1 and RD2 for C_v and γ

Table 5.4 indicates optimal interarrival times in presence of γ and C_v. It is obvious that when C_v increases, the difference between first and last interarrival times also increases. When γ decreases from 10 to 0.1, the interarrival times become larger.

Figure 5.4 illustrates interarrival times in respect to γ and C_v. It can be observed from this figure that when *gamma* increases, the interarrival times decreases. When C_v increases, variance of the interarrival times increases. Dome shape structure deteriorates when C_v increases.



(a) $C_v = 0.01$

(b) $C_v = 0.07$

(c) $C_v = 0.15$

(d) $C_v = 0.23$

(e) $C_v = 0.30$

Figure 5.4: Optimal solution of RAASP for different C_v levels

Table 5.4: Optimal solution of RAASP ($Z_i \sim \mathcal{N}(7, 0.07), d = 70$)

$\gamma = 10$					
Interarrival Times	$C_v = 0.01$	$C_v = 0.08$	$C_v = 0.15$	$C_v = 0.23$	$C_v = 0.30$
x_1	6.9280	6.4790	6.5290	5.6820	4.6800
x_2	6.9890	6.8610	6.6870	6.1790	7.6610
x_3	6.9920	6.8340	6.4390	6.3760	5.8350
x_4	6.9990	7.1660	6.4610	6.9580	7.1690
x_5	7.0000	6.9620	7.0710	6.4730	6.4420
x_6	6.9820	6.9900	6.8870	6.9740	6.9010
x_7	7.0090	6.8020	6.7810	7.3900	6.8730
x_8	7.0200	7.1450	7.7470	7.3760	7.7750
x_9	7.0030	7.0260	7.4570	8.1000	8.3330
$\gamma = 1$					
Interarrival Times	$C_v = 0.01$	$C_v = 0.08$	$C_v = 0.15$	$C_v = 0.23$	$C_v = 0.30$
x_1	7.0220	7.1580	7.3800	7.5450	7.7750
x_2	7.0180	7.2520	7.5930	7.3460	7.8150
x_3	7.0390	7.2440	7.3650	7.5630	7.5340
x_4	7.0470	7.2220	7.5520	7.8860	8.4490
x_5	7.0450	7.2300	7.3700	7.2220	8.2610
x_6	7.0430	7.2200	7.3250	7.3220	7.4660
x_7	7.0340	7.0210	7.5110	7.9370	7.8670
x_8	7.0320	7.2230	7.6530	8.0110	8.2160
x_9	7.0360	7.1310	7.4860	7.6960	7.8960
$\gamma = 0.1$					
Interarrival Times	$C_v = 0.01$	$C_v = 0.08$	$C_v = 0.15$	$C_v = 0.23$	$C_v = 0.30$
x_1	7.1110	7.8950	8.6770	9.3620	9.7930
x_2	7.1020	7.6800	8.6110	9.2580	9.7380
x_3	7.1090	7.6990	8.3770	9.6110	10.4020
x_4	7.1230	7.8060	8.3330	9.4130	10.4900
x_5	7.1220	7.8800	8.6130	9.2600	9.7910
x_6	7.1270	7.9920	8.6150	8.9770	10.2320
x_7	7.1010	7.6600	8.6970	9.6030	10.6260
x_8	7.1230	7.8110	8.8610	9.7020	10.5070
x_9	7.1070	7.9080	8.6710	9.8580	10.3700

5.4 RAWASP Results

We begin with demonstration of RAWASP with price discrimination (WPD) and without price discrimination solutions (PD). The change in the ratio, more precisely the ratio (γ) of a patient's waiting cost to the doctor overtime cost (c^0/c^w) is also investigated. In the model without price discrimination, revenue acquired from all patients is 7. In the model with price discrimination, revenue of the appointed patients equal to 7 and revenue of the walkin patients is equal to 10. Ten data sets are generated with each including 30 scenarios and the problem associated with each data set is solved with risk averse and risk neutral approaches.

To validate our approach, we compare the results of the risk neutral (RNM) or expected cost model with the results of the risk averse (RAM) or CVaR based model. Let us denote the expected total cost obtained by solving RNM and RAM optimally as ERNM and ERAM respectively. Similarly, let the CVaR value obtained by solving RNM and RAM optimally be CRNM and CRAM respectively. Then, we set RD1 = (ERAM-ERNM)/ERNM and RD2 = (CRAM-CRNM)/CRNM. RD1 and RD2 values in respect to with and without price discrimination are provided in Table 5.5 and 5.6

respectively.

Table 5.5: RD1 values with respect to model parameters

Model without Price Discrimination									
γ/b_1	1	2	3	4	5	6	7	8	9
0.1	-0.680	-0.686	-0.469	-0.394	-0.716	-0.527	-0.624	-0.637	-0.501
1	-8.562	-9.394	-8.339	-8.814	-8.345	-8.724	-8.222	-7.296	-6.905
Model with Price Discrimination									
γ/b_1	1	2	3	4	5	6	7	8	9
0.1	-0.518	-0.620	-0.610	-0.736	-0.825	-0.745	-0.863	-0.384	-0.501
1	-6.824	-7.698	-7.633	-7.467	-7.965	-8.086	-8.351	-7.774	-6.905

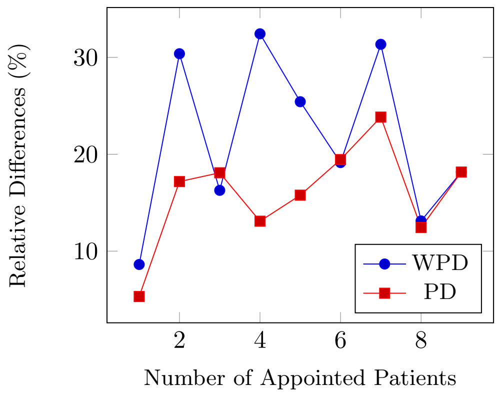
RD1 ranges between -0.394% and -9.394% while RD2 ranges between 5.314% and 32.431%. When absolute values are compared, it is clear that the solutions of the proposed RAM is effective in avoiding large costs for ASP. It is obvious that when c^o increases increment of number of allocated places for walkin patients improves our solution.

Table 5.6: RD2 values with respect to model parameters

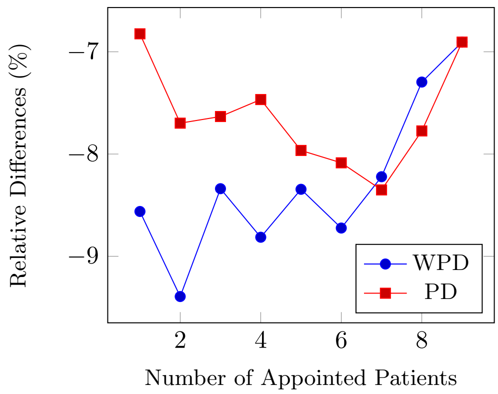
Model without Price Discrimination									
γ/b_1	1	2	3	4	5	6	7	8	9
0.1	9.550	16.138	18.069	15.208	18.849	19.425	22.641	22.641	23.306
1	8.620	30.376	16.273	32.431	25.422	19.139	31.344	13.118	18.163
Model with Price Discrimination									
γ/b_1	1	2	3	4	5	6	7	8	9
0.1	10.137	13.825	15.713	14.745	17.305	19.573	25.746	24.806	23.306
1	5.314	17.180	18.087	13.091	15.776	19.446	23.835	12.434	18.163

Figure 5.5 illustrates how RD1 and RD2 values change when model parameter γ vary

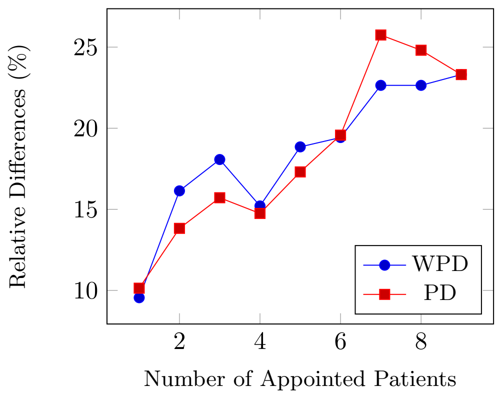
in respect to WPD and PD. When *gamma* decreases, difference between WPD and PD becomes meaningless. According to results of PD, best solutions of the all trials include that 7 of 10 patients are appointed patients. According to results of WPD, best solutions of the all trials include that 4 of 10 patients are appointed patients. It is obvious that price discrimination becomes meaningful when *gamma* increases.



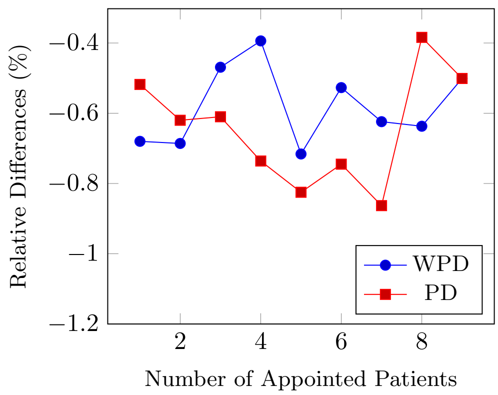
(a) Change in RD2 values ($\gamma = 1$)



(b) Change in RD1 values ($\gamma = 1$)



(c) Change in RD2 values ($\gamma = 0.1$)



(d) Change in RD1 values ($\gamma = 0.1$)

Figure 5.5: Optimal solution of RAWASP

6 CONCLUSION

In this study, we solve ASP with stochastic examination times, walk-in and no show patients. To take into account risks, we model the ASP with mean-risk approach. This problem is commonly faced in healthcare facilities. In this thesis, we reformulate ASP as a risk averse two stage stochastic programming model. The model objective is find optimal appointment times while minimizing the total cost related to the waiting times of patients and overtime of doctors. To model the risk averse behavior, CVaR risk measure is used. Different scenarios where the patient examination times and no-shows vary randomly are taken into account. We then reformulate this model by including differentiated revenues from ordinary patients and walk-in patients.

We assume queue discipline is FIFO, all patients are punctual and decision maker wants to minimize a convex cost function. Our assumptions are widely addressed in scheduling systems at service environment. However, the proposed model has also some limitations. First, FIFO queue discipline is commonly used in service scheduling but this assumption is not appropriate when priority levels of patients differ. Secondly, the punctual arrival of every patient is not a very realistic assumption but is almost a must to control the complexity of the model. Finally, the appointment times are decided in a static instead of a dynamic fashion, and this may limit the application of the proposed approach.

REFERENCES

Artzner, P., Delbaen, F., Eber, J.-M. and Heath, D. (1999). Coherent measures of risk, *Mathematical Finance* **9**(3) : p. 203–228.

Bailey, N. T. J. (1952). Study of queues and appointment systems in outpatient departments, with special reference to waiting-times., *Journal of the Royal Statistical Society : Series B* **14**(2) : p. 185–199.

Benders, J. (2005). Partitioning procedures for solving mixed-variables programming problems, *Computational Management Science* **2**(1) : p. 3–19.

Birge, J. and Louveaux, F. (2011). *Introduction to Stochastic Programming*, second edn, Springer New York.

Brahimi, M. and Worthington, D. J. (1991). Queueing models for out-patient appointment systems – a case study, *The Journal of the Operational Research Society* **42**(9) : p. 733–746.

Cai, X. and Li, K. (2000). A genetic algorithm for scheduling staff of mixed skills under multi-criteria, *European Journal of Operational Research* **125**(2) : p. 359 – 369.

Cayirli, T. and Veral, E. (2003). Outpatient scheduling in health care : A review of literature, *Production and Operations Management* **12**(4) : p. 519–549.

Chern, C.-C., Chien, P.-S. and Chen, S.-Y. (2008). A heuristic algorithm for the hospital health examination scheduling problem, *European Journal of Operational Research* **186**(3) : p. 1137 – 1157.

Choi, S. (2009). *Risk-Averse Newsvendor Models*, PhD thesis, Rutgers, The State University of New Jersey.

Dantzig, G. B. (1955). Linear programming under uncertainty, *Management Science* **1**(3/4) : p. 197–206.

Dantzig, G. B. and Wolfe, P. (1960). Decomposition principle for linear programs, *Operations Research* **8**(1) : p. 101–111.

Denton, B. and Gupta, D. (2003). A sequential bounding approach for optimal appointment scheduling, *IIE Transactions* **35**(11) : p. 1003–1016.

Deyo, R. A. and Inui, T. S. (1980). Dropouts and broken appointments : A literature review and agenda for future research, *Medical Care* **18**(11) : p. 1146–1157.

Erdogan, S. A. (2010). *Optimization of Appointment Based Scheduling Systems*, PhD thesis, North Carolina State University.

Erdogan, S. A. and Denton, B. (2013). Dynamic appointment scheduling of a stochastic server with uncertain demand, *INFORMS J. on Computing* **25**(1) : p. 116–132.

Fetter, R. B. and Thompson, J. D. (1966). Patients' waiting time and doctors' idle time in the outpatient setting., *Health Services Research* **1**(1) : p. 66–90.

Fries, B. E. and Marathe, V. P. (1981). Determination of optimal variable-sized multiple-block appointment systems, *Operations Research* **29**(2) : p. 324–345.

Gocgun, Y., Bresnahan, B. W., Ghate, A. and Gunn, M. L. (2011). A markov decision process approach to multi-category patient scheduling in a diagnostic facility, *Artificial Intelligence in Medicine* **53**(2) : p. 73 – 81.

Green, L. V. and Savin, S. (2008). Reducing delays for medical appointments : A queueing approach., *Operations Research* **56**(6) : p. 1526–1538.

Green, L. V., Savin, S. and Wang, B. (2006). Managing patient service in a diagnostic medical facility, *Operations Research* **54**(1) : p. 11–25.

Gupta, D. and Denton, B. (2008). Appointment scheduling in health care : Challenges and opportunities, *IIE Transactions* **40**(9) : p. 800–819.

Ho, C.-J. and Lau, H.-S. (1992). Minimizing total cost in scheduling outpatient appointments, *Management Science* **38**(12) : p. 1750–1764.

Jansson, B. (1966). Choosing a good appointment system-a study of queues of the type (d, m, 1), *Operations Research* **14**(2) : p. 292–312.

Jerbi, B. and Kamoun, H. (2011). Multiobjective study to implement outpatient appointment system at hedi chaker hospital, *Simulation Modelling Practice and Theory* **19**(5) : p. 1363–1370.

Jonas, S. (1971). Appointment-breaking in a general medical clinic, *Medical Care* **9**(1) : p. 82–88.

Kaandorp, G. C. and Koole, G. (2007). Optimal outpatient appointment scheduling, *Health Care Management Science* **10**(3) : p. 217–229.

Kolisch, R. and Sickinger, S. (2008). Providing radiology health care services to stochastic demand of different customer classes, *OR Spectrum* **30**(2) : p. 375–395.

Lindley, D. V. (1952). The theory of queues with a single server, *Mathematical Proceedings of the Cambridge Philosophical Society* **48** : p. 277–289.

Lowery, J. C. (1992). Simulation of a hospital's surgical suite and critical care area, *Proceedings of the 24th conference on Winter simulation*, WSC '92, ACM, New York, NY, USA, pp. p. 1071–1078.

Lu, L., Li, J. and Gisler, P. (2011). Improving financial performance by modeling and analysis of radiology procedure scheduling at a large community hospital, *Journal of Medical Systems* **35**(3) : p. 299–307.

Maisonneuve, C. d. l. and Martins, J. O. (2013). Public spending on health and long-term care : A new set of projections, *Technical Report 06*, OECD, New York.

Mercer, A. (1960). A queueing problem in which the arrival times of the customers are scheduled, *Journal of the Royal Statistical Society. Series B (Methodological)* **22**(1) : p. 108–113.

Mercer, A. (1973). Queues with scheduled arrivals : A correction, simplification and extension, *Journal of the Royal Statistical Society. Series B (Methodological)* **35**(1) : p. 104–116.

Muthuraman, K. and Lawley, M. (2008). A stochastic overbooking model for outpatient clinical scheduling with no-shows, *IIE Transactions* **40**(9) : p. 820–837.

Naidu, K. D., Sulllivan, K. M., Wang, P. P. and Yang, Y. (2000). Managing personnel through staff scheduling algorithms, *Proceedings of the Fifth Joint Conference on Information Sciences*, Vol. 5, pp. p. 829–835.

Noyan, N. (2012). Risk-averse two-stage stochastic programming with an application to disaster management, *Computers and Operations Research* **39**(3) : p. 541 – 559.

Patrick, J., Puterman, M. L. and Queyranne, M. (2008). Dynamic multipriority patient scheduling for a diagnostic resource, *Operations Research* **56**(6) : p. 1507–1525.

Pflug, G. (2000). Some remarks on the value-at-risk and the conditional value-at-risk, *in* S. Uryasev (ed.), *Probabilistic Constrained Optimization*, Vol. 49 of *Nonconvex Optimization and Its Applications*, Springer US, pp. p. 272–281.

Robinson, L. W. and Chen, R. R. (2003). Scheduling doctors' appointments : optimal and empirically-based heuristic policies, *IIE Transactions* **35**(3) : p. 295–307.

Rockafellar, R. and Uryasev, S. (1969). Optimization of conditional value-at-risk, *The Journal of Risk* **2**(3) : p. 21–41.

Rust, C., Gallups, N., Clark, W., Jones, D. and Wilcox, W. (1995). Patient appointment failures in pediatric resident continuity clinics, *Archives of Pediatrics and Adolescent Medicine* **149**(6) : p. 693–695.

Schroeder, S. A. (1973). Lowering broken appointment rates at a medical clinic, *Medical Care* **11**(1) : p. 75–78.

Slyke, R. M. V. and Wets, R. (1969). L-shaped linear programs with applications to optimal control and stochastic programming, *SIAM Journal on Applied Mathematics* **17**(4) : p. 638–663.

Turkcan, A., Zeng, B., Muthuraman, K. and Lawley, M. (2011). Sequential clinical scheduling with service criteria, *European Journal of Operational Research* **214**(3) : p. 780 – 795.

Vanden Bosch, P. and Dietz, D. (2000). Minimizing expected waiting in a medical appointment system, *IIE Transactions* **32**(9) : p. 841–848.

Vanden Bosch, P. M., Dietz, D. C. and Simeoni, J. R. (1999). Scheduling customer arrivals to a stochastic service system, *Naval Research Logistics (NRL)* **46**(5) : p. 549–559.

Vissers, J. and Wijngaard, J. (1979). The outpatient appointment system : Design of a simulation study, *European Journal of Operational Research* **3**(6) : p. 459–463.

Walter, S. D. (1973). A comparison of appointment schedules in a hospital radiology department, *British Journal of Preventive and Social Medicine* **27**(3) : p. 160–167.

Wang, P. P. (1993). Static and dynamic scheduling of customer arrivals to a single-server system, *Naval Research Logistics (NRL)* **40**(3) : p. 345–360.

Weiss, E. N. (1990). Models for determining estimated start times and case orderings in hospital operating rooms, *IIE Transactions* **22**(2) : p. 143–150.

Zeng, B., Turkcan, A., Lin, J. and Lawley, M. (2010). Clinic scheduling models with overbooking for patients with heterogeneous no-show probabilities, *Annals of Operations Research* **178**(1) : p. 121–144.

BIOGRAPHICAL SKETCH

Mr. Sener was born in 1989, in Van. He attended Bursa Ali Osman Sönmez Science High School. After completing high school, He started to study Industrial Engineering at Dokuz Eylül University. Subsequently, he started to working in University of Yalova. He completed language course at METU. He obtained abroad graduate education scholarship of Ministry of Education. Currently, he is charged with a duty in Galatasaray University as research assistant.

PUBLICATIONS

- **Sener, N**, Feyzioglu, O. ; A Risk Averse Stochastic Programming Approach for the Appointment Scheduling Problem, Proceedings of GOR/DASIG 2013, September 23-24, 2013, Hamburg, Germany.

- Turan, H. H., **Sener, N.** ; An Analysis of Profit Maximization Problem of Telecommunications Intermediaries under Uncertainty, The International IIE Conference, June 26-28, 2013, Istanbul, Turkey.

- Turan, H. H., **Sener, N.** ; A Possibilistic Mathematical Programming Approach for Backbone Selection and Demand Allocation Problem in Telecom Networks, Proceedings of the 10th International FLINS Conference on Uncertainty Modelling in Knowledge Engineering and Decision Making (FLINS 2012), August 26-29, 2012, Istanbul, Turkey.

- **Sener, N.**, Turan, H. H. ; An Integration of TOPSIS Method and Interval Type II Fuzzy Sets to Evaluate Municipality Investment Alternatives : The Case of Istanbul Municipality, The 25th European Conference on Operational Research (EURO XXV), July 8-11, 2012, Vilnius, Lithuania.