Modelling activities in a Critical Care Unit

Mari Jones

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Summary

The Critical Care Unit (CCU) is the sector of the hospital where, as the name suggests, critically ill patients receive treatment. The main aim of this research is to identify and apply suitable Operational Research techniques to model patient flow in the CCU at the University Hospital of Wales, Cardiff. The Operational Research techniques employed in this thesis include queueing theory and simulation. These methods have been utilised previously in the field of healthcare with much success.

The thesis begins by considering two aspects of queueing theory, namely batch service queueing theory and batch arrival queueing theory. The latter of these is utilised to model patient flow within the CCU. Although queueing theory may be used as a good approximation to activities in the Unit, it does not incorporate all aspects of real-life. Thus discrete-event simulation is suggested as an alternative approach.

Two types of statistical analysis, CART and Regression, are applied to both length of stay and mortality variables. The results from these statistical tests are compiled and investigated in more depth.

Finally, a discrete event simulation model is built in Visual Basic for Applications, for Microsoft Excel. This simulation model incorporates many of the complexities of a CCU, such as patient priority and cancellation of scheduled patients if all beds on the Unit are occupied. The model is then used to test various "what-if type" scenarios, including the possibility of funding additional beds, the concept of ring-fencing of beds for different levels of care, and the likely effect of reducing the impact of bed-blocking.

Acknowledgements

Firstly, I would like to thank Professor Jeff Griffiths and Dr Janet Williams for their invaluable help and support throughout the past three years. Their supervision has been a great source of encouragement to me during my time as a Doctoral student, and for this I am very grateful.

I would also like to thank Dr Martyn Read for his enthusiasm and expert guidance through the course of this project. I would also like to thank Dr Read for supplying the data that proved central to this PhD thesis.

Finally, I would like to thank my husband, Ben, and my parents, Ceri and Shan, for the hours of encouragement, prayers and support given to me, as well as a huge amount of proof-reading! Diolch yn fawr i chi gyd.

"You see, at just the right time, when we were still powerless, Christ died for the ungodly. Very rarely will anyone die for a righteous man, though for a good man someone might possibly dare to die. But God demonstrates his own love for us in this: While we were still sinners, Christ died for us." Romans 5:6-8

Publications and Presentations

Publications

Accepted:

Griffiths JD, Jones M, Read M, Williams JE. (February 2009) A Simulation Model of Bed-occupancy in a Critical Care Unit. *Journal of Simulation.*

Submitted:

Burton DAM, Blundell N, Elwyn G, Fraser AG, Jones M. (November 2008) Shared clinical decision-making: do patients really want it and are doctors prepared to provide it? *Heart.*

In draft format:

Griffiths JD, Jones M, Williams JE. (February 2009) Batch arrival queues with Hyperexponential service times and c service channels. *Journal of the Operational Research Society.*

Griffiths JD, Jones M, Williams JE. (February 2009) Factors affecting length of stay and mortality in a Critical Care Unit. *Health care management science.*

Evans R, Jones M, Griffiths J, Williams J, Melia J, Kynaston H, Cochlin D, Joseph-Williams N, Edwards A, Newcombe R, Kinnersley P, Wright P, Grol R, Elwyn G (April 2009) A mathematical simulation model of the effect of a web-based Prostate Specific Antigen decision support technology on urological services.

Evans R, Joseph-Williams N, Edwards A, Newcombe RG, Kinnersley P, Wright P, Griffiths JD, Jones M, Williams JE, Grol R, Elwyn G. A randomised controlled trial of the effect of a web-based PSA decision-support technology, *Prosdex*, on informed decision making.

Presentations

Griffiths JD, Jones M, Read M, Williams JE. (September and March 2007) A Simulation Model of Bed-occupancy in a Critical Care Unit. *OR49, Edinburgh and YOR15, Bath*

Griffiths JD, Jones M, Williams JE. (May, July and September 2008) Batch arrival queues with Hyperexponential service times. *Applied Mathematics Colloquium, Gregynog, Welsh Institute of Mathematical and Computational Sciences, Swansea and OR50, York*

Table of Contents

Chapter 1: Introduction

1.1. Introduction

In the United Kingdom the National Health Service (NHS) is a very important provision which affects the lives of all residents. Yet despite substantial increases in healthcare expenditure during the past decade (OECD 2006), problems such as long waiting times for treatment continue to be reported on a regular basis. It is essential therefore that the NHS works to its optimum capacity in order to meet the ever increasing demand for healthcare services in the UK.

Healthcare in the UK is administered through a devolved government department so that each of the constituent countries (England, Northern Ireland, Scotland and Wales) has its own targets and policies. The waiting time target for Wales, where this study is based, as set by the National Assembly for Wales, was that by the end of March 2008 no patient should have been waiting for admission as an in-patient or day case for over 22 weeks. In August 2008, there were 21,290 people on the waiting list for inpatient appointment or day case care. Of these 67 patients had been waiting for 40 weeks or more already, with approximately 2.5% of patients already past the 22 week waiting time target (WAG 2007b). The situation is improving but there is still a long way to go. The corresponding waiting time targets for England were that all patients experienced a wait of less than 18 weeks from referral to treatment (DOH 2008).

One factor which can contribute to waiting time problems is the cancellation of Elective surgery. These cancellations occur for many reasons but can generally be classified into one of three groups, namely hospital reasons, clinical reasons or patient reasons. Clinical reasons (worsening of health, for example) and patient reasons (such as patients rethinking their treatment options) are largely out of the control of hospital managers. However, hospital reasons such as insufficient beds available on the Critical Care Unit (CCU) for recovering patients, or that no surgeon is available, can be addressed by managers and can therefore be

optimised. This research will focus on one of the hospital reasons, that is ensuring that there are an optimal number of beds available in CCU to minimise cancellations of Elective surgery.

1.1.1. Critical Care Unit at the University Hospital of Wales

The CCU is the sector of the hospital where, as the name suggests, critically ill patients receive treatment. In many hospitals, the CCU is split into two Units, the Intensive Care Unit (ICU) and the High Dependency Unit (HDU). In such hospitals, the HDU acts as an intermediate venue for care between the ICU and the ward. The nursing requirements for ICU and HDU are different; patients in the ICU require 1:1 nursing care, whereas patients in the HDU require 1:2 nursing care.

This thesis investigates activities at the CCU at the University Hospital of Wales, Cardiff (UHW). This Unit consists of 24 beds (with 5 additional beds that can be utilised in periods of peak demand) and is the largest in Wales. Since it is the largest Unit, very often patients from all over Wales are treated at this Unit. The CCU in UHW is the amalgamation of the previous ICU and HDU. This amalgamation occurred in 2003 and has been running as a combined Unit ever since. The 24 beds in the Unit can either be utilised as HDU beds or ICU beds. The nursing requirement in the Unit is thus 1 nurse to 2 beds or 1 nurse per bed depending on the severity of illness of the patient.

The Critical Care Unit nurses are specially trained, in short supply and expensive. An important factor is the cost of beds in CCU. In 2005-2006 it was estimated by the Department of Health (DOH 2006) that each CCU bed costs the NHS £1,716 per day, which includes nursing costs. However, in 2006-2007 the Department of Health (DOH 2007) changed their policy on costing and now calculate the cost per patient in CCU according to the number of organ failures they have rather than the average cost of a bed.

1.1.2. Patients

The patients who are treated at the CCU of any hospital are diverse in nature. The patients at the CCU at UHW are especially so since it is the largest Unit in Wales and thus accounts for much of the specialist care for the Welsh population. Patients are admitted from 6 different sources onto the Unit; Accident and Emergency, the Wards, Emergency Surgery, Elective Surgery, X-Ray and Other hospitals. The largest source of admission is Elective surgery which accounts for approximately 28% of admissions. The smallest source of admission is X-Ray which accounts for only 1% of admissions. The source of admission can be sub-classified into two groups; planned and unplanned admissions. The planned admissions are simply those which the hospital has control over which is only Elective surgery patients. The remaining sources of admission are unplanned in nature, whereby the hospital very often has no idea that the patients will be arriving at the Unit. The Clinical Director of the CCU at UHW attempted previously to control the number of admissions from other hospitals (which presently account for approximately 6% of admissions) but the number of admissions has been fairly stable year on year and this source is thus accepted as being uncontrollable also (or unplanned).

The length of stay of patients at the CCU is also varied in nature. Many patients only require a stay of 1 or 2 days in the Unit, until their condition stabilises. However, a small proportion become long stay patients and can stay for many months. In the data set under investigation, the longest length of stay was 374 days.

A large proportion of patients admitted from Elective surgery have a short length of stay. The reason for this is that following some major operations patients require forms of life support such as ventilation, for a short time and once their bodies have sufficiently recovered from the surgery they require no further treatment and are thus discharged to a ward.

The unplanned admissions have a more varied length of stay profile. Very often they are admitted onto the CCU as they are acutely ill and have one or more failed organ system. Others are admitted as a result of major trauma. Intuitively, it is clear that the unplanned admissions will have a diverse length of stay profile.

1.1.3. Data

The UHW is the largest hospital in Wales. It has an average of 1064 beds available each day with 86.1% occupied (WAG 2007a). The CCU is also large, with 24 beds currently employed, but with space available for a further five beds, which may be utilised at times of peak demand.

The data set used in this study, which is routinely collected, has fairly complete records between April 2004 and May 2007. It contains a vast amount of information about each patient who is admitted into the CCU. Firstly, it has information about source of arrival of the patient, and at what time and date they arrived. Secondly, it has discharge data, which is the patient's destination and the time at which they left the Unit. Thirdly, it includes over a hundred physiological factors such as body temperature or blood pressure, measured on each day the patient is in the CCU. Fourthly, it has information about the treatment received; the primary reason a patient was admitted into the CCU, and several other diagnostic variables. Finally, it has information specifying the medical specialty of the patient's consultant.

Some interesting variables worthy of mention are the Acute Physiology and Chronic Health Evaluation score (APACHE), the Simplified Acute Physiology score (SAPS) and the Therapeutic Intervention Scoring System (TISS score). The APACHE score ranges from 0 to 71 and is calculated based on 12 physiological scores. The higher the APACHE score, the more severely ill the patient. The SAPS score is similar to APACHE and is calculated by a score between 0 and 163 based on 12 physiological factors. The TISS score is an integer value between 0 and 200 and is calculated from a set of 76 variables. Many of the variables, in contrast to APACHE and SAPS, relate to the therapies

a patient has received rather than physiological factors. The TISS score will be discussed in more depth in Chapter 4.

In the UHW there is a separate Paediatric CCU, so only data on patients aged sixteen or over were used in this analysis. With these restrictions, the data set available for use in the analysis has 4226 patients.

1.2. Objectives of thesis

The Critical Care Unit is highly resource intensive and deals with a variety of patient groups. Also, poor management of this facility can result in many problems for the NHS as a whole such as the cancellation of Elective surgeries. This thesis proposes several models which address these issues and predicts improvements based on varying the parameters.

The main aim of this study is to produce several complementary models using suitable Operational Research techniques which can model patient flow at the CCU at UHW. More specifically, four objectives can be identified:

- Identify and investigate appropriate queueing systems which can model activities at the Critical Care Unit
- Understand the factors which affect length of stay and outcome at the Critical Care Unit
- Use the insight gained from the previous objectives to build a simulation model of the Critical Care Unit.
- Analyse the results from the simulation model to show how varying some parameters will affect cancellations and cost.

These objectives are addressed in the thesis in the following way:

Objective 1 is addressed in Chapters 2 and 3 where two queueing models are discussed namely batch service queues and batch arrival queues. The novel model, $M^{(b)}/He/c$ is presented and analysed.

Objective 2 is addressed in Chapters 4 and 5. Both Chapters begin by considering two statistical analysis techniques, CART and Regression, applied to length of stay and outcome. The influential variables in both models are discussed and analysed in greater depth. Finally, a novel approach, using five logistic regression equations is suggested which can predict the daily probability of a patient remaining in the system until the next day. Once again, the influential variables are highlighted.

Objectives 3 and 4 are addressed in Chapter 6 whereby a discrete event simulation model is built in VBA which seeks to imitate the real-life practice of the Critical Care Unit. This simulation model is then adjusted to account for factors such as bed-blocking and a change of working pattern in the Unit. The results are presented and analysed.

The remainder of this Chapter acts as a thorough review of the literature. The first two sections, 1.3 and 1.4, review the literature from the middle of the last century to the present day, relating to batch queueing theory, whereas section 1.5 concentrates on advances made in the past 10 years alone in the field of modelling length of stay, mortality and patient flow, due to the way that medical practice has developed recently.

1.3. Batch service queueing theory

This section aims to chart the evolution of the theory and application of batch service queuing systems from their earliest roots in the middle of the last century to the present day. In order to achieve this, a thorough literature search was undertaken and the relevant papers are summarised below.

1.3.1. 1950s

Much of the foundational work of queuing theory was undertaken in the 1950s and one of the earliest pieces was by Kendall (1951). His paper outlined some of the key features of queuing theory and considered a system with Poisson arrivals with a single server and a General service–time distribution; the M/G/1 queue. He also introduced the concept that even though fluctuations in queue size are not usually Markovian in nature, an imbedded Markovian technique can be adopted. In a later paper Kendall (1953) extended his analysis to include regular arrivals with up to three servers.

Shortly after the publication of the second of Kendall's papers, the concept of batch (or bulk) service from a queue was introduced by Bailey (1954). In this influential seminal study, he assumed that the inter-arrival rate followed the χ^2 distribution and utilised imbedded Markov chains to find the solution. He was also the first to suggest an application of this theory in the medical field. He studied waiting times for an out-patient appointment with a hospital consultant and concluded that if a clinic were held once per week and the consultant were prepared to see at least one more patient than the average demand per week, then the average waiting time for an appointment would not be greater than one week.

During the following year Downton (1955) published results which were complementary to, and added to, the work of Bailey. He calculated the variance of the waiting time distribution and was the first to publish tables of summary statistics for waiting times (for values of ρ of 0.8 and 0.9). Shortly afterwards, in

an extension to his own work, Downton (1956) considered the case of Negative Exponential service time distributions and the limiting properties of batch service queues. He noted the similarity of these limiting results to the solution for multi-server queues with random arrivals and regular service.

Again in 1956, in a transport application, Luchak (1956) investigated Poisson arrivals and General service times (which he called holding times). This work was investigatory in nature and examined what happened if traffic intensity was a continuous function of time. He considered the system for various holding time distributions.

The end of this decade saw early consideration of the impact on waiting time distributions of batch arrivals to a queue (Gaver 1959) as well as batch arrivals and batch service (Miller 1959). These papers will not be considered in this section since they are largely focussed on batch arrivals as well as batch services.

1.3.2. 1960s

The 1960s saw the start of the move from steady state solutions to time-dependent solutions. As one of the early researchers in this area, Jaiswal (1960a) utilised the phase method and differential-difference equations to find the time-dependent solution to the queuing system presented by Bailey in 1954. He considered particular cases for the service distribution, including Erlang. In the same year Jaiswal (1960b) again extended Bailey's work to encompass the situation where the maximum number that can be served at one time is dependent on the number of units already present within the server, as well as the capacity. In this paper he compared two methods of solving this system - the imbedded Markov chain technique and the phase method utilised in his earlier work. He argued that although the phase method is simpler and useful for certain cases, the imbedded Markov model is statistically more sound and of greater general use.

Fabens (1961) used the theory of semi-Markov processes to further investigate the batch service queueing system. Two basic theorems were developed; the first

found a relationship between the probability of the Markov chain reaching a given state, the mean of the waiting time distribution and the mean recurrence time; the second looked at the limiting behaviour of the Markov chain transition probabilities. An equivalence theorem was found indicating that the batch service queue is equivalent to a single server queue with Gamma inter-arrival distribution and a General service distribution. He also published the waiting time distribution and many other summary measures. Two years later, Fabens and Perera (1963) submitted a correction to the previous work which suggested that there was a small error in the waiting time distribution that Fabens found earlier, simply as a result of an incorrect assumption regarding independence.

During this period the concept of multiple servers was introduced. Arora (1964) developed an extension to Bailey's seminal work incorporating two-servers with different capacities. He also presented results for the mean queue length for a special case (where each server has the same mean service rate) and found the distribution of the busy period. In an extension to the work of Arora (1964), Ghare (1968) generalised the system to incorporate c ($c > 2$) service channels.

The work of Roes (1966) further extended the system to consider n service channels, each of which was a batch service facility. In this study, arrivals were not assumed to be Markovian while service time was assumed to have a Negative Exponential distribution. Roes found waiting time distributions for a number of scenarios including service in order of arrival (FIFO), service in random order (RIRO) and service in inverse order (LIFO).

In the previous year Neuts (1965) studied the busy period of a batch queue. He discovered that the busy period was equal to the time between successive visits to the state zero in an imbedded semi Markov process. He also obtained the transform of the distribution of the busy period.

1.3.3. 1970s

The 1970s saw consideration of special cases of the queueing systems already derived. Love (1970) considered the system where there were S servers with only batches of size Q being served. Arrivals were assumed to be Erlangian in nature and service Markovian. Love presented summary measures including the mean waiting times in the system and in the queue.

In 1972, Bahary and Kolesar (1972) used the imbedded Markov chain technique to study bulk service queueing systems. The system they chose to analyse had a variable service capacity and a variable service mechanism which depended upon the number of customers waiting in the queue when service began. In addition, they developed the concept of 'scale' which allowed the characteristics of small systems to be used to develop the characteristics for large systems.

Chaudhry and Lee (1972) derived probabilities and calculated the transient steady state solutions for a bulk service queue with Poisson arrivals and alternate Exponentially distributed service intervals with different means. On a similar theme, three years later Medhi (1975) considered a queue with Poisson input and an Exponential service distribution with the number served being greater than a and less than b, the acceptance threshold.

In 1975, Griffiths and Cresswell (1976) applied the theory of batch service queues to a Pelican crossing. In this, the first application in this domain, several different aspects of the system mechanism, were considered such as the distribution of the queue of cars and pedestrians.

1.3.4. 1980s

The 1980's saw the emphasis shift onto a general service time distribution with Markovian arrivals.

In 1981, Chaudhry and Templeton (1981) studied the M/G(B)/1 queueing system. They assumed Poisson arrivals and a General batch service distribution with maximum batch size B. They aimed to generalise Jaiswal's queueing system ($M/E_k(b)/1$), derived twenty years previously, to find a relationship between the probability generating functions of the number of customers in the system at a random instant of time and the number in the system at post-departure instants. They discovered a simple derivation of the waiting time distribution having utilised the supplementary variable technique. They were also able to demonstrate that Little's result holds for this queueing system. Graphical representations of the roots of the characteristic equations were included and these are complementary to ones found in Chapter 2 of this thesis.

In a traffic application, Griffiths (1981) extending his previous work, applied batch service queueing theory to pedestrian crossings. The system assumed that vehicles arrived at a crossing at mean rate λ and pedestrians arrived at the crossing with mean rate μ. Car arrivals were described by a displaced Exponential distribution and groups of pedestrians took time α to cross the road (α was fixed for each individual pedestrian crossing). When vehicles were present in the queue, they dispersed at constant time intervals β (once the crossing is empty). Distributions of effective red and green periods were considered as well as probabilities regarding the number of vehicles in the queue at regeneration points. Pedestrian queues were also discussed.

Neuts & Nadarajan (1982) studied the (M/M(a,b)/N) queue, considering the marginal density of the number of customers waiting, the stationary probability that there were j servers free and the waiting time distribution.

In 1983, a book was written by Chaudhry and Templeton (1983) which discussed both batch arrival and batch service queues. Much of their previous work was summarised in this volume, as well as some interesting special cases, such as infinite time-dependent queues.

In 1985, Madill, Chaudhry and Buckholtz (1985) investigated the solution of a deterministic queueing system. In an extension of Medhi (1975) the system had a single server providing service in batches with an acceptance threshold (minimum batch size a, maximum batch size b). Analytic results were found for the steady state probabilities at both random and pre-arrival instants. The solution was not trivial and therefore graphical solutions were discussed but the system required the traffic intensity (ρ) to be rational.

Two years later Chaudhry, Madill and Briere (1987) considered the (M/G(a,b)/1) queueing system. They employed the supplementary variable technique to find the post-departure and arbitrary time instant probabilities and then used the results to find various summary measures including the mean queue length and the mean waiting time in the system.

Liu, Kashyap and Templeton (1987) introduced the concept of allowing the customer to choose which of the queues to join when each queue had a different service distribution and the number of servers was infinite. In this piece of research, the transient results for the mean and variance for the number in the system were obtained, in addition to the waiting time distribution and steady state results.

Once again in 1987, Ozekici (1987) considered the case where arrival and service rates were not independent. He analysed and exploited the relationship between the arrival and service processes with emphasis on the impact of this relationship on mean times. He applied the work to the transportation arena.

In 1988, Jacob and colleagues (1988) explored the impact of a limited waiting area. They considered a single server bulk service queueing system where

arrivals were assumed to be Markovian and any arrival which occurred when the waiting area was full was lost.

Later in that same year, Bertsimas and Papaconstantinou (1988) considered the system where the service distribution was assumed to be Coxian-2 with s service facilities. Batches were served with a minimum size a and a maximum size b. The method of stages was used to find the steady state probabilities. They found the probabilities of each state where all servers were busy and found waiting time, idle server and service batch distributions.

Finally, in 1989 Briere and Chaudhry (1989) constructed algorithms to find the solutions to the queue M/G[Y]/1 using transform methods. They found the steady-state probabilities and moments of the number of customers in the system at three instants of time: post-departure, pre-arrival and random. Then, they presented numerical results for the deterministic, Erlang, Hyperexponential and Uniform distributions.

1.3.5. 1990s

The early 1990s heralded the introduction of more flexibility into queueing systems. For example Reddy, Nadadajan and Kandasamy (1991) considered the situation where additional servers were available if the queue length exceeded a value Y with server vacations, and Zili, Wang and Li (1991) investigated a novel idea of a non-symmetric cyclic queueing system. Briefly, a cyclic queueing system consists of N stations with k queues at each station and a single server who moves along the cycle. A little later Lee and colleagues (1992) examined the batch service queueing system with a single vacation policy. In this instance if no customers were present the server remained idle.

In a technical method paper Chaudhry, Gupta and Madill (1991) undertook a study comparing two analysis techniques: the Jacobi and the root-finding method, and reported steady state probabilities, means and variances for the number in the system. They also considered a special case for a specific batch

size b and published a number of performance measures for this queue including blocking probabilities. Numerical results were obtained for several service time distributions including Erlang, Deterministic and Hyperexponential. They concluded that the Jacobi method is less cumbersome than the root-finding method.

Griffiths and colleagues (1991) applied batch service queueing theory to the Channel Tunnel. The main focus of the study was to investigate whether the planned provision of freight service in the Tunnel was sufficient to deal with the demand. Specifically, an approximation to the time-dependent solution of a bulk service queue was utilised.

Next the emphasis moved to consideration of a constant service time distribution. Ryden (1993) looked at the system where the time between services was constant. Arrivals were assumed to be Markovian, and batches were served with a maximum batch size of n. A buffer of size L was included in the system. As an extension to the work of Roes (1966) the authors analysed this system looking at various queue disciplines including FIFO and LIFO. They also introduced the concept of age-related service where if the buffer was full, then an arriving customer could either be lost or could replace the customer who had been waiting in the queue for the longest length of time. Sample results (for the waiting time probability distribution functions) were included for various scenarios. Alfa and Yannopoulos (1993) and Sivasamy and Senthamariakannan (1994) extended this work by including the idea of renewal processes.

In 1994, an extension of the work on server vacations by Reddy and colleagues (1991) was undertaken by Choi and Han (1994), who introduced multiple vacations with vacation times which followed the Negative Exponential distribution. The authors used the supplementary variable technique to obtain queue length probabilities, and the shift operator method to solve the resulting equations. Results were verified by comparing them to the GI/M/1 queueing system with multiple vacations with single service.

Griffiths (1995) applied batch service queueing theory to the Suez Canal in a case study paper. The capacity of the Suez Canal was increased by 44%, by simply changing the cycle times from 24 hours to 48 hours.

A further development occurred in 1996 when Baba (1996) introduced the concept that service rate could depend upon the number in the batch being served. He showed that the queue size and service batch size at arrival instants form imbedded Markov chains and found the steady-state probabilities.

In 1997 Selim (1997) considered a traffic applications system where customers were served in batches of maximum size n. Time-dependent probability distributions for the number of customers in the system were obtained and the solution was used to predict the optimal service rates.

Reddy and Anitha (1998) also considered the possibility of multiple vacations, but the focus of their work was on determining the stationary distribution of the number of customers in the queue and the waiting time distribution of an arriving customer. Various numerical examples and expected length of queue were reported.

The end of this decade saw an extension to the work of Ryden (1993) when Laxmi and Gupta (1999) considered a buffer system with Exponential arrival and service times. Distributions for the number of customers in the queue were obtained at pre-arrival instants and arbitrary times. Analysis of waiting time in the queue was also carried out. In that same year Hebuterne and Rosenberg (1999) found a relationship between steady-state probability distributions of the buffer occupancy at arrival and departure instants.

In another technical method paper Chaudhry and Gupta (1999) put forward an alternative approach for solving the finite capacity bulk service system. Arrivals were assumed to be Markovian and service times had an arbitrary distribution. The queue was analysed using the imbedded Markov chain technique and the

supplementary variables technique. Relations between state probabilities at departure instants and arbitrary instants were presented.

Holland and Griffiths (1999) developed an approximate solution to the mean number of customers in a queue for the time-dependent M/M(1,s)/c queue. The Coordinate transformation solution method was utilised where steady state queueing results were transformed to deal with traffic intensities of 1.

Finally, Worthington and Hall (1999) built on previous works by Worthington and colleagues (Omosigho and Worthington, 1985, Brahiami and Worthington, 1991) to consider the time-inhomogeneous batch queueing system, M(t)/G(0,c)/1. They utilised discrete time modelling methods to evaluate this queue. They discovered that although solution algorithms were developed, they proved to be less good than for the single service case.

1.3.6. 2000s

The new millennium heralded the beginning of investigations into more complex queueing systems. For example, in 2000 Krishnamoorthy and Ushakumari (2000) considered the system whereby customers were served in batches but departed individually. Also, the concept of accessibility came into the forefront. In this paper, the batches were accessible which means that if an arriving customer found service underway but the maximum batch size had not been reached, the arriving customer could join the service facility to complete the remaining service time. The authors found the number in system probabilities in both transient and steady-state conditions. They also found the waiting time distribution, the busy period distribution and other summary measures as well as the optimum values for maximum batch size and system capacity. In addition, they showed that Little's formula holds for this queueing system.

In the same year Adan and Resing (2000) considered a specific multi-server batch service queueing model. Arrivals followed a Poisson process, service batch size must be between a and b and there were two types of service distribution: Coxian-2 and Erlang-r. Equilibrium probabilities for states with all servers busy

were expressed as a sum of geometric terms. The waiting time distribution was also deduced and summary measures computed.

In 2002, Xia and colleagues (2002) considered the single server batch processor. The system under consideration had a set of parallel queues and only customers from the same queue could be included in a batch. The batch service distribution was Exponential. When both finite buffers of equal size and infinite buffers were present, the server was allocated to the longest queue to maximise the throughput of the system. The most interesting results occurred when there were unequal buffer sizes. Here simulation methods were used to determine the best system for service.

Again in 2002 Tadj and Tadj (2002) examined the single-channel bulk service queueing system with the added dimension of a specified accumulation level. Given the accumulation level r, the server stopped processing customers when there were fewer than r customers present, then resumed when there were r present. They developed a recursive numerical procedure which calculated the steady state probabilities. To assess the accuracy and efficacy of the recursive procedure they compared it to an exact solution procedure. Effectiveness and computation time were compared with other approximation techniques.

A year later, Tadj (2003) investigated the bulk queueing system under D-policy. Customers were served in batches of r units if queue length was larger than r. Otherwise service was delayed until the cumulative service time of the customers in the queue first reached D (hence the label D-policy). Steady state probabilities were then derived at departure instants. Special cases were considered and examples were provided.

In that same year Tadj and Sarhan (2003) developed an extension to the earlier work of Tadj and Tadj (2002) and evaluated the optimal control of a single channel bulk service queueing system with random service capacity with an accumulation level R. A search procedure was designed to determine the value of R that yielded the minimum expected cost per unit time. Sensitivity analysis was

then conducted to assess the validity of the results. This was the first time that the cost dimension was added to batch queueing systems.

In 2004, Chaudhry and Chang (2004) considered the discrete time bulk service queue with variable capacity and finite waiting space - an extension of the work of Jacob et al (1988). Arrivals were assumed to follow a Bernoulli distribution and the service distribution was General. Analytic and computational aspects of the distributions for the number of customers in the queue (at departure instants, arrival instants and random instants) were discussed.

Again, that same year, Gupta and Sikdar (2004) investigated the single server finite-buffer bulk service queue. They considered the arrival distribution to be Exponential and the service distribution arbitrary with maximum batch size w and minimum batch size v. A single vacation was taken when less than v customers were present and the server was available. The distribution of the number in the queue at various time instants was calculated and summary measures were obtained.

The following year saw the introduction of a new approach – the use of fuzzy logic. Chen (2005) considered a queueing system using fuzzy logic where arrival and service rates were deemed to be fuzzy numbers. They transformed a fuzzy queue into a family of crisp queues by using the alpha-cut approach. Using these alpha-cut representations, two parametric nonlinear programs were formulated to describe the family of crisp queues.

Again in 2005, Denteneer and Van Leeuwaarden (2005) considered the system where the arrival process was linked to the number in the queue, assuming that a fixed minimum delay exists between arrival instant and service commencement. This is called the delayed bulk service queue. Higher dimension Markov chains were required to characterise this system. Approximate bounds for the mean queue size were derived and simulations showed interesting results.

Alfa (2005) introduced the GI/G[y]/1 queuing system where both arrivals and service were assumed to follow a General distribution in discrete time.

In 2006, Goswami, Mohanty and Samanta (2006) once again considered a queueing system in discrete time and introduced the concept of accessibility in discrete time. Arrivals and service times were assumed to be geometrically distributed and a buffer was present in this system. Late entries could join the service part way through as long as there were fewer than d customers in the server (where d is less than the maximum batch size). If there were fewer than d customers in a batch when the late entry arrived, then the batch was labelled 'accessible', if not it was labelled 'non-accessible'.

2006 saw the introduction of another new approach with Armero and Conesa (2006) using Bayesian statistics as well as queueing theory to analyse the congestion of bulk service queues.

In 2007, Yi and colleagues (2007) developed an extension to the work of Chaudhry and Chang (2004) and considered a system with Bernoulli arrivals and batch service, where service commenced when the number in the queue reached a threshold value. The authors derived queue length distributions just after a service was completed and deduced a relationship for the queue length distribution at a random moment in time. They then evaluated the mean queue length and other summary measures.

Finally, Claeys and colleagues (2007) considered the system where a batch server operated with a minimum batch size l and a maximum batch size c. They developed the probability generating function for the number in the system when there was a dependency between the number in the batch and the service time of a batch.

1.4. Batch arrival queueing theory

The next section of the literature review will consider the development of the batch arrivals queueing system. The roots of this system are again found in the middle part of the last century and one of the first researchers to study batch arrivals was Donald Gaver. He considered the system whereby groups of customers arrived at a single service facility according to a stationary compound Poisson process and utilised imbedded Markov chains to investigate the busy period (1959).

1.4.1. 1960s

Developmental work on the batch arrivals queueing system continued into the 1960s and one of the first papers published during that decade was by Conolly (1960). He presented an extension to the work on batch service queues undertaken earlier by Bailey (1954), and considered the bulk arrival single service system where customers arrived according to a General distribution and are served according to the Negative Exponential distribution. He also used the imbedded Markov chain technique.

A few years later Hawkes (1965) was the first to consider the time-dependent solution of a queue with bulk arrivals operating with a priority setting. He used Laplace transforms to consider a case for two classes of arrivals (priority and non-priority) and derived the equilibrium distribution (steady-state) for both classes of arrivals, as well as the distribution of the number of customers remaining in the system immediately after a departure occurred. The mean queueing times were also calculated for the case where the service distribution was Negative Exponential.

Later that year Gupta and Goyal (1965) considered the queueing system M(x)/He/1, that is, Markovian batch arrivals (x is a random variable here) with one Hyperexponential service facility. The authors derived the Laplace

Transform for the number in the system using the imbedded Markov chain method adopted by Bailey (1954). This work is extended upon in Chapter 3.

Abolniko (1967) was the first to consider multi-server batch arrival queues and the concept was further developed by Reynolds (1968). Reynolds investigated bulk arrival queues with an infinite number of servers and Poisson arrivals. The number of customers within a batch was random. The time-dependent queue length was investigated and a generating function expression was derived. This distribution was used to show that the limiting distribution is compound Poisson. Interestingly, Reynolds showed that the regression model of the queue length at time t initially is linear, and this expression was used to find the autocorrelation function.

In the same year Kabak (1968) was the first to consider c service facilities when he developed the M(n)/M/c batch arrival queue (with n arrivals in each batch). He examined the blocking probabilities for a loss system and a delay system, and calculated the mean and variance of the delay time along with other numerical results.

1.4.2. 1970s

Two years later Kabak (1970) extended his earlier (Kabak 1968) work to look at the M(x)/M/c queueing system where x is a random variable. He calculated the steady state probabilities, the blocking probabilities and the mean and variance for the delay time.

In that same year Harris (1970) extended the current theoretical base to consider the system where the service time was state-dependent. The single service facility system was deemed to have a service distribution which was dependent upon the number of customers in the queue. The imbedded Markov chain technique was utilised and numerical results for steady-state probabilities and expected queue sizes were included.

The issue of the time-dependent solution of a queueing problem with correlated batch arrivals and a general service time distribution was first considered by Murari (1972). In this paper the author found Laplace transforms of a variety of probability generating functions of queue lengths and used these to derive special cases.

Later in that same year Halfin and Segal (1972) returned to the issue of priority (this time defined as primary and secondary groups) within the queue thereby extending the work of Hawkes (1965). The primary group of customers immediately entered service if there was a facility available, if not, they were sent away. Their service time was Negative Exponential. The secondary group had a General service time distribution and entered service if there was a service facility available, if not, they were sent to the buffer. The moments for the number of customers in the buffer were derived for steady state conditions.

During the mid 1970s Dagsvik wrote two papers (1975a, b). In the first of these, he outlined the waiting time process of the single server bulk queue and found a corresponding waiting time equation. The second paper was an extension to this whereby the inter-arrival or service time distributions were a linear combination of Erlang distributions. In order to solve this queueing system he used algebraic methods (Wiener-Hopf). He also noted that the solution of the waiting time equation may be reduced to the problem of fitting a Hyperexponential solution in a modified equation when the service distribution is Hyper-Erlang.

One year later Cohen (1976) investigated the GI/G/1 queueing system with batch arrivals and individual service. He considered the steady state results for the waiting time distribution using the theory of regenerative processes.

1.4.3. 1980s

The early part of this decade saw the introduction of the use of algorithms into modelling methodology and the emergence of case-study techniques to highlight issues surrounding very specific queueing problems.

In 1981, Van Hoorn (1981) used algorithms to find the state probabilities in a single server queueing system with batch arrivals. The author developed this theory for finite or infinite capacity models where arrival batches could be totally or partially rejected. Baily and Neuts (1981) also derived an algorithm for computing the steady state probabilities for the c-server queue with Exponential service times and bounded group arrivals. Probability densities for queue length were found at various time instants. Algorithmic methods for the waiting time distribution for a customer in a batch were derived.

That same year saw one of the first explicit case study papers using the batch arrival queueing in a clinical setting (Lopezsoriano et al. 1981). Different hospital departments release their staff at different times for their lunch, enforcing a batch arrival queueing scenario. The authors sought to optimise the system such that long queues and excessive waiting times for customers during the lunch period were minimised. What-if type scenarios were tested and the performance of the system was evaluated.

A few years later Yao, Chaudhry and Templeton (1984) produced bounds for the mean waiting time and the mean queue length for bulk arrival queues by utilising the established results for the bound of single arrival queues. The two main forms of bulk arrival queues GI(x)/G/1 and GI(x)/G/c were discussed. At around the same time Kimura and Ohsone (1984) used a diffusion approximation to calculate the steady state distribution of the number of customers in the system.

In 1986 Kulkarni (1986) investigated a very specific but interesting situation whereby two types of customers (for example, customers with different priority) arrived in batches at a single server queueing station with no waiting room. If a

customer arrived and found the server busy, the customer would immediately enter a Negative Exponential holding time and would then try to enter service again. The expected waiting time for the customers was derived. A little later Falin (1988) developed this work further in order to derive the waiting time for two types of customer arriving in batches. The author suggested a different method for solving this problem which extended Kulkarni's work to look at the case where there are greater than two types of customer.

The previous year, in a more technical method paper, Briere and Chaudhry (1987) looked at the computations required for the bulk arrival queueing model GI(x)/M/1. Firstly, the roots of the characteristic equation were found, then the roots were utilised to find the moments and the steady state distribution of the number of customer in the system. Briere and Chaudhry (1988) extended this work in 1988 to look at the queueing system M(x)/G/1 with batch arrivals. They consider 4 types of service distribution: Hyperexponential, Erlang, Deterministic and Uniform for which they found the limiting distribution of the number of customers in the system at a random instant in time.

In an extension to a previous study looking at finite or limited waiting areas, Jacob and colleagues (1988) discussed the queueing system with General inter-arrival and service time, one service facility and finite waiting space. The authors suggested two different rejection strategies which come into force when a batch arrives while there is not enough space in the buffer: the entire batch is rejected or only the excess is rejected. The rejection probabilities were calculated and the waiting time distribution was considered.

At the end of this decade Lee (1989) was the first author to consider the batch arrival queueing system where the server had vacations. He considered two cases of server vacation: if the server finds no customer present in the system when the server returns from vacation, another is immediately taken, or if the server finds no customer present when the server returns from vacation, the server waits until the next group of customers arrive. The server idle probabilities and waiting time

distributions were derived as well as the mean number of customers in the system.

1.4.4. 1990s

This decade saw the blossoming of interest and research in the field of queueing theory. It was also the time when the impact of queueing systems on cost was beginning to be acknowledged as a factor for consideration. It began by the introduction of the systems where priorities could be set in anticipation of certain queueing events occurring. Takahashi and Takagi (1990) discussed a single server priority system where the arriving batch contains customers from many different classes. Preemptive and non-preemptive priority rules were discussed and the supplementary variable technique was utilised to find the queue length and waiting time distributions. In an extension to their previous paper, Takahashi and Takagi (1991) considered the M(x)/G/1 priority queue with single or multiple vacations. Again preemptive and non-preemptive priority rules were discussed. Also in that same year Pechinkin (1990) discussed a batch arrival queueing system with absolute continuous priorities. The steady-state distributions for the waiting time and the number of customers in the system were found. In a further extension relating to priority queues, Towsley and Tripathi (1991) considered a priority queue with two classes of customers. This time, the server could fail and when it did, the system emptied. The queueing system was described as M(x)/M/1 with General repair times and Exponential inter-failure times.

In the previous year Van Ommeren (1990) considered approximations for the waiting time probabilities of individual customers for the M(x)/G/1 queueing system. A variety of batch and service time distributions were examined yielding accurate results.

In a development of the work of Kabak (1968), Fakinos (1990) considered the equilibrium behaviour of the M(x)/G/k loss system. If an arriving batch did not find enough available service facilities for the whole batch, the excess were lost.

The author began by deriving explicit results for particular service time distributions and concluded by discussing specific results for the situation where the service distribution was general and there were two service facilities.

The following year Lee and Lee (1991) introduced the concept of multiple server vacations to the batch arrival queueing system, in which the vacations were differently distributed. The distributions of the number in the system and the waiting time in the queue were derived.

Also at that time, Fakinos (1991) introduced the concept of group departure after a bulk arrival queueing system.

During that time a number of papers were published which adopted new methods of solving batch arrival queueing problems. Firstly, Stanford and Pagurek (1992) utilised generating functions for the serial covariances for number in the system for the GI/M/1 queue with a fixed batch size. Next Moustafa and Elsayed (1992) examine the matrix-geometric solution for the M(x)/C2/s queueing system (C2 refers to the Coxian-2 distribution). Analytic expressions for the mean number of customers in the system and the mean queue length were found. A year later Ferrandiz (1993) used Palm-Martingale calculus to find the queue length moment generating function for the BMAP/GI/1 queueing system (Batch Markovian Arrival Process) with server vacations. A new notation was introduced for the Batch Markovian Arrival Process queue – the BMAP queue. He also extended the concept of vacation to include the time when an arriving customer finds an empty service facility. In this case, a set up time was required before the server could commence with the next service.

In 1992 members of the same research group – Chaudhry and colleagues – published widely on the M(x) system of queues. Firstly Chaudhry, Templeton and Medhi (1992b) extended the work of Briere and Chaudhry (1988) to look at the queueing system M(x)/D/c with batch arrivals. Numerical results and graphs were included for interest. Next in that same year, again as an extension to the work of Briere and Chaudhry, Chaudhry and Gupta (1992) studied the queueing

system M(x)/G/1 with batch arrivals. They found the waiting time distribution for the first and a random customer in an arriving group. They considered different service distributions, including Hyperexponential, Erlang and Deterministic. Finally in that year Chaudhry, Gupta and Agarwal (1992a) suggested an alternative method for deriving the limiting distribution for the number of customers in the system for the M(x)/G/1 queueing system. The roots of the characteristic equation proved to be useful here.

1994 saw the introduction of a number of new ideas. Initial investigations of queues in series was undertaken by Zhu (1994), while Vinck and Bruneel (1994) were the first to publish work on the G(G)/GEO/1 queueing system (General inter-arrival distribution, General batch size distribution and Geometric service time distribution). The probability generating function of the number of customers in the system was found at various time instants. The method of solution is suggested also for the G(G)/G/1 queueing system. The BMAP queue was considered by Schellhaas (1994). This study investigated the system where the server could take two types of vacation. The steady state equations were derived using the imbedded Markov chain technique. It is also one of the first studies to include cost considerations. Early investigation of the generalised switch batch Bernoulli arrival and general service time process (Generalised SBBP/G/1) was undertaken by Ishizaki and colleagues (1994) . The batch size and the service time distributions followed a discrete-time alternating renewal process with states 1 and 2. The main focus of the research was to derive analytic results for discrete time queues and to show possible applications of this system.

Stadje (1994) considered the busy period of the M(x)/M/1 queueing system. The main object of interest was the queue size distribution.

Shanthikumar (1994) examined the convexity of the waiting time in a batch arrival queue. The author demonstrated that the number of customers in the G(x)/GI/1 queueing system and the G(x)/M/c queueing system was component-wise convex in x.

Moustafa and Elsayed (1994) extended their previous work (1992) to include a finite buffer. Expressions for the mean number of customers in the system and the mean queue length were found.

In 1994, Falin (1994) developed his own previous work (1988) to look at the problem of M/G/infinity queues with batch arrivals where there were k types of arrivals. Falin also developed the transient solution to this problem.

Cong (1994) considered the M(x)/G/infinity queue (as Falin above) and derived an expression for the joint probability generating function of the number of customers of type i being served at a fixed time *t*.

In 1994 Bocharov and Yakoutina (1994) considered the GI/G/1/r queueing system with batch arrivals. They particularly considered the case where the service time was of phase-type. The Laplace-Stieltjes transformation of the steady state waiting-time distribution was found.

The following year saw the publication of a paper which acknowledged the effect of queueing on the cost of running a system. Lee and colleagues (1995) examined the M(x)/G/1 queueing system with N-policy and a single vacation. When the server is empty, the server went on vacation of a random length. If there were more than N customers present in the system when the server completed the vacation period, the server began to serve the customers, otherwise the server remained idle until there were N customers present. The distribution of the number of customers in the system was derived and also the distribution of the waiting time. A procedure for finding the optimal stationary operating policy under a linear cost structure was also outlined.

That same year, a further study was published on the impact of preemptive priorities. Langaris and Moutzoukis (1995) examined a retrial queue which accepted two types of customers with correlated batch arrivals and preemptive priorities. The service distribution was arbitrary and different for each customer type. When the server became free, a single vacation was taken. Transient and

steady state probabilities were obtained and the virtual waiting time for a customer was deduced.

Sharma and Sharma (1996) investigated the time-dependent queueing system where the batch arrival rate depended upon the service type. Each customer must go through one stage of service; after this, if the number in the system was low, they could go through a second optional service stage. The researchers then derived Laplace transforms of the probability generating functions for the number of customers in the system.

A little later in the decade Chaudhry and Gupta (1997), in an extension to Vinck and Bruneel (1994), considered the discrete time GI(X)/Geom/1 queueing system. The supplementary variable technique was utilised for early and late arrival systems. The distribution of the number of customers in the system was derived at pre-arrival instants. Numerical results were included for different inter-arrival distributions and batch size distributions. The following year Chaudhry and Gupta (1998) extended this work by considering the discrete time GI(X)/Geom/1 queueing system with a finite buffer. The supplementary variable and imbedded Markov chain techniques were used to find the queue size for early and late arrivals. Loss probabilities for batches and customers were also discussed as well as a waiting time analysis.

Also in that year the M(X)/M/1 queueing system with bilevel control was considered by Lee and colleagues (1998). Once the system was empty, the server was idle until there were m or more customers present. Once there were m customers present, the server began a start-up process and then would commence service when there were at least n customers in the system. The distribution of the number of customers in the system and the waiting time distribution were derived. The same system was studied a year later by Liu and Tseng (1999). In this paper there were two control threshold values of interest, k and n. If the number in the system was more than n, the service rate was switched from u to tu. If the number in the system dropped off to k (k is smaller than n) then the service rate was switched back from tu to u. The steady state probabilities for the

number in the system were calculated. The concept of a cost model was introduced where there was a service cost, a queueing cost and a switching cost.

A year earlier Choudhury (1998) studied the steady state behaviour of a batch arrival Poisson queue with a random set-up time and a vacation period. Before the first customer from each new batch was served, the server undergoes a set-up period. Once complete, service could begin. Once the system was empty, the server began a vacation and continued to do so until there were customers present in the system. The author derived an expression for the probability of the number of customers in the system and system performance measures.

Shin and Pearce (1998) considered the single server vacation queue where the queue length was dependent upon the vacation schedules and with BMAP arrivals. The Laplace-Stieltjes transform of the queue length distribution (transient) was found.

Takahashi, Osawa and Fujisawa (1999) considered the batch arrival retrial queue where non-preemptive priority existed among two types of customer. The joint generating function was derived for the number of customers in the priority queue and in the retrial group at an arbitrary instant of time. The mean number of customers in the system was found. The relationship between the discrete time and continuous time system was discussed.

Nobel and Tijms (1999) investigated the M(x)/G/1 queueing system with a controllable service rate. There was one server present but the server could take two modes. If the server was required to switch modes, a switch-over time was required. A set of switch-over rules was considered and an algorithm was derived which minimised the long-run average number of customers in the system. This work was an extension to the work of Liu and Tseng (1999).

1.4.5. 2000s

The new millennium brought with it a further increase in interest in the theory of batch arrival queues once more. Several strides were made in investigating vacations for the server as well as multiple servers and buffer systems.

In 2000, Tang and Tang (2000) examined the M(x)/G/1 queueing system with a single server vacation. They derived a recursion expression of the Laplace transform of the transient queue length distribution. The result was confirmed by looking at special cases including the M/G/1 queue with single arrivals and the M(x)/G/1 queue without vacations.

Later on in the same year, Laxmi and Gupta (2000) analysed the GI(x)/M/c/N queueing system, that is General batch arrivals, Markovian service, c channels and a finite buffer of size N. They discussed the model in two ways; when a batch arrival occurs and finds the buffer full (or too full to take the entire batch), either the whole batch is rejected or the excess in the batch will be rejected thus extending the work of Nobel (1989). The probability distributions of the number of customers in the system at pre-arrival and random instants of time were derived. Also blocking probabilities were found and waiting time analysis was undertaken.

Kawasaki and colleagues (2000) studied the M(x)/G/1 queueing system by considering the order of service to be random. Laplace-Stieltjes transforms were developed of the waiting time distributions and the first two moments were calculated. The relationship found by Takacs (1963) and Fuhrmann (1991) relating the second moments under random order of service to the second moments found for the FIFO queue discipline, for single arrival queues was confirmed for batch arrival queues.

Finally in 2000, Armero and Conesa (2000) dealt with the statistical analysis of bulk arrival queues using Bayesian statistics. The main focus of the research was to discover the performance measures of the system in equilibrium.

Moving into the first year of the new millennium, Chaudhry and colleagues (2001) considered the multi-server queue with batch arrivals, a General arrival distribution and a geometric service distribution in discrete time. This development built on the previous work of the research group in the 1990s (1997, 1998). The supplementary variable technique and imbedded Markov chain technique were utilised to analyse the queue. Chaudhry and Gupta (2001) continued to work in this field and discussed the discrete time bulk arrival queues with a general inter-arrival and service distribution. A simple procedure was described for computing the waiting time probabilities of the first and a random customer in a batch. In 2004, Chaudhry and Kim (2004) examined the same system but with an infinite number of servers. System size distributions were derived at pre-arrival instants and random instants. Numerical results were also provided for special cases.

In 2002 Masuyama and Takine (2002) considered the infinite server queue with multiple batch Markovian arrival streams. Customers arriving from different streams may have different service time distributions. A system of ordinary differential equations was derived for the time-dependent matrix joint generating function of the number of customers in the system. The service distributions were assumed to be phase-type and they derived the time-dependent and limiting joint binomial moments. Numerical examples were provided which illustrated the influence of parameters on the performance of the queueing system.

Kumar and colleagues (2002), investigated the single-server batch arrival retrial queue with Bernoulli vacation schedules and general retrial times. The queue length distribution was derived and other summary measures were found.

An extension to the previous vacation systems was developed in 2002. The M(x)/G/1 queueing system was considered by Choudhury (2002) whereby the server took a vacation between two successive busy periods. Steady state queue size distributions were derived for this model.

In 2003, Chaudhry and Kim (2003) investigated the system size of a discrete time multi-server queue with batch arrivals. The service distribution was Deterministic and the distribution of the waiting time in the queue was derived.

Bratiychuk and Kempa (2003) considered the G(x)/G/1 batch arrival system once again. They concentrated on the non-steady state characteristics such as the first busy period and the first idle time.

In 2003, Ke and Wang (2003) examined the M(x)/M/1 queueing system where breakdowns could occur and where a start-up time was required, under N policy. The arrival rate also varied depending upon the status of the server. Breakdowns to the server occurred following a Poisson process and the repair distribution was Negative Exponential. The steady state system size distribution was derived at a departure instant. Also costs were taken into account to optimise the system from an economic perspective.

The following year, Choudhury and Krishnamoorthy (2004) built on Choudhury's previous work (1998, 2000) and investigated the M(x)/G/1 queueing system with a random set-up time. At the commencement of each busy period, the server needed to be set-up, then service could commence. The busy period distribution was derived.

Takagi and Wu (2004) considered the multiserver case with semi-Markovian batch arrivals, Exponential service time. The theory of piecewise Markov processes was utilised to analyse this queue.

Choudhury and Madan (2004) considered the system whereby batch arrivals occurred and where the server provided two stages of heterogeneous service under Bernoulli schedule vacation (similar to (Kumar et al 2002)). Each customer undergoes two stages of service then once these are complete, the server takes a vacation with probability r. Queue size distributions were derived for steady state conditions. Later that same year, Choudhury and Paul (2004) considered the same system but with service under N-policy. Until there were N

customers present in the system, the server was idle. Once N were present, the server commenced service and each customer goes through at least one stage of service. The customer will undergo the second phase of service in a second service channel with probability θ. Queue size distributions were again derived for steady state conditions for this queue. In 2005, Choudhury and Madan (2005) extended this system. This time, each customer had to go through two phases of service. Once the second phase was complete, the server could take a vacation or begin the next service. Queue size distributions were derived for steady state conditions. In 2007 Choudhury (2007) extended the theory to a control admission policy on arrival. The model under consideration generalised the M/G/1 system with retrial policy. Steady state analysis was undertaken and many statistical distributions were found using the imbedded Markov chain technique. Later that year, Choudhury and colleagues (2007) extended this work to include a multiple vacation policy. The way that vacation policies worked in this setting was until there was a batch present in the system the server could go on multiple vacations. Steady state analysis of this system was undertaken and many statistical distributions were found using the imbedded Markov chain technique.

Chen (2004) considered the batch arrival queueing system using fuzzy logic. Arrival and service rates were deemed to be fuzzy numbers. They transformed a fuzzy queue into a family of crisp queues by using the alpha-cut approach. Using these alpha-cut representations, two parametric nonlinear programs were formulated to describe the family of crisp queues.

Hur and Ahn (2005) investigated the system where the arrival stream was compound Poisson and the service times followed a General distribution. The authors considered three types of idle period: threshold, multiple vacations, and single vacation. After an idle period, the server needed to undergo a setup time before service could commence once more. Steady state distributions were found for system size and waiting time.

Franx (2005) found the waiting time distribution for the M(x)/D/c queueing system without using probability generating functions of Laplace transforms.

Numerical calculations were shown to be simple, even when the traffic intensity was high.

2005 saw the first consideration of a control of admissions. Artalejo and colleagues (2005) investigated the Geo/G/1 retrial queue with batch arrivals in discrete time. The underlying Markov chains were studied. The results for various special cases were verified using known theory.

Ke and Chu (2006) investigated the M(x)/G/1 queueing system with a modified vacation policy, that is, at most J vacations could be taken in a row. The system size distributions were derived and also the expected busy period. These results were shown to be a generalisation of the multiple and single vacation policy for the same queueing system. A cost model was developed to determine the optimum value of J to achieve minimum cost.

Chydzinski (2006) looked at the buffer overflow period (i.e. the service time remaining when a customer arrives and finds a full buffer).

Choudhury and Paul (2006) considered the M(x)/G/1 queueing system with a second optional service channel which behaves under N-policy. Once there were N present, the server began to serve the most essential of the customers present. Some of the customers go through a second, optional, service. They derived the queue size distribution at a random instant of time and look at some costs. Numerical examples were also considered.

The transient departure process was considered by Tang (2007). The author uses the M(x)/G/1 queueing system with single vacations. A probability decomposition method was used to derive the expected number of departures in a finite time interval from any initial state. Tang then looked at practical results for some special cases.

An unreliable server in an M/M/1 queueing system was considered by Wang and colleagues in 2007 (2007). The server also took multiple vacations of

Exponential time when the system was empty. The server could breakdown and these times (as well as repair times) followed the Negative Exponential distribution. Also, the arrival rates were dependent on the status of the server. The maximum entropy principle was used to develop approximate expressions for the probability distributions of the number of customers in the system. Later that year, Wang et al (2007) extended this work to encompass continuous time loss behaviours in a space priority queue rather than discrete time.

The year 2007 sees the introduction of balking into the batch arrival queueing theory. Balking is simply the phenomena when a customer arrives in the system, they choose not to join the queue for some reason. The queuing system under consideration from Ke (2007b) looked at the M(x)/G/1 queueing system with server vacations and balking. At most J vacations could occur in a row and an arriving batch would balk with a probability (1 – b) when the server was either operating or on vacation. System size and waiting time distributions were derived and numerical illustrations were given.

Ke (2007a) examined the queueing system with Markovian batch arrivals and a General service distribution (with one server) whereby vacations occurred and also where start-up and closedown times were also taken into consideration. When all customers in the system had been served, the server undergoes a closedown time. Once closed, the server could take one or more vacations. Once an arrival occurred, the system needed to undergo a start-up time before service could commence. Also, the server could break down at any point while working according to a Poisson process. Ke developed characteristics of the system.

Banik and Gupta (2007) investigated the system whereby customers arrived in batches to a finite buffer single server queue. The time between batch arrivals had a General distribution and the batch size was Random. The service process was more complicated in nature and was described as a Markovian service process which was correlated. This model was then used to analyse two customer rejection policies namely partial batch rejection and total batch rejection. Steady state distributions were developed at specific and arbitrary time instants. They

obtained performance measures including the blocking probabilities and summary measures. Numerical results were presented graphically showing the affect of the model parameters on the performance measures.

In 2007, the authors Baek and Lee (2007) investigated the waiting time of the M(x)/G/1 queue under workload control. This workload control was simply the server does not begin service until the number in the queue reaches a threshold value. They derived the Laplace-Stieltjes transform of the waiting time of an arbitrary customer which was then used to derive the mean waiting time. Two systems were considered as special cases, the M/G/1 queue and the M(x)/M/1 queue using the same workload control. The mean waiting time was then discussed in relation to the importance of the workload control.

Chydzinski and Winiarczyk (2008) consider the blocking probability in a finite-buffer queue with arrivals following a batch Markovian process (BMAP). Firstly the authors gave a comprehensive description of the BMAP under consideration. They then derived an expression for the transform of the blocking probability and demonstrated time-dependent and steady state characteristics from this expression. Numerical results were provided for two different types of BMAP.

1.5. Modelling of length of stay, mortality and patient flow in CCU

Operational research (OR) techniques have been widely utilised in the field of healthcare. Some examples of these are queueing theory (Cooper and Corcoran 1974, Gornescu et al. 2002, Griffiths et al. 2006, Worthington 1987), Markov modelling (Kapadia et al. 2000) and simulation modelling (Ashton et al. 2005, Griffiths et al. 2005a, Harper and Shahani 2002, Moore 2003, Pilgrim and Chilcott 2008, Ridge et al. 1998, Su and Shih 2003). In fact, a survey was undertaken by Jun et al (1999) of applications of discrete event simulation in health care clinics from 1979 – 1999. The review was subdivided into two sections, patient flow and allocation of resources. Very little was included on the use of simulation in an Intensive Care Unit. However, Costa and colleagues

(2003) applied OR techniques directly to the CCU but used CART analysis to generate similar patient groups.

The purpose of this review is to ascertain the factors which affect Critical Care Unit mortality and Critical Care length of stay by highlighting the results of research undertaken by others. The section will be divided into two parts; the influence of clinical factors, and the influence of organisational factors. It is worth noting that the terms ICU and CCU were utilised in the literature search. A number of studies were found which could easily be eliminated on the basis that they compared the effectiveness of existing tools such as SAPS, APACHE and MPM to predict length of stay and mortality. Other studies were excluded on the basis that they focussed on Paediatric Intensive Care Units. The main methodological techniques used to determine length of stay, mortality and bed-occupancy were: simulation, queueing theory, multiple linear/logistic regression, Artificial Neural Networks and CART analysis. The concluding section describes some methodological articles which compare and contrast these techniques.

1.5.1. Clinical factors

In a study based in Australia and New Zealand, Moran and colleagues (2008) found that similar variables proved significant in predicted mortality and length of stay in a logistic or linear regression model. Specifically, the factors affecting both length of stay and mortality were age, gender, APACHE III score, mechanical ventilation and Elective or Emergency surgery.

In a study to determine a multiple regression equation for length of stay in the ICU of patients with an intra-cerebral haemorrhage, Ohawki and colleagues (2008) found once again that gender, age and surgical intervention were influencing factors, along with the Glasgow Coma Scale score on admission and complicating infection.

Using a multiple logistic regression model, concentrating on patients with acute lung injury, Cooke and colleagues (2008) found that an array of measures were

independently predictive of in-hospital death. These factors were modified acute physiology score, age, co-morbidities, arterial pH, minute ventilation, PaCO2, PaO2/FIO2 ratio, intensive care unit admission source, and number of Intensive Care unit days before onset of acute lung injury. Patients within the general Intensive Care Unit population were shown to have a similar profile to those with acute lung injury.

Using a logistic regression, Gomes and colleagues (2007), considered mortality of patients awaiting cardiac surgery both pre-operatively and post-operatively. They found that the important pre and intra operative factors were age, left atrial diameter, creatinine, and cardiopulmonary bypass time. In the post-operative period, the significant variables were found to be PaO(2)/FiO(2), epinephrine or norepinephrine dose and mechanical ventilation time for longer than 12 hours.

In a study in the Netherlands Van Houdenhoven and colleagues (2007) developed three models to predict length of stay in an ICU – pre-operative, post-operative and intra-ICU. The models for pre-operative and post-operative periods did not explain a sufficiently large proportion of the variance of the data. However, that for intra-ICU, was able to explain 45% of the variation. The factors which were found to affect intra-ICU length of stay were patient age, co-morbidity, type of surgical approach, intra-operative respiratory minute volume and complications occurring within 72 hours in the ICU.

In a study limited to patients undergoing coronary artery bypass surgery (CABG), Rosenfeld and colleagues (2006) found that factors increasing length of stay in ICU were age, increased pump time, chronic obstructive pulmonary disease, and Emergency surgery.

Friere and colleagues (2002) investigated the use of the APACHE II score and the Logistic Organ Dysfunction System as predictors of prolonged ICU length of stay in patients with Diabetic Ketoacidosis (DKA). It was concluded that patients with DKA are less severely ill and have a lower mortality risk than those without

DKA and that severity of illness scores do not predict length of stay well in this group of patients.

Nierman and colleagues (2001) fitted an ordinal logistic regression model to predict discharge location – discharge to home, rehabilitation or death. The factors of influence were found to be age, gender, baseline support level, type of ICU, heart rate at ICU admission, use of mechanical ventilation, vasopressors or a pulmonary artery catheter during the ICU stay, and the development of respiratory, Neurologic or haematologic failure or sepsis while in the ICU.

1.5.2. Organisational changes

There is some evidence to suggest that the time of day that a patient is admitted or discharged from ICU can impact on mortality rates. Laupland and colleagues (2008) considered this issue and using logistic regression concluded that admission or discharge on the weekend was not associated with increased mortality, while admission and discharge during the night-time period were both independently associated with mortality.

In contrast, Wunsch and colleagues (2004) found that there was no significant difference in hospital mortality, once adjusted for individual components of the APACHE II score (which are confounding variables), between patients admitted on the weekend compared to mid-week, nor during the night compared with the day.

In another study considering the influence of time of admission to ICU, Luyt and colleagues (2007) concluded that patients admitted outside the normal working week were less critically ill than those admitted during normal working hours, had fewer failed organs, required fewer support procedures, and had a lower death rate. There are some differences between these outcomes and those found by Laupland and colleagues (2008) and Wunch and colleagues (2004).

Zare and colleagues (2007) found that, when considering Elective surgery that was undertaken on Fridays rather than earlier in the week, there were no significant differences in 30-day mortality.

Duke and colleagues (2004) investigated the link between the time of discharge from ICU and mortality. Using Logistic regression, they found that patients discharged during the night had a higher mortality rate.

Another investigation into time of discharge at an intensive care unit was undertaken by Goldfrad and Rowan (2000). They found that the percentage of patients discharged at night was increasing, and that these patients did not fare so well as those discharged during the day. Also, discharges which occurred at night were more likely to be premature.

A literature review undertaken by Coombs and Lattimer (2007), focussing on organisational issues, found that staffing levels and skill mix within the CCU team can have an impact on patient outcome. They called for long-term evaluation using a whole systems approach in order to ensure efficiency within workforce modelling.

Several studies including Keenan and colleagues (2007) investigated the link between hospital site and size to ICU length of stay. While this study is located in a single site, and therefore cannot investigate the possible disparity between various Critical Care Units, this may be an area of possible future research.

A study undertaken by Friedrich and colleagues (2006) considered patient outcomes after very long stays in Intensive Care. Using multivariate logistic regression, they found that predictors of hospital mortality were age, immunosuppression, mechanical ventilation for longer than 90 days, treatment with inotropes or vasopressors for more than 3 days at or after day 30 in the ICU, and acute renal failure requiring dialysis at or after day 30 in the ICU.

Ranhoff and colleagues (2006) investigated the possibility of treating severely ill elderly patients in a sub-intensive care setting. They found that patients treated in the SICU were less likely to die than those in the Acute Care for the Elderly Unit. They argued that this approach could avoid overcrowding in an ICU.

Shahani and colleagues (2008) used CART analysis to populate a simulation model of patient flow in a Critical Care Unit. The main aim of the study was to investigate the role of mathematical modelling in the Critical Care setting. They used a case-study approach modelling 5 different scenarios (several of which will be considered in Chapter 6 of this thesis). Patient characteristics were not considered; the study focussed on organisational changes only.

A simulation model was built using PASCAL by Ridge and colleagues (1998) to aid in bed-occupancy planning in an Intensive Care Unit. A simplified version of the model was developed analytically and summary measures for the queue were derived. Several what-if type scenarios were investigated such as, varying bed-numbers, changing the deferral period of a planned admission and changing the number of beds reserved for emergency admissions. The authors suggest that to improve the model, CART analysis should be employed to create homogeneous patient groups.

A simulation model was built by Griffiths and colleagues (2005b) to model patient flow in a CCU. The main focus of the research was to optimise the number of rostered nursing staff. The purpose of the study was to minimise cost by reducing the use of supplementary nurses.

In 2004, McManus and colleagues (2004) constructed a model of an ICU in a children's hospital. The ICU was considered as an M/M/c/c queueing system. What if type scenarios tested included a change in bed numbers. Although interesting, the assumption of Markovian service times for hospital length of stay data is not very useful.

Kim and colleagues (1999) utilised a simulation model and queueing theory to describe activities in an ICU. The admission and discharge process was of particular interest to the researchers. Four sources of admission were included in the model; Ward, A&E, Emergency surgery and Elective surgery and inter-arrival and length of stay distributions were fitted for these sources. The model was utilised to determine whether the 14 bed capacity of the ICU was sufficient.

1.5.3. Methodological comparisons

An illustrative paper into methods of modelling hospital length of stay was developed by Vasilakis and Marshall (2005). The authors compared modelling techniques such as survival analysis, mixed-exponential distributions, phase type distributions, simulation modelling and compartmental modelling. A comparison was drawn between the parameter estimates calculated for the phase type distribution and compartmental modelling and were found to be very similar. It was concluded that all of these techniques were valuable in the prediction of length of stay for stroke patients. It was also suggested that their implementation could be extended to cover many other patient groups.

Harper (2002) developed a framework for modelling of hospital resources. The author discussed several issues regarding the complexity of a hospital system such as uncertainty in demand, variability in factors such as length of stay and limited resources and suggested that any simulation model should take these issues into account. A generic hospital model, PROMPT, is described in the paper. The author concludes that it is imperative to capture the variability of the data using classification tools such as CART analysis, and then to use this information to populate detailed simulation models.

Lin and colleagues (2006) undertook a comparison of logistic regression and artificial neural networks in the prediction of survival in an ICU. They found that the neural network model was better able to predict survival than the logistic regression model.

Another comparison of Artificial Neural Networks and Logistic regression in predicting mortality, (Clermont et al. 2001), concluded that these techniques have similar performance when sample size is at least 800.

Chapter 2: Batch service queues

2.1. Introduction

This Chapter investigates the theory and practical applications of batch service queues and forms a contribution to the first objective highlighted in section 1.2, to identify and investigate appropriate queueing systems which can model activities at the Critical Care Unit.

A batch service queue is a queueing system whereby customers arrive (singly for the sake of this Chapter) and are served in batches. Practically, batch service queues crop up frequently. For example, a lift is a batch service queue (people arrive singly and are served in a batch). Also, many different forms of transport such as an aeroplane and a train are examples of batch service queues. For the sake of this thesis the main practical applications considered will be those in the field of healthcare. A day surgery unit can be considered as an example of a batch service queue. People arrive singly first thing in the morning and are served (that is they undergo treatments) as a group (or a batch) during the day. This work has been investigated in much depth in the past (section 1.3 highlights many of the key papers) and acts as a stepping stone to the more sophisticated and difficult models investigated in Chapter 3.

2.2. The $M/G^{(0,s)}/1$ queueing system

The $M/G^{(0,s)}/1$ queue (i.e. Markovian arrivals, general batch service times with a minimum batch size of zero and a maximum batch size of s, with one server) will be discussed in this Chapter, with consideration given to different service time distributions.

In many queueing systems, s (the maximum batch size) is pre-defined and fixed and the minimum batch size is zero. The minimum batch size is called the

quorum. This Chapter will only consider systems with a quorum of zero and a fixed maximum batch size of s. An assumption of this queueing system is that the server will begin service regardless of the number of customers waiting, including the case when there are no customers waiting.

Assume that customers arrive singly and at random at a mean rate λ, and that batches of customers are served at a mean rate μ with maximum capacity s, then

$$\rho = \frac{\text{Demand on system over long period of time T}}{\text{Capacity of system over long period of time T}}$$

The total demand on the system during a long time period T will be equal to the average rate at which the customers arrive (λ) multiplied by the time period T.

The total capacity of the system over a long period of time T will be equal to the average service rate (μ), multiplied by the maximum number that could be served (s), multiplied by the time period T.

$$\rho = \frac{\lambda T}{s\mu T} = \frac{\lambda}{s\mu}$$

Hence, $\frac{\lambda}{\mu} = s\rho$

To analyse this queue, the imbedded Markov chain technique is used and since we have random arrivals, it is appropriate to use the departure instants as regeneration points. However, since a new service starts immediately after the previous one has finished, the instants at which we link the probabilities need to be clearly defined.

For clarity the procedure is magnified at the instants where one batch service ends and the next begins. We will consider the process at points such as A and B

(as indicated in Figure 2.1). However, it should be remembered that all three event instants at A and B in fact correspond to one time instant.

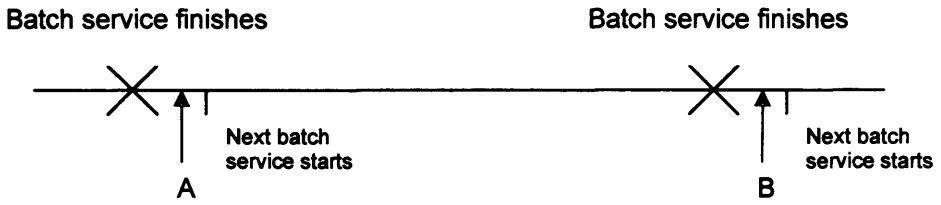


Figure 2.1: Regeneration points for the M/G$^{(0,s)}$/1 system

Let p_n be the probability that there are n customers in the system just after a batch service finished (e.g. at A or B). The interval AB represents the service time of a batch.

Let k_j be the probability that j customers arrive during the batch service time AB.

The probabilities, p_n are expressed as follows:

$$p_0 = k_0 \sum_{i=1}^{s} p_i \text{ and } p_n = k_n \sum_{i=1}^{s} p_i + \sum_{j=1}^{n} p_{s+j} k_{n-j}$$

Define the probability generating functions, $G(z)$ and $K(z)$ to be

$$G(z) = p_0 + zp_1 + z^2 p_2 + \ldots = \sum_{n=0}^{\infty} z^n p_n$$

$$K(z) = k_0 + zk_1 + z^2 k_2 + \ldots = \sum_{n=0}^{\infty} z^n k_n$$

Through multiplication of appropriate powers of z, and summation, the following expression for $G(z)$ is found:

$$G(z)=\frac{K(z)\left[\sum_{i=0}^{s-1}\left(z^s-z^i\right)p_i\right]}{\left[z^s-K(z)\right]}$$

$$G(z)=\frac{\left[\sum_{i=0}^{s-1}\left(z^s-z^i\right)p_i\right]}{\left[\frac{z^s}{K(z)}-1\right]}$$

We note that the numerator contains s unknown constants,

p_i, where $i = 0,1,\ldots (s-1)$

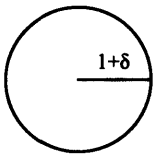
Rouché's theorem can be applied to the denominator of the above expression for $G(z)$, since $G(z)$ is an analytic function of z within and on the unit circle. We show that the denominator of $G(z)$ has exactly s zeros within and on the unit circle.

Denote the s zeros of the denominator within and on the unit circle as 1, z_1, z_2, ..., z_{s-1} (1 will always be a root of $G(z)$). For $G(z)$ to remain analytic the numerator must also equal zero for these values of z. When each of the z_i's are individually substituted, s simultaneous equations are produced that can be solved for the unknowns, $p_0, p_1, \ldots, p_{s-1}$.

Rouché's theorem

If $f(z)$ and $g(z)$ are analytic within and on a closed contour, C, and if $|f(z)| > |g(z)|$ on C, then $f(z)$ and $f(z)+g(z)$ have the same number of zeros within C.

Take as the closed contour, a circle slightly bigger than the unit circle.



$x = r\cos\theta \qquad y = r\sin\theta$

$C : z = re^{i\theta} = r(\cos\theta + i\sin\theta)$ where $r = 1+\delta$, $\delta > 0$ but small.

Choose $f(z) = z^s$ and $g(z) = -K(z)$. It is clear that both f and g are analytic within and on the unit circle. Thus the first condition of Rouché's theorem is satisfied.

We now need to show that $|f(z)| > |g(z)|$ on C.

On C,

$|f(z)| = |z^s| = |z|^s = (1+\delta)^s = 1+s\delta$ to first order in δ.

$$\begin{aligned}
|g(z)| &= |-K(z)| \\
&= |k_0 + zk_1 + z^2k_2 + \ldots| \\
&\le |k_0| + |zk_1| + |z^2k_2| + \ldots \\
&= k_0 + k_1|z| + k_2|z^2| + \ldots \\
&= k_0 + k_1(1+\delta) + k_2(1+\delta)^2 + \ldots \\
&= K(1+\delta)
\end{aligned}$$

By Taylor's theorem:

$$K(1+\delta) = K(1) + \delta K'(1) + \frac{\delta^2}{2!}K''(1) + \ldots$$

$K(1+\delta) = K(1) + \delta K'(1)$ to first order in δ.

But $K(1) = 1$ since it is a PGF, and

$K'(1) = \frac{\lambda}{\mu}$, since $K'(1)$ is the mean number of arrivals in a service time.

$= s\rho$

So

$|g(z)| < 1 + s\rho\delta$ on C

$|f(z)| > |g(z)|$ on C

if $1 + s\delta > 1 + s\rho\delta$

i.e. if $\rho < 1$

So Rouché's theorem is satisfied if $\rho < 1$. Hence $f(z)$ has the same number of zeros as $f(z) + g(z)$ within and on C.

Clearly z^s has s zeros within and on the unit circle (all at $z = 0$) therefore $z^s - K(z)$ has s zeros within and on the unit circle.

Previously, we had

$$G(z) = \frac{\left[\sum_{i=0}^{s-1}(z^s - z^i)p_i\right]}{\left[\frac{z^s}{K(z)} - 1\right]}$$

We noted that if $G(z)$ was to remain analytic on the unit circle then the s zeros of the denominator 1, z_1, z_2, ..., z_{s-1}, must also be zeros of the numerator.
If we write the numerator out in full, we have:

$$(z^s - 1)p_0 + (z^s - z)p_1 + (z^s - z^2)p_2 + \ldots + (z^s - z^{s-1})p_{s-1}$$

We see that this is a polynomial in z of degree s. Thus, the numerator will have exactly s zeros and we have already seen that these must give the factors

$(z-1)(z-z_1)(z-z_2)...(z-z_{s-1})$. However the numerator could also have a constant multiplier apart from its factors, i.e. numerator = $A(z-1)(z-z_1)(z-z_2)...(z-z_{s-1})$ where A is constant.

Thus we may write $G(z)$ in the form:

$$G(z)=\frac{A(z-1)\prod_{i=1}^{s-1}(z-z_i)}{\frac{z^s}{K(z)}-1} \qquad (2.1)$$

The problem has now been simplified – only the constant A needs to be found.

To find A, straightforward substitution of $z = 1$ is not possible since both numerator and denominator become zero. By using differentiation, the solution can be found.

$$\left[\frac{z^s}{K(z)}-1\right]G(z)=A(z-1)\prod_{i=1}^{s-1}(z-z_i)$$

Differentiating both sides with respect to z gives:

$$\left[\frac{z^s}{K(z)}-1\right]G'(z)+\left[\frac{K(z)sz^{s-1}-z^sK'(z)}{\left[K(z)^2\right]}\right]G(z)=A\left[(z-1)\frac{d\prod}{dz}+\prod_{i=1}^{s-1}(z-z_i)\right]$$

Setting $z = 1$ gives

$$\left[\frac{K(1)s-K'(1)}{\left[K(1)^2\right]}\right]G(1)=A\left[\prod_{i=1}^{s-1}(1-z_i)\right]$$

Using $K(1)=1$ and $G(1)=1$ and $K'(1)=\frac{\lambda}{\mu}=s\rho$

$$s - s\rho = A\left[\prod_{i=1}^{s-1}(1-z_i)\right]$$

$$\therefore A = \frac{s(1-\rho)}{\prod_{i=1}^{s-1}(1-z_i)}$$

Hence

$$G(z) = \frac{s(1-\rho)(z-1)}{\frac{z^s}{K(z)} - 1} \prod_{i=1}^{s-1}\left(\frac{z-z_i}{1-z_i}\right) \qquad (2.2)$$

To complete the picture for *G(z)*, the following observations relating to *K(z)* may be made.

K(z) is the PGF of the k_j probabilities that *j* arrivals occur during a service time.

K(z) is related to the Laplace transform of the PDF of the service time distribution.

$$K(z) = \mathcal{L}_{\lambda(1-z)}\{f(t)\}$$

Substitution of the appropriate form of *K(z)* into *G(z)* gives an expression for *G(z)* that can be expanded in ascending powers of *z* to pick out the probabilities $p_0, p_1, \ldots$. However, the summary measures are usually required and the mean number of customers in the system at A, B (i.e. immediately after a service time has been completed) is given as $L^{+d} = G'(1)$. Therefore differentiation of *G(z)* is required.

Manipulation of Equation 2.2 gives:

$$G(z)\left(\frac{z^s}{K(z)} - 1\right) = s(1-\rho)(z-1)\prod_{i=1}^{s-1}\left(\frac{z-z_i}{1-z_i}\right)$$

Differentiation gives:

$$G'(z)\left(\frac{z^s}{K(z)}-1\right)+\frac{K(z)sz^{s-1}-z^sK'(z)}{K(z)^2}=s(1-\rho)\left((z-1)\frac{d\Pi}{dz}+\prod_{i=1}^{s-1}\left(\frac{z-z_i}{1-z_i}\right)\right)$$

Setting $z = 1$ in the above equation causes $G'(z)$ to disappear. Thus differentiation a second time is required:

$$G''(z)\left(\frac{z^s}{K(z)}-1\right)$$
$$+G(z)\frac{\left(K(z)^2\left[s\left(K'(z)z^{s-1}+K(z)(s-1)z^{s-2}\right)-sz^{s-1}K'(z)-z^sK''(z)\right]\right)}{K(z)^4}$$
$$-G(z)\frac{\left(\left[K(z)sz^{s-1}-z^sK'(z)\right]2K(z)K'(z)\right)}{K(z)^4}$$
$$+\frac{2\left[K(z)sz^{s-1}-z^sK'(z)\right]}{K(z)^2}G'(z)$$
$$=s(1-\rho)\left((z-1)\frac{d^2\Pi}{dz^2}+2\frac{d\Pi}{dz}\right)$$

Now, setting $z = 1$ gives:

$$\frac{\left(\left[s\left(K'(1)+(s-1)\right)-sK'(1)-K''(1)\right]\right)}{1}-\frac{\left(\left[s-K'(1)\right]2K'(1)\right)}{1}+\frac{2\left[s-K'(1)\right]}{1}G'(1)$$
$$=s(1-\rho)\left(2\frac{d\Pi}{dz}\right)$$

The final terms that are required are $\frac{d\Pi}{dz}$ and $K''(1)$.

Firstly, $\frac{d}{dz}\left(\prod_{i=1}^{s-1}\left(\frac{z-z_i}{1-z_i}\right)\right)=\sum_{i=1}^{s-1}\frac{1}{1-z_i}$.

Therefore, substituting in the appropriate expressions gives:

$$s\left(s\rho+(s-1)\right)-s^2\rho-K''(1)-2s^2\rho(1-\rho)+2s(1-\rho)G'(1)=2s(1-\rho)\sum_{i=1}^{s-1}\frac{1}{1-z_i}$$

Rearrangement then finally gives an expression for L^{+d}:

$$\begin{aligned} G'(1) &= \sum_{i=1}^{s-1}\frac{1}{1-z_i}+s\rho+\frac{s^2\rho+K''(1)-s\left(s\rho+(s-1)\right)}{2s(1-\rho)} \\ &= L^{+d} \end{aligned} \tag{2.3}$$

As previously noted, the values of z which are zeros of the denominator of $G(z)$ need to be found. The only zeros of interest are the ones which lie within the unit circle. Within the unit circle there should be $s-1$ roots. If s is an even number, one root will be real (less than 1) and the remaining roots will be complex conjugate pairs. If s is an odd number then all roots will be complex conjugate pairs. When considering a service time distribution from the Exponential family, and when considering small batch sizes, finding the zeros of the denominator of $G(z)$ is a fairly simple task, but if large batch sizes are considered or if the service time distribution differs from Exponential, an iterative process is used to find the roots.

Equation 2.2 showed that:

$$G(z)=\frac{s(1-\rho)(z-1)}{\frac{z^s}{K(z)}-1}\prod_{i=1}^{s-1}\left(\frac{z-z_i}{1-z_i}\right)$$

Therefore, we need to find the roots of the equation:

$$\frac{z^s}{K(z)}-1=0 \tag{2.4}$$

The computer software package, MAPLE, was used to solve (2.4) for various values of ρ and s (batch size).

Once the zeros were found, L^{+d} could be easily found. A simulation model was built in Visual Basic (in Excel) to simulate the batch service queue system as a check. The summary measures L, Lq, W and W_q were calculated from the output. For the sake of comparison, and for completeness, the values L, Lq, W and W_q needed to be derived theoretically and compared with the simulation output. This derivation may be found in (Holland 1991):

$$L = L^{+d} + \left(\frac{1+c_s^2}{2}\right)s\rho$$

$$L_q = L^{+d} - \left(1 - \frac{1+c_s^2}{2}\right)s\rho$$

2.3. Service time distributions

Different service time distributions will be considered individually. The first distribution considered will be the Negative Exponential. After this the E_2 distribution will be considered, extending to the E_{10}. Following this, the k parameter in the Erlang distribution will be taken to tend to infinity, thus giving the constant service distribution. Finally, the 2 phase Hyperexponential distribution will be considered.

2.3.1. Negative Exponential distribution

The probability density function of the Negative Exponential distribution is:

$$f(t) = \mu e^{-\mu t}, \quad t \geq 0,$$

To find the probability generating function, $K(z)$, the following expression needs to be evaluated:

$$K(z) = \mathcal{L}_{\lambda(1-z)}\{f(t)\}$$
$$= \mathcal{L}_{\lambda(1-z)}\{\mu e^{-\mu t}\}$$

Initially, the Laplace transform using the variable x will be taken and then $\lambda(1-z)$ will be substituted for x.

$$\mathcal{L}_x\{\mu e^{-\mu t}\} = \int_0^\infty e^{-xt}\mu e^{-\mu t}dt$$
$$= \frac{\mu}{x+\mu}$$

Substitution of $\lambda(1-z)$ for x gives:

$$K(z) = \frac{\mu}{\lambda(1-z)+\mu}$$
$$= \left[1+\frac{\lambda(1-z)}{\mu}\right]^{-1}$$

Very often, when considering a queueing system, the quantity ρ is of most interest rather than the individual arrival rate (λ) and service rate (μ). Substitution of $\frac{\lambda}{\mu} = s\rho$, where s = maximum batch size, gives,

$$K(z) = \left[1+s\rho(1-z)\right]^{-1}$$

To complete the expression previously derived for $G'(1)$, that is (2.3), $K''(1)$ is required:

$$K'(z) = s\rho\left[1+s\rho(1-z)\right]^{-2}$$

Setting $z = 1$ gives:

$K'(1) = s\rho$ as expected.

Differentiation once again gives:

$$K''(z) = (s\rho)^2 (2)\left[1 + s\rho(1-z)\right]^{-3}$$

Setting $z = 1$ gives:

$$K''(1) = 2(s\rho)^2$$

Substitution of these expressions back into (2.3) gives:

$$L^{+d} = \sum_{i=1}^{s-1} \frac{1}{1-z_i} + s\rho + \frac{2s\rho^2 - (s-1)}{2(1-\rho)} \qquad (2.5)$$

To verify the above expression, a simulation model was built in Simul8. Simul8 has the capability of producing summary measures for any queue that is built, but it will not give a definitive value for L, the mean number of customers in the system. However Little's result can be utilised to calculate L from W, the mean time spent in the system.

For this queueing situation, $M/M^{(0,s)}/1$, the summary measures are defined below. It is worth noting that the coefficient of variation for the Negative Exponential distribution is 1:

$$L = \sum_{i=1}^{s-1} \frac{1}{1-z_i} + 2s\rho + \frac{2s\rho^2 - (s-1)}{2(1-\rho)}$$

$$L_q = \sum_{i=1}^{s-1} \frac{1}{1-z_i} + s\rho + \frac{2s\rho^2 - (s-1)}{2(1-\rho)}$$

$$W = \frac{1}{\lambda}\sum_{i=1}^{s-1} \frac{1}{1-z_i} + \frac{2}{\mu} + \frac{2s\rho^2 - (s-1)}{2\lambda(1-\rho)}$$

$$W_q = \frac{1}{\lambda}\sum_{i=1}^{s-1} \frac{1}{1-z_i} + \frac{1}{\mu} + \frac{2s\rho^2 - (s-1)}{2\lambda(1-\rho)}$$

Table 2.1 summarises the simulated values compared with the theoretical values obtained from the above expressions. As can be seen, the theoretical values correspond to the simulated values very well, even for large batch sizes.

Table 2.1: Theoretical and simulated summary measures of the $M/M^{(0,s)}/1$ system, $s = 3, 5, 10, 100$

BATCH SIZE	RHO	THEORETICAL VALUE	SIMULATED VALUE
3	0.1	$Lq = 0.30$	$Lq = 0.30$
		$L = 0.60$	$L = 0.60$
	0.5	$Lq = 2.24$	$Lq = 2.26$
		$L = 3.74$	$L = 3.76$
	0.9	$Lq = 18.32$	$Lq = 18.33$
		$L = 21.02$	$L = 21.03$
5	0.1	$Lq = 0.50$	$Lq = 0.50$
		$L = 1.00$	$L = 1.00$
	0.5	$Lq = 3.49$	$Lq = 3.51$
		$L = 5.99$	$L = 6.01$
	0.9	$Lq = 27.64$	$Lq = 27.79$
		$L = 32.14$	$L = 32.30$
10	0.1	$Lq = 1.00$	$Lq = 1.01$
		$L = 2.00$	$L = 2.01$
	0.5	$Lq = 6.62$	$Lq = 6.64$
		$L = 11.62$	$L = 11.65$
	0.9	$Lq = 50.94$	$Lq = 51.15$
		$L = 59.94$	$L = 60.16$
100	0.1	$Lq = 10.00$	$Lq = 10.02$
		$L = 20.00$	$L = 20.02$
	0.5	$Lq = 63.09$	$Lq = 63.11$
		$L = 113.09$	$L = 113.16$
	0.9	$Lq = 470.41$	$Lq = 466.40$
		$L = 560.41$	$L = 556.53$

Table 2.1 demonstrates that as ρ increases, the values of L and Lq increase. As ρ tends to 1 the value of L tends to infinity. Figure 2.2 demonstrates the value of L as ρ tends to 1 for a maximum batch size of 10:

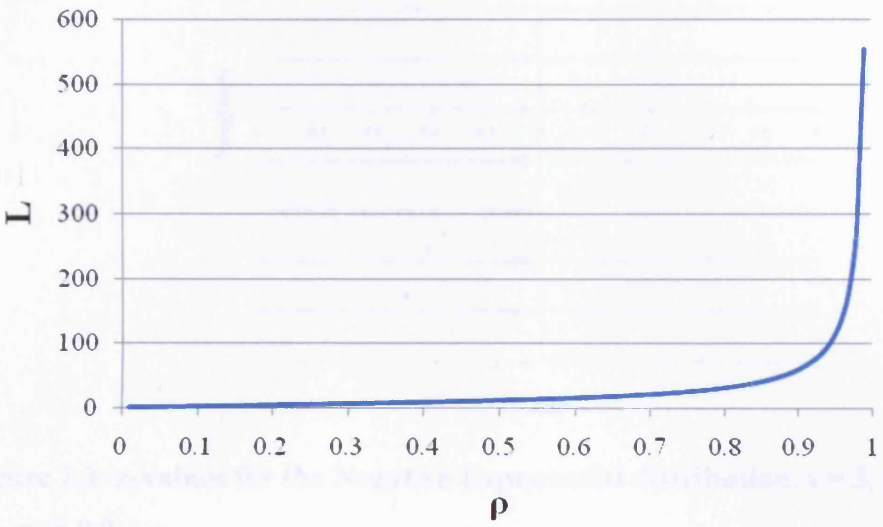


Figure 2.2: Mean number of customers in the system (L) as ρ increases – Negative Exponential service

Graphical representation of z_i

The following graphs display the roots within the unit circle, of (2.4), where the service time is Negative Exponential. As ρ increases, the modulii of the roots decrease, irrespective of the batch size. As the batch size, s, increases, the roots become closer together. The roots are conjugate pairs, apart from one real root when s is even, and as s increases they tend towards forming a circle.

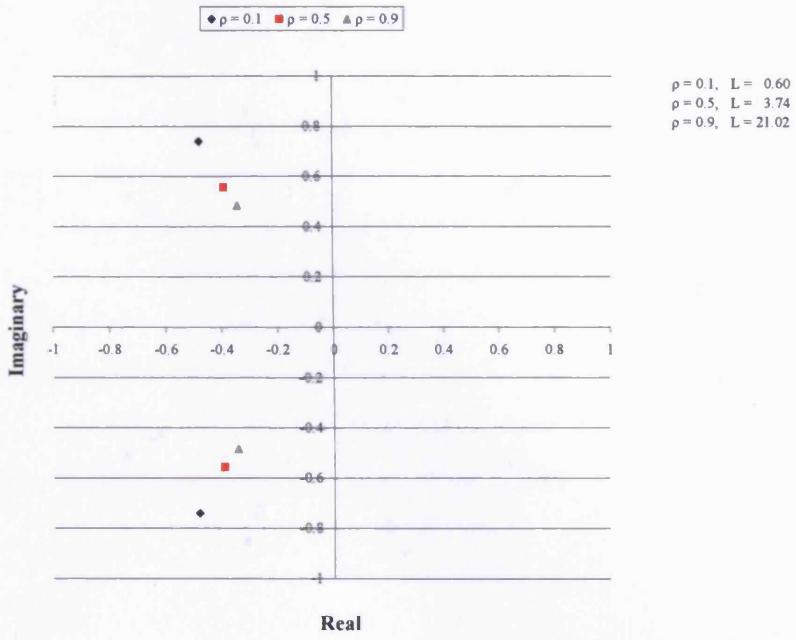


Figure 2.3: z_i values for the Negative Exponential distribution, $s = 3$, $\rho = 0.1$, 0.5, and 0.9

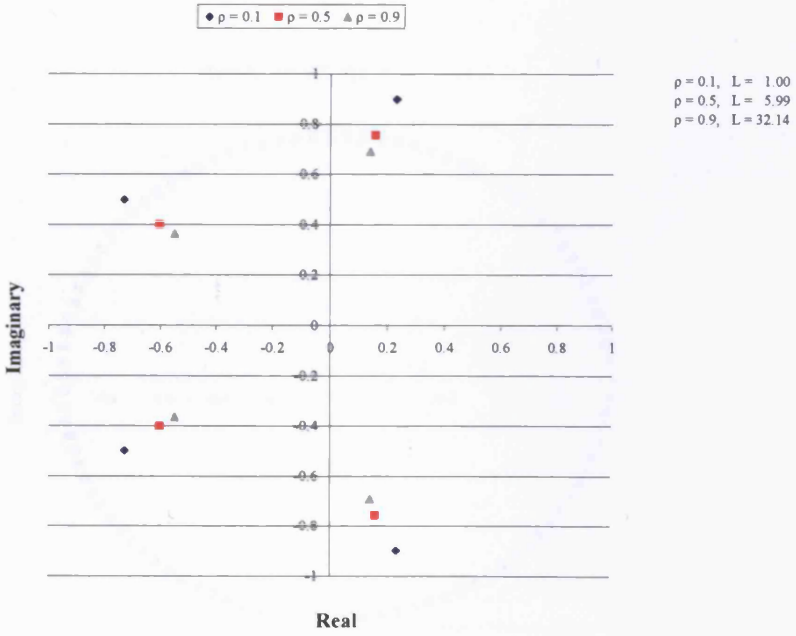


Figure 2.4: z_i values for the Negative Exponential distribution, $s = 5$, $\rho = 0.1$, 0.5, and 0.9

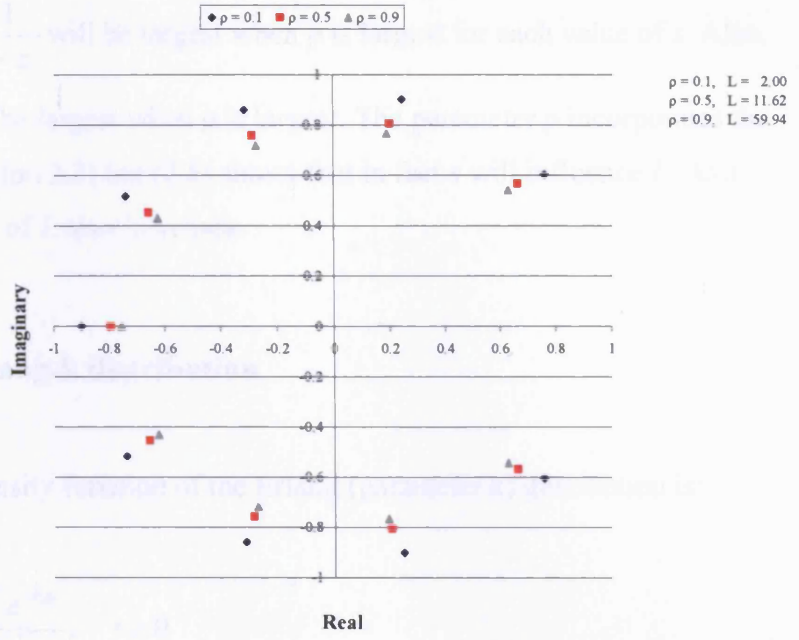


Figure 2.5: z_i values for the Negative Exponential distribution, s = 10, ρ = 0.1, 0.5, and 0.9

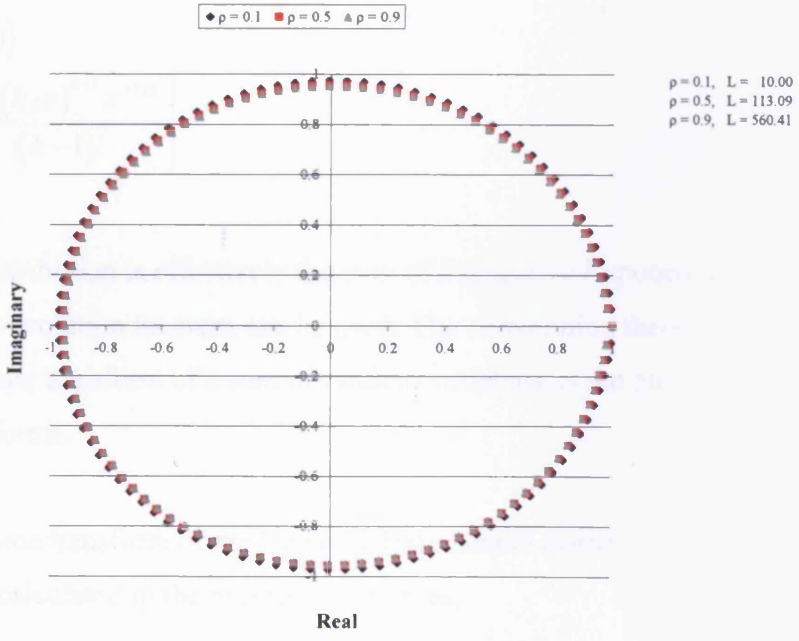


Figure 2.6: z_i values for the Negative Exponential distribution, s = 100, ρ = 0.1, 0.5, and 0.9

The value of $\sum_{i=1}^{s-1} \frac{1}{1-z_i}$ will be largest when ρ is largest for each value of s. Also, the value of L will be largest when ρ is largest. The parameter ρ incorporates the batch size (see section 2.3) but (2.5) shows that in fact s will influence L. As s increases the value of L also increases.

2.3.2. Erlang-k distribution

The probability density function of the Erlang (parameter k) distribution is:

$$f(t) = \frac{k\mu(k\mu t)^{k-1} e^{-k\mu t}}{(k-1)!}, \quad t \geq 0$$

To find the probability generating function, the following expression needs to be evaluated:

$$K(z) = \mathcal{L}_{\lambda(1-z)}\{f(t)\}$$
$$= \mathcal{L}_{\lambda(1-z)}\left\{\frac{k\mu(k\mu t)^{k-1} e^{-k\mu t}}{(k-1)!}\right\}$$

Since the Erlang distribution is effectively the sum of k Negative Exponential distributions, the convolution theorem can be used. The convolution theorem states that the Laplace transform of a sum of random variables, is the product of their Laplace transforms.

Therefore, the Laplace transform of the Negative Exponential distribution is required. This was calculated in the previous section as,

$$\mathcal{L}_x\{\mu e^{-\mu t}\} = \frac{\mu}{x+\mu}$$

Substitution of $\lambda(1-z)$ for z gives:

$$\mathcal{L}_x\left\{\mu e^{-\mu t}\right\} = \left[1+\frac{\lambda(1-z)}{\mu}\right]^{-1}$$

From the convolution theorem we have:

$$\mathcal{L}_x\{h(t)\} = \mathcal{L}_x\{f_1(t)\}\mathcal{L}_x\{f_2(t)\}....\mathcal{L}_x\{f_k(t)\} \quad \text{where } h(t) = (f_1 \cdot f_2 \cdot .... \cdot f_k)(t)$$

Therefore

$$K(z) = \mathcal{L}_{\lambda(1-z)}\{f(t)\}$$
$$= \left[1+\frac{\lambda(1-z)}{\mu}\right]^{-k}$$

Substitution of $\frac{\lambda}{\mu} = s\rho$, where s = batch size, and gives,

$$K(z) = \left[1+\frac{s\rho(1-z)}{k}\right]^{-k}$$

To complete the expression previously derived for L^{+d}, (2.3), $K''(1)$ is required:

$$K'(z) = \frac{s\rho k}{k}\left[1+\frac{s\rho(1-z)}{k}\right]^{-k-1}$$

Differentiation again gives:

$$K''(z) = \frac{(k+1)}{k}(s\rho)^2\left[1+\frac{s\rho(1-z)}{k}\right]^{-k-2}$$

Setting $z = 1$ gives:

$$K''(1) = \frac{(k+1)}{k}(s\rho)^2\left[1+\frac{s\rho(1-1)}{k}\right]^{-k-2}$$

Therefore $K''(z) = (s\rho)^2 + \frac{(s\rho)^2}{k}$

Substitution of these expressions back into (2.3) gives:

$$L^{+d} = \sum_{i=1}^{s-1}\frac{1}{1-z_i} + s\rho + \frac{s\rho^2 + \frac{s\rho^2}{k} - s + 1}{2(1-\rho)} \qquad (2.6)$$

Using this expression for L^{+d}, the mean number of customers in the system and the mean number of customers in the queue, at an arbitrary instant in time, are given by:

$$L = \sum_{i=1}^{s-1} \frac{1}{1-z_i} + s\rho\left(\frac{3k+1}{2k}\right) + \frac{s\rho^2 + \frac{s\rho^2}{k} - (s-1)}{2(1-\rho)}$$

$$L_q = \sum_{i=1}^{s-1} \frac{1}{1-z_i} + s\rho\left(\frac{k+1}{2k}\right) + \frac{s\rho^2 + \frac{s\rho^2}{k} - (s-1)}{2(1-\rho)}$$

Corresponding waiting time expressions can be found simply by utilising Little's result.

The results were verified numerically using the simulation model for $k = 2$ and $k = 10$:

Table 2.2 Theoretical and simulated summary measures of the $M/E_k^{(0,s)}/1$ system, $s = 3, 5, 10, 100$ and $k = 2, 10$

BATCH SIZE	ρ	K = 2		K = 10	
		THEORETICAL VALUE	SIMULATED VALUE	THEORETICAL VALUE	SIMULATED VALUE
3	0.1	$Lq = 0.23$	$Lq = 0.23$	$Lq = 0.17$	$Lq = 0.17$
		$L = 0.53$	$L = 0.49$	$L = 0.47$	$L = 0.46$
	0.5	$Lq = 1.54$	$Lq = 1.56$	$Lq = 1.01$	$Lq = 1.03$
		$L = 3.04$	$L = 3.06$	$L = 2.51$	$L = 2.53$
	0.9	$Lq = 11.65$	$Lq = 11.69$	$Lq = 6.37$	$Lq = 6.42$
		$L = 14.35$	$L = 14.39$	$L = 9.07$	$L = 9.12$
5	0.1	$Lq = 0.38$	$Lq = 0.39$	$Lq = 0.28$	$Lq = 0.28$
		$L = 0.88$	$L = 0.78$	$L = 0.78$	$L = 0.78$
	0.5	$Lq = 2.36$	$Lq = 2.38$	$Lq = 1.52$	$Lq = 1.54$
		$L = 4.86$	$L = 4.88$	$L = 4.02$	$L = 4.04$
	0.9	$Lq = 16.57$	$Lq = 16.57$	$Lq = 7.87$	$Lq = 7.90$
		$L = 21.07$	$L = 21.07$	$L = 12.37$	$L = 12.40$
10	0.1	$Lq = 0.75$	$Lq = 0.77$	$Lq = 0.55$	$Lq = 0.56$
		$L = 1.75$	$L = 1.65$	$L = 1.55$	$L = 1.56$
	0.5	$Lq = 4.41$	$Lq = 4.44$	$Lq = 2.85$	$Lq = 2.87$
		$L = 9.41$	$L = 9.44$	$L = 7.85$	$L = 7.87$
	0.9	$Lq = 28.89$	$Lq = 28.91$	$Lq = 11.69$	$Lq = 11.72$
		$L = 37.89$	$L = 37.92$	$L = 20.69$	$L = 20.72$
100	0.1	$Lq = 7.50$	$Lq = 7.52$	$Lq = 5.50$	$Lq = 5.52$
		$L = 17.50$	$L = 17.52$	$L = 15.50$	$L = 15.52$
	0.5	$Lq = 41.68$	$Lq = 41.66$	$Lq = 27.58$	$Lq = 27.60$
		$L = 91.68$	$L = 91.65$	$L = 77.58$	$L = 77.60$
	0.9	$Lq = 250.78$	$Lq = 254.65$	$Lq = 81.74$	$Lq = 81.88$
		$L = 340.78$	$L = 344.72$	$L = 171.74$	$L = 171.89$

As can be seen from the above table, the simulated values correspond well to the theoretical values for different values of k. Another interesting point is the way in which as k increases, the values of L and Lq decrease. As ρ tends to 1, L will tend to infinity but more quickly when k is small. The graph below demonstrates this

fact. Note that the Negative Exponential distribution is simply the Erlang distribution with $k = 1$.

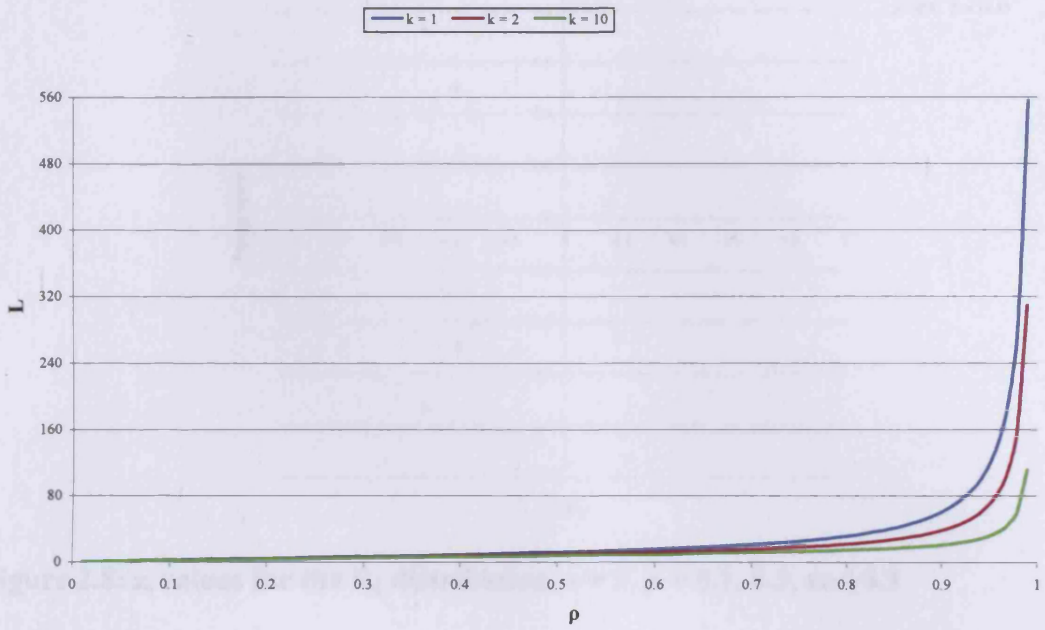


Figure 2.7: Mean number of customer in the system as ρ tends to 1 (Negative Exponential, E_2 and E_{10})

Graphical representation of z_i

The following graphs represent the values of z_i found for various batch sizes, values of ρ and k = 2:

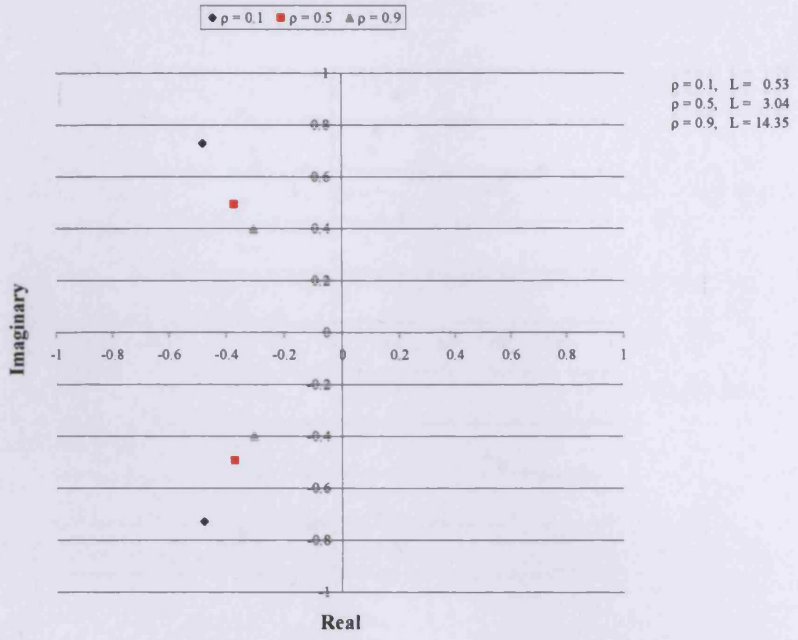


Figure 2.8: z_i values for the E_2 distribution, $s = 3$, $\rho = 0.1$, 0.5, and 0.9

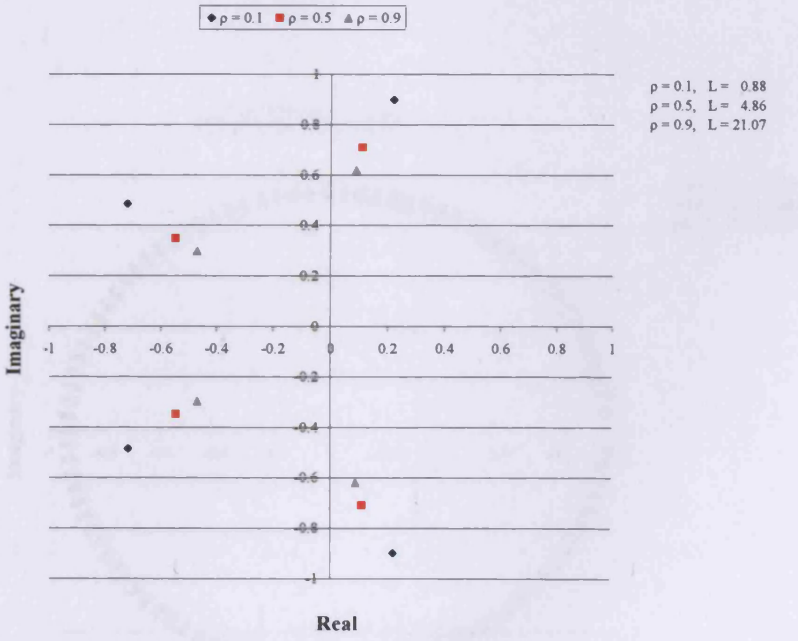


Figure 2.9: z_i values for the E_2 distribution, $s = 5$, $\rho = 0.1$, 0.5, and 0.9

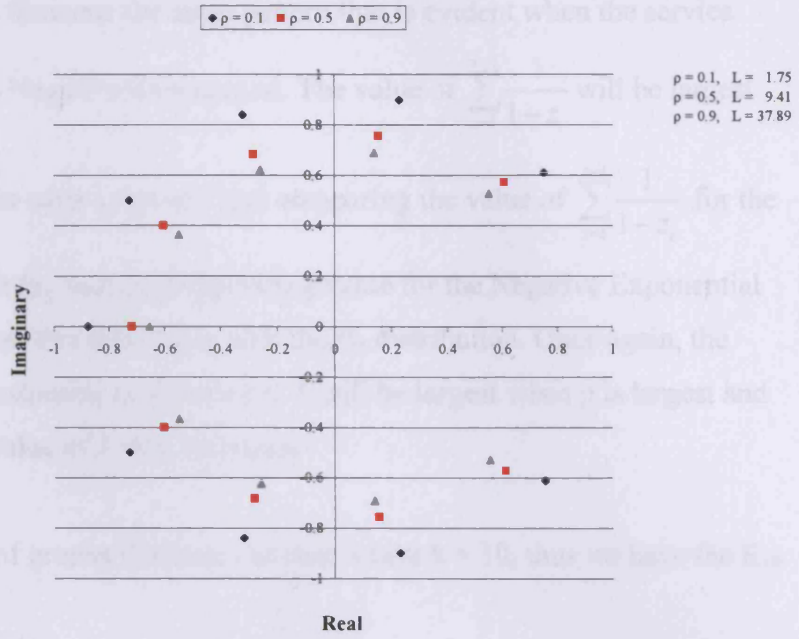


Figure 2.10: z_i values for the E_2 distribution, s = 10, ρ = 0.1, 0.5, and 0.9

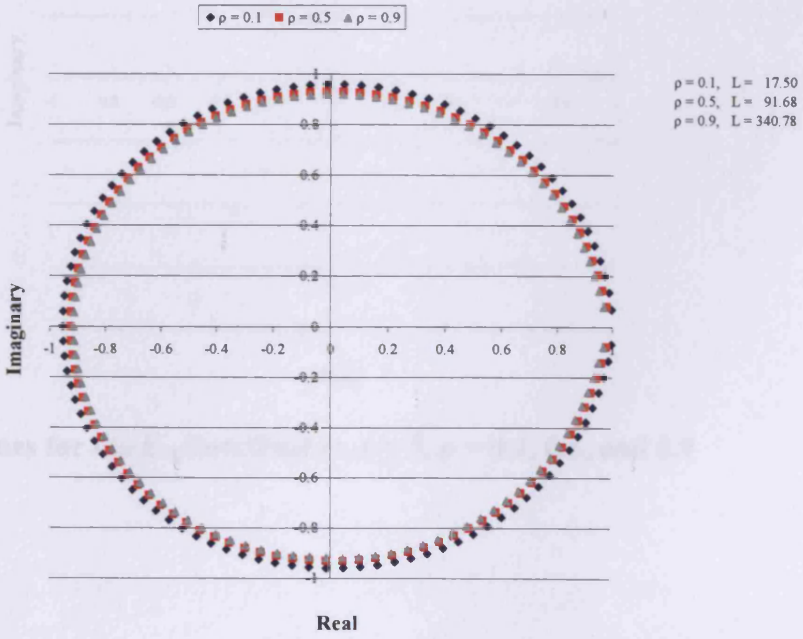


Figure 2.11: z_i values for the E_2 distribution, s = 100, ρ = 0.1, 0.5, and 0.9

Figures 2.8 to 2.11 illustrate the same pattern that is evident when the service time distribution is Negative Exponential. The value of $\sum_{i=1}^{s-1} \frac{1}{1-z_i}$ will be largest when ρ is largest for each value of s and comparing the value of $\sum_{i=1}^{s-1} \frac{1}{1-z_i}$ for the E_2 distribution with the same corresponding value for the Negative Exponential distribution indicates that it is larger with the E_2 distribution. Once again, the mean number of customers in the system, L will be largest when ρ is largest and as s increases the value of L also increases.

The following set of graphs illustrate the case where k = 10, thus we have the E_{10} distribution:

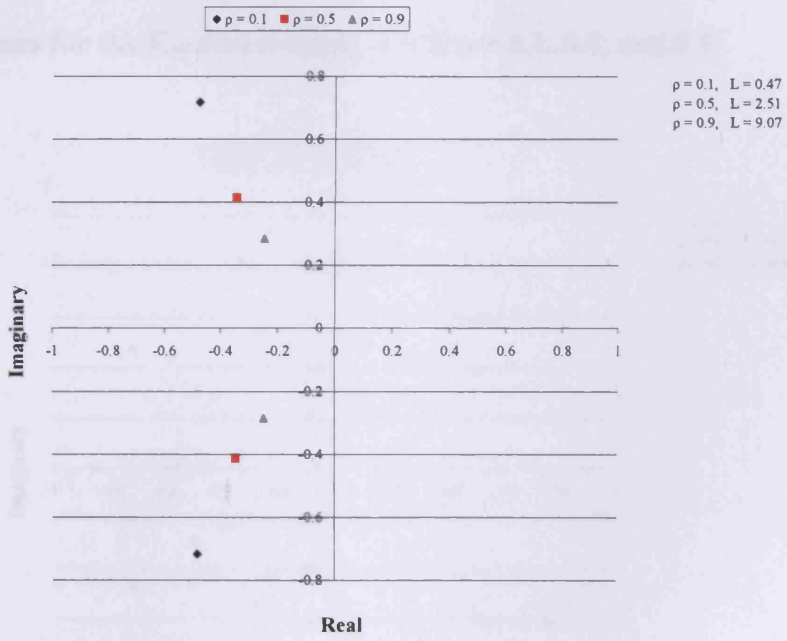


Figure 2.12: z_i values for the E_{10}distribution, s = 3, ρ = 0.1, 0.5, and 0.9

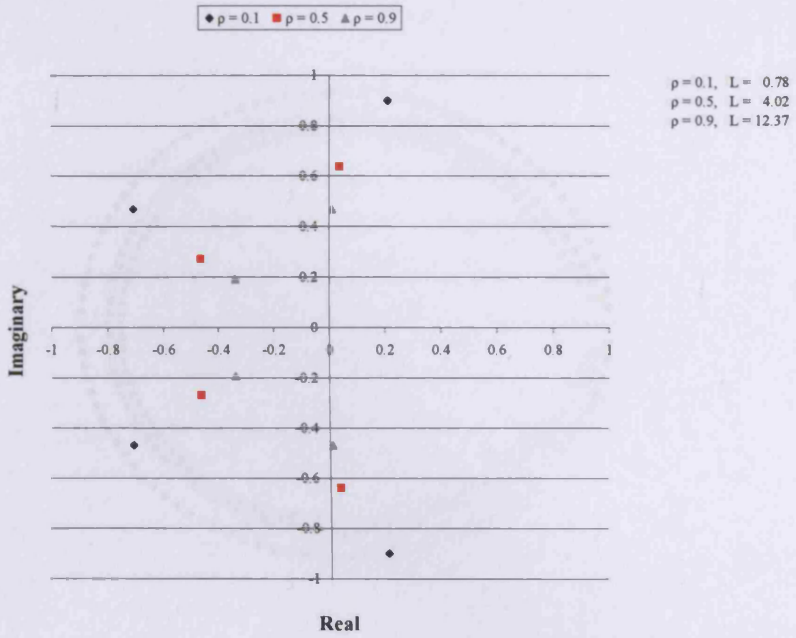


Figure 2.13: z_i values for the E_{10} distribution, $s = 5$, $\rho = 0.1$, 0.5, and 0.9

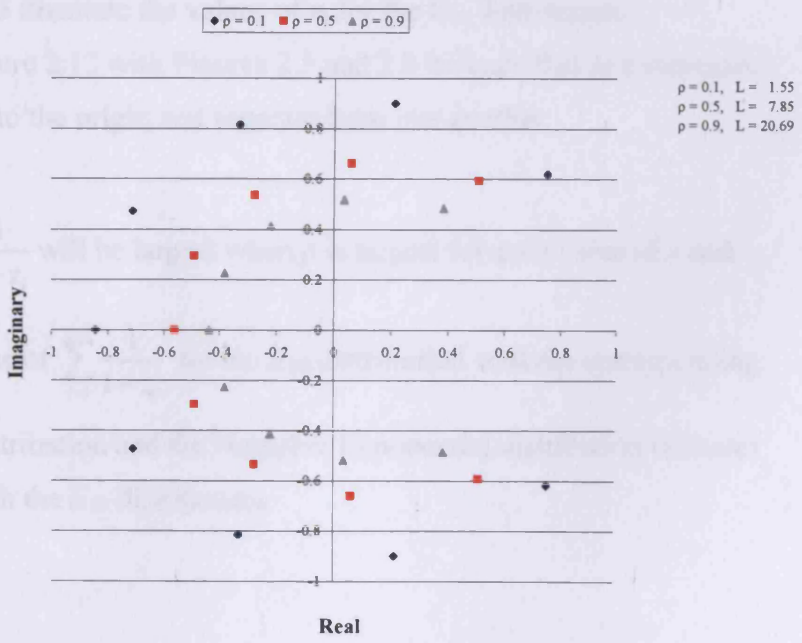


Figure 2.14: z_i values for the E_{10} distribution, $s = 10$, $\rho = 0.1$, 0.5, and 0.9

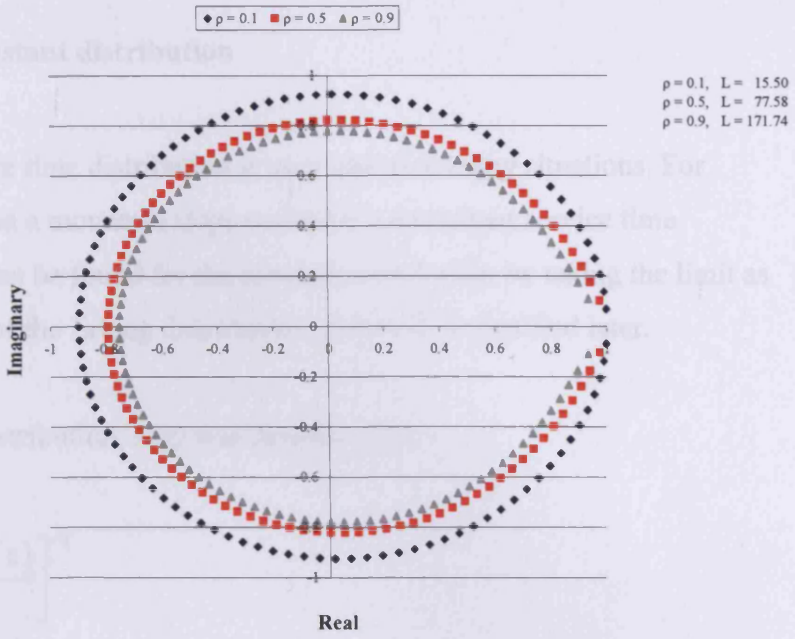


Figure 2.15: z_i values for the E_{10} distribution, s = 100, ρ = 0.1, 0.5, and 0.9

Figures 2.12 to 2.15 illustrate the values of z_i for the E_{10} distribution. Comparison of Figure 2.12 with Figures 2.3 and 2.8 indicate that as k increases, the z_i move nearer to the origin and separate from one another.

The value of $\sum_{i=1}^{s-1} \frac{1}{1-z_i}$ will be largest when ρ is largest for each value of s and comparing the value of $\sum_{i=1}^{s-1} \frac{1}{1-z_i}$ for the E_{10} distribution with the corresponding value for the E_2 distribution and the Negative Exponential distribution indicates that it is largest with the E_{10} distribution.

2.3.3. Constant distribution

The constant service time distribution is very useful in many situations. For example a ski-lift on a mountain slope could have a constant service time distribution. *K(z)* can be found for the constant distribution by taking the limit as k tends to infinity of the Erlang distribution. This will be justified later.

For the Erlang k distribution, *K(z)* was derived above:

$$K(z)=\left[1+\frac{s\rho(1-z)}{k}\right]^{-k}$$

Rearranging this formula gives:

$$K(z)=\frac{1}{\left[1+\frac{s\rho(1-z)}{k}\right]^{k}}$$

Taking limits gives:

$$\lim_{k\to\infty}\frac{1}{\left[1+\frac{s\rho(1-z)}{k}\right]^{k}}=e^{-s\rho(1-z)},\ \text{from the definition of } e^{x}$$

Therefore $K(z)=e^{-s\rho(1-z)}$

Differentiation of the above gives the following:

$$K'(z)=s\rho e^{-s\rho(1-z)}$$

Differentiation again and setting $z = 1$ gives:

$$K''(1)=(s\rho)^{2}$$

Substitution of this value into (2.3) gives:

$$L^{+d} = \sum_{i=1}^{s-1} \frac{1}{1-z_i} + s\rho + \frac{s\rho^2 - s + 1}{2(1-\rho)} \qquad (2.7)$$

Using the definitions for L and Lq previously noted:

$$L = \sum_{i=1}^{s-1} \frac{1}{1-z_i} + \frac{3s\rho}{2} + \frac{s\rho^2 - s + 1}{2(1-\rho)}$$

$$L_q = \sum_{i=1}^{s-1} \frac{1}{1-z_i} + \frac{s\rho}{2} + \frac{s\rho^2 - s + 1}{2(1-\rho)}$$

The values of z_i are a little more difficult to find for this queueing system. An iterative solution is required by equating real and imaginary parts. This method was devised by Griffiths, Williams and Holland (1991) and a brief outline of the procedure is below:

Equation (2.3) is:

$$\frac{z^s}{K(z)} - 1 = 0$$

$$\frac{z^s}{e^{-s\rho(1-z)}} - 1 = 0$$

$$z^s = e^{s\rho(1-z)}$$

Simplification and substitution of well known representations of complex numbers gives:

$$re^{i\theta} = e^{\rho(1-r(\cos(\theta)+i\sin(\theta)))}$$

Equating real and imaginary parts gives:

Real:

$$r = e^{\rho(1-r\cos(\theta))}$$

Imaginary:

$$\theta = -\rho r \sin(\theta) + \frac{2\pi k}{s}, \quad k = 0, 1, 2, \ldots, (s-1)$$

These equations were then solved iteratively to find the values of r and theta, then these values were substituted into the expression below:

$$z = r(\cos(\theta) + i\sin(\theta))$$

The table below shows the theoretical and simulated summary measures for the queueing system. As can be seen, the simulated values agree very well with the theoretical values thus adding confidence that both methods are working correctly.

Table 2.3: Theoretical and simulated summary measures of the $M/D^{(0,s)}/1$ system, $s = 3, 5, 10, 100$

BATCH SIZE	P	THEORETICAL VALUE	SIMULATED VALUE
3	0.1	$Lq = 0.15$	$Lq = 0.15$
		$L = 0.45$	$L = 0.45$
	0.5	$Lq = 0.88$	$Lq = 0.88$
		$L = 2.38$	$L = 2.38$
	0.9	$Lq = 5.07$	$Lq = 5.08$
		$L = 7.77$	$L = 7.78$
5	0.1	$Lq = 0.25$	$Lq = 0.25$
		$L = 0.75$	$L = 0.75$
	0.5	$Lq = 1.33$	$Lq = 1.35$
		$L = 3.83$	$L = 3.85$
	0.9	$Lq = 5.75$	$Lq = 5.78$
		$L = 10.25$	$L = 10.28$
10	0.1	$Lq = 0.5$	$Lq = 0.5$
		$L = 1.5$	$L = 1.5$
	0.5	$Lq = 2.52$	$Lq = 2.55$
		$L = 7.52$	$L = 7.55$
	0.9	$Lq = 7.60$	$Lq = 7.65$
		$L = 16.60$	$L = 16.65$
100	0.1	$Lq = 5.00$	$Lq = 5.00$
		$L = 15.00$	$L = 15.00$
	0.5	$Lq = 25.00$	$Lq = 25.00$
		$L = 75.00$	$L = 75.00$
	0.9	$Lq = 46.11$	$Lq = 46.21$
		$L = 136.11$	$L = 136.21$

The batch size, once again, dictates the number of roots to be found. If s is the batch size, then $s - 1$ roots need to be found within the unit circle. If s is an odd

number, there will be $\frac{s}{2}$ conjugate pairs. If s is an even number, there will be one real number then $\frac{s-1}{2}$ conjugate pairs.

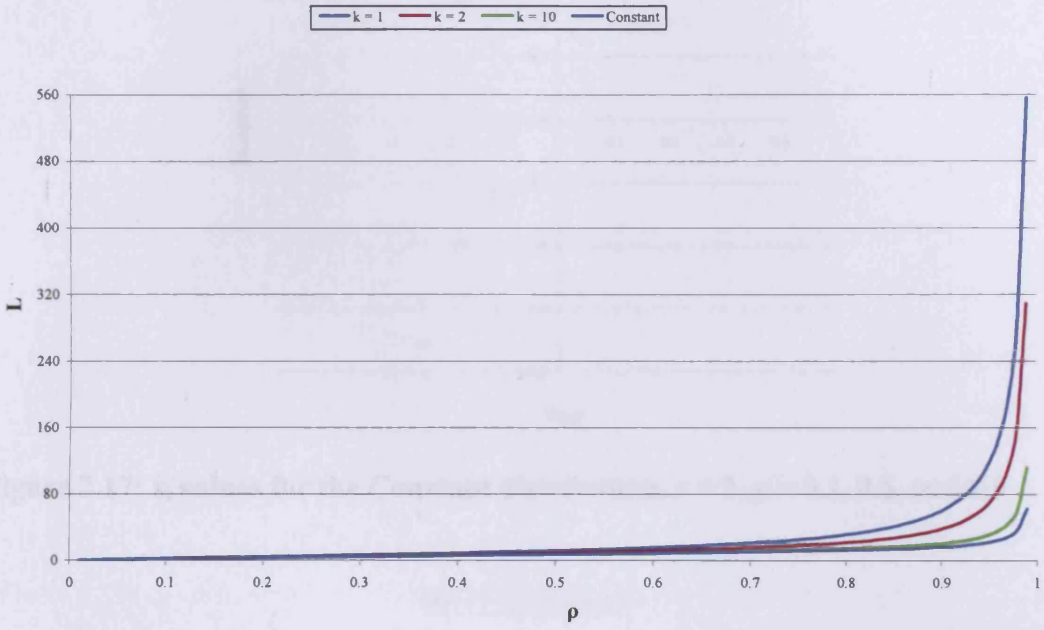


Figure 2.16: Mean number of customers in the system against ρ (Negative Exponential, E_2, E_{10} and Constant)

Figure 2.16 displays the mean number of customers in the system for values of ρ for four different service time distributions; the Negative Exponential, the E_2, the E_{10} and the Constant distribution. It is evident that as k increases, L decreases. The largest mean occurs when the service time distribution is Negative Exponential and the smallest when we have constant service times. The reason for this is that the coefficient of variation is largest for the Negative Exponential distribution and smallest (i.e. 0) for the constant distribution. However, service time distributions exist which have coefficients of variation greater than 1, so it is possible to find service time distributions which will produce greater values of L (e.g. Hyperexponential distribution).

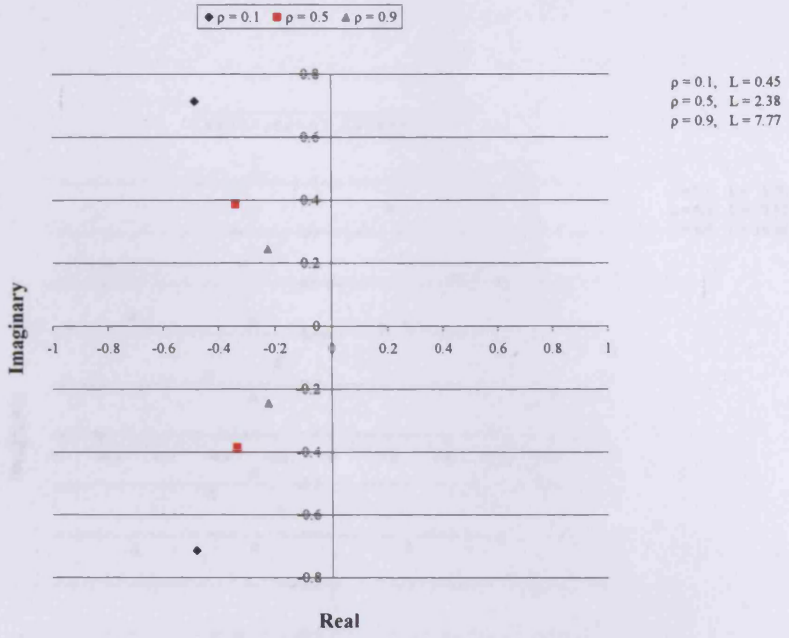


Figure 2.17: z_i values for the Constant distribution, s = 3, ρ = 0.1, 0.5, and 0.9

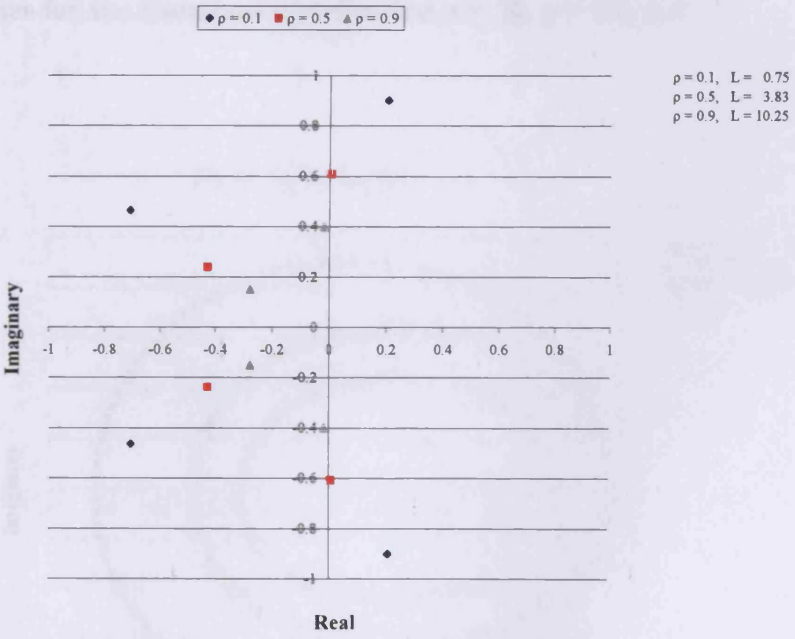


Figure 2.18: z_i values for the Constant distribution, s = 5, ρ = 0.1, 0.5, and 0.9

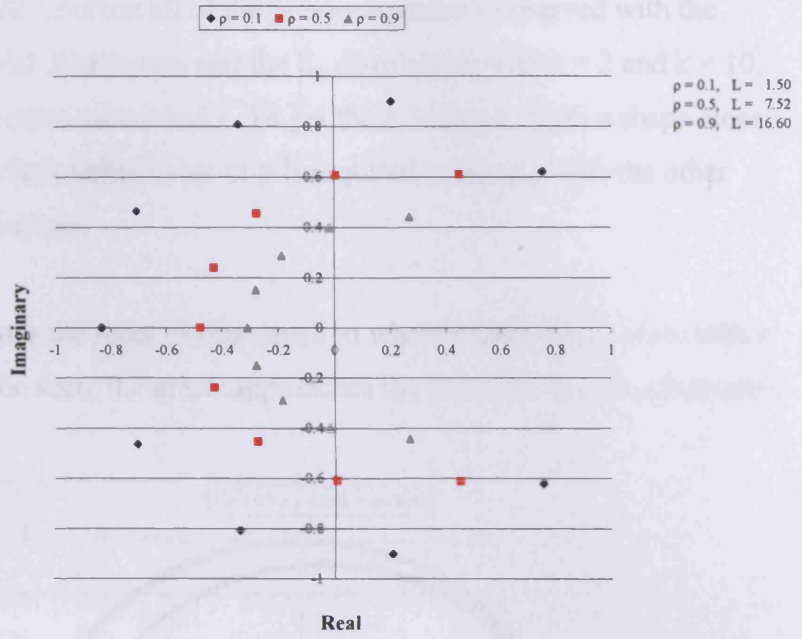


Figure 2.19: z_i values for the Constant distribution, s = 10, ρ = 0.1, 0.5, and 0.9

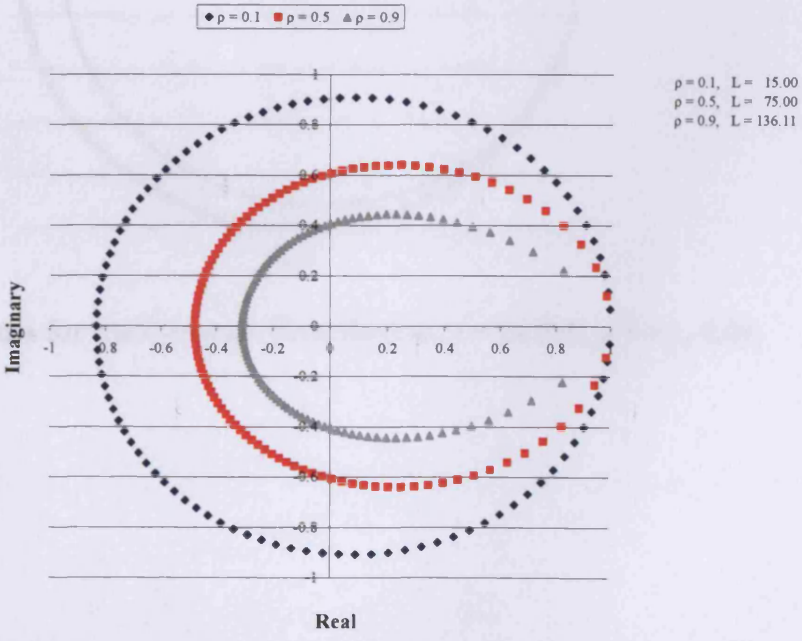


Figure 2.20: z_i values for the Constant distribution, s = 100, ρ = 0.1, 0.5, and 0.9

As has been mentioned several times already, the constant distribution is simply the E_k distribution with limits taken as k tends to infinity. The above graphs (Figures 2.17 to 2.20) confirm all of the previous patterns observed with the Negative Exponential distribution and the E_k distribution with k = 2 and k = 10. The value of L increases with ρ and s. To get the z_i values to form a shape close to the unit circle, a far smaller value of ρ is required compared with the other service time distributions.

Figure 2.21 illustrates the roots of this equation where ρ approaches zero (with s = 10,000). As can be seen, the graph approaches the unit circle as ρ tends to zero.

Figure 2.21: z_i values for the Constant distribution, s = 10,000, ρ = 0.1, 0.01, and 0.001

2.3.4. Hyperexponential distribution

Another service time distribution of interest is the Hyperexponential distribution. This distribution deals adequately with data which has a significant probability of obtaining a small value and a small probability of obtaining a large value (i.e. long tails).

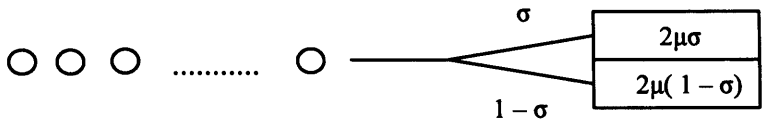


Figure 2.22: The M/H$^{(0,s)}$/1 queueing system

The Hyperexponential distribution is illustrated in Figure 2.22. Customers arrive singly and form a queue. With a probability σ, they enter the top branch of the service facility where they will be served according to the Negative Exponential distribution with a mean rate 2μσ.With a probability (1 – σ) they will enter the bottom branch of service where, once again, they will be served according to the Negative Exponential distribution with a mean rate 2μ(1 – σ). The service rate for each branch has been chosen in order to ensure that the overall service rate is μ.

The PDF of the above Hyperexponential distribution is as follows:

$$f(t)=\sigma\left(2\mu\sigma e^{-2\mu\sigma t}\right)+(1-\sigma)\left(2\mu(1-\sigma)e^{-2\mu(1-\sigma)t}\right), \text{ where } t\geq 0$$

From the above PDF, we can calculate the expected value and the variance of this distribution:

$$E(t)=\int_0^{\infty}\left[\sigma t\left(2\mu\sigma e^{-2\mu\sigma t}\right)+(1-\sigma)t\left(2\mu(1-\sigma)e^{-2\mu(1-\sigma)t}\right)\right]dt$$

$$=\frac{1}{\mu}$$

$$E(t^2) = \int_0^\infty \left[\sigma t^2 \left(2\mu\sigma e^{-2\mu\sigma t} \right) + (1-\sigma) t^2 \left(2\mu(1-\sigma) e^{-2\mu(1-\sigma)t} \right) \right] dt$$

$$= \frac{1}{2\mu^2\sigma(1-\sigma)}$$

Therefore

$$Var(t) = \frac{1}{2\mu^2\sigma(1-\sigma)} - \frac{1}{\mu^2}$$

$$= \frac{1-2\sigma+2\sigma^2}{2\mu^2\sigma(1-\sigma)}$$

The coefficient of variation is defined as $\frac{\text{standard deviation}}{\text{mean}}$.

$$CV^2(t) = \frac{\frac{1-2\sigma+2\sigma^2}{2\mu^2\sigma(1-\sigma)}}{\frac{1}{\mu^2}}$$

$$= \frac{1-2\sigma+2\sigma^2}{2\sigma(1-\sigma)}$$

$$= \frac{1}{2\sigma(1-\sigma)} - 1$$

Hence

$$CV(t) = \sqrt{\frac{1}{2\sigma(1-\sigma)} - 1}$$

Differentiation of the above expression gives:

$$\frac{dCV^2(t)}{d\sigma} = \frac{2\sigma-1}{2\sigma^2(1-\sigma)^2}$$

This is zero when $\sigma = 0.5$. This is a minimum point since

$\frac{d^2CV^2(t)}{d\sigma^2} = 16$ when σ = 0.5. The minimum value of $CV^2(t) = 1$. Therefore,

$CV(t) \geq 1$

Figure 2.23 illustrates the behaviour of the coefficient of variation as σ ranges from near zero to 0.5:

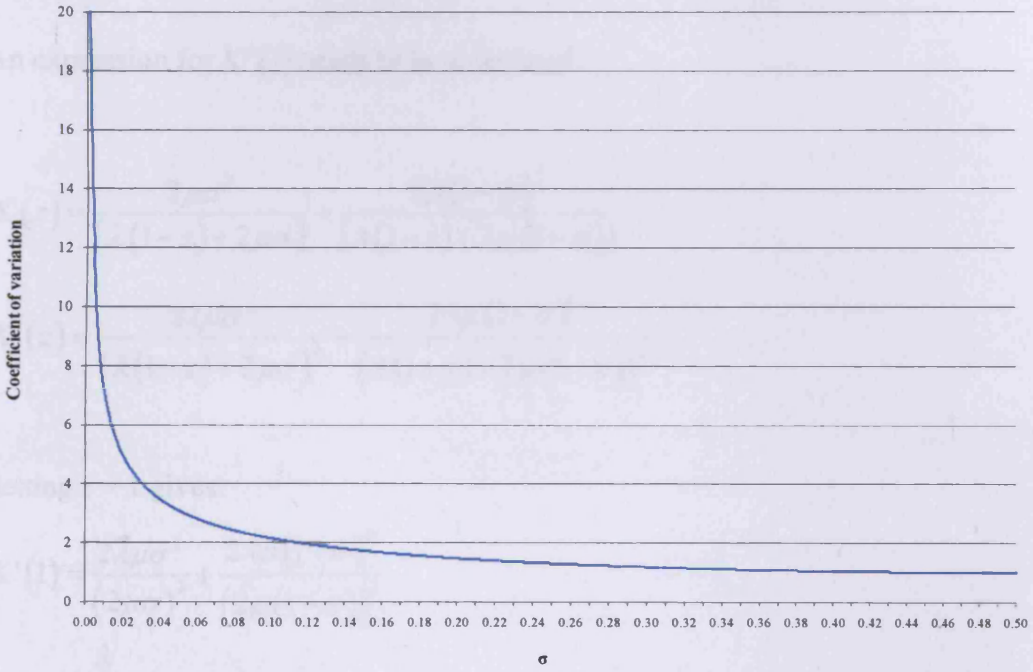


Figure 2.23: Coefficient of variation for the Hyperexponential distribution

The coefficient of variation has a value greater than 1 for $0 < \sigma < 0.5$, but for $\sigma = 0.5$ the coefficient of variation has a value of 1.

In order to solve the batch service queueing system, we need to find *K(z)*, i.e. the Laplace transform of the service time distribution.

$$\mathcal{L}_x\{f(t)\} = \int_0^\infty e^{-xt} f(t)dt = \int_0^\infty e^{-xt}\left(2\mu\sigma^2 e^{-2\mu\sigma t} + 2\mu(1-\sigma)^2 e^{-2\mu(1-\sigma)t}\right)dt$$

$$= \frac{2\mu\sigma^2}{(x+2\mu\sigma)} + \frac{2\mu(1-\sigma)^2}{(x+2\mu(1-\sigma))}$$

Substitution of $x = \lambda(1-z)$ into the above equation gives the following:

$$\mathcal{L}_{\lambda(1-z)}\{f(t)\} = \frac{2\mu\sigma^2}{(\lambda(1-z)+2\mu\sigma)} + \frac{2\mu(1-\sigma)^2}{(\lambda(1-z)+2\mu(1-\sigma))}$$

$$= K(z)$$

An expression for $K''(1)$ needs to be calculated:

$$K(z) = \frac{2\mu\sigma^2}{(\lambda(1-z)+2\mu\sigma)} + \frac{2\mu(1-\sigma)^2}{(\lambda(1-z)+2\mu(1-\sigma))}$$

$$K'(z) = \frac{2\lambda\mu\sigma^2}{(\lambda(1-z)+2\mu\sigma)^2} + \frac{2\lambda\mu(1-\sigma)^2}{(\lambda(1-z)+2\mu(1-\sigma))^2}$$

Setting $z = 1$ gives:

$$K'(1) = \frac{2\lambda\mu\sigma^2}{(2\mu\sigma)^2} + \frac{2\lambda\mu(1-\sigma)^2}{(2\mu(1-\sigma))^2}$$

$$= \frac{\lambda}{\mu}$$

$$= s\rho$$

as expected.

Then,

$$K''(z) = \frac{4\lambda^2\mu\sigma^2}{(\lambda(1-z)+2\mu\sigma)^3} + \frac{4\lambda^2\mu(1-\sigma)^2}{(\lambda(1-z)+2\mu(1-\sigma))^3}$$

$$K''(1) = \frac{4\lambda^2\mu\sigma^2}{(2\mu\sigma)^3} + \frac{4\lambda^2\mu(1-\sigma)^2}{(2\mu(1-\sigma))^3}$$

$$= \frac{\lambda^2}{2\mu^2\sigma} + \frac{\lambda^2}{2\mu^2(1-\sigma)}$$

$$= \frac{\lambda^2}{2\mu^2\sigma(1-\sigma)}$$

Hence, (2.3) becomes

$$G'(1)=\sum_{i=1}^{s-1}\frac{1}{1-z_i}+s\rho+\frac{s^2\rho+\dfrac{\lambda^2}{2\mu^2\sigma(1-\sigma)}-s\left(s\rho+(s-1)\right)}{2s(1-\rho)} \qquad (2.8)$$
$$=\sum_{i=1}^{s-1}\frac{1}{1-z_i}+s\rho-\frac{(s-1)}{2(1-\rho)}+\frac{s\rho^2}{4\sigma(1-\sigma)(1-\rho)}$$

To check this expression, the trivial case $\sigma=0.5$ is tested. This should yield the same results for $G'(1)$ as the eEponential distribution version of this formula.

For the Exponential case, $G'(1)=\sum_{i=1}^{s-1}\frac{1}{1-z_i}+\frac{2s\rho-s+1}{2(1-\rho)}$.

Substitution of $\sigma=0.5$ into the above equation gives:

$$G'(1)=\sum_{i=1}^{s-1}\frac{1}{1-z_i}+s\rho-\frac{(s-1)}{2(1-\rho)}+\frac{s\rho^2}{4(0.5)(0.5)(1-\rho)}$$
$$=\sum_{i=1}^{s-1}\frac{1}{1-z_i}+\frac{2s\rho-s+1}{2(1-\rho)}$$

To find the summary measures of the batch service queue, the zeros of the denominator of *G(z)* need to be found.

$$G(z)=\frac{s(1-\rho)(z-1)\prod_{i=1}^{s-1}\left(\dfrac{z-z_i}{1-z_i}\right)}{\dfrac{z^s}{K(z)}-1}$$

where $K(z)=\dfrac{2\mu\sigma^2}{(\lambda(1-z)+2\mu\sigma)}+\dfrac{2\mu(1-\sigma)^2}{(\lambda(1-z)+2\mu(1-\sigma))}$

Therefore it is necessary to solve

$$\frac{z^s}{K(z)}-1=0$$

$$\therefore z^s=\frac{2\mu\sigma^2}{\left(\lambda(1-z)+2\mu\sigma\right)}+\frac{2\mu(1-\sigma)^2}{\left(\lambda(1-z)+2\mu(1-\sigma)\right)} \tag{2.9}$$

Special case, $s = 1$

Firstly, considering the special case where $s = 1$, i.e. a batch size of 1. The following equation needs to be solved:

$$z=\frac{2\mu\sigma^2}{\left(\lambda(1-z)+2\mu\sigma\right)}+\frac{2\mu(1-\sigma)^2}{\left(\lambda(1-z)+2\mu(1-\sigma)\right)}$$

Rearranging gives:

$$z=\frac{\left(4\mu\sigma^2+2\mu-4\mu\sigma\right)\lambda(1-z)+\left(\sigma-\sigma^2\right)4\mu^2}{z^2\lambda^2+\left(-2\lambda^2-2\mu\sigma\lambda-2\mu(1-\sigma)\lambda\right)z+\left(\lambda^2+2\mu\sigma\lambda+2\mu(1-\sigma)\lambda+4\mu^2\sigma(1-\sigma)\right)}$$

Now, collecting terms gives:

$$\begin{aligned}0=&\lambda^2z^3+\\&\left(-2\lambda(\lambda+\mu)\right)z^2+\\&\left(4\mu\sigma^2\lambda+2\mu\lambda-4\mu\sigma\lambda+\lambda^2+2\mu\lambda+4\mu^2\sigma(1-\sigma)\right)z+\\&\left(2\mu\left(2\sigma(1-\sigma)(\lambda-\mu)-\lambda\right)\right)\end{aligned}$$

A solution is known to exist at $z = 1$. After long division, the above equation factorises to:

$$0=(z-1)\left(\lambda^2z^2-\left(\lambda^2+2\mu\lambda\right)z+\left(2\mu\left(\lambda-2\sigma(1-\sigma)(\lambda-\mu)\right)\right)\right)$$

This quadratic equation can easily be solved using the quadratic formula to find the remaining roots:

$$z_{1,2} = \frac{(\lambda^2 + 2\mu\lambda) \pm \sqrt{(\lambda^2 + 2\mu\lambda)^2 - 4\lambda^2 (2\mu(\lambda - 2\sigma(1-\sigma)(\lambda - \mu)))}}{2\lambda^2}$$

Which simplifies to

$$z_{1,2} = \frac{(\lambda + 2\mu) \pm \sqrt{(\lambda - 2\mu)^2 + 16\mu\sigma(1-\sigma)(\lambda - \mu)}}{2\lambda}$$

This equation can be verified easily for the special case where $\sigma = 0.5$ i.e. when we have a Negative Exponential server.

Using the above equation,

$$z_{1,2} = \frac{(\lambda + 2\mu) \pm \sqrt{(\lambda - 2\mu)^2 + 4\mu(\lambda - \mu)}}{2\lambda}$$

$$= \frac{(\lambda + 2\mu) \pm \sqrt{\lambda^2}}{2\lambda}$$

$$= \frac{(\lambda + 2\mu) \pm \lambda}{2\lambda}$$

Giving:

$$z_1 = \frac{(\lambda + 2\mu) + \lambda}{2\lambda}$$

$$= \frac{\lambda + \mu}{\lambda}$$

$$= 1 + \frac{1}{\rho}, \text{ a root outside the unit circle.}$$

$$z_2 = \frac{(\lambda + 2\mu) - \lambda}{2\lambda}$$

$$= \frac{\mu}{\lambda}$$

$$= \frac{1}{\rho}, \text{another root outside the unit circle.}$$

Therefore, there are no roots within the unit circle, so $G'(1)$ collapses to be:

$$G'(1) = s\rho - \frac{(s-1)}{2(1-\rho)} + \frac{s\rho^2}{4\sigma(1-\sigma)(1-\rho)}$$

$$= \rho + \frac{\rho^2}{(1-\rho)}$$

$$= \frac{\rho}{(1-\rho)}$$

when $\sigma = 0.5$

This is expected since this is the M/M/1 queue.

General batch size, *s*

Extending the theory further to a batch size of *s* gives:

$$z^s = \frac{2\mu\sigma^2}{\lambda(1-z)+2\mu\sigma} + \frac{2\mu(1-\sigma)^2}{\lambda(1-z)+2\mu(1-\sigma)}$$

Rearrangement gives:

$$\lambda^2 z^{s+2} - 2\lambda(\lambda+\mu)z^{s+1} + \left(\lambda^2 + 2\mu\lambda + 4\mu^2\sigma(1-\sigma)\right)z^s$$
$$+2\mu\lambda\left(1-2\sigma(1-\sigma)\right)z + 2\mu\left(2\sigma(1-\sigma)(\lambda-\mu)-\lambda\right) = 0$$

Factorisation gives:

$$(z-1)\left(\lambda^2 z^{s+1} - \lambda(\lambda+2\mu)z^s + 4\mu^2\sigma(1-\sigma)z^{s-1} + \ldots + 4\mu^2\sigma(1-\sigma)z + 2\mu\left(\lambda - 2\sigma(1-\sigma)(\lambda-\mu)\right)\right)$$

This will not readily factorise, and solutions were found using Maple.

Table 2.4 summarises the values of *L* and *Lq* found when looking at a variety of batch sizes and values of ρ and σ:

Table 2.4: Theoretical summary measures of the M/He$^{(0,s)}$/1 system, s = 3, 5, 10, 100 and σ = 0.1, ... 0.4

BATCH SIZE	ρ	σ = 0.1	σ = 0.2	σ = 0.3	σ = 0.4
3	0.1	*Lq* =0.88	*Lq* = 0.48	*Lq* = 0.36	*Lq* = 0.32
		L = 1.18	*L* = 0.78	*L* = 0.66	*L* = 0.62
	0.5	*Lq* = 7.52	*Lq* = 3.90	*Lq* = 2.80	*Lq* = 2.36
		L = 9.02	*L* = 5.40	*L* = 4.30	*L* = 3.86
	0.9	*Lq* = 66.26	*Lq* = 33.47	*Lq* = 23.44	*Lq* = 19.44
		L = 68.96	*L* = 36.17	*L* = 26.14	*L* = 22.14
5	0.1	*Lq* = 1.45	*Lq* = 0.80	*Lq* = 0.60	*Lq* = 1.02
		L = 1.95	*L* = 1.30	*L* = 1.10	*L* = 0.52
	0.5	*Lq* = 12.27	*Lq* = 6.24	*Lq* = 4.41	*Lq* = 3.69
		L = 14.77	*L* = 8.74	*L* = 6.91	*L* = 6.19
	0.9	*Lq* = 107.52	*Lq* = 52.88	*Lq* = 36.18	*Lq* = 29.51
		L = 112.02	*L* = 57.38	*L* = 40.68	*L* = 34.01
10	0.1	*Lq* = 2.88	*Lq* = 1.58	*Lq* = 1.20	*Lq* = 1.04
		L = 3.88	*L* = 2.58	*L* = 2.20	*L* = 2.04
	0.5	*Lq* = 24.16	*Lq* = 12.10	*Lq* = 8.46	*Lq* = 7.02
		L = 29.16	*L* = 17.10	*L* = 13.46	*L* = 12.02
	0.9	*Lq* = 210.68	*Lq* = 101.41	*Lq* = 68.01	*Lq* = 54.67
		L = 219.68	*L* = 110.41	*L* = 77.01	*L* = 63.67
100	0.1	*Lq* = 28.63	*Lq* = 15.73	*Lq* = 11.92	*Lq* = 10.42
		L = 38.63	*L* = 25.73	*L* = 21.92	*L* = 20.42
	0.5	*Lq* = 238.19	*Lq* = 117.74	*Lq* = 81.39	*Lq* = 67.07
		L = 288.19	*L* = 167.74	*L* = 131.39	*L* = 117.07
	0.9	*Lq* = 2067.68	*Lq* = 975.01	*Lq* = 641.06	*Lq* = 507.71
		L = 2157.68	*L* = 1065.01	*L* = 731.06	*L* = 597.71

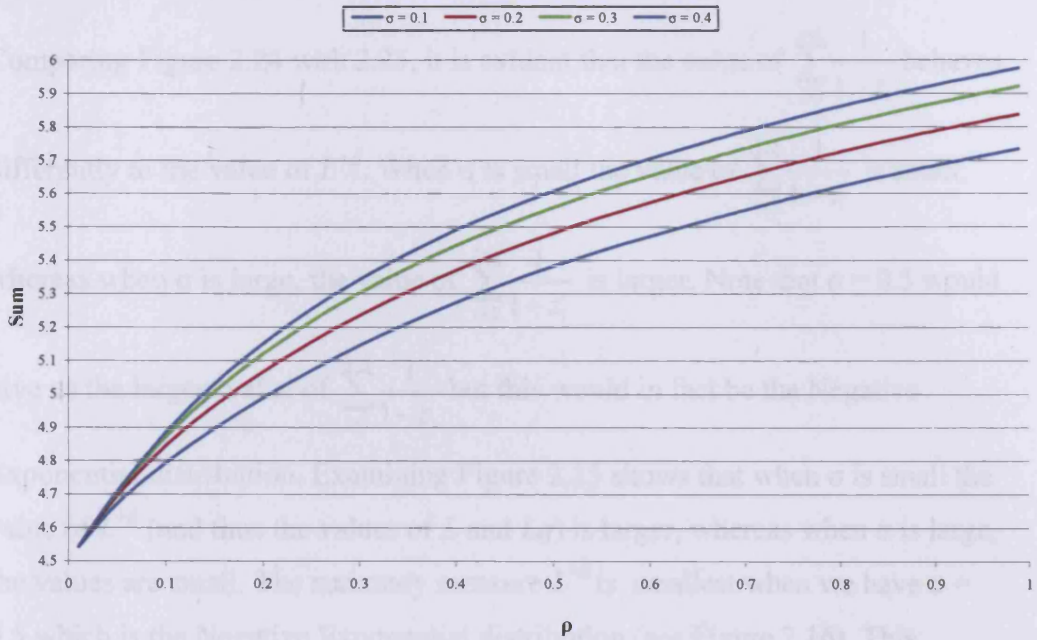


Figure 2.24: Value of $\sum_{i=1}^{s-1} \frac{1}{1-z_i}$ against ρ, σ = 0.1, ..., 0.4

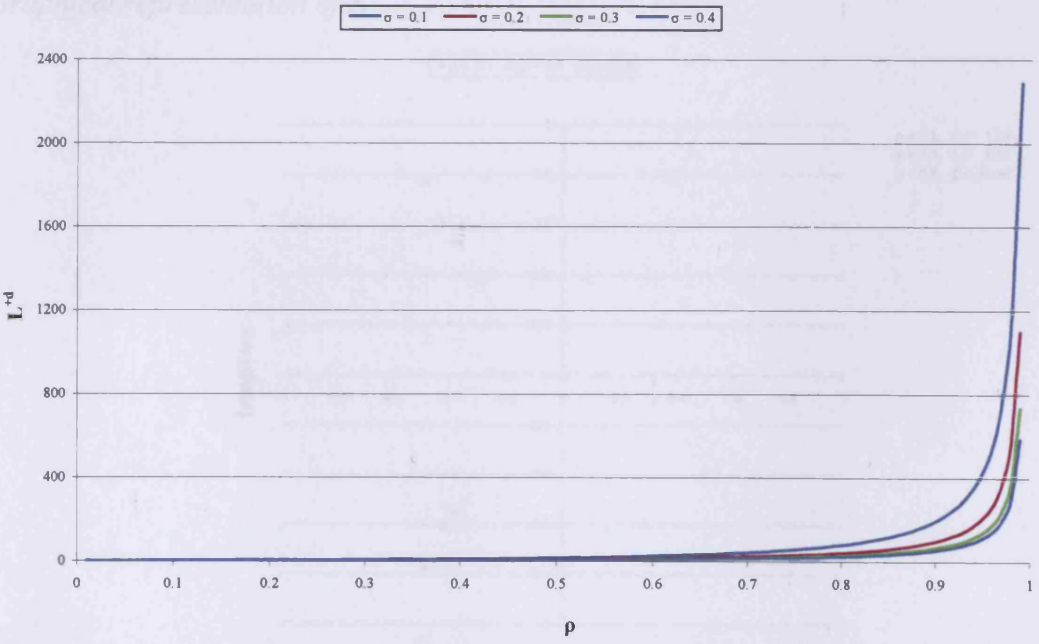


Figure 2.25: Value of L^{+d} against ρ, σ = 0.1, ..., 0.4

Comparing Figure 2.24 with 2.25, it is evident that the value of $\sum_{i=1}^{s-1} \frac{1}{1-z_i}$ behaves differently to the value of L^{+d}. When σ is small the value of $\sum_{i=1}^{s-1} \frac{1}{1-z_i}$ is small, whereas when σ is large, the value of $\sum_{i=1}^{s-1} \frac{1}{1-z_i}$ is larger. Note that σ = 0.5 would give us the largest value of $\sum_{i=1}^{s-1} \frac{1}{1-z_i}$ but this would in fact be the Negative Exponential distribution. Examining Figure 2.25 shows that when σ is small the value of L^{+d} (and thus the values of L and Lq) is larger, whereas when σ is large, the values are small. The summary measure L^{+d} is smallest when we have σ = 0.5 which is the Negative Exponential distribution (see Figure 2.16). This implies, that of the five distributions that have been considered in this Chapter, the Hyperexponential distribution with a small value of σ gives the highest values for the summary measures.

Graphical representation of z_i

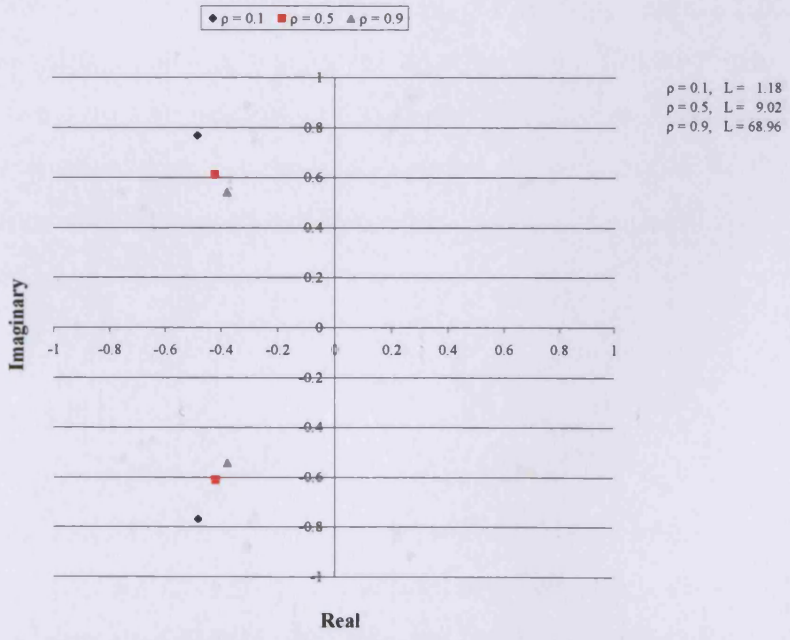


Figure 2.26: z_i values for the Hyperexponential distribution, s = 3, σ = 0.1, ρ = 0.1, 0.5, and 0.9

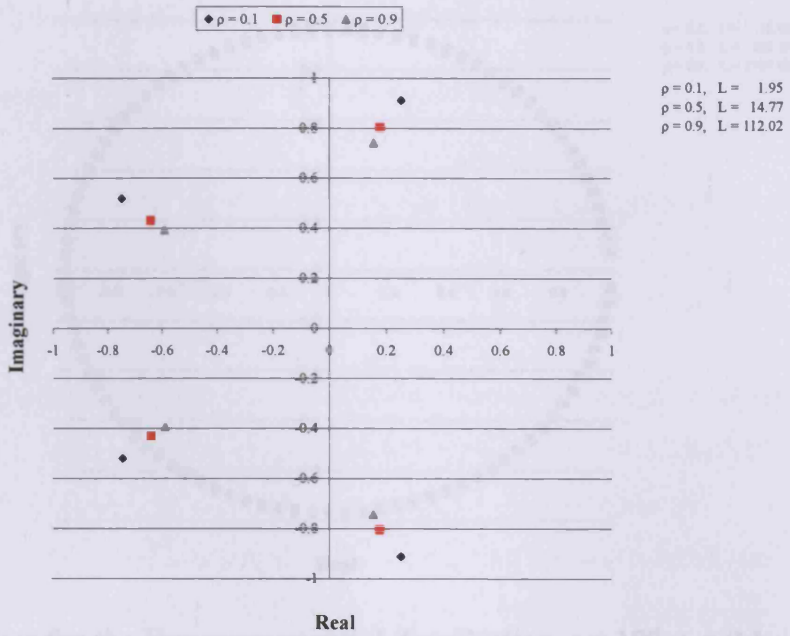


Figure 2.27: z_i values for the Hyperexponential distribution, $s = 5$, $\sigma = 0.1$, $\rho = 0.1, 0.5$, and 0.9

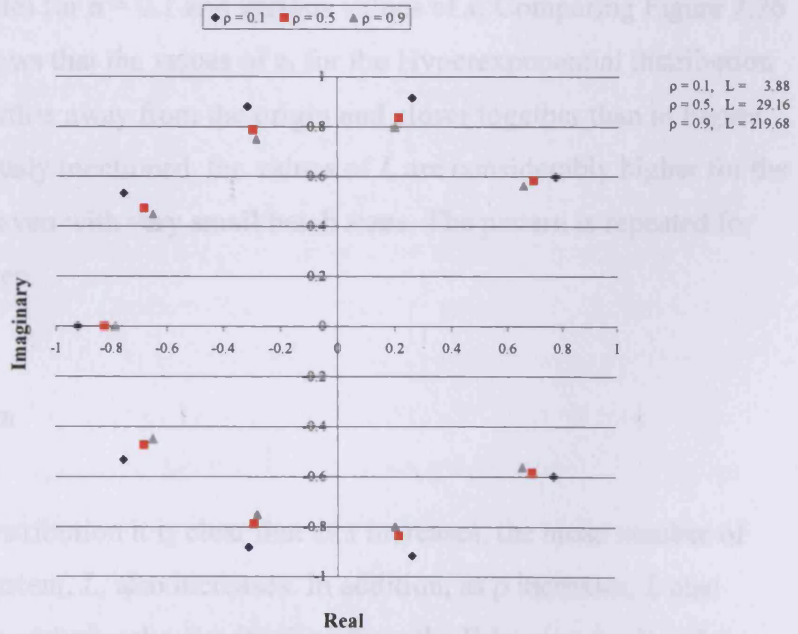


Figure 2.28: z_i values for the Hyperexponential distribution, $s = 10$, $\sigma = 0.1$, $\rho = 0.1, 0.5$, and 0.9

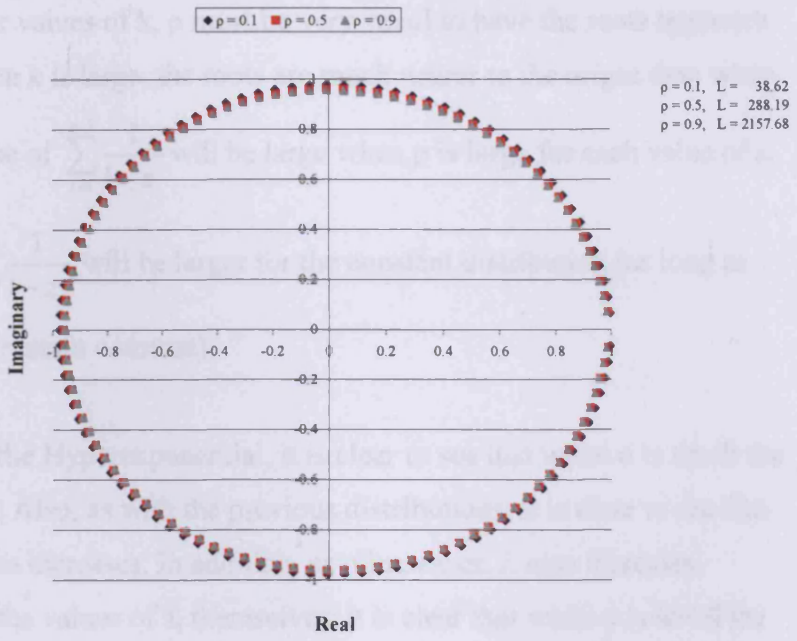


Figure 2.29: z_i values for the Hyperexponential distribution, s = 100, σ = 0.1, ρ = 0.1, 0.5, and 0.9

Figures 2.26 to 2.29 illustrate the values of z_i (i.e. the roots of (2.8) which are within the unit circle) for $\sigma = 0.1$ and various values of s. Comparing Figure 2.26 with Figure 2.3 shows that the values of z_i for the Hyperexponential distribution with $\sigma = 0.1$ are further away from the origin and closer together than in Figure 2.3. As was previously mentioned, the values of L are considerably higher for the Hyperexponential even with very small batch sizes. The pattern is repeated for each value of s given.

2.4. Conclusion

For each service distribution it is clear that as s increases, the mean number of customers in the system, L, also increases. In addition, as ρ increases, L also increases. When considering the distributions from the Erlangian family (the Negative Exponential, E_2, E_{10} and the Constant) it can be seen that as k increases, L decreases. Looking at the values of z_i, it is evident that when ρ is small, the z_i values, when plotted on an Argand diagram, are nearer to the unit circle. Also,

when s is large and ρ is small, these roots tend quickly towards the unit circle. Although, for larger values of k, ρ must be very small to have the roots approach the unit circle. When k is large, the roots are much nearer to the origin than when k is small. The value of $\sum_{i=1}^{s-1} \frac{1}{1-z_i}$ will be large when ρ is large for each value of s, but the value of $\sum_{i=1}^{s-1} \frac{1}{1-z_i}$ will be larger for the constant distribution (as long as all other quantities remain constant).

When considering the Hyperexponential, it is clear to see that when σ is small the value of L is larger. Also, as with the previous distributions, it is clear to see that as s increases, L also increases. In addition, as ρ increases, L also increases. When considering the values of z_i themselves, it is clear that when σ is small the roots are further away from the origin and closer together compared to when σ is large. Moreover, when σ is small the value of $\sum_{i=1}^{s-1} \frac{1}{1-z_i}$ is small, whereas when σ is large, the value of $\sum_{i=1}^{s-1} \frac{1}{1-z_i}$ is larger.

In this Chapter, the theory has been presented. This theoretical work can be applied to a variety of real life situations, many of which are in the medical field. As well as the applications mentioned in Section 2.1, the theory could be applied to a blood bank situation, where blood is taken from a patient singly and then a batch of blood is "served" together in a screening process. Also, the theory, with a slight modification, could be applied to operating theatres, whereby patients who require surgery "arrive" (or become ill) individually, and are then "served" in a batch in parallel operating theatres, with one surgical team serving the entire batch. Research has been done in this field but without utilising this batch service queueing theory.

In light of these conclusions, the work in this Chapter acts as a foundation to the work found in Chapter 3 where the first objective is met in its fullness.

Chapter 3: Batch arrival queues

3.1. Introduction

Batch arrival queues occur in many places in the real world. For example, batches of people will arrive each day at a hospital and will then be served singly or batches of mail will arrive at a sorting office and will be sorted singly. This Chapter, which seeks to address the first objective found in section 1.2, will commence by considering the $M^{(b)}/M/1$ queue, that is, Markovian batch arrivals of size b and Markovian single service with one server. Then, this work will be extended to investigate the Hyperexponential service time distribution, with one server. Finally, the steady state queueing equations of the $M^{(b)}/He/c$ queue will be derived and this theory will be applied to a real life situation, namely the Critical Care Unit at the University Hospital of Wales, Cardiff. This queueing system has not been developed in the literature and thus is one of the most prominent research contributions in this thesis.

3.2. The $M^{(b)}/M/1$ queue

Let $P_n(t)$ denote the probability of having n customers in the system at time t.

Let arrivals occur in batches at a mean rate λ according to a Negative Exponential distribution and let batches of size i arrive where

$P(\text{Batch size} = i) = b_i,\ i = 0, 1, \ldots$

It is assumed that batches of size zero are allowed in this Chapter. It is assumed that customers are served singly according to the Negative Exponential distribution with a mean rate μ.

Let $\rho = \frac{\lambda \bar{b}}{\mu}$ be the traffic intensity of this system, where $\bar{b}$ is the mean batch size and μ is the mean service rate.

The following equations can be derived for this situation:

$$P_0(t+\delta t) = P_0(t)\left[1-\lambda\left(\sum_{i=0}^{\infty} b_i\right)\delta t + o(\delta t)\right] + P_1(t)\left[1-\lambda\left(\sum_{i=0}^{\infty} b_i\right)\delta t + o(\delta t)\right]\left[\mu\delta t + o(\delta t)\right] + o(\delta t)$$

Before solving this equation, it is worth noting that $\sum_{i=0}^{\infty} b_i = 1$ and hence will not be included in any further expressions.

After rearranging and taking limits, the following differential difference equation is obtained:

$$\frac{dP_0(t)}{dt} = -\lambda P_0(t) + \mu P_1(t)$$

Equating to zero gives the following steady state equation:

$$\lambda P_0 = \mu P_1$$

In the same manner, the equations for $n > 0$ are obtained:

$$P_n(t+\delta t) = P_n(t)\left[1-\lambda\delta t + o(\delta t)\right]\left[1-\mu\delta t + o(\delta t)\right] + P_{n+1}(t)\left[1-\lambda\delta t + o(\delta t)\right]\left[\mu\delta t + o(\delta t)\right] + \sum_{k=1}^{n} P_{n-k}(t)\left[\lambda b_k \delta t + o(\delta t)\right]\left[1-\mu\delta t + o(\delta t)\right]$$

After rearranging and taking limits, the following differential difference equation is obtained:

$$\frac{dP_n(t)}{dt} = -(\lambda+\mu)P_n(t) + \mu P_{n+1}(t) + \lambda\sum_{k=1}^{n} P_{n-k}(t)b_k$$

Equating to zero gives the following set of steady state equations:

$$(\lambda+\mu)P_n = \mu P_{n+1} + \lambda\sum_{k=1}^{n} P_{n-k}b_k$$

To calculate summary measures for a queue, $G(z)$, the probability generating function is required.

Let $G(z)=\sum_{n=0}^{\infty} z^n P_n$ and $B(z)=\sum_{n=0}^{\infty} z^n b_n$. Multiplying the steady state equations by ascending powers of z and summing by column gives the following equation:

$$G(z)\left(\lambda+\mu-\frac{\mu}{z}-\lambda B(z)\right) = \mu P_0\left(1-\frac{1}{z}\right) \qquad (3.1)$$

Rearranging gives:

$$G(z) = \frac{\mu P_0(1-z)}{\left(\mu(1-z)-\lambda z(1-B(z))\right)}$$

To complete this expression for $G(z)$, P_0 must be evaluated. Simple use of $G(1)=1$ will not suffice, as the right hand side of this equation will produce $\frac{0}{0}$.

Differentiation of Equation 3.1 gives:

$$G'(z)\left(\lambda+\mu-\frac{\mu}{z}-\lambda B(z)\right) + G(z)\left(\frac{\mu}{z^2}-\lambda B'(z)\right) = \frac{\mu P_0}{z^2} \qquad (3.2)$$

Setting $z = 1$ gives:

$$\mu P_0 = \mu - \lambda B'(1)$$

$$P_0 = 1 - \frac{\lambda B'(1)}{\mu}$$

Now, by definition, $B'(1)$ is the mean batch size. Therefore,

$$P_0 = 1 - \frac{\lambda \bar{b}}{\mu}$$
$$= 1 - \rho$$

Hence, the final expression for $G(z)$ is

$$G(z) = \frac{\mu(1-\rho)(1-z)}{\left(\mu(1-z) - \lambda z(1-B(z))\right)} \tag{3.3}$$

The next step in the derivation of summary measures is to find an expression for $G'(z)$. To do this, (3.2) has to be differentiated again.

$$G''(z)\left(\lambda + \mu - \frac{\mu}{z} - \lambda B(z)\right) + 2G'(z)\left(\frac{\mu}{z^2} - \lambda B'(z)\right) + G(z)\left(-\frac{2\mu}{z^3} - \lambda B''(z)\right) = -\frac{2\mu(1-\rho)}{z^3}$$

Setting $z = 1$ gives:

$$G'(1) = \frac{\lambda B''(1) + 2\rho}{2\left(\mu - \lambda B'(1)\right)}$$

Division of top and bottom by μ gives:

$$G'(1) = \frac{\left(\frac{\rho}{\bar{b}} B''(1)\right) + 2\rho}{2(1-\rho)} \tag{3.4}$$

To complete this expression, $B''(1)$ is required.

3.2.1. Batch size distribution

Batches of fixed size

Let $\overline{b}$ be the number in a batch. Therefore $P\left(B=\overline{b}\right)=1$

Recall, the probability generating function, B(z) is defined as $B(z)=\sum_{i=0}^{\infty}b_i z^i$

Therefore

$$B(z)=\sum_{i=0}^{\infty}b_i z^i$$
$$=z^{\overline{b}}$$

Differentiation of this expression gives:

$$B'(z)=\overline{b}z^{\overline{b}-1}$$

Setting $z = 1$ gives $B'(1)=\overline{b}$ as required

Differentiation again gives:

$$B''(z)=\overline{b}\left(\overline{b}-1\right)z^{\overline{b}-2}$$

Setting $z = 1$ gives:

$$B''(1)=\overline{b}\left(\overline{b}-1\right)$$

Substitution into (3.4) gives:

$$G'(1)=\frac{\left(\frac{\lambda}{\mu}\overline{b}\left(\overline{b}-1\right)\right)+2\rho}{2(1-\rho)}$$
$$=\frac{\rho\left(\overline{b}+1\right)}{2(1-\rho)}$$

A simulation model was built in Simul8 as a check, and the results are outlined in the table below (simulation was run 20 times):

Table 3.1 Summary measures for batch arrivals where batches are of Constant size and service is Exponential

BATCH SIZE 3						BATCH SIZE 10					
ρ	W	W_Q	*L*	*Lq*	$G'(1)$	ρ	W	W_Q	*L*	*Lq*	$G'(1)$
0.1	0.07	0.04	0.22	0.12	0.22	0.1	0.06	0.05	0.61	0.51	0.61
0.2	0.17	0.1	0.5	0.3	0.5	0.2	0.14	0.12	0.37	1.17	1.38
0.3	0.29	0.19	0.86	0.56	0.86	0.3	0.24	0.21	2.36	2.06	2.36
0.4	0.45	0.31	1.34	0.94	1.33	0.4	0.37	0.33	3.68	3.28	3.67
0.5	0.67	0.50	2	1.50	2	0.5	0.55	0.50	5.52	5.02	5.5
0.6	1.00	0.80	3.01	2.41	3	0.6	0.83	0.77	8.28	7.68	8.25
0.7	1.56	1.33	4.68	3.98	4.67	0.7	1.29	1.22	12.85	12.15	12.83
0.8	2.67	2.41	8.01	7.21	8	0.8	2.21	2.13	22.06	22.26	22
0.9	6.02	5.72	18.06	17.16	18	0.9	4.99	4.90	49.89	48.99	49.5

As may be seen, there is good agreement between the simulated values of *L* and the theoretical values for $G'(1)$.

Batches of a variable size (following a Poisson distribution)

Let $\bar{b}$ be the mean number in a batch. Therefore

$$P(B=i)=\frac{(\bar{b})^i e^{-\bar{b}}}{i!}=b_i, \quad i=0, 1, ...$$

Recall, the probability generating function, B(*z*) is defined as:

$$B(z)=\sum_{i=0}^{\infty} b_i z^i$$

Therefore

$$
\begin{aligned}
B(z) &= \sum_{i=0}^{\infty} b_i z^i \\
&= \sum_{i=0}^{\infty} \frac{\left(\bar{b}\right)^i e^{-\bar{b}}}{i!} z^i \\
&= e^{-\bar{c}} \sum_{i=0}^{\infty} \frac{\left(\bar{b}z\right)^i}{i!} \\
&= e^{\bar{b}(z-1)}
\end{aligned}
$$

Differentiation of this expression gives:

$$B'(z) = \bar{b}e^{\bar{b}(z-1)}$$

Setting $z = 1$ gives

$B'(1) = \bar{b}$ as required

Differentiation again gives:

$$B''(z) = \bar{b}^2 e^{\bar{b}(z-1)}$$

Setting $z = 1$ gives:

$$B''(1) = \bar{b}^2$$

Substituting into (3.4) gives:

$$G'(1) = \frac{\left(\frac{\lambda}{\mu}\overline{b}^2\right) + 2\rho}{2(1-\rho)}$$
$$= \frac{\rho(\overline{b}+2)}{2(1-\rho)}$$

Again, a simulation model was built in Simul8 to compare the results, as outlined in the table below (simulation was run 20 times):

Table 3.2: Summary measures for batch arrivals where batch size follows a Poisson distribution and service is Exponential

MEAN BATCH SIZE = 3						MEAN BATCH SIZE = 10					
ρ	W	W_Q	*L*	*Lq*	$G'(1)$	ρ	W	W_Q	*L*	*Lq*	$G'(1)$
0.1	0.09	0.06	0.27	0.17	0.28	0.1	0.07	0.06	0.66	0.56	0.66
0.2	0.21	0.14	0.63	0.43	0.63	0.2	0.15	0.13	1.48	1.28	1.5
0.3	0.36	0.26	1.07	0.77	1.07	0.3	0.26	0.23	2.56	2.26	2.57
0.4	0.56	0.42	1.67	1.27	1.67	0.4	0.4	0.36	4.01	3.61	4
0.5	0.83	0.67	2.5	2	2.5	0.5	0.6	0.55	6.03	5.53	6
0.6	1.25	1.05	3.76	3.16	3.75	0.6	0.91	0.85	9.06	8.46	9
0.7	1.95	1.71	5.85	5.15	5.83	0.7	1.42	1.35	14.15	13.45	14
0.8	3.33	3.07	10.01	9.21	10	0.8	2.43	2.35	24.34	23.54	24
0.9	7.5	7.2	22.53	21.63	22.5	0.9	5.49	5.40	54.96	54.06	54

Once more, we have close agreement between the simulated values of *L* and the theoretical values for $G'(1)$.

3.3. The M$^{(b)}$/He/1 queue

Let $P_{n,1}(t)$ denote the probability of having n customers in the system at time t, with the customer in service in the first branch of the service facility. Let $P_{n,2}(t)$ denote the probability of having n customers in the system at time t, with the customer in service is in the second branch of the service facility. Let arrivals occur in batches at mean rate λ according to a Negative Exponential distribution and let batches of size i arrive where $P(\text{Batch size} = i) = b_i,\ i = 0, 1, ...$

For a more detailed description of the Hyperexponential distribution, see Chapter 2.

Let $\rho = \frac{\lambda \bar{b}}{\mu}$ be the traffic intensity of this system, where $\bar{b}$ is the mean batch size and the overall mean service rate is μ.

The following equation can be derived for this situation:

$$\begin{aligned} P_0(t+\delta t) = {} & P_0(t)\left[1-\lambda\delta t+o(\delta t)\right] \\ & +P_{1,1}(t)\left[1-\lambda\delta t+o(\delta t)\right]\left[2\mu\sigma\delta t+o(\delta t)\right] \\ & +P_{1,2}(t)\left[1-\lambda\delta t+o(\delta t)\right]\left[2\mu(1-\sigma)\delta t+o(\delta t)\right] \\ & +o(\delta t) \end{aligned}$$

After rearranging and taking limits, the following differential difference equation is obtained:

$$\frac{dP_0(t)}{dt} = -\lambda P_0(t) + 2\mu\sigma P_{1,1}(t) + 2\mu(1-\sigma)P_{1,2}(t)$$

Equating to zero gives the following steady state equation:

$$\lambda P_0 = 2\mu\sigma P_{1,1} + 2\mu(1-\sigma)P_{1,2} \qquad (3.5)$$

In the same manner, the equations for $n > 0$ are obtained:

$$\begin{aligned}P_{n,1}(t+\delta t) &= P_{n,1}(t)\left[1-\lambda\delta t+o(\delta t)\right]\left[1-2\mu\sigma\delta t+o(\delta t)\right]\\ &+P_{n+1,1}(t)\left[1-\lambda\delta t+o(\delta t)\right]\left[2\mu\sigma\delta t+o(\delta t)\right]\sigma\\ &+P_{n+1,2}(t)\left[1-\lambda\delta t+o(\delta t)\right]\left[2\mu(1-\sigma)\delta t+o(\delta t)\right]\sigma\\ &+\sum_{j=1}^{n-1}P_{n-j,1}(t)\left[\lambda b_j\delta t+o(\delta t)\right]\left[1-2\mu\sigma\delta t+o(\delta t)\right]\\ &+P_0(t)\left[\lambda b_n\delta t+o(\delta t)\right]\sigma\\ &+o(\delta t)\end{aligned}$$

After rearranging and taking limits, the following differential difference equation is obtained:

$$\frac{dP_{n,1}(t)}{dx} = -(\lambda+2\mu\sigma)P_{n,1}(t)+(2\mu\sigma^2)P_{n+1,1}(t)+(2\mu(1-\sigma)\sigma)P_{n+1,2}(t)+\lambda\sum_{j=1}^{n-1}b_jP_{n-j,1}(t)+\lambda\sigma b_nP_0(t)$$

Equating to zero gives the following steady state equations:

$$(\lambda+2\mu\sigma)P_{n,1} = (2\mu\sigma^2)P_{n+1,1}+(2\mu(1-\sigma)\sigma)P_{n+1,2}+\lambda\sum_{j=1}^{n-1}b_jP_{n-j,1}+\lambda\sigma b_nP_0 \qquad (3.6)$$

Similarly, for the second branch of the service facility, by symmetry,

$$\begin{aligned}(\lambda+2\mu(1-\sigma))P_{n,2} &= (2\mu(1-\sigma)\sigma)P_{n+1,1}+(2\mu(1-\sigma)^2)P_{n+1,2}\\ &+\lambda\sum_{j=1}^{n-1}b_jP_{n-j,2}+\lambda b_n(1-\sigma)P_0\end{aligned} \qquad (3.7)$$

To calculate summary measures for a queue, $G(z)$, the probability generating function is required. It is useful first to define sub-probability generating functions,

Define $G_1(z)=\sum_{j=1}^{\infty}z^jP_{j,1}$, $G_2(z)=\sum_{j=1}^{\infty}z^jP_{j,2}$ and $B(z)=\sum_{j=1}^{\infty}z^jb_j$

Then $G(z)=P_0+G_1(z)+G_2(z)$

Multiplying (3.6) by z^n and summing by column gives:

$$\begin{aligned}\sigma\left(2\mu\sigma P_{1,1}+2\mu(1-\sigma)P_{1,2}\right)=G_1(z)\left[\lambda B(z)+\frac{2\mu\sigma^2}{z}-(\lambda+2\mu\sigma)\right]\\ +G_2(z)\left[\frac{2\mu\sigma(1-\sigma)}{z}\right]\\ +\lambda\sigma B(z)P_0\end{aligned} \tag{3.8}$$

Substitution of (3.5) into (3.8) and rearranging gives:

$$\sigma\lambda P_0\left(1-B(z)\right)=G_1(z)\left[\lambda B(z)+\frac{2\mu\sigma^2}{z}-(\lambda+2\mu\sigma)\right]+G_2(z)\left[\frac{2\mu\sigma(1-\sigma)}{z}\right] \tag{3.9}$$

Multiplying (3.7) by z^n and summing by column then using (3.5) once more gives:

$$\begin{aligned}(1-\sigma)\lambda P_0\left(1-B(z)\right)=G_2(z)\left[\lambda B(z)+\frac{2\mu(1-\sigma)^2}{z}-\left(\lambda+2\mu(1-\sigma)\right)\right]\\ +G_1(z)\left[\frac{2\mu\sigma(1-\sigma)}{z}\right]\end{aligned} \tag{3.10}$$

This gives two simultaneous equations involving $G_1(z)$ and $G_2(z)$. Solving these gives:

$$G_1(z)=\frac{-\sigma\lambda P_0 z\left(-2\mu(1-\sigma)\left(B(z)-1\right)+\lambda\left(B(z)-1\right)^2\right)}{Q(z)}$$

$$G_2(z)=\frac{\lambda P_0 z\left(2\mu\sigma(1-\sigma)\left(B(z)-1\right)-\lambda(1-\sigma)\left(B(z)-1\right)^2\right)}{Q(z)}$$

Where

$$Q(z)=-4\mu\sigma(1-\sigma)\big(\mu(1-z)+\lambda(B(z)-1)\big)+\lambda(B(z)-1)\big(\lambda z(B(z)-1)+2\mu(1-z)\big)$$

Note, $Q(1)=0$.

$G(z)=P_0+G_1(z)+G_2(z)$:

$$G(z)=\frac{\lambda z(B(z)-1)\big(4\mu\sigma(1-\sigma)-\lambda(B(z)-1)\big)P_0}{Q(z)}+P_0 \qquad (3.11)$$

To evaluate the summary measures for this queueing system, it is required to differentiate $G(z)$

$$Q(z)G(z)=\lambda z(B(z)-1)\big(4\mu\sigma(1-\sigma)-\lambda(B(z)-1)\big)P_0+Q(z)P_0 \qquad (3.12)$$

Differentiating requires the following component parts to be evaluated:

$$Q'(z)G(z)+Q(z)G'(z)=\big(\lambda z(B(z)-1)\big(4\mu\sigma(1-\sigma)-\lambda(B(z)-1)\big)P_0+Q(z)P_0\big)'$$

The derivative of $Q(z)$ is:

$$\begin{aligned}Q'(z)&=4\mu\sigma(1-\sigma)(\mu-\lambda B'(z))\\&\quad+\lambda B'(z)\big(\lambda z(B(z)-1)+2\mu(1-z)\big)\\&\quad+\lambda(B(z)-1)\big(\lambda(B(z)-1)+\lambda zB'(z)-2\mu\big)\end{aligned}$$

Setting $z=1$ gives:

$$Q'(1)=4\mu\sigma(1-\sigma)\big(\mu-\lambda\bar{b}\big)$$

The derivative of the right hand side is:

$$\left(\lambda z\left(B(z)-1\right)\left(4\mu\sigma(1-\sigma)-\lambda\left(B(z)-1\right)\right)P_0+Q(z)P_0\right)'=$$
$$\lambda\left(B(z)-1\right)\left(4\mu\sigma(1-\sigma)-\lambda\left(B(z)-1\right)\right)P_0$$
$$+\lambda zB'(z)\left(4\mu\sigma(1-\sigma)-\lambda\left(B(z)-1\right)\right)P_0$$
$$-\lambda^2 z\left(B(z)-1\right)B'(z)P_0$$

Setting $z=1$ gives:

$$4\lambda\bar{b}\mu\sigma(1-\sigma)P_0$$

Substituting for the derivatives,

$$4\mu\sigma(1-\sigma)(\mu-\lambda b)G(1)=4\lambda\bar{b}\mu\sigma(1-\sigma)P_0+4\mu\sigma(1-\sigma)(\mu-\lambda b)P_0$$
$$(\mu-\lambda b)=\lambda\bar{b}P_0+(\mu-\lambda b)P_0$$
$$P_0=1-\frac{\lambda\bar{b}}{\mu}$$

Therefore $P_0=1-\rho$

To find $G'(z)$, differentiate (3.12) again:

$$Q''(z)G(z)+2Q'(z)G'(z)+Q(z)G''(z)=\left(\lambda z\left(B(z)-1\right)\left(4\mu\sigma(1-\sigma)-\lambda\left(B(z)-1\right)\right)P_0+Q(z)P_0\right)''$$

The term $Q(z)G''(z)$ will vanish as $Q(1)=0$. Rearrangement of the above gives:

$$G'(z)=\frac{\left(\lambda z\left(B(z)-1\right)\left(4\mu\sigma(1-\sigma)-\lambda\left(B(z)-1\right)\right)P_0+Q(z)P_0\right)''-Q''(z)G(z)}{2Q'(z)}$$

The second derivative of Q(z) is:

$$
\begin{aligned}
Q''(z) = & -4\lambda\mu\sigma(1-\sigma)B''(z) \\
& +\lambda B''(z)\left(\lambda z\left(B(z)-1\right)+2\mu(1-z)\right) \\
& +2\lambda B'(z)\left(\lambda\left(B(z)-1\right)+\lambda zB'(z)-2\mu\right) \\
& +\lambda\left(B(z)-1\right)\left(2\lambda B'(z)+\lambda zB''(z)\right)
\end{aligned}
$$

Setting $z = 1$ gives:

$$
Q''(1) = -4\lambda\mu\sigma(1-\sigma)B''(1)+2\lambda\bar{b}\left(\lambda\bar{b}-2\mu\right)
$$

The second derivative of the right hand side is:

$$
\begin{aligned}
& \left(\lambda z\left(B(z)-1\right)\left(4\mu\sigma(1-\sigma)-\lambda\left(B(z)-1\right)\right)(1-\rho)\right)'' = \\
& \lambda(1-\rho)\left(4\mu\sigma(1-\sigma)-\lambda\left(B(z)-1\right)\right)\left(2B'(z)+zB''(z)\right) \\
& -\lambda^2\left(B(z)-1\right)(1-\rho)\left(2B'(z)+zB''(z)\right) \\
& -4\lambda\mu\sigma(1-\sigma)B''(z)(1-\rho) \\
& -2\lambda^2 z\left(B'(z)\right)^2(1-\rho)
\end{aligned}
$$

Setting $z = 1$ gives:

$$
4\mu\sigma(1-\sigma)(1-\rho)\left(2\lambda\bar{b}+\lambda B''(z)\right)
$$

Completing the expression for $G'(1)$ and simplifying gives:

$$
G'(1) = \frac{4\lambda\mu\sigma(1-\sigma)(1-\rho)\left(2\bar{b}+B''(1)\right)-\rho\left(2\lambda\bar{b}\left(\lambda\bar{b}-2\mu\right)-4\lambda\mu\sigma(1-\sigma)B''(1)\right)}{8\mu^2\sigma(1-\sigma)(1-\rho)}
$$

Further simplification yields:

$$
\begin{aligned}
G'(1) &= \rho+\frac{\lambda B''(1)}{2\mu(1-\rho)}+\frac{\rho^2}{4\sigma(1-\sigma)(1-\rho)} \\
&= L
\end{aligned}
\qquad (3.13)
$$

To check this result several steps will be taken; firstly, comparison with known theoretical results for the M/He/1 queue with single arrivals and secondly, comparison with a simulation model built in Simul8.

As was shown in the previous section, if batches are of a constant size, the quantity $B''(1)$ is simply $\bar{b}(\bar{b}-1)$, where $\bar{b}$ is the mean batch size.

Therefore (3.13) becomes:

$$\begin{aligned} G'(1) &= \rho + \frac{\lambda\bar{b}(\bar{b}-1)}{2\mu(1-\rho)} + \frac{\rho^2}{4\sigma(1-\sigma)(1-\rho)} \\ &= \rho + \frac{\rho(\bar{b}-1)}{2(1-\rho)} + \frac{\rho^2}{4\sigma(1-\sigma)(1-\rho)} \\ &= L \end{aligned} \tag{3.14}$$

Letting $\bar{b}=1$, $G'(1)$ becomes:

$$\begin{aligned} G'(1) &= \rho + \frac{\rho^2}{4\sigma(1-\sigma)(1-\rho)} \\ &= L \end{aligned}$$

which is the same as the result for M/He/1.

As an aside, it is always interesting to see which value of σ will produce the minimum value of L. Differentiation of (3.13) gives:

$$L'(\sigma) = \frac{\rho^2}{4(1-\rho)}\left(-\frac{1}{\sigma^2} + \frac{1}{(1-\sigma)^2}\right)$$

Equating to zero gives $\sigma = 0.5$, which is of course the Negative Exponential distribution.

Differentiating $L'(\sigma)$

$$L''(\sigma)=\frac{\rho^2}{4(1-\rho)}\left(\frac{2}{\sigma^3}+\frac{2}{(1-\sigma)^3}\right)$$

Since $0<\sigma\leq 0.5$ this quantity is always positive, confirming that L is a minimum at $\sigma=0.5$.

Hence

$L_{\min}=\frac{\rho}{1-\rho}$ as for M/M/1.

3.3.1. Batch size distribution

Batches of fixed size

A simulation model was built in Simul8 to confirm these equations. The model was run for various values of ρ, σ and batch size. The results below hold for the constant batch case, where batch size = 2:

Table 3.3: Summary measures from simulation package Simul8 for batch arrivals where batches are of Constant size 2 and service is Hyperexponential

	$\sigma=0.1$				$\sigma=0.2$				$\sigma=0.3$				$\sigma=0.4$			
ρ	*Lq*	W_q	*L*	W	*Lq*	W_q	*L*	W	*Lq*	W_q	*L*	W	*Lq*	W_q	*L*	W
0.1	0.09	0.04	0.18	0.09	0.07	0.04	0.18	0.09	0.07	0.03	0.16	0.08	0.07	0.03	0.16	0.08
0.2	0.27	0.13	0.46	0.23	0.20	0.10	0.4	0.20	0.18	0.09	0.38	0.19	0.18	0.09	0.38	0.19
0.3	0.58	0.29	0.88	0.44	0.42	0.21	0.72	0.36	0.37	0.18	0.66	0.33	0.35	0.17	0.64	0.32
0.4	1.07	0.54	1.48	0.74	0.75	0.38	1.16	0.58	0.65	0.33	1.06	0.53	0.61	0.31	1.02	0.51
0.5	1.89	0.94	2.38	1.19	1.28	0.64	1.78	0.89	1.10	0.55	1.6	0.80	1.02	0.51	1.52	0.76
0.6	3.25	1.62	3.86	1.93	2.17	1.08	2.76	1.38	1.83	0.92	2.44	1.22	1.70	0.85	2.3	1.15
0.7	5.71	2.85	6.4	3.20	3.75	1.87	4.44	2.22	3.17	1.58	3.86	1.93	2.90	1.45	3.6	1.80
0.8	10.94	5.46	11.74	5.87	7.07	3.53	7.88	3.94	5.96	2.98	6.76	3.38	5.46	2.73	6.26	3.13
0.9	27.07	13.52	27.94	13.97	17.00	8.49	17.88	8.94	14.33	7.13	15.22	7.61	13.13	6.56	14.02	7.01

Table 3.4: Theoretical solutions for L, where batches are of Constant size 2 and service is Hyperexponential

$G'(1)=L$					
ρ	$\sigma=0.1$	$\sigma=0.2$	$\sigma=0.3$	$\sigma=0.4$	$\sigma=0.5$
0.1	0.19	0.17	0.17	0.17	0.17
0.2	0.46	0.40	0.38	0.38	0.38
0.3	0.87	0.72	0.67	0.65	0.64
0.4	1.47	1.15	1.05	1.01	1.00
0.5	2.39	1.78	1.60	1.52	1.50
0.6	3.85	2.76	2.42	2.29	2.25
0.7	6.40	4.42	3.81	3.57	3.50
0.8	11.69	7.80	6.61	6.13	6.00
0.9	27.90	18.06	15.04	13.84	13.50

It is worth noting here that Little's result will hold for the above queueing system. When considering a queueing system whereby customers arrive in batches, it is clear that the overall arrival rate is equal to the arrival rate of batches multiplied by the mean number in a batch, or algebraically $\lambda\bar{b}$.

Therefore, for a queueing system where arrivals occur in batches, Little's result is:

$$L = \lambda\bar{b}W$$

The theoretical solutions for L, or $G'(1)$ are found in Table 3.4.

When $\lambda = 1$, and the batch size is 2, it is clear that Little's result holds. It is worth noting that other results from queueing theory hold for this queueing system.

For example, when comparing the mean wait in the queue, W_q, to the mean wait in the system, W, it is clear that the difference is simply $\frac{1}{\mu}$. This is to be

expected. Also, when comparing the mean number in the queue, *Lq*, to the mean number in the system, *L*, the difference is simply ρ.

Thus the following expressions hold also:

$$W = W_q + \frac{1}{\mu}$$

$$L = L_q + \rho$$

It is imperative that we ensure that *L* remains stable for larger batches. At a first glance, the Simul8 model showed that in fact as the batch size increased, the summary measures obtained from the model differed from the theoretical values. Running the Simul8 model for a considerable number of runs takes much time. Thus a simulation model was built in VBA for Excel which ran very much faster.

The following table includes the simulated summary measures for a batch size of 10 and λ equal to 2:

Table 3.5: Summary measures from VBA simulation for batch arrivals where batches are of Constant size 10 and service is Hyperexponential

	$\sigma = 0.1$				$\sigma = 0.2$				$\sigma = 0.3$				$\sigma = 0.4$			
ρ	*Lq*	W_q	W	*L*	*Lq*	W_q	W	*L*	*Lq*	W_q	W	*L*	*Lq*	W_q	W	*L*
0.1	0.53	0.03	0.03	0.63	0.52	0.03	0.03	0.62	0.51	0.03	0.03	0.62	0.51	0.03	0.03	0.61
0.2	1.26	0.06	0.07	1.46	1.20	0.06	0.07	1.40	1.18	0.06	0.07	1.40	1.18	0.06	0.07	1.38
0.3	2.28	0.11	0.13	2.58	2.13	0.11	0.12	2.43	2.08	0.10	0.12	2.43	2.08	0.10	0.12	2.36
0.4	3.74	0.19	0.21	4.14	3.42	0.17	0.19	3.82	3.32	0.17	0.19	3.82	3.32	0.16	0.18	3.68
0.5	5.89	0.29	0.32	6.39	5.28	0.26	0.29	5.78	5.09	0.25	0.28	5.78	5.09	0.25	0.28	5.52
0.6	9.23	0.46	0.49	9.83	8.15	0.41	0.44	8.75	7.82	0.39	0.42	8.75	7.82	0.38	0.41	8.27
0.7	15.00	0.75	0.78	15.70	13.06	0.65	0.69	13.76	12.43	0.62	0.66	13.76	12.43	0.61	0.64	12.88
0.8	26.81	1.34	1.38	27.61	22.95	1.15	1.19	23.75	21.70	1.08	1.12	23.75	21.70	1.07	1.11	22.16
0.9	62.78	3.14	3.18	63.68	52.45	2.62	2.67	53.35	50.02	2.50	2.55	53.35	50.02	2.44	2.49	49.77

Comparison of Table 3.5 with the theoretical values for *L* (Table 3.6), yields a good result. Therefore it is clear that the expression derived for *L* is satisfied for large and small batch sizes.

Table 3.6: Theoretical solutions for *L*, where batches are of Constant size 10 and service is Hyperexponential

$G'(1)=L$					
ρ	$\sigma=0.1$	$\sigma=0.2$	$\sigma=0.3$	$\sigma=0.4$	$\sigma=0.5$
0.1	0.63	0.62	0.61	0.61	0.61
0.2	1.46	1.40	1.38	1.38	1.38
0.3	2.59	2.43	2.38	2.36	2.36
0.4	4.14	3.82	3.72	3.68	3.67
0.5	6.39	5.78	5.60	5.52	5.50
0.6	9.85	8.76	8.42	8.29	8.25
0.7	15.74	13.75	13.14	12.90	12.83
0.8	27.69	23.80	22.61	22.13	22.00
0.9	63.90	54.06	51.04	49.84	49.50

Before moving on to consider batches of variable size, it is worth examining what happens to the summary measures as batch size increases (keeping all other values constant). The following set of graphs demonstrate this for different values of ρ and σ:

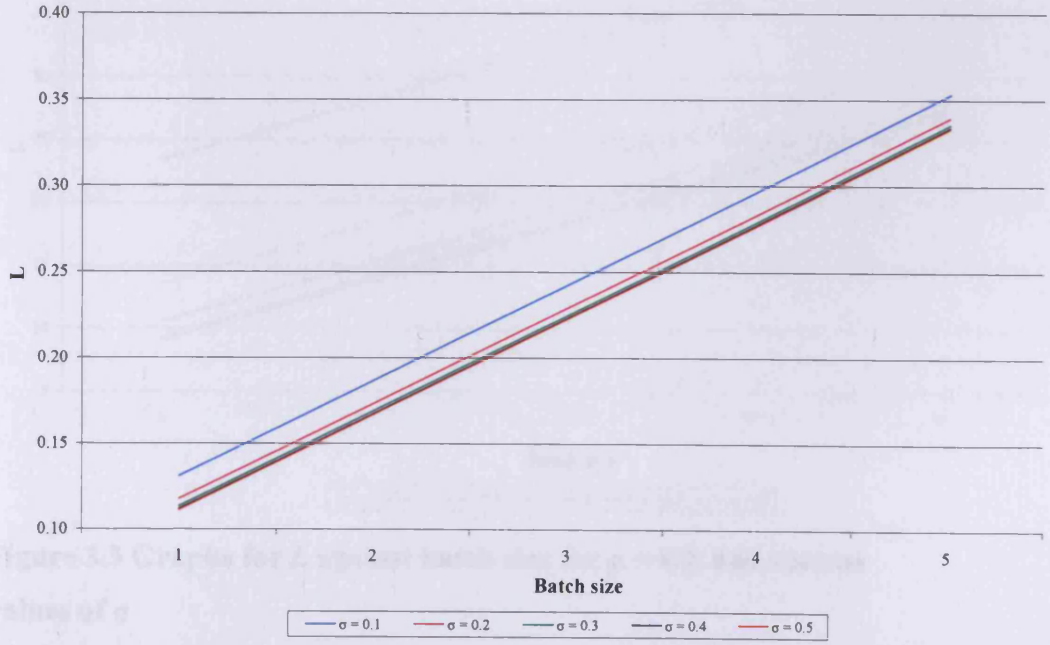


Figure 3.1: Graphs for *L* against batch size for ρ = 0.1, and various values of σ

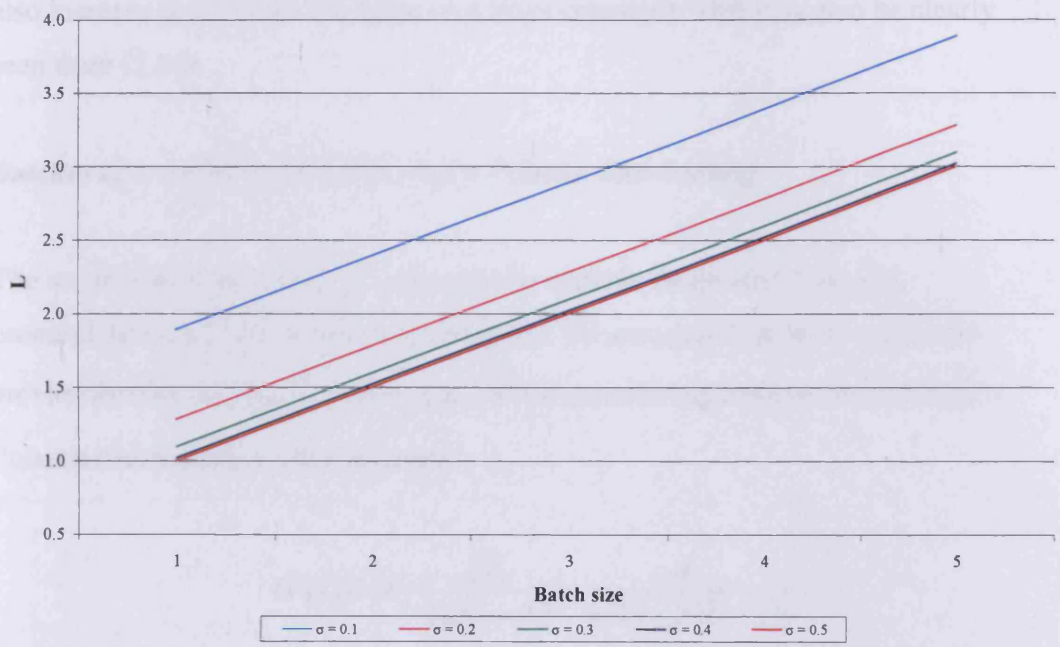


Figure 3.2: Graphs for *L* against batch size for ρ = 0.5, and various values of σ

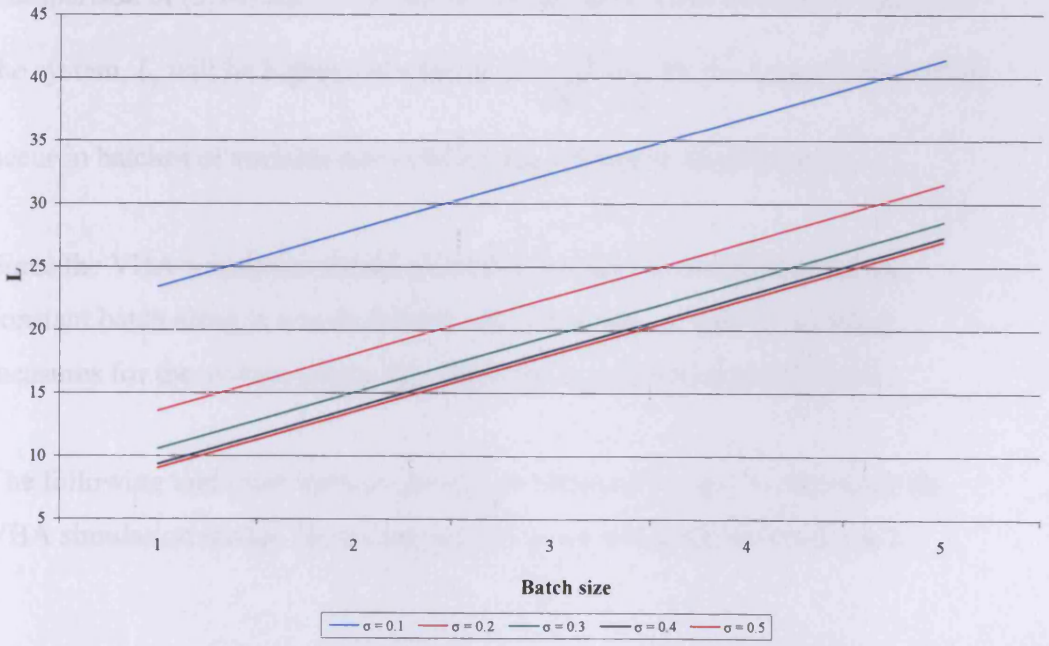


Figure 3.3 Graphs for *L* against batch size for ρ = 0.9, and various values of σ

From the above graphs it is clear that as batch size increases, the value of L will also increase (even when the value of ρ stays constant). This may also be clearly seen from (3.14).

Batches of a variable size (following a Poisson distribution)

The model was then altered to encompass a variable batch size. The most probable batch size distribution would be the Poisson distribution. It was shown previously that $B''(1) = \bar{b}^2$ in this case. When considering batches which follow a Poisson distribution, (3.13) becomes:

$$\begin{aligned} G'(1) &= \rho + \frac{\lambda \bar{b}^2}{2\mu(1-\rho)} + \frac{\rho^2}{4\sigma(1-\sigma)(1-\rho)} \\ &= \rho + \frac{\rho \bar{b}}{2(1-\rho)} + \frac{\rho^2}{4\sigma(1-\sigma)(1-\rho)} \\ &= L \end{aligned} \tag{3.15}$$

Comparison of (3.14) and (3.15) show that the mean time a customer spends in the system, L, will be higher, by a factor of $\frac{\rho}{2(1-\rho)}$ for the case where arrivals occur in batches of variable sizes (following a Poisson distribution).

Since the VBA simulation model proved to be more time efficient for the constant batch sizes, it was decided to use this model to gather summary measures for the system where arrivals occur in batches of variable size.

The following table summarises the results obtained for this system from the VBA simulation model. Note here that the mean batch size is 2 and λ is 1:

Table 3.7: Summary measures from VBA simulation model for batch arrivals where batch sizes follow a Poisson distribution with mean 2 and service is Hyperexponential

	$\sigma = 0.1$				$\sigma = 0.2$				$\sigma = 0.3$				$\sigma = 0.4$			
ρ	Lq	W_q	W	L	Lq	W_q	W	L	Lq	W_q	W	L	Lq	W_q	W	L
0.1	0.14	0.07	0.12	0.24	0.14	0.07	0.12	0.24	0.12	0.06	0.11	0.22	0.12	0.06	0.11	0.22
0.2	0.39	0.20	0.30	0.60	0.34	0.17	0.27	0.54	0.31	0.16	0.26	0.52	0.3	0.15	0.25	0.50
0.3	0.79	0.39	0.55	1.10	0.64	0.32	0.47	0.94	0.58	0.29	0.44	0.88	0.56	0.28	0.43	0.86
0.4	1.41	0.71	0.91	1.82	1.08	0.54	0.74	1.48	0.98	0.49	0.69	1.38	0.94	0.47	0.67	1.34
0.5	2.39	1.19	1.44	2.88	1.78	0.89	1.14	2.28	1.6	0.8	1.05	2.10	1.52	0.76	1.01	2.02
0.6	3.99	1.99	2.29	4.58	2.92	1.46	1.76	3.52	2.57	1.29	1.59	3.18	2.44	1.22	1.52	3.04
0.7	6.86	3.43	3.78	7.56	4.90	2.45	2.80	5.60	4.28	2.14	2.49	4.98	4.04	2.02	2.37	4.73
0.8	12.85	6.42	6.82	13.64	9.00	4.50	4.90	9.80	7.81	3.91	4.31	8.62	7.34	3.67	4.07	8.13
0.9	31.12	15.55	16.15	32.30	21.66	10.83	11.28	22.56	18.64	9.32	9.77	19.54	17.44	8.72	9.17	18.34

Values of L and Lq were calculated using Little's result. Comparison of the above table with the theoretical values in the table below provides confirmation that the derived expression for variable batch sizes (following a Poisson distribution), as well as the constant batches shown previously, is correct.

Table 3.8: Theoretical solutions for L, where batch sizes follow a Poisson distribution with mean 2 and service is Hyperexponential

$G'(1) = L$					
ρ	$\sigma = 0.1$	$\sigma = 0.2$	$\sigma = 0.3$	$\sigma = 0.4$	$\sigma = 0.5$
0.1	0.24	0.23	0.22	0.22	0.22
0.2	0.59	0.53	0.51	0.50	0.50
0.3	1.09	0.93	0.88	0.86	0.86
0.4	1.81	1.48	1.38	1.34	1.33
0.5	2.89	2.28	2.10	2.02	2.00
0.6	4.60	3.51	3.17	3.04	3.00
0.7	7.57	5.59	4.98	4.73	4.67
0.8	13.69	9.80	8.61	8.13	8.00
0.9	32.40	22.56	19.54	18.34	18.00

Next consideration will be given to larger batch sizes. Table 3.9 presents the simulated summary measures for a batch size of 10 and λ equal to 2:

Table 3.9: Summary measures from VBA simulation model for batch arrivals where batch sizes follow a Poisson distribution with mean 10 and service is Hyperexponential

	$\sigma=0.1$				$\sigma=0.2$				$\sigma=0.3$				$\sigma=0.4$			
ρ	*Lq*	W_q	W	*L*	*Lq*	W_q	W	*L*	*Lq*	W_q	W	*L*	*Lq*	W_q	W	*L*
0.1	0.58	0.06	0.07	0.70	0.56	0.06	0.07	0.68	0.56	0.06	0.07	0.68	0.56	0.06	0.07	0.68
0.2	1.38	0.14	0.16	1.62	1.31	0.13	0.15	1.52	1.29	0.13	0.15	1.49	1.28	0.13	0.15	1.50
0.3	2.49	0.25	0.28	2.81	2.34	0.23	0.26	2.64	2.29	0.23	0.26	2.61	2.27	0.23	0.26	2.58
0.4	4.08	0.41	0.45	4.48	3.77	0.38	0.42	4.17	3.66	0.37	0.41	4.08	3.62	0.36	0.40	4.02
0.5	6.40	0.64	0.69	6.90	5.81	0.58	0.63	6.31	5.62	0.56	0.61	5.13	5.55	0.55	0.60	6.04
0.6	10.00	1.00	1.06	10.60	8.93	0.89	0.95	9.53	8.62	0.86	0.92	9.16	8.47	0.84	0.90	9.07
0.7	16.23	1.62	1.69	16.93	14.26	1.42	1.49	14.95	13.67	1.36	1.43	14.34	13.39	1.34	1.41	14.09
0.8	28.94	2.89	2.97	29.70	25.00	2.50	2.58	25.78	23.84	2.38	2.46	24.63	23.32	2.33	2.41	24.12
0.9	65.80	6.57	6.66	66.66	57.32	5.73	5.82	58.23	53.92	5.39	5.48	54.83	52.58	5.25	5.34	53.48

Comparison of the results in Table 3.9 with the theoretical values for L, shown in Table 3.10, gives good agreement. Thus we have evidence that the expression derived for L, (3.15), is satisfied for large and small batch sizes.

Table 3.10: Theoretical solutions for L, where batch sizes follow a Poisson distribution with mean 10 and service is Hyperexponential

$G'(1)=L$					
ρ	$\sigma=0.1$	$\sigma=0.2$	$\sigma=0.3$	$\sigma=0.4$	$\sigma=0.5$
0.1	0.69	0.67	0.67	0.67	0.67
0.2	1.59	1.53	1.51	1.50	1.50
0.3	2.80	2.64	2.60	2.58	2.57
0.4	4.47	4.15	4.05	4.01	4.00
0.5	6.89	6.28	6.10	6.02	6.00
0.6	10.60	9.51	9.17	9.04	9.00
0.7	16.90	14.92	14.31	14.07	14.00
0.8	29.69	25.80	24.61	24.13	24.00
0.9	68.40	58.56	55.54	54.34	54.00

The following set of graphs demonstrate what happens to the value of L as batch size increases.

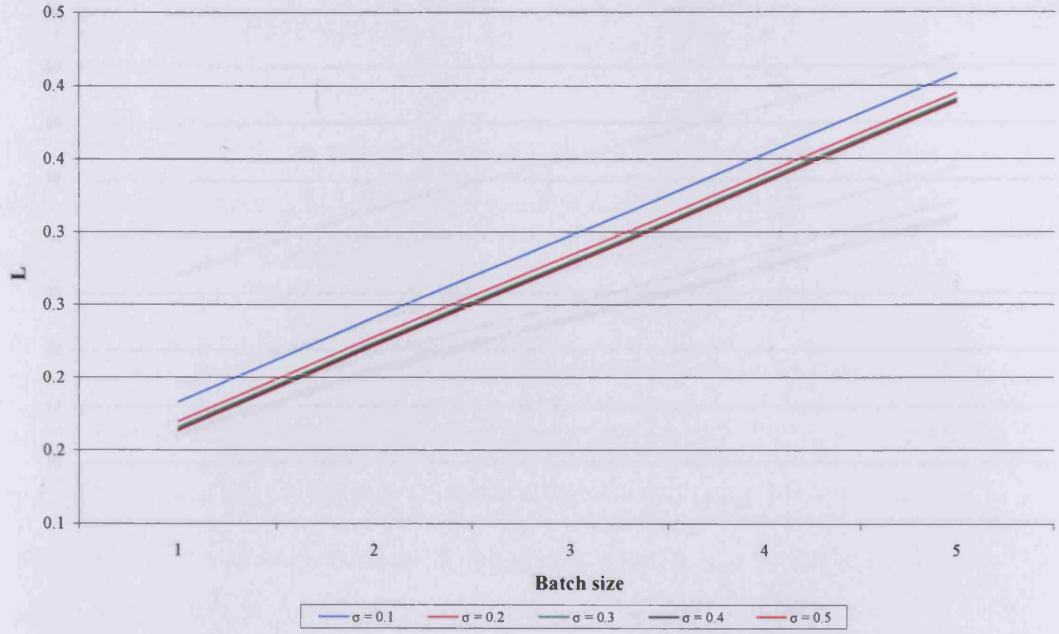


Figure 3.4: Graphs for L against batch size for $\rho = 0.1$, and various values of σ

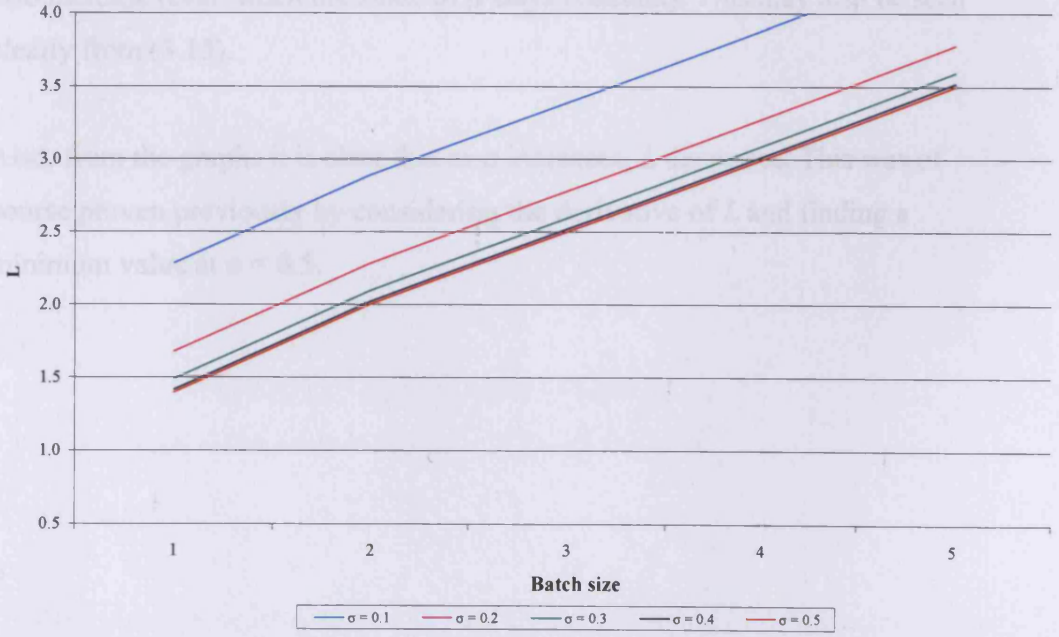


Figure 3.5: Graphs for L against batch size for $\rho = 0.5$, and various values of σ

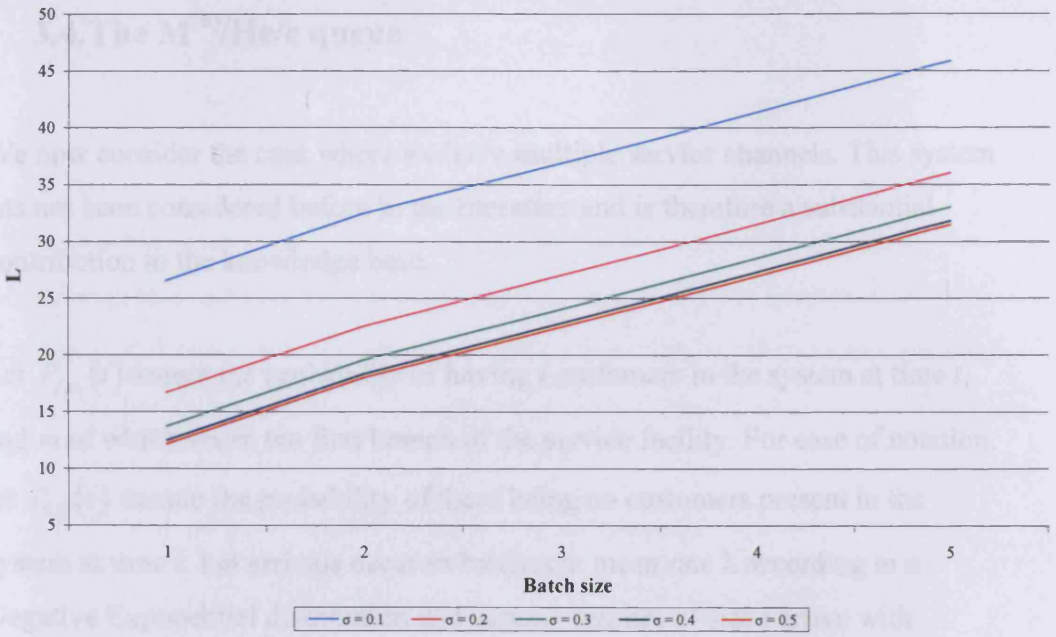


Figure 3.6: Graphs for L against batch size for $\rho = 0.9$, and various values of σ

From the above graphs it is clear that as batch size increases, the value of L will also increase (even when the value of ρ stays constant). This may also be seen clearly from (3.15).

Also, from the graphs it is clear that as σ increases, L decreases. This was of course proven previously by considering the derivative of L and finding a minimum value at $\sigma = 0.5$.

3.4.The $M^{(b)}$/He/c queue

We now consider the case where we have multiple service channels. This system has not been considered before in the literature and is therefore a substantial contribution to the knowledge base.

Let $P_{j,m}(t)$ denote the probability of having j customers in the system at time t, and m of which are in the first branch of the service facility. For ease of notation, let $P_{0,0}(t)$ denote the probability of there being no customers present in the system at time t. Let arrivals occur in batches at mean rate λ according to a Negative Exponential distribution and assume batches of size i arrive with probability $P(\text{Batch size} = i) = b_i \quad i = 0, 1, \ldots.$

When there are fewer than c customers present in the system, the arriving customers enter any of the vacant service channels in random fashion. When there are c service channels occupied, arriving customers form a single file queue.

Firstly, the situation where there are no customers present in the system at time $t+\delta t$ will be considered:

$$\begin{aligned} P_{0,0}(t+\delta t) = {} & P_{0,0}(t)\left[1-\lambda\delta t+o(\delta t)\right] \\ & +P_{1,1}(t)\left[1-\lambda\delta t+o(\delta t)\right]\left[2\mu\sigma\delta t+o(\delta t)\right] \\ & +P_{1,0}(t)\left[1-\lambda\delta t+o(\delta t)\right]\left[2\mu(1-\sigma)\delta t+o(\delta t)\right] \\ & +o(\delta t) \end{aligned}$$

After rearranging and taking limits, the following differential difference equation is obtained:

$$\frac{dP_0(t)}{dt} = -\lambda P_0(t) + 2\mu\sigma P_{1,1}(t) + 2\mu(1-\sigma)P_{1,0}(t)$$

Equating to zero gives the following steady state equation:

$$\lambda P_0 = 2\mu\sigma P_{1,1} + 2\mu(1-\sigma)P_{1,0} \tag{3.16}$$

The following set of three equations, (3.17) to (3.19) describe the situation whereby an arriving customer find at least one service facility available. Equation (3.17) describes the situation where all customers are served in the first branch of the service facility. This equation will hold for $1 \le j \le c-1$

$$\begin{aligned} P_{j,j}(t+\delta t) = {} & P_{j,j}(t)\left[1-\lambda\delta t+o(\delta t)\right]\left[1-(j)2\mu\sigma\delta t+o(\delta t)\right] \\ & +\sum_{i=1}^{j} P_{j-i,j-i}(t)\left[\lambda b_i\delta t+o(\delta t)\right]\left[1-(j-i)2\mu\sigma\delta t+o(\delta t)\right]\sigma^i \\ & +P_{j+1,j+1}(t)\left[1-\lambda\delta t+o(\delta t)\right]\left[(j+1)2\mu\sigma\delta t+o(\delta t)\right] \\ & +P_{j+1,j}(t)\left[1-\lambda\delta t+o(\delta t)\right]\left[2\mu(1-\sigma)\delta t+o(\delta t)\right] \\ & +o(\delta t) \end{aligned}$$

This becomes after some algebraic manipulation:

$$\frac{dP_{j,j}(t)}{dt} = -\left(\lambda+(j)2\mu\sigma\right)P_{j,j}(t)+\lambda\sum_{i=1}^{j}\sigma^i b_i P_{j-i,j-i}(t)+(j+1)2\mu\sigma P_{j+1,j+1}(t)+2\mu(1-\sigma)P_{j+1,j}(t)$$

The set of steady state equations are thus:

$$\left(\lambda+(j)2\mu\sigma\right)P_{j,j} = \lambda\sum_{i=1}^{j}\sigma^i b_i P_{j-i,j-i}+(j+1)2\mu\sigma P_{j+1,j+1}+2\mu(1-\sigma)P_{j+1,j} \tag{3.17}$$

If $(1 \le j \le c-1)$

Equation (3.18) describes the situation where all customers are served in the second branch of the service facility. This equation will hold for $1 \le j \le c-1$

$$P_{j,0}(t+\delta t) = P_{j,0}(t)\left[1-\lambda\delta t+o(\delta t)\right]\left[1-(j)2\mu(1-\sigma)\delta t+o(\delta t)\right]$$
$$+\sum_{i=1}^{j} P_{j-i,0}(t)\left[\lambda b_i\delta t+o(\delta t)\right]\left[1-(j-i)2\mu(1-\sigma)\delta t+o(\delta t)\right](1-\sigma)^i$$
$$+P_{j+1,1}(t)\left[1-\lambda\delta t+o(\delta t)\right]\left[2\mu\sigma\delta t+o(\delta t)\right]$$
$$+P_{j+1,0}(t)\left[1-\lambda\delta t+o(\delta t)\right]\left[(j+1)2\mu(1-\sigma)\delta t+o(\delta t)\right]$$
$$+o(\delta t)$$

This becomes after some algebraic manipulation:

$$\frac{dP_{j,0}(t)}{dt} = -\left(\lambda+(j)2\mu(1-\sigma)\right)P_{j,0}(t)+\lambda\sum_{i=1}^{j}(1-\sigma)^i b_i P_{j-i,0}(t)+2\mu\sigma P_{j+1,1}(t)$$
$$+(j+1)2\mu(1-\sigma)P_{j+1,0}(t)$$

The set of steady state equations are thus

$$\left(\lambda+(j)2\mu(1-\sigma)\right)P_{j,0} = \lambda\sum_{i=1}^{j}(1-\sigma)^i b_i P_{j-i,0}+2\mu\sigma P_{j+1,1}+(j+1)2\mu(1-\sigma)P_{j+1,0}$$
(3.18)

If $(1\le j\le c-1)$

Equation (3.19) describes the situation where *m* customers are served in the first branch of the service facility and (*j* – *m*) customers are served in the second branch of the service facility. This equation will hold for
$1\le j\le c-1$ and $1\le m\le j-1$

$$
\begin{aligned}
P_{j,m}(t+\delta t) = {} & P_{j,m}(t)\left[1-\lambda\delta t+o(\delta t)\right]\left[1-(m)2\mu\sigma\delta t-(j-m)2\mu(1-\sigma)\delta t+o(\delta t)\right] \\
& +P_{j+1,m}(t)\left[1-\lambda\delta t+o(\delta t)\right]\left[(j+1-m)2\mu(1-\sigma)\delta t+o(\delta t)\right]\left[1-m2\mu\sigma\delta t+o(\delta t)\right] \\
& +P_{j+1,m+1}(t)\left[1-\lambda\delta t+o(\delta t)\right]\left[(m+1)2\mu\sigma\delta t+o(\delta t)\right]\left[1-(j-m-1)2\mu(1-\sigma)\delta t+o(\delta t)\right] \\
& +\sum_{i=1}^{j-m} P_{j-i,m}(t)\left[\lambda b_i\delta t+o(\delta t)\right]\left[1-(m)2\mu\sigma-(j-i-m)2\mu(1-\sigma)\delta t+o(\delta t)\right](1-\sigma)^i \\
& +\sum_{i=1}^{m} P_{j-i,m-i}(t)\left[\lambda b_i\delta t+o(\delta t)\right]\left[1-(m-i)2\mu\sigma-(j-m)2\mu(1-\sigma)\delta t+o(\delta t)\right]\sigma^i \\
& +\sum_{n=1}^{m}\sum_{i=1}^{j-m} P_{j-n-i,m-n}(t)\left[\lambda b_{i+n}\delta t+o(\delta t)\right]\left[1-(m-n)2\mu\sigma-(j-i-m)2\mu(1-\sigma)\delta t+o(\delta t)\right]\sigma^n(1-\sigma)^i \\
& +o(\delta t)
\end{aligned}
$$

The differential-difference equation for $P_{j,m}(t)$ is:

$$
\begin{aligned}
\frac{dP_{j,m}(t)}{dt} = {} & -\left(\lambda+(j-m)2\mu(1-\sigma)+(m)2\mu\sigma\right)P_{j,m}(t)+\lambda\sum_{i=1}^{j-m}(1-\sigma)^i b_i P_{j-i,m}(t) \\
& +\lambda\sum_{i=1}^{m}\sigma^i b_i P_{j-i,m-i}(t)+\lambda\sum_{n=1}^{m}\sum_{i=1}^{j-m}(1-\sigma)^i\sigma^n b_{i+n}P_{j-n-i,m-n}(t) \\
& +(j+1-m)2\mu(1-\sigma)P_{j+1,m}(t)+(m+1)2\mu\sigma P_{j+1,m+1}(t)
\end{aligned}
$$

The steady state equations are therefore:

$$
\begin{aligned}
\left(\lambda+(j-m)2\mu(1-\sigma)+(m)2\mu\sigma\right)P_{j,m} = {} & \lambda\sum_{i=1}^{j-m}(1-\sigma)^i b_i P_{j-i,m} \\
& +\lambda\sum_{i=1}^{m}\sigma^i b_i P_{j-i,m-i} \\
& +\lambda\sum_{n=1}^{m}\sum_{i=1}^{j-m}(1-\sigma)^i\sigma^n b_{i+n}P_{j-n-i,m-n} \quad (3.19) \\
& +(j+1-m)2\mu(1-\sigma)P_{j+1,m} \\
& +(m+1)2\mu\sigma P_{j+1,m+1}
\end{aligned}
$$

If $(1\le j\le c-1)$ and $(1\le m\le j-1)$

The following set of three equations, (3.20) to (3.22) describe the situation whereby an arriving customer finds all service facilities occupied, but no queue.

Equation (3.20) describes the situation where all customers are served in the first branch of the service facility:

$$
\begin{aligned}
P_{c,c}(t+\delta t) &= P_{c,c}(t)\left[1-\lambda\delta t+o(\delta t)\right]\left[1-(c)2\mu\sigma\delta t+o(\delta t)\right] \\
&+\sum_{i=1}^{c} P_{c-i,c-i}(t)\left[\lambda b_i\delta t+o(\delta t)\right]\left[1-(c-i)2\mu\sigma\delta t+o(\delta t)\right]\sigma^i \\
&+P_{c+1,c}(t)\left[1-\lambda\delta t+o(\delta t)\right]\left[(c)2\mu\sigma\delta t+o(\delta t)\right]\sigma \\
&+P_{c+1,c-1}(t)\left[1-\lambda\delta t+o(\delta t)\right]\left[2\mu(1-\sigma)\delta t+o(\delta t)\right]\left[1-(c)2\mu\sigma\delta t+o(\delta t)\right]\sigma \\
&+o(\delta t)
\end{aligned}
$$

This becomes after some algebraic manipulation:

$$
\frac{dP_{c,c}(t)}{dt} = -\left(\lambda+(c)2\mu\sigma\right)P_{c,c}(t)+\lambda\sum_{i=1}^{c}\sigma^i b_i P_{c-i,c-i}(t)+(c)2\mu\sigma^2 P_{c+1,c}(t)+2\mu(1-\sigma)\sigma P_{c+1,c-1}(t)
$$

The steady state equations are therefore:

$$
\left(\lambda+(c)2\mu\sigma\right)P_{c,c} = \lambda\sum_{i=1}^{c}\sigma^i b_i P_{c-i,c-i}+(c)2\mu\sigma^2 P_{c+1,c}+2\mu(1-\sigma)\sigma P_{c+1,c-1} \quad (3.20)
$$

Equation (3.21) describes the situation where all customers are served in the second branch of the service facility:

$$
\begin{aligned}
P_{c,0}(t+\delta t) &= P_{c,0}(t)\left[1-\lambda\delta t+o(\delta t)\right]\left[1-(c)2\mu(1-\sigma)\delta t+o(\delta t)\right] \\
&+\sum_{i=1}^{c} P_{c-i,0}(t)\left[\lambda b_i\delta t+o(\delta t)\right]\left[1-(c-i)2\mu(1-\sigma)\delta t+o(\delta t)\right](1-\sigma)^i \\
&+P_{c+1,1}(t)\left[1-\lambda\delta t+o(\delta t)\right]\left[2\mu\sigma\delta t+o(\delta t)\right](1-\sigma) \\
&+P_{c+1,0}(t)\left[1-\lambda\delta t+o(\delta t)\right]\left[(c)2\mu(1-\sigma)\delta t+o(\delta t)\right](1-\sigma) \\
&+o(\delta t)
\end{aligned}
$$

This becomes after some algebraic manipulation:

$$
\begin{aligned}
\frac{dP_{c,0}(t)}{dt} &= -\left(\lambda+(c)2\mu(1-\sigma)\right)P_{c,0}(t)+\lambda\sum_{i=1}^{c}(1-\sigma)^i b_i P_{c-i,0}(t)+2\mu\sigma(1-\sigma)P_{c+1,1}(t) \\
&+(c)2\mu(1-\sigma)^2 P_{c+1,0}(t)
\end{aligned}
$$

The steady state equations are therefore:

$$\begin{aligned}\left(\lambda+(c)2\mu(1-\sigma)\right)P_{c,0} &= \lambda\sum_{i=1}^{c}(1-\sigma)^{i}\,b_i P_{c-i,0} \\ &\quad +2\mu\sigma(1-\sigma)P_{c+1,1} \\ &\quad +(c)2\mu(1-\sigma)^{2}P_{c+1,0}\end{aligned} \tag{3.21}$$

Equation (3.22) describes the situation where m customers are served in the first branch of the service facility and (c – m) customers are served in the second branch of the service facility. This will hold for $1 \le m \le c-1$

$$\begin{aligned}P_{c,m}(t+\delta t) &= P_{c,m}(t)\left[1-\lambda\delta t+o(\delta t)\right]\left[1-(m)2\mu\sigma\delta t-(c-m)2\mu(1-\sigma)\delta t+o(\delta t)\right] \\ &\quad +P_{c+1,m}(t)\left[1-\lambda\delta t+o(\delta t)\right]\left[(m)2\mu\sigma^{2}\delta t+(c-m)2\mu(1-\sigma)^{2}\delta t+o(\delta t)\right] \\ &\quad +P_{c+1,m+1}(t)\left[1-\lambda\delta t+o(\delta t)\right]\left[(m+1)2\mu\sigma\delta t+o(\delta t)\right](1-\sigma) \\ &\quad +P_{c+1,m-1}(t)\left[1-\lambda\delta t+o(\delta t)\right]\left[(c-m+1)2\mu(1-\sigma)\delta t+o(\delta t)\right]\sigma \\ &\quad +\sum_{i=1}^{c-m}P_{c-i,m}(t)\left[\lambda b_i\delta t+o(\delta t)\right]\left[1-(m)2\mu\sigma-(c-i-m)2\mu(1-\sigma)\delta t+o(\delta t)\right](1-\sigma)^{i} \\ &\quad +\sum_{i=1}^{m}P_{c-i,m-i}(t)\left[\lambda b_i\delta t+o(\delta t)\right]\left[1-(m-i)2\mu\sigma-(c-m)2\mu(1-\sigma)\delta t+o(\delta t)\right]\sigma^{i} \\ &\quad +\sum_{n=1}^{m}\sum_{i=1}^{c-m}P_{c-n-i,m-n}(t)\left[\lambda b_{i+n}\delta t+o(\delta t)\right]\left[1-(m-n)2\mu\sigma-(c-i-m)2\mu(1-\sigma)\delta t+o(\delta t)\right]\sigma^{n}(1-\sigma)^{i} \\ &\quad +o(\delta t)\end{aligned}$$

Extending this theory it is possible to see that the differential-difference equation for $P_{c,m}(t)$ is:

$$\begin{aligned}\frac{dP_{c,m}(t)}{dt} &= -\left(\lambda+(m)2\mu\sigma+(c-m)2\mu(1-\sigma)\right)P_{c,m}(t) \\ &\quad +(m)2\mu\sigma^{2}P_{c+1,m}(t)+(c-m)2\mu(1-\sigma)^{2}P_{c+1,m}(t) \\ &\quad +(m+1)2\mu\sigma(1-\sigma)P_{c+1,m+1}(t)+(c-m+1)2\mu\sigma(1-\sigma)P_{c+1,m-1}(t) \\ &\quad +\lambda\sum_{i=1}^{c-m}b_i(1-\sigma)^{i}P_{c-i,m}(t)+\lambda\sum_{i=1}^{m}b_i\sigma^{i}P_{c-i,m-i}(t) \\ &\quad +\lambda\sum_{n=1}^{m}\sum_{i=1}^{c-m}b_{i+n}\sigma^{n}(1-\sigma)^{i}P_{c-n-i,m-n}(t)\end{aligned}$$

The steady state equations are therefore:

$$\begin{aligned}(\lambda+(m)2\mu\sigma+(c-m)2\mu(1-\sigma))P_{c,m} &= (m)2\mu\sigma^2 P_{c+1,m} + \\ &(m+1)2\mu\sigma(1-\sigma)P_{c+1,m+1} \\ &+(c-m+1)2\mu\sigma(1-\sigma)P_{c+1,m-1} \\ &+\lambda\sum_{i=1}^{c-m} b_i(1-\sigma)^i P_{c-i,m} \\ &+\lambda\sum_{i=1}^{m} b_i\sigma^i P_{c-i,m-i} \\ &+\lambda\sum_{n=1}^{m}\sum_{i=1}^{c-m} b_{i+n}\sigma^n(1-\sigma)^i P_{c-n-i,m-n}\end{aligned} \qquad (3.22)$$

If $1 \le m \le c-1$

Finally, equations (3.23) to (3.25) describe the situation whereby an arriving customer finds all service facilities occupied and a queue in place.

Equation (3.23) describes the situation where all customers are served in the first branch of the service facility. This holds if $j \ge c+1$

$$\begin{aligned}P_{j,c}(t+\delta t) &= P_{j,c}(t)[1-\lambda\delta t+o(\delta t)][1-(c)2\mu\sigma\delta t+o(\delta t)] \\ &+\sum_{i=1}^{j-c} P_{j-i,c}(t)[\lambda b_i\delta t+o(\delta t)][1-(c)2\mu\sigma\delta t+o(\delta t)] \\ &+P_{j+1,c}(t)[1-\lambda\delta t+o(\delta t)][(c)2\mu\sigma\delta t+o(\delta t)]\sigma \\ &+P_{j+1,c-1}(t)[1-\lambda\delta t+o(\delta t)][2\mu(1-\sigma)\delta t+o(\delta t)][1-(c-1)2\mu\sigma\delta t+o(\delta t)]\sigma \\ &+\sum_{i=1}^{c} P_{c-i,c-i}(t)[\lambda b_{i+j-c}\delta t+o(\delta t)][1-(c-i)2\mu\sigma\delta t+o(\delta t)]\sigma^i \\ &+o(\delta t)\end{aligned}$$

This becomes after some algebraic manipulation:

$$\begin{aligned}\frac{dP_{j,c}(t)}{dt} &= -(\lambda+(c)2\mu\sigma)P_{j,c}(t)+\lambda\sum_{i=1}^{j-c} b_i P_{j-i,c}(t)+(c)2\mu\sigma^2 P_{j+1,c}(t) \\ &+2\mu(1-\sigma)\sigma P_{j+1,c-1}(t)+\lambda\sum_{i=1}^{c} b_{i+j-c}\sigma^i P_{c-i,c-i}(t)\end{aligned}$$

The steady state equations are therefore:

$$\begin{aligned}(\lambda+(c)2\mu\sigma)P_{j,c} &= \lambda\sum_{i=1}^{j-c} b_i P_{j-i,c} + (c)2\mu\sigma^2 P_{j+1,c} \\ &+2\mu(1-\sigma)\sigma P_{j+1,c-1} \\ &+\lambda\sum_{i=1}^{c} b_{i+j-c}\sigma^i P_{c-i,c-i}\end{aligned} \tag{3.23}$$

If $j \geq c+1$

Equation (3.24) describes the situation where all customers are served in the second branch of the service facility. This holds for $j \geq c+1$

$$\begin{aligned}P_{j,0}(t+\delta t) &= P_{j,0}(t)\left[1-\lambda\delta t+o(\delta t)\right]\left[1-(c)2\mu(1-\sigma)\delta t+o(\delta t)\right] \\ &+\sum_{i=1}^{j-c} P_{j-i,0}(t)\left[\lambda b_i\delta t+o(\delta t)\right]\left[1-(c)2\mu(1-\sigma)\delta t+o(\delta t)\right] \\ &+P_{j+1,1}(t)\left[1-\lambda\delta t+o(\delta t)\right]\left[2\mu\sigma\delta t+o(\delta t)\right]\left[1-(c-1)2\mu(1-\sigma)\delta t+o(\delta t)\right](1-\sigma) \\ &+P_{j+1,0}(t)\left[1-\lambda\delta t+o(\delta t)\right]\left[(c)2\mu(1-\sigma)\delta t+o(\delta t)\right](1-\sigma) \\ &+\sum_{i=1}^{c} P_{c-i,0}(t)\left[\lambda b_{i+j-c}\delta t+o(\delta t)\right]\left[1-(c-i)2\mu(1-\sigma)\delta t+o(\delta t)\right](1-\sigma)^i \\ &+o(\delta t)\end{aligned}$$

This becomes after some algebraic manipulation:

$$\begin{aligned}\frac{dP_{j,0}(t)}{dt} &= -\left(\lambda+(c)2\mu(1-\sigma)\right)P_{j,0}(t)+\lambda\sum_{i=1}^{j-c} b_i P_{j-i,0}(t)+2\mu\sigma(1-\sigma)P_{j+1,1}(t) \\ &+(c)2\mu(1-\sigma)^2 P_{j+1,0}(t)+\lambda\sum_{i=1}^{c} b_{i+j-c}(1-\sigma)^i P_{c-i,0}(t)\end{aligned}$$

Thus the steady state equations are:

$$\begin{aligned}\left(\lambda+(c)2\mu(1-\sigma)\right)P_{j,0} &= \lambda\sum_{i=1}^{j-c} b_i P_{j-i,0} \\ &+2\mu\sigma(1-\sigma)P_{j+1,1} \\ &+(c)2\mu(1-\sigma)^2 P_{j+1,0} \\ &+\lambda\sum_{i=1}^{c} b_{i+j-c}(1-\sigma)^i P_{c-i,0}\end{aligned} \tag{3.24}$$

If $j \geq c+1$

Finally, Equation (3.25) describes the situation where m customers are served in the first branch of the service facility and (c – m) customers are served in the second branch of the service facility. This holds for $j \geq c+1$ and $1 \leq m \leq c-1$

$$
\begin{aligned}
P_{j,m}(t+\delta t) = {} & P_{j,m}(t)\left[1-\lambda\delta t+o(\delta t)\right]\left[1-(m)2\mu\sigma\delta t-(c-m)2\mu(1-\sigma)\delta t+o(\delta t)\right] \\
& +P_{j+1,m-1}(t)\left[1-\lambda\delta t+o(\delta t)\right]\left[(c-m+1)2\mu(1-\sigma)\delta t+o(\delta t)\right]\sigma \\
& +P_{j+1,m}(t)\left[1-\lambda\delta t+o(\delta t)\right]\left[(m)2\mu\sigma^2\delta t+(c-m)2\mu(1-\sigma)^2\delta t+o(\delta t)\right] \\
& +P_{j+1,m+1}(t)\left[1-\lambda\delta t+o(\delta t)\right]\left[(m+1)2\mu\sigma\delta t+o(\delta t)\right](1-\sigma) \\
& +\sum_{i=1}^{j-c} P_{j-i,m}(t)\left[\lambda b_i\delta t+o(\delta t)\right]\left[1-(m)2\mu\sigma\delta t-(c-m)2\mu(1-\sigma)\delta t+o(\delta t)\right] \\
& +\sum_{i=1}^{c-m} P_{c-i,m}(t)\left[\lambda b_{i+j-c}\delta t+o(\delta t)\right]\left[1-(m)2\mu\sigma\delta t-(c-m-i)2\mu(1-\sigma)\delta t+o(\delta t)\right](1-\sigma)^i \\
& +\sum_{i=1}^{m} P_{c-i,m-i}(t)\left[\lambda b_{i+j-c}\delta t+o(\delta t)\right]\left[1-(m-i)2\mu\sigma\delta t-(c-m)2\mu(1-\sigma)\delta t+o(\delta t)\right]\sigma^i \\
& +\sum_{n=1}^{m}\sum_{i=1}^{c-m} P_{c-n-i,m-n}(t)\left[\lambda b_{j-c+i+n}\delta t+o(\delta t)\right]\left[1-(m-n)2\mu\sigma\delta t-(c-m-i)2\mu(1-\sigma)\delta t+o(\delta t)\right]\sigma^n(1-\sigma)^i
\end{aligned}
$$

The differential-difference equations for $P_{j,m}(t)$ are:

$$
\begin{aligned}
\frac{dP_{j,m}(t)}{dt} = {} & -\left(\lambda+(m)2\mu\sigma+(c-m)2\mu(1-\sigma)\right)P_{j,m}(t)+(c-m+1)2\mu(1-\sigma)\sigma P_{j+1,m-1}(t) \\
& +\left((m)2\mu\sigma^2+(c-m)2\mu(1-\sigma)^2\right)P_{j+1,m}(t)+(m+1)2\mu\sigma(1-\sigma)P_{j+1,m+1}(t) \\
& +\lambda\sum_{i=1}^{j-c} b_i P_{j-i,m}(t)+\lambda\sum_{i=1}^{c-m} b_{i+j-c}(1-\sigma)^i P_{c-i,m}(t)+\lambda\sum_{i=1}^{m} b_{i+j-c}\sigma^i P_{c-i,m-i}(t) \\
& +\lambda\sum_{n=1}^{m}\sum_{i=1}^{c-m} b_{j-c+i+n}\sigma^n(1-\sigma)^i P_{c-n-i,m-n}(t)
\end{aligned}
$$

The steady state equations are therefore:

$$\begin{aligned}\left(\lambda+(m)2\mu\sigma+(c-m)2\mu(1-\sigma)\right)P_{j,m} &= (c-m+1)2\mu(1-\sigma)\sigma P_{j+1,m-1} \\ &+\left((m)2\mu\sigma^2+(c-m)2\mu(1-\sigma)^2\right)P_{j+1,m} \\ &+(m+1)2\mu\sigma(1-\sigma)P_{j+1,m+1} \\ &+\lambda\sum_{i=1}^{j-c} b_i P_{j-i,m} \\ &+\lambda\sum_{i=1}^{c-m} b_{i+j-c}(1-\sigma)^i P_{c-i,m} \\ &+\lambda\sum_{i=1}^{m} b_{i+j-c}\sigma^i P_{c-i,m-i} \\ &+\lambda\sum_{n=1}^{m}\sum_{i=1}^{c-m} b_{j-c+i+n}\sigma^n(1-\sigma)^i P_{c-n-i,m-n}\end{aligned}$$

(3.25)

If $j \geq c+1$ and $1 \leq m \leq c-1$

Therefore, the steady state equations are:

$\lambda P_0 = 2\mu\sigma P_{1,1} + 2\mu(1-\sigma)P_{1,0}$
$\left(\lambda+(j)2\mu\sigma\right)P_{j,j} = \lambda\sum_{i=1}^{j}\sigma^i b_i P_{j-i,j-i} + (j+1)2\mu\sigma P_{j+1,j+1} + 2\mu(1-\sigma)P_{j+1,j}$
$1 \leq j \leq c-1$
$\left(\lambda+(j)2\mu(1-\sigma)\right)P_{j,0} = \lambda\sum_{i=1}^{j}(1-\sigma)^i b_i P_{j-i,0} + 2\mu\sigma P_{j+1,1} + (j+1)2\mu(1-\sigma)P_{j+1,0}$
$1 \leq j \leq c-1$
$\left(\lambda+(j-m)2\mu(1-\sigma)+(m)2\mu\sigma\right)P_{j,m} = \lambda\sum_{i=1}^{j-m}(1-\sigma)^i b_i P_{j-i,m} + \lambda\sum_{i=1}^{m}\sigma^i b_i P_{j-i,m-i} + \lambda\sum_{n=1}^{m}\sum_{i=1}^{j-m}(1-\sigma)^i\sigma^n b_{i+n} P_{j-n-i,m-n} + (j+1-m)2\mu(1-\sigma)P_{j+1,m} + (m+1)2\mu\sigma P_{j+1,m+1}$
$1 \leq j \leq c-1$ and $1 \leq m \leq j-1$

$$\left(\lambda+(c)2\mu\sigma\right)P_{c,c}=\lambda\sum_{i=1}^{c}\sigma^{i}b_{i}P_{c-i,c-i}+(c)2\mu\sigma^{2}P_{c+1,c}+2\mu(1-\sigma)\sigma P_{c+1,c-1}$$

$$\begin{aligned}\left(\lambda+(c)2\mu(1-\sigma)\right)P_{c,0}=&\lambda\sum_{i=1}^{c}(1-\sigma)^{i}b_{i}P_{c-i,0}\\&+2\mu\sigma(1-\sigma)P_{c+1,1}\\&+(c)2\mu(1-\sigma)^{2}P_{c+1,0}\end{aligned}$$

$$\begin{aligned}\left(\lambda+(m)2\mu\sigma+(c-m)2\mu(1-\sigma)\right)P_{c,m}=&\left((m)2\mu\sigma^{2}+(c-m)2\mu(1-\sigma)^{2}\right)P_{c+1,m}+\\&(m+1)2\mu\sigma(1-\sigma)P_{c+1,m+1}\\&+(c-m+1)2\mu\sigma(1-\sigma)P_{c+1,m-1}\\&+\lambda\sum_{i=1}^{c-m}b_{i}(1-\sigma)^{i}P_{c-i,m}\\&+\lambda\sum_{i=1}^{m}b_{i}\sigma^{i}P_{c-i,m-i}\\&+\lambda\sum_{n=1}^{m}\sum_{i=1}^{c-m}b_{i+n}\sigma^{n}(1-\sigma)^{i}P_{c-n-i,m-n}\end{aligned}$$

$$1\le m\le c-1$$

$$\begin{aligned}\left(\lambda+(c)2\mu\sigma\right)P_{j,c}=&\lambda\sum_{i=1}^{j-c}b_{i}P_{j-i,c}+(c)2\mu\sigma^{2}P_{j+1,c}\\&+2\mu(1-\sigma)\sigma P_{j+1,c-1}\\&+\lambda\sum_{i=1}^{c}b_{i+j-c}\sigma^{i}P_{c-i,c-i}\end{aligned}$$

$$j\ge c+1$$

$$\begin{aligned}\left(\lambda+(c)2\mu(1-\sigma)\right)P_{j,0}=&\lambda\sum_{i=1}^{j-c}b_{i}P_{j-i,0}\\&+2\mu\sigma(1-\sigma)P_{j+1,1}\\&+(c)2\mu(1-\sigma)^{2}P_{j+1,0}\\&+\lambda\sum_{i=1}^{c}b_{i+j-c}(1-\sigma)^{i}P_{c-i,0}\end{aligned}$$

$$j\ge c+1$$

$$\left(\lambda+(m)2\mu\sigma+(c-m)2\mu(1-\sigma)\right)P_{j,m}=(c-m+1)2\mu(1-\sigma)\sigma P_{j+1,m-1}$$
$$+\left((m)2\mu\sigma^2+(c-m)2\mu(1-\sigma)^2\right)P_{j+1,m}$$
$$+(m+1)2\mu\sigma(1-\sigma)P_{j+1,m+1}$$
$$+\lambda\sum_{i=1}^{j-c}b_iP_{j-i,m}$$
$$+\lambda\sum_{i=1}^{c-m}b_{i+j-c}(1-\sigma)^iP_{c-i,m}$$
$$+\lambda\sum_{i=1}^{m}b_{i+j-c}\sigma^iP_{c-i,m-i}$$
$$+\lambda\sum_{n=1}^{m}\sum_{i=1}^{c-m}b_{j-c+i+n}\sigma^n(1-\sigma)^iP_{c-n-i,m-n}$$

$$j\geq c+1 \text{ and } 1\leq m\leq c-1$$

This infinite set of equations must be solved to individually find $P_{j,m}$ for all j and m. Analytically, this is a very difficult task; therefore the equations will be solved iteratively in order to find the probabilities. Once the probabilities have been found, they can be used to find summary measures for the system, thus completing the picture. Once these summary measures have been calculated, comparison will be made with the Simul8 model. Finally, data from the Critical Care Unit will be used to populate the model.

3.4.1. Solution

A program was written in VBA for Excel in order to solve these equations iteratively. Initially, the probabilities $P_{old}(j, m)$ are estimated using a convergent series which sums to 1. These probabilities are then used in the right hand side of the above equations to generate new probabilities, $P_{new}(j, m)$. As soon as a P_{new} is created, any subsequent equations involving $P_{j,m}$ will use the value $P_{new}(j, m)$. Once each probability required has been generated in this fashion once, $P_{old}(j, m)$ is set equal to $P_{new}(j, m)$ and the process begins again. The iterative procedure is terminated when the difference between $P_{old}(j, m)$ and $P_{new}(j, m)$ is small.

Equations (3.16 – 3.25) were rearranged and solved in a non-intuitive order (for example, $P_{new}(0, 0)$ was generated after $P_{new}(j, 0)$, $P_{new}(j, m)$ and $P_{new}(j, j)$). This was done to ensure that each new probability generated used the previous "new" probabilities, rather than the estimated "old" probabilities. The equations, as they were entered in the VBA program, are found in Appendix 3.2.

In order to gain precise probabilities, the number of iterations required (limit in the above piece of code) is approximately 1,000,000. Due to the complex nature of the equations and the many "for" loops within the "do" loop, running such a large number of times takes a long time, a couple of days at least.

To verify the equations, several steps were taken. To begin with, the parameters were set in the equations such that the equations would describe the M/He/1 queueing system and could thus be verified with known theoretical results. With c = 1 and a fixed batch size of 1, the equations become:

$$\lambda P_0 = 2\mu\sigma P_{1,1} + 2\mu(1-\sigma)P_{1,0}$$

$$(\lambda + 2\mu\sigma)P_{1,1} = \lambda\sigma b_1 P_0 + 2\mu\sigma^2 P_{2,1} + 2\mu(1-\sigma)\sigma P_{2,0}$$

$$(\lambda + 2\mu(1-\sigma))P_{1,0} = \lambda(1-\sigma)b_1 P_0 + 2\mu\sigma(1-\sigma)P_{2,1} + 2\mu(1-\sigma)^2 P_{2,0}$$

$$(\lambda + 2\mu\sigma)P_{j,1} = \lambda\sum_{i=1}^{j-1} b_i P_{j-i,1} + 2\mu\sigma^2 P_{j+1,1} + 2\mu(1-\sigma)\sigma P_{j+1,0} + \lambda b_j \sigma P_0$$

$$(\lambda + 2\mu(1-\sigma))P_{j,0} = \lambda\sum_{i=1}^{j-1} b_i P_{j-i,0} + 2\mu\sigma(1-\sigma)P_{j+1,1} + 2\mu(1-\sigma)^2 P_{j+1,0} + \lambda b_j (1-\sigma)P_0$$

Comparison with (3.5 – 3.7) show that the equations do in fact collapse to give the M/He/1 equations.

Next, c was set equal to 6, σ was set equal to 0.5 and the batch distribution was set to be fixed and constant with a batch size of 1. This situation is directly comparable to well-known M/M/c queueing system and the steady state probabilities and the summary measures are well documented.

The probability of there being *n* customers in the system is given by:

$$P_n = \begin{cases} \dfrac{(c\rho)^n P_0}{n!}, & 1 \le n \le c-1 \\ \dfrac{c^c \rho^n P_0}{c!}, & n \ge c \end{cases}$$

Where $P_0 = \dfrac{1}{\left(\sum_{n=0}^{c-1} \dfrac{(c\rho)^n}{n!}\right) + \left(\dfrac{c^c \rho^c}{c!(1-\rho)}\right)}$

The summary measures, *L*, *Lq*, W and W_q are given by the following expressions:

$$L = \frac{c^c \rho^{c+1} P_0}{c!(1-\rho)^2} + c\rho, \; L_q = \frac{c^c \rho^{c+1} P_0}{c!(1-\rho)^2}, \; W = \frac{c^c \rho^{c+1} P_0}{\lambda c!(1-\rho)^2} + \frac{1}{\mu} \text{ and}$$

$$W_q = \frac{c^c \rho^{c+1} P_0}{\lambda c!(1-\rho)^2}$$

The M/He/c equations contain probabilities in the form $P_{j,m}$, whereas the M/M/c equations do not have this second suffix. Some facts concerning the probabilities of the M/He/c queue when σ = 0.5 need to be established before the equations are manipulated.

1. $P_{0,0}$ can simply be changed to P_0
2. Since σ = 0.5 $P_{j,0} = P_{j,j}$ and thus only one of these equations need to be considered.
3. In fact, each probability $P_{j,m}$ can be expressed in terms of $P_{j,0}$ using the following expression:

$$P_{j,m} = \binom{j}{m} P_{j,0} \quad j \le c$$

4. If $j > c$ then

$$P_{j,m} = \binom{c}{m} P_{j,0} \quad j \le c$$

5. As a result of the above observations, it is clear that

$$P_j = \sum_{m=0}^{c} \binom{j}{m} P_{j,0}$$
$$= 2^j P_{j,0} \qquad j \le c$$
or
$$P_j = \sum_{m=0}^{c} \binom{c}{m} P_{j,0}$$
$$= 2^c P_{j,0} \qquad j > c$$

The final probabilities can be expressed as the following, using the M/He/c expressions as a starting point:

The equation for P_0 becomes:

$$\lambda P_0 = \mu\left(P_{1,1} + P_{1,0}\right)$$
$$= \mu P_1$$
$$\Rightarrow P_1 = \frac{\lambda}{\mu} P_0$$

The equation for $P_{1,0}$ becomes:

$$(\lambda+\mu)P_{1,0}=\frac{\lambda}{2}P_0+4\mu P_{2,0}$$

$$(\lambda+\mu)\frac{P_1}{2}=\frac{\lambda}{2}P_0+4\mu\frac{P_2}{4}$$

$$P_2=\left(\frac{\lambda}{\mu}+1\right)\frac{P_1}{2}-\frac{\lambda}{2\mu}P_0$$

$$=\left(\frac{\lambda}{\mu}+1\right)\frac{\lambda}{2\mu}P_0-\frac{\lambda}{2\mu}P_0$$

$$\Rightarrow P_2=\frac{\lambda^2}{2\mu^2}P_0$$

The equation for $P_{2,0}$ becomes:

$$(\lambda+2\mu)P_{2,0}=\frac{\lambda}{2}P_{1,0}+\mu P_{3,1}+3\mu P_{3,0}$$

$$(\lambda+2\mu)\frac{P_2}{4}=\frac{\lambda}{2}\frac{P_1}{2}+6\mu\frac{P_3}{8}$$

$$\frac{3}{4}\mu P_3=(\lambda+2\mu)\frac{P_2}{4}-\frac{\lambda}{2}\frac{P_1}{2}$$

$$3\mu P_3=(\lambda+2\mu)P_2-\lambda P_1$$

$$3\mu P_3=(\lambda+2\mu)\frac{\lambda^2}{2\mu^2}P_0-\lambda\frac{\lambda}{\mu}P_0$$

$$\Rightarrow P_3=\frac{\lambda^3}{3!\mu^3}P_0$$

In the same manner, P_4, P_5 and P_6 can be derived. Finally, considering the equation for P_7 we have:

$$(\lambda+6\mu)P_{6,0}=\frac{\lambda}{2}P_{5,0}+\frac{\mu}{2}P_{7,1}+6\frac{\mu}{2}P_{7,0}$$

$$(\lambda+6\mu)\frac{P_6}{64}=\frac{\lambda}{2}\frac{P_5}{32}+6\mu\frac{P_7}{64}$$

$$(\lambda+6\mu)P_6=\lambda P_5+6\mu P_7$$

$$6\mu P_7=(\lambda+6\mu)\frac{\lambda^6}{6!\mu^6}P_0-\lambda\frac{\lambda^5}{5!\mu^5}P_0$$

$$\Rightarrow P_7=\frac{6^6\rho^7}{6!}P_0$$

It is clear therefore, that for the case σ = 0.5 and $b_1=1,\ b_i=0\ \forall\ i\neq 1$, the $M^{(b)}$/He/c equations collapse to give the known theoretical equations for M/M/c.

Next comparison with the $M^{(b)}$/He/1 is required, this will be done in using two different batch size distribution, firstly the fixed distribution, then the Poisson distribution. For the sake of this piece of verification, σ will be kept constant at 0.2 and ρ will be kept constant at 0.5:

Batches of fixed size

Table 3.11: Iterative and theoretical values for *L* and L_q for fixed batches

	ITERATIVE		THEORETICAL	
BATCH SIZE	*L*	L_Q	*L*	L_Q
1	1.28125	0.78125	1.28125	0.78125
2	1.78125	1.28125	1.78125	1.28125
3	2.28125	1.78125	2.28125	1.78125
4	2.78125	2.28125	2.78125	2.28125
5	3.28125	2.78125	3.28125	2.78125
10	5.78125	5.28125	5.78125	5.28125
15	8.28125	7.78125	8.28125	7.78125

Table 3.11 displays the results for *L* and *Lq* for the iterative procedure as well as the theoretical solution found in section 3.3. The results from the iterative

procedure concord well with the theoretical results, but for larger batch sizes, the iterative procedure required a long run time.

Batches of a variable size (following a Poisson distribution)

When the batch distribution is changed to Poisson, the summary measures, found in Table 3.12, are slightly less accurate than with the fixed batch size. But when batches are larger than 3, the results, for all practical purposes, are close enough.

Table 3.12: Iterative and theoretical values for *L* and L_q for Poisson batches

	ITERATIVE		THEORETICAL		PERCENTAGE DIFFERENCE IN *L*
MEAN BATCH SIZE	*L*	L_Q	*L*	L_Q	
1	1.936	1.393	1.78125	1.28125	8%
2	2.449	1.925	2.28125	1.78125	7%
3	2.873	2.363	2.78125	2.28125	3%
4	3.324	2.821	3.28125	2.78125	1%
5	3.800	3.298	3.78125	3.28125	< 1%
10	6.281	5.781	6.28125	5.78125	< 1%
15	8.781	8.281	8.78125	8.28125	< 1%

If the batch size distribution is changed to Poisson, the results for *L* and *Lq* are larger. As the mean batch size increases, the iterative results approach the theoretical results.

The iterative procedure needs to be verified for two final cases, where c is larger than 1 and batch size is of fixed size, or where batch size follows the Poisson distribution. Table 3.13 contains the summary measures produced by the iterative procedure and also, for comparison, the simulated values produced by the Simul8 model, for the case c = 2.

Table 3.13 Summary measures for the $M^{(b)}/He/2$ queue

Distribution	σ	ρ	Iterative		Simul8	
			L	L_Q	L	L_Q
Fixed, mean = 5	0.1	0.1	0.30	0.11	0.33	0.13
		0.5	3.67	2.67	3.55	2.54
		0.9	41.38	39.60	40.45	38.64
	0.3	0.1	0.30	0.10	0.32	0.12
		0.5	3.26	2.26	3.15	2.14
		0.9	29.90	28.10	29.21	27.41
Poisson, mean = 5	0.1	0.1	0.31	0.12	0.34	0.14
		0.5	4.10	3.09	4.00	3.01
		0.9	43.34	41.55	43.08	41.28
	0.3	0.1	0.31	0.11	0.32	0.12
		0.5	3.72	2.72	3.62	2.61
		0.9	32.82	31.01	32.65	30.85

Table 3.13 demonstrates that the equations for the $M^{(b)}/He/c$ queue are valid for $c > 1$ and for the two batch distributions, constant and Poisson. It is clear from the table that as ρ increases the accuracy of the results are slightly less good, but for all practical purposes, they are close enough.

The main purpose of deriving the above queueing equations is to apply them to a real life situation. As was explained in Chapter 1, the Critical Care Unit (CCU) at the University Hospital of Wales, Cardiff (UHW) have given us a wealth of data referring to the number of patients present in the Unit on each day and also data related to each patient. The following section seeks to model bed-occupancy in the CCU by utilising the above queueing theory.

3.5. Application to CCU data

Since the above equations have proven to be valid when compared with simulation, the next step is to populate the equations with parameters which reflect the conditions of arrival and service in the Critical Care Unit.

3.5.1. Arrival distribution

Since the theory developed in this Chapter relates to batches of customers, the first step in generating suitable parameters for this model is to consider the arrival process.

The number of patients that arrive at the CCU each day was calculated. Stat::Fit was utilised to obtain the statistical distribution of the number in a batch. Figure 3.7 displays a histogram of batch size along with the fitted Poisson distribution. The Poisson distribution provides a good fit with mean value 3.61. This will correspond to the batch size distribution which is utilised in the above set of expressions.

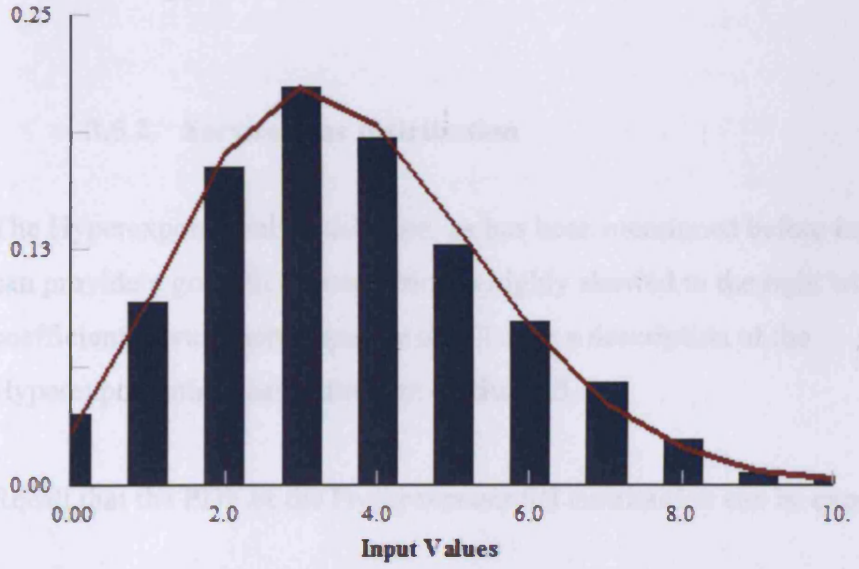


Figure 3.7: Histogram of batch size with fitted Poisson distribution

Table 3.14 contains summary measures for the number of patients in a batch.

Table 3.14: Summary statistics of the number of patients in a batch

Mean	Std Dev	Variance	Minimum	Median	Maximum
3.61	1.95	3.79	0	3	10

As can be seen from the summary statistics the theoretical mean and variance (both 3.61) are a very good approximation of the actual mean and variance.

One point worthy of consideration is the duration between each batch arrival. The theory assumes that the time between each batch arrival follows a Negative Exponential distribution. The batch distribution calculated above relates to the number of patients who arrive at the CCU each day. The time between each batch arrival from the data is thus Deterministic. The Deterministic batch arrivals case will be considered briefly in Chapter 7, but its solution is beyond the scope of this research.

Therefore, the arrival process utilised in the equations developed in this Chapter will act as an approximation to the real-life arrival process at the CCU and hence we have $\lambda = \frac{1}{24}$.

3.5.2. Service time distribution

The Hyperexponential distribution, as has been mentioned before in this thesis, can provide a good fit to data which is highly skewed to the right where the coefficient of variation is greater than 1. For a description of the Hyperexponential distribution see section 2.3.4.

Recall that the PDF of the Hyperexponential distribution can be expressed as

$$f(t) = 2\mu\sigma e^{-2\mu\sigma t} + 2\mu(1-\sigma)e^{-2\mu(1-\sigma)t} \quad t \geq 0$$

In order to calculate suitable values for the parameters μ and σ, the method of least squares was utilised. Firstly, from the CCU data, the probability of a patient having a length of stay of 0 – 24 hours, 24 – 48 hours, etc, was calculated. Corresponding Hyperexponential probabilities for these intervals were also calculated by integrating the PDF across the appropriate limits. The squared difference of these quantities was calculated for each 24 hour interval, and the sum of the squared differences was calculated. The optimisation tool, Solver which uses the Simplex algorithm to find the optimal value, was used to find values of μ and σ such that the sum of the squared differences was minimised.

Table 3.15 contains the summary measures for the length of stay of patients at the CCU. Note that patients with a length of stay greater than 3000 hours, which is 125 days, were excluded as they account for only 0.3% of the data and cause a large skew in the mean.

Table 3.15: Summary statistics for length of stay of patients, in hours

N	Mean	Std Dev	Min	Median	Max
4212	121.51	203.32	0	50	2170

Figure 3.8 presents the histogram of length of stay of patients at the CCU with the fitted Hyperexponential overlaid. The values of μ and σ are 0.0083 and 0.073 respectively and thus the estimated mean and standard deviation is 121.51 and 306.62 hours. The mean, as can be seen from Table 3.15, is a very good reflection of reality, but the standard deviation is not as accurate. If the number of phases of the Hyperexponential was extended from 2 to 4, the fit provided is much better. Future research is required in order to develop similar queueing equations for the four phase Hyperexponential distribution.

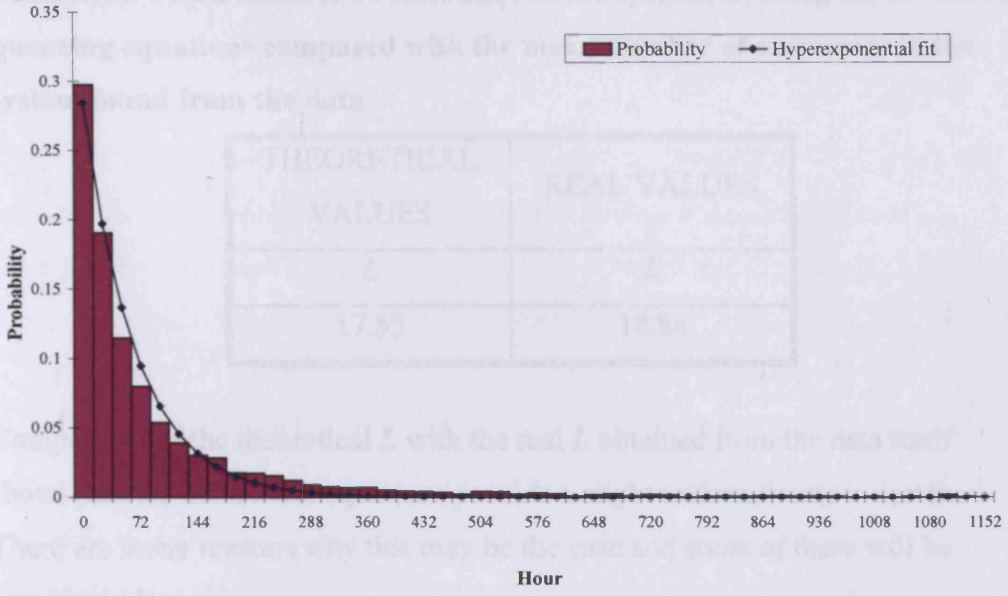


Figure 3.8: Length of stay distribution for all patients

3.5.3. Number of beds (c)

The final quantity of interest is c, the number of service channels or beds. The CCU, at present, has 24 funded beds. Therefore, c = 24.

3.6. Results

Table 3.16 contains the theoretical value for L calculated using the iterative method and also the actual mean number of customers in the system calculated from the data.

Table 3.16: Mean number of customers in the system, L, using the $M^{(b)}$/He/c queueing equations compared with the mean number of customers in the system found from the data

THEORETICAL VALUES	REAL VALUES
L	L
17.55	18.86

Comparison of the theoretical L with the real L obtained from the data itself shows that the theoretical equations provide a slight underestimate in results. There are many reasons why this may be the case and some of these will be considered later.

Figure 3.9 contains the probability density function for the real data and the theoretical model using the above parameters and limiting the number of admissions to 24:

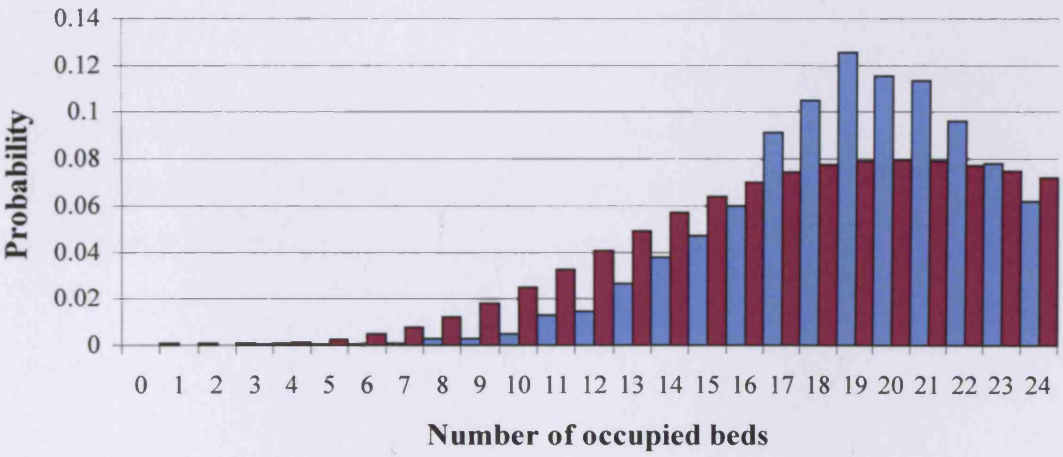


Figure 3.9: Probability density function of the number of occupied beds

It is clear from Figure 3.9 that the $M^{(b)}$/He/c queue acts as an approximation to the queueing system in the Critical Care Unit. It is evident that compared to the smooth distribution obtained from the theory, the actual occupancy profile in the

Unit is more jagged in nature. Also, the actual occupancy profile is more peaked than the theoretical model.

There are many reasons why there is a difference in distribution. Firstly, as has been mentioned already, the time between successive batch arrivals in the theoretical work is assumed to be Markovian, whereas the way in which the data was considered assumed that batches arrived daily.

Secondly, the CCU itself has priority set in place. For example, if a patient who has undergone Emergency surgery requires the use of a CCU bed, they will have precedence over any Elective surgery case. Thus, Elective surgery can be cancelled due to lack of CCU bed space (for further consideration of this point, see Chapter 6).

Finally, the Hyperexponential distribution with two phases is a reasonable approximation to the length of stay distribution of patients in the CCU but a better fit is found when the number of phases is extended to 4. The theoretical work requires extending to incorporate these additional phases.

3.7. Conclusion

This Chapter has introduced a novel queueing system, $M^{(b)}$/He/c, which has not been considered before in the Literature. A solution has been suggested and an application to a real life situation has been made.

To conclude, it is worthy to note that the $M^{(b)}$/He/c queueing system proves to be a good approximation to the CCU thus addressing the first objective of this thesis. However, notable improvements could be made by considering constant time-intervals between batch arrivals, extension to different service time distributions (such as additional phases in the Hyperexponential distribution) and also introducing an element of priority to the queueing system to account for the priority present in the CCU.

There are many avenues for further work to investigate, for example, a loss system, whereby customers of certain characteristics are "lost" if there is insufficient space for them. This system would correspond well to the CCU also and would possibly give simpler solutions than the queueing system considered.

One final system that would be worth considering would be the M/He/c queueing system (the same system but with single arrivals). This queueing system has been considered before in the literature but could prove to be another useful approximation to the CCU.

Chapter 4: Length of stay analysis

4.1. Introduction

In this Chapter, a discussion follows into the factors which affect length of stay in the Critical Care Unit at the University Hospital of Wales, Cardiff. A description of the data set used can be found in Chapter 1. In a clinical setting, it can help in the management of patients if the clinicians have an estimate of patients' likely length of stay and if a tool is available which can predict the length of stay of a patient based on a set of patient characteristics relating to the first day of admission into the hospital. Also, from a modelling perspective, this type of analysis can be utilised in future pieces of work to suggest the most influential variables in the prediction of length of stay and thereby constructing more intuitive models. This Chapter seeks to identify the main factors which affect length of stay at the CCU and thus helps to address the second objective of this thesis, which is to understand the factors which affect length of stay and outcome at the Critical Care Unit.

The statistical analysis techniques used to analyse the influence of different factors upon length of stay include CART analysis and Regression analysis. Both techniques provide a predictive tool for length of stay. Using these two initial analyses, the important variables in deducing length of stay will be highlighted. Mann-Whitney tests and the Kruskal-Wallis test follow to investigate the influential variables in more depth.

4.2. CART analysis

CART analysis, undertaken using the TreeWorks (Harper and Leite 2008) program in this study, takes the data set as a whole and systematically subdivides the data set until homogenous groups are formed. The end product is a "tree" consisting of nodes. The aim of this splitting procedure is to create homogenous

groups (nodes) thus reducing the overall variance of the data. Typically, seventy percent of the data is used to create the splitting rules (this is called the learning sample) and the final 30%, the testing sample, is then passed through the tree and is used to validate the final nodes (Harper and Leite 2008).

4.2.1. Data

The data set used to perform the CART analysis consists of 2448 observations (note, the original CCU data set consisted of 4226 observations but TreeWorks requires a data set which is complete therefore some observations could not be used). The learning sample had 1713 observations and the testing sample had 735 observations.

To ensure that the final nodes were helpful from a modelling perspective and appropriate for this study, the first split by the source of admission variable (Elective/Non-elective) was undertaken manually. The remaining procedure was purely statistical in nature, free from researcher intervention. It was specified that final nodes should have no fewer than 50 observations. The rules used to determine the final nodes are listed below (the actual tree appears on the next page) and the validated final nodes have been shaded (20 in total).

Table 4.1: Split specifications – with final nodes shaded

NODE ID	SPLIT VARIABLE	SPLIT SPECIFICATION
0	NONE	None
1	ADMSOURCE	Elective
2	ADMSOURCE	A+E, Emergency, Ward, Other Hosp, X-Ray
3	SDCODE	47, 22, 36, 50, 34, 13, 41, 52, 28, 43, 1, 23, 31, 46, 21, 20, 49, 19, 42, 16, 25, 24, 15, 48, 30, 9, 44
4	SDCODE	4, 10, 6, 29, 51, 45, 37, 5, 3, 14, 2, 39, 12, 32, 26, 7, 27, 11, 18, 8, 38
5	TISSPOINT	<= 56.5
6	TISSPOINT	> 56.5
7	HCO3	<= 32.8
8	HCO3	> 32.8
9	ORGSCORE	<= 10.1
10	ORGSCORE	> 10.1
11	PRECIP	29, 1, 32, 10, 4, 26, 22, 27, 24, 5, 23, 15, 12
12	PRECIP	7, 6, 20, 3, 28, 34, 37, 35, 19
13	PO	<= 81.4
14	PO	> 81.4
15	SDCODE	50, 47, 52, 41, 34, 22, 48, 25, 15, 16, 19
16	SDCODE	43, 1, 21, 30, 23, 42, 20, 49, 24, 13, 28
17	PO	<= 85.9
18	PO	> 85.9
19	VENT	N
20	VENT	Y
21	PRECIP	12, 4, 5, 2, 25, 16, 8, 23, 7, 15, 27, 26, 36, 13, 6, 34, 22, 1, 37, 24
22	PRECIP	20, 17, 30, 35, 3
23	PRECIP	29, 19, 4, 5, 13, 37, 9, 35, 10, 7, 32, 6, 27, 34
24	PRECIP	28, 24, 3, 1, 23, 15, 22, 20, 26
25	NVVERBAL	0
26	NVVERBAL	1
27	ALBUMIN	<= 2.35
28	ALBUMIN	> 2.35
29	SDCODE	47, 34, 31, 36, 19, 22, 16, 24, 42, 25, 41
30	SDCODE	49, 46, 21, 48, 20, 28, 43, 15, 30, 44
31	DIALYSED	N
32	DIALYSED	Y
33	ALBUMIN	<= 2.55
34	ALBUMIN	> 2.55
35	TPN	T, N
36	TPN	E
37	HCO3	<= 18.95
38	HCO3	> 18.95

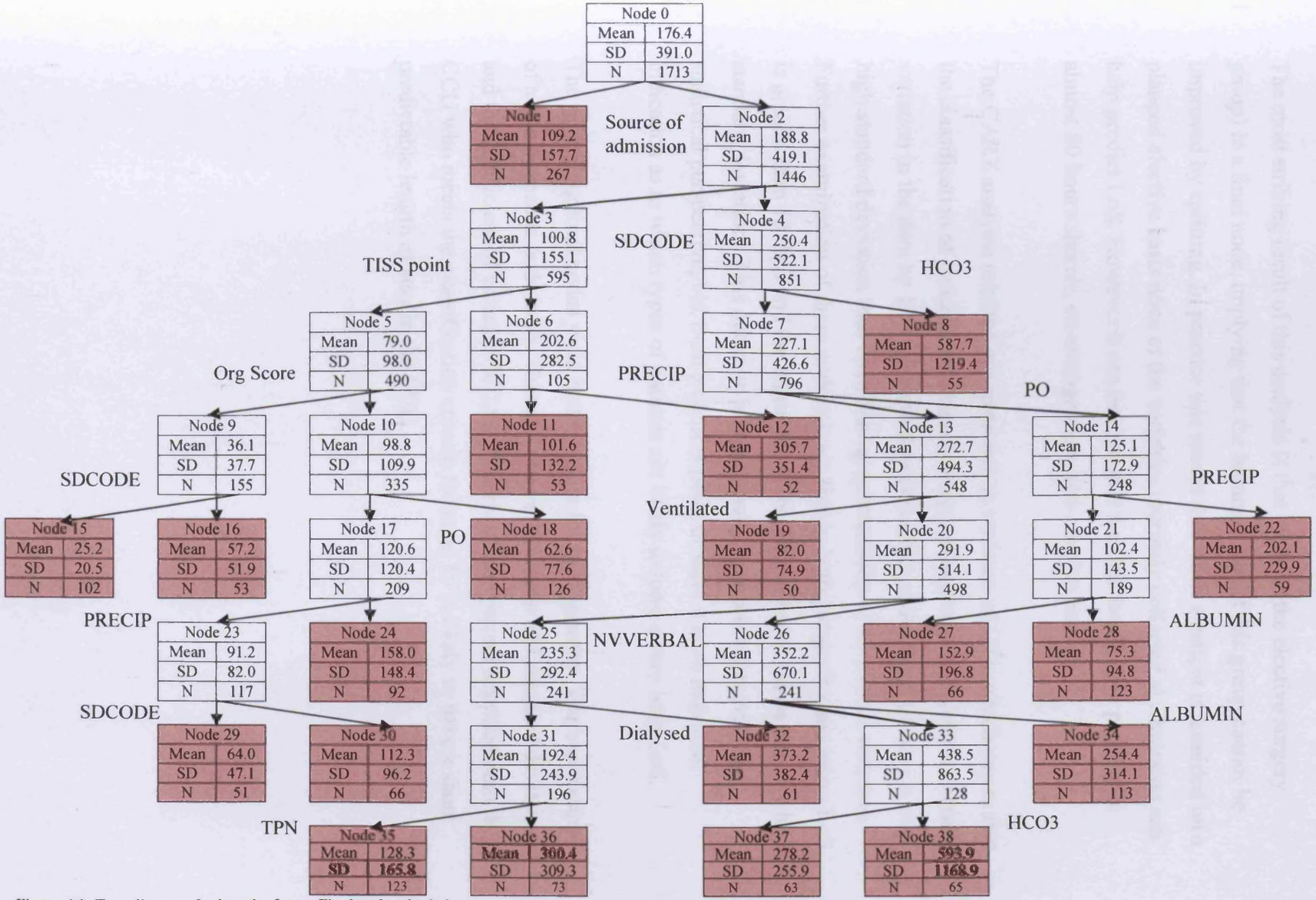


Figure 4.1: Tree diagram for length of stay. Final nodes shaded

The most striking result of this analysis is that Node 1 (the elective surgery group) is a final node, implying that the homogeneity of this group cannot be improved by splitting. In practice this means that when a patient is admitted on a planned elective basis none of the variables routinely collected at admission can help predict LoS. However it can be seen that the LoS for elective patients is almost 80 hours shorter, on average, than non-elective patients.

The CART analysis resulted in a reduction in variance of 12%, which means that the classification of patients into more homogeneous groups reduced the overall variation in the data by 12%. It should be noted that nodes 8 and 38 have a very high standard deviation thus contributing substantially to the overall variation. Further examination of these nodes shows that in both groups the maximum LoS is greater than 40 *days* and the minimum is less than 6 *hours,* hence the very high standard deviation. This outcome is problematic from a homogeneity and statistical perspective, but from a clinical point of view it is an important indication as to which types of patients are likely to have a very long LoS.

The remaining final nodes yield better results for homogeneity. The best example of a homogeneous node is 15. This node has a low mean and standard deviation and the coefficient of variation is 0.81. Therefore, any patient admitted onto the CCU who meets the classification criteria for node 15 is likely to have a short predictable length of stay in the CCU.

The CART analysis results in 38 splitting rules depending upon the following 12 distinct variables:

- Source of admission
- SDCODE – Specific Diagnostic Code
- TISS point – Daily TISS point
- HCO3 - The plasma bicarbonate variable
- Org Score – Knauss's organ score
- PRECIP – Precipitating factor
- Po – $Pa0_2$ - measurement of arterial partial pressure of oxygen
- Ventilated – was the patient ventilated?
- NVVERBAL - verbal response in intubated patients
- Dialysed – was the patient dialysed?
- Albumin - concentration of Albumin in plasma
- TPN – Total Parenteral Nutrition

For a full definition of the above variables, refer to Appendix 4.1. The influence of these variables will be discussed in greater detail later on in this Chapter.

To make this information useful from a modelling perspective, it is important to fit statistical distributions to the length of stay and inter-arrival time data within each node. Using Stat::Fit the LoS distributions were found and are presented in Table 4.2. Note that these distributions were found to be significant at the 95% level.

Table 4.2: Length of stay distributions for CART final nodes

Node	Data				Theoretical		
	N	Mean (hrs)	Std Dev	Median	Distribution	Mean (hrs)	SD
1	394	104.41	151.65	48.00	-----	-----	-----
8	78	558.01	1094.04	244.00	Lognormal	563	1163
11	75	99.85	128.51	57.00	Exponential	99.9	99.9
12	68	283.71	318.50	211.50	Exponential	284	277
15	141	29.60	48.08	21.00	Exponential	29.6	24.6
16	73	54.74	48.02	41.00	Exponential	54.7	48.7
18	190	64.41	75.92	38.50	Exponential	64.4	64.4
19	73	75.45	67.91	59.00	Exponential	75.5	63.5
22	92	208.59	233.22	142.50	Exponential	209	208
24	127	154.08	154.01	120.00	Exponential	154	153
27	93	158.66	206.61	87.00	Exponential	159	157
28	174	94.60	160.07	46.50	Lognormal	111	298
29	74	79.04	93.50	45.00	Exponential	79	73
30	95	109.73	110.94	76.00	Exponential	110	107
32	82	353.60	359.07	265.50	Exponential	354	348
34	172	269.41	375.61	133.00	Lognormal	277	469
35	182	134.55	185.73	61.00	Lognormal	160	397
36	99	371.13	774.15	199.00	Lognormal	356	543
37	93	297.74	289.81	233.00	Exponential	298	293
38	104	484.24	958.19	238.50	Lognormal	475	872

The distribution fitting exercise gave reasonably accurate results for the first and second moments. For the mean, the majority of nodes are within a margin of 5% of the actual mean, apart from two nodes (28 and 35). When considering the standard deviation, the theoretical fit is not as accurate but according to the chi - square goodness of fit test (which Stat::Fit utilises), these distributions cannot be rejected. The Elective patients, node 1, have a time dependent length of stay distribution.

Next, the inter-arrival distribution for each node was found, again using Stat::Fit. The results which are presented in Table 4.3 show that, apart from the elective patients (node 1) which is time-dependent in nature and discussed in depth in Chapter 6, the Negative Exponential distribution is a good fit to the data. Nodes 16 and 36 do not have as good a fit as the others but the Negative Exponential distribution is still not rejected at the 95% level.

Table 4.3: Inter-arrival distribution for CART analysis nodes

Node	Data				Theoretical
	N	Mean	Std Dev	Median	Mean
1	394	69.09	69.20	48.00	-----
8	78	345.95	386.50	264.50	346
11	75	354.76	337.60	280.00	355
12	68	391.00	317.36	339.50	391
15	141	194.13	200.37	164.00	193
16	74	341.09	282.93	303.00	335
18	190	141.79	136.17	96.00	142
19	73	373.37	356.65	252.00	368
22	94	269.55	260.13	244.50	270
24	127	211.62	232.27	117.00	212
27	93	282.88	307.30	186.00	283
28	174	155.78	141.56	119.00	156
29	74	343.54	346.38	203.00	344
30	95	283.53	285.35	216.00	284
32	82	328.60	359.57	201.00	329
34	172	158.75	156.74	120.00	159
35	182	150.36	164.55	95.00	150
36	99	266.83	302.50	178.00	267
37	93	292.25	298.94	196.00	292
38	106	253.19	238.35	187.50	253

The length of stay and inter-arrival distributions can be used to populate a simulation model based upon these nodes. Each node would represent an arrival source and then the service time would be dependent upon the arrival source. This is beyond the scope of this thesis since a different simulation model is built in Chapter 6, but would certainly be an area for further work.

4.3. Linear Regression

A linear regression analysis was undertaken in order to consider which variables influence LoS, from a different perspective.

Linear regression is a statistical technique which seeks to model a dependent variable using a linear combination of independent variables. The parameters found will be deemed as significant or not using the *t*-test. The model itself will be deemed as significant or not using the F-test. It is important to note here that fitting the model is a fairly trivial task; the more difficult task is to ensure that the correct model is fitted.

Linear regression has various assumptions that need to be satisfied before the results can be used with any confidence.

Firstly, it is assumed that the independent variables have no error attached to them, i.e. they are measured accurately and correctly each time. It is also assumed that the independent variables are linearly independent. The remaining assumptions are related to the errors associated with the dependent variable – the errors must be normally distributed with a mean of zero and a constant variance.

The computer program SAS 9.1.3 was used to undertake this analysis. SAS can test the independence assumption as well as the assumptions associated with the errors. SAS will flag up any variables which are a linear combination of others. The assumptions associated with the errors are simply tested by looking at a plot of residuals against predicted values. If this plot represents a random scatter about the point zero which is fairly symmetrical about zero then the error assumptions are satisfied.

Once the assumptions have been met, it is possible to use the findings of the regression analysis to predict the outcome of the dependent variable using values of the independent variables.

It became evident that a number of the variables in the original data set were linear combinations of others and were therefore excluded from the regression analysis. Also, some variables contained many missing values or were inappropriate for regression, such as details concerning GP practices. The remainder were used as independent variables for this regression analysis.

Initially, the variable LoS (Length of stay) was used as the dependent variable but it was found that the errors associated with LoS did not satisfy the assumptions of normality of errors and therefore a log transformation was made.

Given that the CART analysis indicated that the elective group of patients was homogenous, and that the regression was undertaken in order to compare the outcomes of the regression with those of CART, only non-elective patient data (n = 2102) were included in this part of the analysis. The analysis was run using an option in SAS which would search for the model (from all independent variables) that would maximise R^2.

Figure 4.2 illustrates the residual against predicted plot (N.B. this must be a random scatter about the line residual = 0 to satisfy the assumptions).

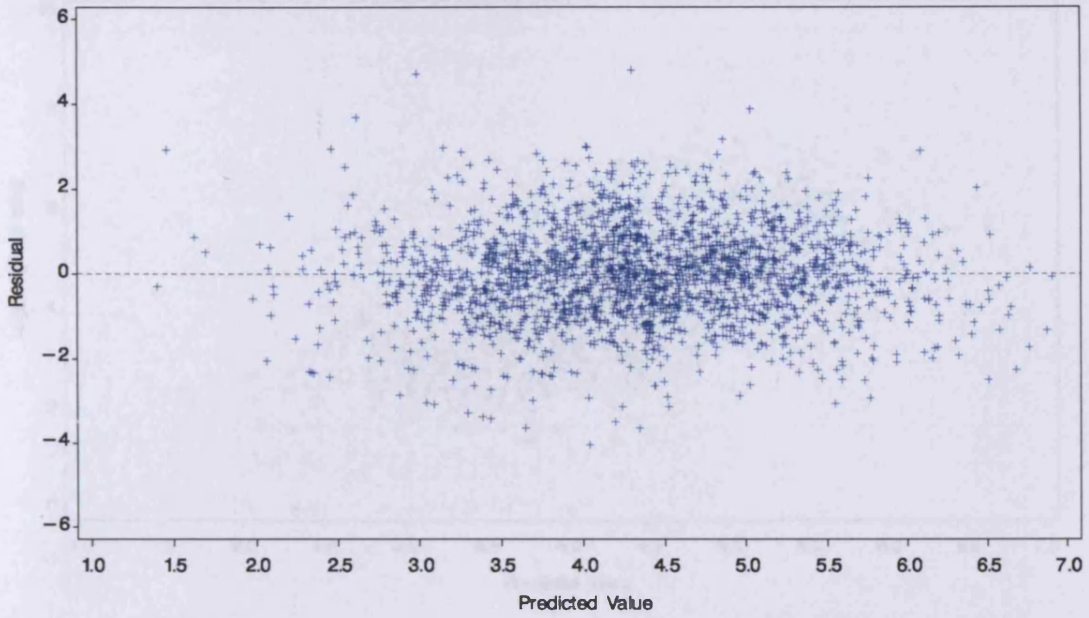


Figure 4.2: Residual vs. predicted plot – non-elective patients

This is clearly a random scatter which is fairly symmetrical above and below the predicted axis.

Figure 4.3 represents a plot of the observed data against the predicted data. If the model corresponded to a perfect fit, this plot would be a straight line.

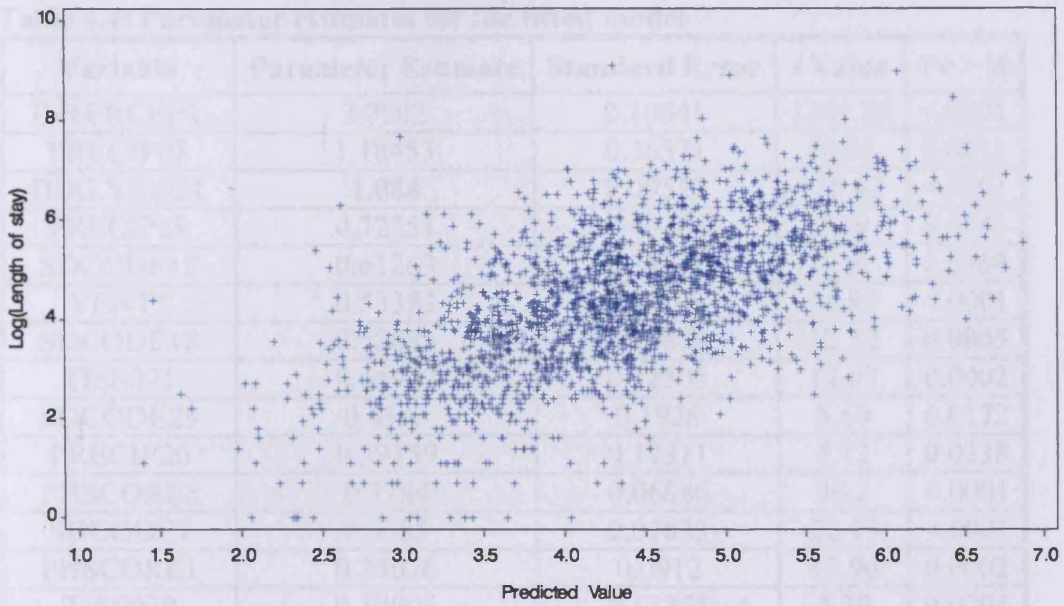


Figure 4.3: Observed vs. Predicted values – non-elective patients

There is an obvious positive trend in these points. Therefore these two figures indicate that the results of the regression procedure may be considered.

The ANOVA results show that the model is significant ($p < 0.00010$) and Table 4.4 shows the parameter estimates generated by the regression procedure.

Table 4.4: Parameter estimates for the fitted model

Variable	Parameter Estimate	Standard Error	*t* Value	Pr > |*t*|
INTERCEPT	3.7502	0.10641	1241.96	<.0001
PRECIP08	1.18453	0.36371	10.61	0.0011
DIALYSED1	1.088	0.09527	130.42	<.0001
PRECIP29	0.72251	0.40464	3.19	0.0743
SDCODE12	0.61263	0.29341	4.36	0.0369
VENT1	0.53383	0.0856	38.89	<.0001
SDCODE18	0.52081	0.14839	12.32	0.0005
TISS071	0.46907	0.12505	14.07	0.0002
SDCODE29	0.4593	0.1926	5.69	0.0172
PRECIP20	0.39159	0.17311	5.12	0.0238
PHSCORE3	0.3784	0.06886	30.2	<.0001
SDCODE7	0.3685	0.07833	22.13	<.0001
PHSCORE1	0.34076	0.0912	13.96	0.0002
TISS039	0.30995	0.13355	5.39	0.0204
EYESCORE3	0.30153	0.12494	5.82	0.0159
PHSCORE2	0.26291	0.06198	17.99	<.0001
TEMPSCORE3	0.24185	0.10217	5.6	0.018
RESPSCORE3	0.23037	0.09985	5.32	0.0211
MBPSCORE2	0.22319	0.05173	18.62	<.0001
TISS063	0.12208	0.05659	4.65	0.0311
RESPSCORE1	0.11804	0.06429	3.37	0.0665
TEMPSCORE1	0.10204	0.05413	3.55	0.0595
HCO3	0.09721	0.0279	12.14	0.0005
VSYSBP	0.09016	0.02679	11.32	0.0008
ORGSCORE	-0.0997	0.03077	10.5	0.0012
ALBUMIN	-0.15064	0.02811	28.72	<.0001
PO	-0.24373	0.03353	52.83	<.0001
PRECIP24	-0.27334	0.10464	6.82	0.0091
SDCODE14	-0.29195	0.18114	2.6	0.1072
SDCODE27	-0.30826	0.14399	4.58	0.0324
SDCODE19	-0.38929	0.13853	7.9	0.005
PRECIP05	-0.40645	0.19819	4.21	0.0404
PRECIP06	-0.45656	0.1312	12.11	0.0005
TISS065	-0.47285	0.15054	9.87	0.0017
TISS068	-0.47809	0.12137	15.52	<.0001
TPNN	-0.60006	0.05787	107.51	<.0001
SDCODE41	-0.61393	0.24777	6.14	0.0133
HAEMCAN	-0.70235	0.18878	13.84	0.0002
SDCODE22	-0.70991	0.10223	48.22	<.0001
TISS064	-1.08599	0.16485	43.4	<.0001

The majority of the parameters in Table 4.4 are significant at the 95% level. Those which are not have been included as they contribute sufficiently to R^2. The scale variables were standardised in order to make their interpretation simpler. Appendix 4.1 contains definitions for each of these variables.

It is important to remember that the dependent variable in this case is the log of length of stay rather than the length of stay itself. Therefore, when interpreting the parameter estimates it is important to bear in mind that the Exponential of the parameter estimate is the factor that the independent variable should be multiplied with, rather than the parameter value itself, that is if

$$\log(LOS) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + + \beta_k X_k + \varepsilon ,$$

then

$$\begin{aligned} LOS &= e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + + \beta_k X_k + \varepsilon} \\ &= e^{\beta_0} e^{\beta_1 X_1} e^{\beta_2 X_2} ... e^{\beta_k X_k} e^{\varepsilon} \end{aligned}$$

Therefore, if all variables are kept constant, but X_1 increases by 1 unit, $\log(LOS)$ will increase by β_1 and LOS will increase by a factor of e^{β_1}

Referring back to Table 4.4, the variables in the model have been included in the table in descending order of parameter estimate value. Therefore, it is clear that the variables which contribute the most to log length of stay are PRECIP08 and DIALYSED1. The prefix PRECIP refers to the precipitating factor. This is simply the factor that precipitated the primary system failure or insufficiency. A PRECIP value of 08 refers to those who have neuromuscular failure. The other important variable, DIALYSED1 refers to patients who require Kidney Dialysis in the Critical Care Unit.

Conversely, the variable TISS064 contributes negatively to log length of stay (thus reducing the predicted length of stay of a patient). This TISS score refers to patients who have suffered a cardiac arrest or been defibrillated within the past

24 hours. This seems counter-intuitive; however, on closer inspection; over 70% of non-elective patients who have TISS064 on the first day of admission onto the CCU do not survive their stay in the CCU. Since the parameter estimate for this variable is negative, this implies that the patients who experience arrest/defibrillation are likely to die quickly after arriving at the CCU.

The R^2 value for this model is 0.3614. This suggests that approximately 36% of the variation in the log of the original data is accounted for by the model. Ideally, a higher R^2 value would be helpful but even with this value it is possible to get some information from the model.

The final model can then be written as:

$$\begin{aligned}\log(LOS) = {} & 3.8 + 1.2(\text{PRECIP08}) + 1.1(\text{DIALYSED1}) + 0.7(\text{PRECIP29}) \\ & + 0.6(\text{SDCODE12}) + 0.5(\text{VENT1}) + 0.5(\text{SDCODE18}) + 0.5(\text{TISS071}) \\ & + 0.5(\text{SDCODE29}) + 0.4(\text{PRECIP20}) + 0.4(\text{PHSCORE3}) + 0.4(\text{SDCODE7}) \\ & + 0.3(\text{PHSCORE1}) + 0.3(\text{TISS039}) + 0.3(\text{EYESCORE3}) + 0.3(\text{PHSCORE2}) \\ & + 0.2(\text{TEMPSCORE3}) + 0.2(\text{RESPSCORE3}) + 0.2(\text{MBPSCORE2}) + 0.1(\text{TISS063}) \\ & + 0.1(\text{RESPSCORE1}) + 0.1(\text{TEMPSCORE1}) + 0.1(\text{HCO3}) + 0.1(\text{VSYSBP}) \\ & - 0.1(\text{ORGSCORE}) - 0.2(\text{ALBUMIN}) - 0.2(\text{PO}) - 0.3(\text{PRECIP24}) \\ & - 0.3(\text{SDCODE14}) - 0.3(\text{SDCODE27}) - 0.4(\text{SDCODE19}) - 0.4(\text{PRECIP05}) \\ & - 0.5(\text{PRECIP06}) - 0.5(\text{TISS065}) - 0.5(\text{TISS068}) - 0.6(\text{TPNN}) \\ & - 0.6(\text{SDCODE41}) - 0.7(\text{HAEMCAN}) - 0.7(\text{SDCODE22}) - 1.1(\text{TISS064})\end{aligned}$$

Therefore

$$LOS = \frac{e^{\begin{array}{l}3.8+1.2(\text{PRECIP08})+1.1(\text{DIALYSED1})+0.7(\text{PRECIP29}) \\ +0.6(\text{SDCODE12})+0.5(\text{VENT1})+0.5(\text{SDCODE18})+0.5(\text{TISS071}) \\ +0.5(\text{SDCODE29})+0.4(\text{PRECIP20})+0.4(\text{PHSCORE3})+0.4(\text{SDCODE7}) \\ +0.3(\text{PHSCORE1})+0.3(\text{TISS039})+0.3(\text{EYESCORE3})+0.3(\text{PHSCORE2}) \\ +0.2(\text{TEMPSCORE3})+0.2(\text{RESPSCORE3})+0.2(\text{MBPSCORE2})+0.1(\text{TISS063}) \\ +0.1(\text{RESPSCORE1})+0.1(\text{TEMPSCORE1})+0.1(\text{HCO3})+0.1(\text{VSYSBP})\end{array}}}{e^{\begin{array}{l}0.1(\text{ORGSCORE})+0.2(\text{ALBUMIN})+0.2(\text{PO})+0.3(\text{PRECIP24}) \\ +0.3(\text{SDCODE14})+0.3(\text{SDCODE27})+0.4(\text{SDCODE19})+0.4(\text{PRECIP05}) \\ +0.5(\text{PRECIP06})+0.5(\text{TISS065})+0.5(\text{TISS068})+0.6(\text{TPNN}) \\ +0.6(\text{SDCODE41})+0.7(\text{HAEMCAN})+0.7(\text{SDCODE22})+1.1(\text{TISS064})\end{array}}}$$

The above expression for length of stay, as was previously mentioned, has an R^2 value of 0.3614. This is achieved by including 39 variables. From a practical perspective, this large number of variables in the regression equation could prove impractical for clinicians to use. Also, at this stage, it is important to acknowledge the dangers of overfitting. It may be that some of the variables in this model are included to describe noise rather than a true relationship between the variables.

Therefore, an alternative model was constructed with a maximum of 12 variables, and these variables were selected based upon their significance in the previous model. The parameter estimates are indicated in Table 4.5.

Table 4.5: Parameter estimates for the reduced variable model

Variable	Parameter Estimate	Standard Error	*t* Value	Pr > |*t*|
INTERCEPT	4.04132	0.09731	41.53	<.0001
DIALYSED1	1.03821	0.09753	10.64	<.0001
VENT1	0.58408	0.08124	7.19	<.0001
PHSCORE3	0.25911	0.06906	3.75	0.0002
SDCODE7	0.37488	0.07143	5.25	<.0001
PHSCORE2	0.20434	0.06222	3.28	0.0010
MBPSCORE2	0.17002	0.05149	3.30	0.0010
ALBUMIN	-0.13744	0.02720	-5.05	<.0001
PO	-0.25245	0.03413	-7.40	<.0001
TISS068	-0.66345	0.12268	-5.41	<.0001
TPNN	-0.70106	0.05820	-12.05	<.0001
SDCODE22	-0.59908	0.10081	-5.94	<.0001
TISS064	-1.33234	0.16501	-8.07	<.0001

The above model yields an R^2 value of 0.2955. Therefore, the reduction in variables causes a decrease in R^2, or a decrease in the proportion of variation accounted for by the regression model. Thus, for the sake of this Chapter, the first model, using all 39 variables will be compared to the CART analysis.

4.4. Comparison between CART analysis and Regression analysis

Given that both the CART and regression analyses are valid methods for predicting factors which are likely to affect LoS, it was considered important to compare the outcomes of these analyses.

All variables from the CART analysis are included in the regression analysis apart from one, NVVERBAL. NVVERBAL is the verbal response in intubated patients. Several variables which appear in the linear regression do not appear in the CART analysis. Many of these are all physiological scoring variables such as temperature, respiratory rate and blood pressure. The other variable included is HAEMCAN which is a binary variable referring to whether the patient has a history of haematological cancer.

It emerges that there are 11 variables which are influential in both analyses. These are:

- Albumin – concentration of Albumin in plasma
- Dialysed – was the patient dialysed?
- HCO3 – Plasma bicarbonate
- ORGSCORE – Knauss's organ failure score
- PO – $Pa0_2$ – measurement of arterial partial pressure of oxygen
- PRECIP – Precipitating Factor
- SDCODE – Specific Diagnostic Code
- Source of admission
- TISS score – Daily TISS score
- TPN – Total Parenteral Nutrition
- Ventilated – was the patient ventilated?

Each of these variables will be considered in more detail in the following section.

4.4.1. Albumin

The variable Albumin, which is essentially the concentration of Albumin in plasma, appears twice on the tree diagram, once to create nodes 27 and 28 and once to create nodes 33 and 34. To create node 27, an Albumin level of less than 2.35 is required. If the Albumin level is greater than 2.35 then node 28 is created. The mean LoS in 27 is greater than in 28, thus an inverse relationship. A similar pattern is evident in nodes 33 and 34. This inverse relationship implies that patients with a lower albumin score have a longer length of stay. The regression parameter ($\beta = -0.2$) agrees with the conclusions drawn from the CART analysis.

4.4.2. Dialysed

The next variable which appears in both regression and CART analyses is kidney dialysis. Approximately ten percent of Critical Care admissions are dialysed at some point during their stay. The following table presents the summary statistics for length of stay of patients dialysed compared to those who are not dialysed:

Table 4.6: Summary statistics of length of stay for patients who were dialysed during their stay

Dialysed	Percentage	Mean	SD	Median
No	89.56%	109	301.64	45
Yes	10.44%	297.23	323.40	191

The length of stay distributions are significantly different ($p < 0.0001$) and Figure 4.4 highlights this difference:

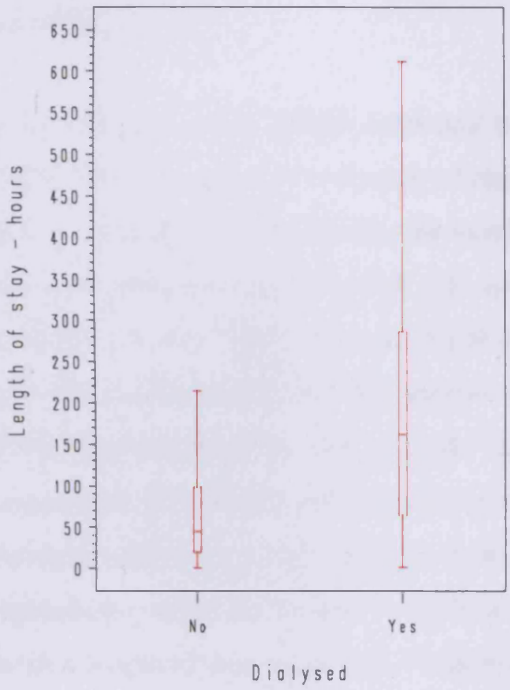


Figure 4.4: length of stay distribution of patients who are dialysed

The CART analysis tree indicates that nodes 31 and 32 are created in order to differentiate between the patients who are dialysed compared to those who are not. It is clear that node 31 has a lower mean length of stay, thus emphasising that if a patient is dialysed in the CCU then expected length of stay is longer.

The regression parameter of 1.04 strengthens this argument.

4.4.3. HCO3

The plasma bicarbonate variable (HCO3) has a positive parameter in the regression and thus has a positive contribution to length of stay; the higher the HCO3 level, the longer the length of stay. Also, looking at the CART analysis tree it is clear to see that the construction of nodes 7, 8, 37 and 38 rely upon the value of HCO3 registered. Nodes 7 and 37 have a lower mean length of stay than nodes 37 and 38 respectively, thus confirming the findings from the regression analysis.

4.4.4. ORGSCORE

This variable refers to the Knauss's organ failure score and is a measurement of organ failure. To calculate this daily score, the number of organs failed on the day in question is counted, and is then multiplied by the number of days that these organs have failed for. In this analysis, since all data considered is on day 1 alone, Knauss's organ failure score is simply the number of organ's failed on day 1. In the CART analysis tree diagram, ORGSCORE appears once, to create nodes 9 and 10. Node 9 is reserved for patients with fewer than 10.1 organ failures on day 1, whereas node 10 includes patients with more than 10.1 organ failures on day 1. Patients in node 9 have a shorter length of stay than patients in node 10. The linear regression coefficient for this variable is -0.0997. The negative sign indicates that length of stay is inversely linked to the number of organ failures. Note that for the linear regression, ORGSCORE was standardised to have a mean of zero and a standard deviation of 1. Therefore, patients with an ORGSCORE larger than the average will have a shorter length of stay than patients with an ORGSCORE smaller than the average.

4.4.5. $Pa0_2$

$Pa0_2$ (PO) is a measurement of arterial partial pressure of oxygen. Further investigation of both CART tree and regression parameters indicates that $Pa0_2$ is inversely related to length of stay. For example, when looking at the decision used to create nodes 13 and 14, patients with $Pa0_2$ levels less than 81.4 go to node 13, whereas patients with $Pa0_2$ values higher than this cut-off go to node 14. Node 14 has the shortest mean length of stay. The regression parameter being negative ($\beta = -0.2$) also confirms this to be the case.

4.4.6. Precipitating factor

The precipitating factor is simply the factor that precipitated the primary system failure or insufficiency. Figure 4.5 illustrates the length of stay distribution for each of the factors:

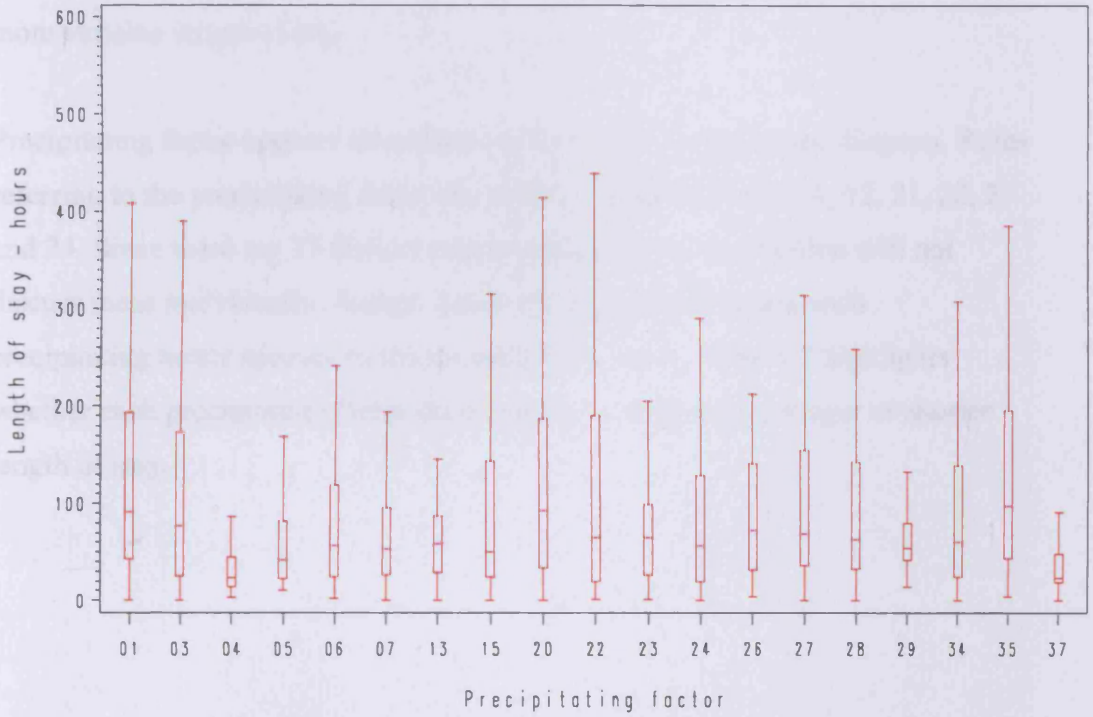


Figure 4.5: Length of stay distribution by precipitating factor

LoS is very much influenced by the specific nature of the factors. For example, precipitating factors 04 (Self-intoxication) and 37 (planned post-operative monitoring) have a fairly consistent length of stay profile without long whiskers. By contrast factors 01 (infection), 03 (trauma), 20 (septic shock) and 22 (haemorrhagic shock) have a much more varied length of stay profile.

After consultation with the clinicians at the CCU, they confirmed that this pattern in length of stay distribution was to be expected. For example, patients with a precipitating factor of 37 tend to be elective surgery patients who are sent to the

CCU for interventions such as ventilation but are then discharged to the ward quickly provided that there are no complications. Also patients with precipitating factor 04 tend to undergo a rapid cleansing procedure and are then very quickly fully recovered and able to be discharged to a ward.

Similarly, it is not surprising that patients with the precipitating factors 01, 03, 20 and 22, which are critical illnesses and notoriously difficult to control, have a more variable length of stay.

Precipitating factor appears three times in the CART analysis tree diagram. Rules referring to the precipitating factor are used to construct nodes 11, 12, 21, 22, 23 and 24. Since there are 37 distinct precipitating factors, this section will not discuss these individually. Rather, Table 4.7 summarises where each precipitating factor appears in the splitting rules. Also, Table 4.7 highlights whether each precipitating factor contributes to a node with a longer or shorter length of stay.

Table 4.7: the number of times the precipitating factor contributes to nodes with short or long LoS

PRECIP	DEFINITION	NO. OF TIMES THE FACTOR CONTRIBUTES TO SHORT LOS	NO. OF TIMES THE FACTOR CONTRIBUTES TO LONG LOS
4	Self-intoxication (overdose)	3	0
5	Intracerebral haemorrhage	3	0
27	Coma/mental derangement (metabolic)	3	0
10	Myocardial infarction (documented)	2	0
12	Peripheral vascular disease	2	0
13	Embolus (localised)	2	0
29	Diabetic ketoacidosis	2	0
32	Haematologic insufficiency/crisis	2	0
2	Neoplasm	1	0
8	Neuromuscular failure	1	0
9	Coronary artery disease	1	0
16	Hypertension	1	0
25	Allergic reaction	1	0
36	Toxic/chemical poisoning	1	0
1	Infection	2	1
6	Extracerebral (Subdural/arachnoid haemorrhage)	2	1
7	Seizures	2	1
15	Congestive heart failure/pulmonary edema	2	1
22	Haemorrhagic/hypovolaemic shock	2	1
23	Bleeding (significant but not shocked)	2	1
24	Post arrest (cardiac and/or respiratory)	2	1
26	Obstruction/perforation	2	1
34	Unplanned post-op ventilation	2	1
37	Planned post operative monitoring	2	1
35	Acute-on-chronic end stage disease	1	2
19	Cardiogenic	1	1

	shock/myocardiopathy		
17	Rhythm disturbance	0	1
30	Endocrine emergency	0	1
28	Electrolyte/Acid-base disturbance	0	2
3	Trauma	0	3
20	Septic shock/sepsis	0	3

The remaining precipitating factors were not used in the formation of any nodes. Comparing the green shaded section of the table, i.e. the factors which create nodes with a short length of stay, with the regression analysis variables shows that the precipitating factor 05 creates nodes with short lengths of stay, 3 times, and appears with a negative regression coefficient (-0.40645). Looking at the red shaded section of the table, a precipitating factor of 20 creates nodes with a long length of stay, in each of the three instances which it appears and has a positive regression coefficient (0.39159). Precipitating factors 24 and 06 appear in the yellow shaded section of Table 4.7 as they create nodes with a short length of stay twice and a long length of stay once. The regression coefficients for these nodes are -0.27334 and -0.45656 respectively. These regression coefficients imply that these precipitating factors contribute negatively to length of stay, but in one instance, the CART analysis implies that they contribute positively to length of stay (in the creation of nodes 12 and 24). This can be explained, since both of these nodes appear some way down the tree and thus do not include the entire data set in their creation.

Finally, precipitating factors of 08 and 29 appear to have positive regression coefficients (1.18453 and 0.72251 respectively) but create nodes with comparatively short lengths of stay. Again, the explanation is the same as for the previous precipitating factors.

4.4.7. SDCODE

The next variable which appears both in the CART analysis and the regression analysis is the specific diagnostic code. Figure 4.6 demonstrates the length of stay distribution for the different specific diagnostic codes. The definitions corresponding to these codes can be found in Appendix 4.2 and 4.3.

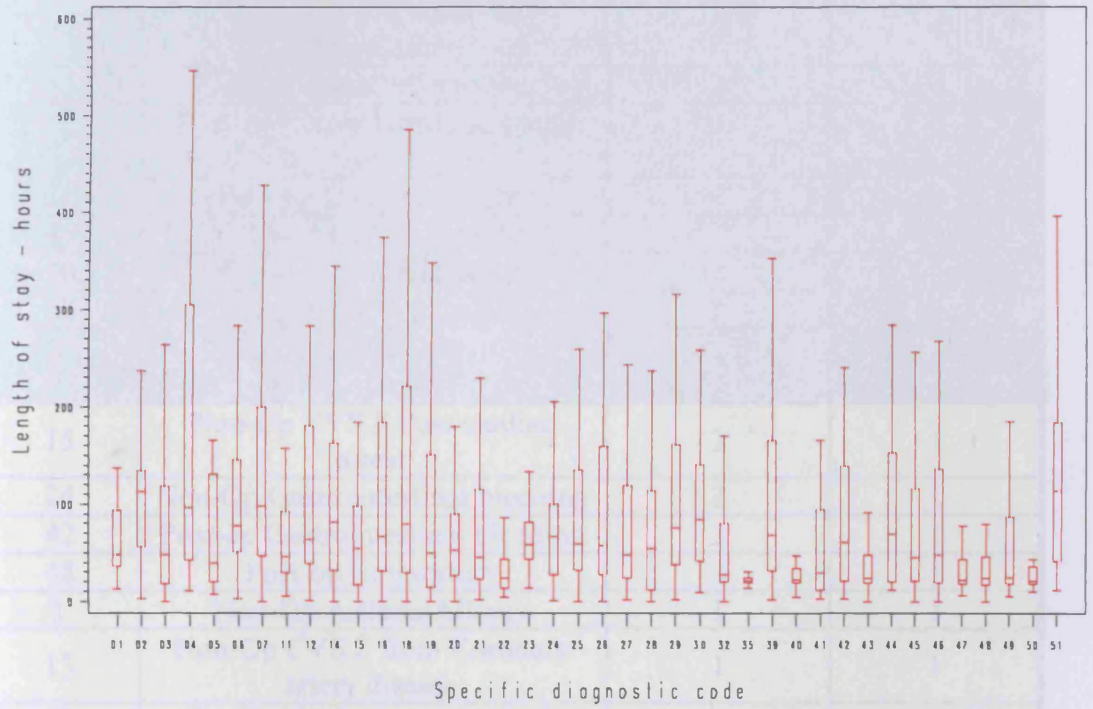


Figure 4.6: length of stay distribution by Specific Diagnostic code

There are a number of points to note here. Firstly, codes 35 (Post-op Renal surgery for Neoplasm) and 50 (Post-op Metabolic/Renal) have the least varied length of stay distributions. By contrast the most varied distribution corresponds to code 04, which is Non-Op Post respiratory arrest.

Table 4.8: the number of times the specific diagnostic code contributes to nodes with short or long LoS

SDCODE	Definition	No. of times the code contributes to Short LoS	No. of times the code contributes to Long LoS
16	Non-Op CVS F Cardiogenic shock	3	0
19	Non-Op Head Trauma	3	0
22	Non-Op Drug Overdose	3	0
25	Non-Op Metabolic/Renal	3	0
34	Post-op Craniotomy for Neoplasm	3	0
41	Post-op Haemorrhagic shock	3	0
31	Post-op Chronic cardiovascular disease	2	0
36	Post-op Renal transplant	2	0
47	Post-op Cardiovascular	2	0
50	Post-op Metabolic/Renal	2	0
52	Post-op Cardiac Arrest	2	0
9	Non-Op CVS failure from Hypertension	1	0
15	Non-Op CVS F Post cardiac arrest	2	1
24	Non-Op Gastrointestinal bleeding	2	1
42	Post-op Gastrointestinal bleeding	2	1
48	Post-op Respiratory	2	1
1	Non-Op Asthma/Allergy	1	1
13	Non-Op CVS F from Coronary artery disease	1	1
23	Non-Op Diabetic ketoacidosis	1	1
44	Post-op Respiratory insufficiency	1	1
46	Post-op Neurologic	1	1
20	Non-Op Neurologic Seizure disorder	1	2
21	Non-Op Neurologic ICH/SDH/SAH	1	2
28	Non-Op Cardiovascular	1	2
30	Post-op Multiple Trauma	1	2
43	Post-op GI surgery for Neoplasm	1	2
49	Post-op Gastrointestinal	1	2
2	Non-Op Chronic Obstructive Pulmonary Disease	0	1
3	Non-Op Pulmonary oedema (non cardiogenic)	0	1

4	Non-Op Post respiratory arrest	0	1
5	Non-Op Aspiration/poisoning/toxic	0	1
6	Non-Op Pulmonary embolus	0	1
7	Non-Op Respiratory failure from Infection	0	1
8	Non-Op Respiratory failure from Neoplasm	0	1
10	Non-Op CVS F from Rhythm disturbance	0	1
11	Non-Op Congestive heart failure	0	1
12	Non-Op CVS F Haemorrhagic/hypovolaemic shock	0	1
14	Non-Op CVS F from Sepsis	0	1
18	Non-Op Multiple Trauma	0	1
26	Non-Op Respiratory	0	1
27	Non-Op Neurologic	0	1
29	Non-Op Gastrointestinal	0	1
32	Post-op Peripheral vascular surgery	0	1
37	Post-op Head Trauma	0	1
38	Post-op Thoracic surgery for Neoplasm	0	1
39	Post-op Craniotomy for ICH/SDH/SAH	0	1
45	Post-op GI Perforation/Obstruction	0	1
51	Post-op Sepsis	0	1

Again comparing the CART results in Table 4.8 with the regression coefficients indicates a level of consistency between the two sets of analyses. For example, in regression and CART analysis, the specific diagnostic codes 7, 12, 18 and 29 all produce long lengths of stay. Whereas, the specific diagnostic codes 19, 22 and 41 always produce short lengths of stay. The only discrepancy occurs with the specific diagnostic codes of 14 and 27. In the CART analysis, they appear once and aid in the construction of nodes with long lengths of stay, whereas they have a negative regression coefficient (-0.292 and -0.3083 respectively).

4.4.8. Source of admission

This variable categorises patients from their source of arrival onto the critical care unit and takes the following values: Accident and Emergency, the Wards, X-Ray, Elective Surgery, Emergency Surgery and Other Hospitals. Note the numbers in the Table 4.9 differ from those found in previous tables since the source of admission data was complete for almost all admissions.

Table 4.9: Summary statistics for length of stay in hours by source of admission

Source of admission	N	Mean	Std Dev	Minimum	Median	Maximum
A+E	1134	123.59	304.40	0	56	7971
Elective	1182	57.34	99.78	0	23	1318
Emergency	715	153.45	396.48	1	65	8973
Other Hosp	235	210.13	431.48	2	88	4746
Ward	916	189.63	362.80	0	92	7378
X-Ray	44	91.27	124.33	1	49	622

The Kruskal-Wallis test was performed to ascertain whether there were significant differences between the distributions of the above groups. It emerged that there were some significant differences between the groups (see Table 4.10).

Table 4.10: Significant differences in length of stay, between sources of admission

	A+E	ELECTIVE	EMERGENCY	OTHER HOSP	WARD	X-RAY
A+E	---------	$p < 0.05$	$p < 0.05$	$p < 0.05$	$p < 0.05$	$p = 0.6820$
ELECTIVE	$p < 0.05$	---------	$p < 0.05$	$p < 0.05$	$p < 0.05$	$p < 0.05$
EMERGENCY	$p < 0.05$	$p < 0.05$	---------	$p < 0.05$	$p < 0.05$	$p = 0.2052$
OTHER HOSP	$p < 0.05$	$p < 0.05$	$p < 0.05$	---------	$p = 0.6378$	$p < 0.05$
WARD	$p < 0.05$	$p < 0.05$	$p < 0.05$	$p = 0.6378$	---------	$p < 0.05$
X-RAY	$p = 0.6820$	$p < 0.05$	$p = 0.2052$	$p < 0.05$	$p < 0.05$	---------

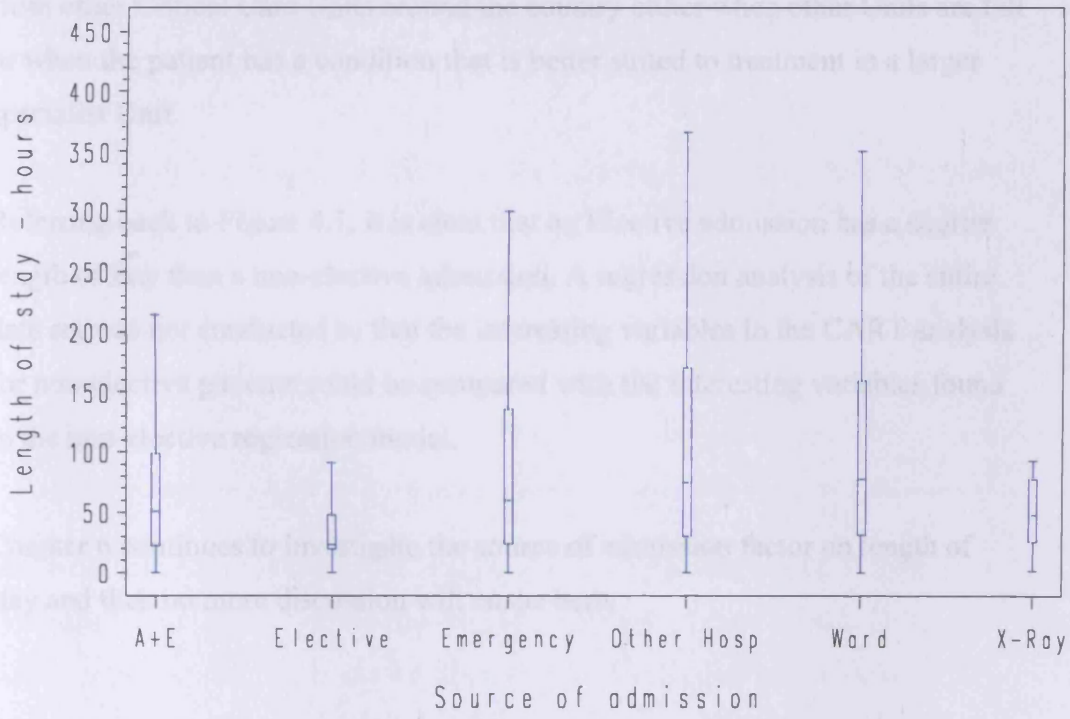


Figure 4.7: Box-plot showing length of stay by source of admission

Figure 4.7 shows the extent of the variation in the length of stay of patients. Patients arriving from Elective Surgery have a small amount of variation between their lengths of stay. The patients from Accident and Emergency have a moderate variation but less that those arriving from Emergency Surgery, Other

Hospitals or the Ward. There are too few observations in the X-Ray category to conclude anything from these. Figure 4.7 seems intuitively correct since, if patients arrive from Elective Surgery, generally it is because they need ventilation or other forms of life support while they recover and are thus present in the CCU for a short period of time until the life-support interventions are not required any longer. Also, patients from Emergency Surgery have been rushed into Hospital and are probably very sick thus needing to spend a long time in the CCU. The same logic applies to patients who have arrived from the Wards; their condition will have deteriorated resulting in their need for Critical Care. It is also intuitive that patients from Other Hospitals have a varied length of stay. Since the Critical Care Unit at the University Hospital of Wales is the largest Unit in Wales and also treats many serious cases each year, very often patients are transferred from other Critical Care Units around the country either when other Units are full or when the patient has a condition that is better suited to treatment in a larger specialist Unit.

Referring back to Figure 4.1, it is clear that an Elective admission has a shorter length of stay than a non-elective admission. A regression analysis of the entire data set was not conducted so that the interesting variables in the CART analysis for non-elective patients could be compared with the interesting variables found in the non-elective regression model.

Chapter 6 continues to investigate the source of admission factor on length of stay and thus no more discussion will ensue here.

4.4.9. Therapeutic Intervention Scoring System (TISS score)

In 1974 Cullen and colleagues (Cullen et al. 1974) developed the Therapeutic Intervention Scoring System (TISS score). This scoring system sought to compare levels of patient care within an Intensive Care Unit. This TISS score is based on 57 interventions that a patient could experience in the Intensive Care

Unit. In 1983, Keene and Cullen (Keene and Cullen 1983) updated this scoring system to include new techniques, thus increasing the number of interventions to 76. Guidelines for the use of the TISS score suggest that data should be collected at the same time each day and by the same observer. A TISS item should be checked if the intervention has been administered within the previous twenty four hours. The therapies are sub-classified into 4 categories and each category has a score 1 to 4 associated with it. The more severe or invasive the therapy, the higher the category score. For example, ECG monitoring has a score of 1 whereas intracranial pressure monitoring has a score of 4. The number of interventions checked in each category should be counted and the weighted sum should be computed. This final weighted sum, known as the TISS score, can then be used to classify a patient according to the level of care required. The following table summarises the nurse provision required for patients with various TISS scores:

Table 4.11: Nurse Provision required for a patient according to TISS score

CLASS	NURSE: PATIENT	TISS SCORE
1	1:4 or greater	<10
2	1 experienced nurse, 1 nursing aide : 4 patients	10-19
3	1:2	20-39
4	1:1	≥ 40

In the Critical Care Unit at UHW, they classify patients according to their professional judgement rather than strictly applying the above rules. There are eleven interventions that cause a patient to have a 1:1 nurse ratio and they are listed in Table 4.12 below along with the score associated with them:

Table 4.12: Therapies making 1:1 nursing care a requirement

TISS code	Definition	Score
064	Arrest/defibrillation (24h)	4
046	Cardioversion	3
049	Cardiac output measurement	3
073	Intra-aortic balloon pump	4
068	CAVHD/CVVHD	2/4
071	Intracranial pressure monitor	4
065	Pulmonary artery (SG) catheter	4
063	Controlled ventilation CMV	4
038	IMV or assisted ventilation	3
039	CPAP	3
077	>1 vasoactive infusions	4

Note CAVHD/CVVHD has a score of 2 or 4 depending upon whether the patient is in a stable condition or not respectively.

TISS point appears once on the CART analysis tree diagram (Figure 4.1). It simply states that if the TISS point on the day of admission is less than 56.5 then a shorter length of stay is expected. A Mann-Whitney test was conducted on the data using the grouping variable TISS point less than 56.5 or TISS point greater than 56.5.The result was significant ($p < 0.0001$) implying that there is a significant difference in length of stay depending upon TISS point.

TISS features 6 times in the regression model, but not as a total score. Rather, it appears as 6 distinct variables which reflect the therapies which require 1:1 nursing care. Table 4.13 contains summary statistics for the length of stay of patients who have received the intervention compared to those who have not.

Also, a Mann-Whitney test was undertaken to see whether this difference was significant at the 95% level, and the final column contains this result:

Table 4.13: Summary statistics for length of stay by TISS code

Variable	Regression Parameter	Intervention received?	N	Mean (hrs)	Std	Median	Sig. diff?
TISS071	------	No	1982	191.34	426.84	76	Yes
	0.46907	Yes	103	173.34	150.71	127	
TISS039	------	No	2011	187.77	408.14	77	No
	0.30995	Yes	79	258.34	608.73	90	
TISS063	------	No	922	162.09	382.86	64	Yes
	0.12208	Yes	1168	212.84	441.84	95	
TISS065	------	No	2026	191.18	422.14	78.5	No
	-0.47285	Yes	59	165.12	202.28	65	
TISS068	------	No	1878	187.08	430.92	76	Yes
	-0.47809	Yes	207	221.00	265.54	129	
TISS064	------	No	2034	190.98	417.40	80	Yes
	-1.08599	Yes	56	171.21	424.78	12	

The regression parameter column has been ordered by the size of the parameter. The positive regression coefficients should result in the patient group having a longer length of stay. For TISS071 (Intracranial pressure monitor), the mean length of stay for patients who receive the intervention is shorter than those who do not, whereas the median length of stay for the intervention group is much longer. The difference in median is significant at the 95% level. The effect is different for the TISS039 (CPAP) intervention; both mean and median for the intervention group are longer than the non-intervention group. The difference in median is not significant at the 95% level. Considering the TISS063 (Controlled ventilation CMV) intervention, again, both mean and median length of stay for the intervention group are higher than the non-intervention group (the difference in median is significant at the 95% level).

The negative regression coefficients should result in the intervention group having a shorter length of stay. This is true in two cases, although not for the CAVHD/CVVHD intervention (TISS068). The difference is not significant for the Pulmonary artery (SG) catheter intervention (TISS065).

4.4.10. TPN (Total Parenteral Nutrition)

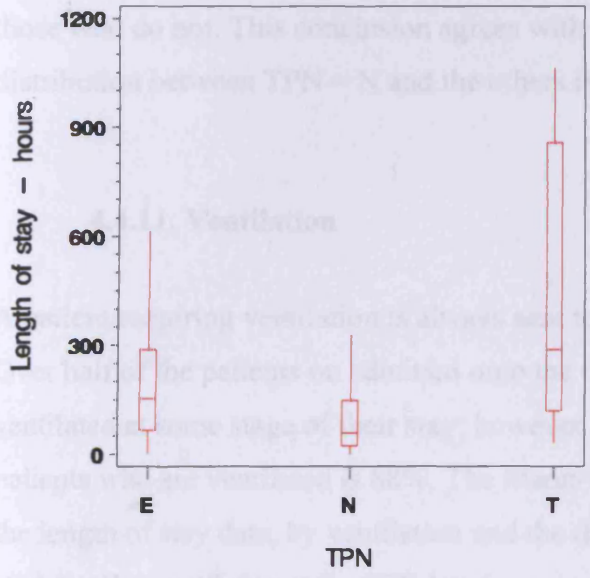


Figure 4.8: Length of stay distribution by TPN

TPN is simply a medical term for intravenous feeding. The value T indicates that the patient is receiving Total Parenteral Nutrition. This type of intervention can cause dangerous side effects and is thus only used when necessary. The value E stands for enteral nutrition, which is simply when a patient is fed directly through a tube into the stomach. N stands for no feeding and this is generally not a long term option. If a patient arrives and received no feeding initially, they will progress onto E or T when it becomes apparent that they will remain in the unit for more than a few days. Patients who receive total parenteral nutrition are usually fairly sick and have gastrointestinal problems which tend to result in many operations and hence long lengths of stay.

TPN appears once in Figure 4.1 and rules concerning it are used to create nodes 35 and 36. If TPN is N or T then the patient is categorised as node 35, else if TPN in E the patient is classified as node 36. Node 35 has a significantly shorter length of stay than node 36 ($p < 0.0001$).

The linear regression model features TPN once, and more specifically where TPN = N (the name of the variable is TPNN). The regression coefficient is -0.60 which implies that patients with a TPN of N have a shorter length of stay than those who do not. This conclusion agrees with Figure 4.8 since the difference in distribution between TPN = N and the others is significant at the 95% level.

4.4.11. Ventilation

A patient requiring ventilation is always sent to the CCU for this intervention. Over half of the patients on admitted onto the Critical Care Unit will be ventilated at some stage of their stay; however the percentage of non-elective patients who are ventilated is 88%. The Mann-Whitney test was performed on the length of stay data, by ventilation and the difference was found to be statistically significant at the 95% level.

Table 4.14: Summary statistics of length of stay for patients who were ventilated during their stay – non-elective patients

Ventilated	Percentage	Mean	SD	Median
No	11.77	58.29	28.31	39
Yes	88.23	208.04	440.97	89

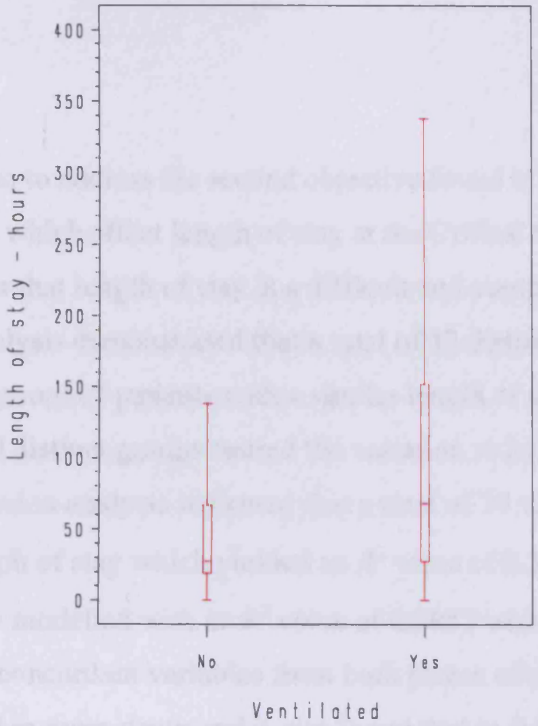


Figure 4.9: length of stay distribution of patients who are ventilated

Figure 4.9 indicates that as well as having a vastly different median length of stay the ventilated patients also have a far more variable length of stay.

Ventilation appears in Figure 4.1 once to create nodes 19 (if the patient is not ventilated) and 20 (if the patient is ventilated). Node 19 has a significantly shorter length of stay than node 20 ($p<0.0001$). This agrees with the evidence found in the above graph.

Finally, considering the regression coefficient of 0.53383, it is clear that ventilation causes an increase in length of stay.

4.5. Conclusion

This Chapter has begun to address the second objective found in section 1.2, to understand the factors which affect length of stay at the Critical Care Unit. It is clear from this Chapter that length of stay is a difficult and complex quantity to model. The CART analysis demonstrated that a total of 12 distinct variables were required to create 20 groups of patients with a similar length of stay profile. Treating the data as 20 distinct groups caused the variation to be reduced by 12%. The linear regression analysis indicated that a total of 39 variables were required to model length of stay which yielded an R^2 value of 0.3614. However, length of stay could be modelled with an R^2 value of 0.2955 with only 12 variables. Finally, the concordant variables from both pieces of statistical analysis were analysed in more depth and it was found that in the majority of cases the CART analysis and the linear regression analysis gave similar results.

The following table indicates the direction of the relationship between the concordant variables and length of stay.

Table 4.15: Contribution of variables to length of stay

HIGH LOS	LOW LOS
Higher than average HCO3	Higher than average Albumin
Dialysed	Higher than average Po
Ventilated	TPN = N

The remaining variables, Precipitating factors, Specific diagnostic codes, TISS codes and ORGSCORE have differing effects, depending on the values taken.

The nodes constructed from the CART analysis could be utilised as Arrival sources for a simulation model. Alternatively, the regression model could be utilised, also in a simulation model, to determine a patient's length of stay in the Critical Care Unit.

Chapter 5: Mortality analysis

5.1. Introduction

In this Chapter, the findings of an investigation of the factors that affect the mortality of a patient in the Critical Care Unit are presented and thus address the final points in objective 2. A patient's stay in the Critical Care Unit can be traumatic for family and friends; therefore an indication of survival could serve the purpose of preparing the family for a possible death.

The analysis, similar to that considered in Chapter 4, takes several forms. Firstly, a CART analysis is undertaken and the variables which contribute to the splitting rules are highlighted. Secondly, a logistic regression analysis is undertaken which seeks to develop an equation to predict the mortality outcome of a patient. The significant variables in the logistic regression are also highlighted. Next, the variables of note from both analyses will be investigated more thoroughly using statistical methods such as the chi-square contingency tables, odds ratios and relative risks.

Finally, a novel use of logistic regression is considered. Five logistic regression equations were formulated to predict the probability that a patient would be present in the Unit on the next day, given their current physiological state. This tool would be very useful for clinicians and managers at the CCU, since it would allow for planning future admissions to the Unit.

5.2. CART analysis

In the context of continuous dependent variables, such as length of stay in the previous Chapter, the aim of CART analysis is to reduce the overall variation in a data set by creating homogeneous nodes. When the dependent variable is binary as in the case of "outcome", the theory of CART analysis seeks to reduce

the overall impurity in a data set by creating homogeneous nodes of data. To create homogeneous nodes in this case, a procedure called the Gini Index was used. Again, more information about this can be found in Harper & Evandro Jr. (2008).

The termination rules for this CART analysis were slightly different from the case in Chapter 4. The minimum number of observations in each node was reduced to 30 to attempt to capture as much detail from the model as possible. The learning sample that was used was 70% of the original data leaving 30% available for testing the validity of the tree. In addition only variables which are possible to generalise were used in the model. For example, variables referring to which consultant the patient was seen by were not included as these results would not be transferrable across hospitals and could be corrupted, for example, by a consultant retiring.

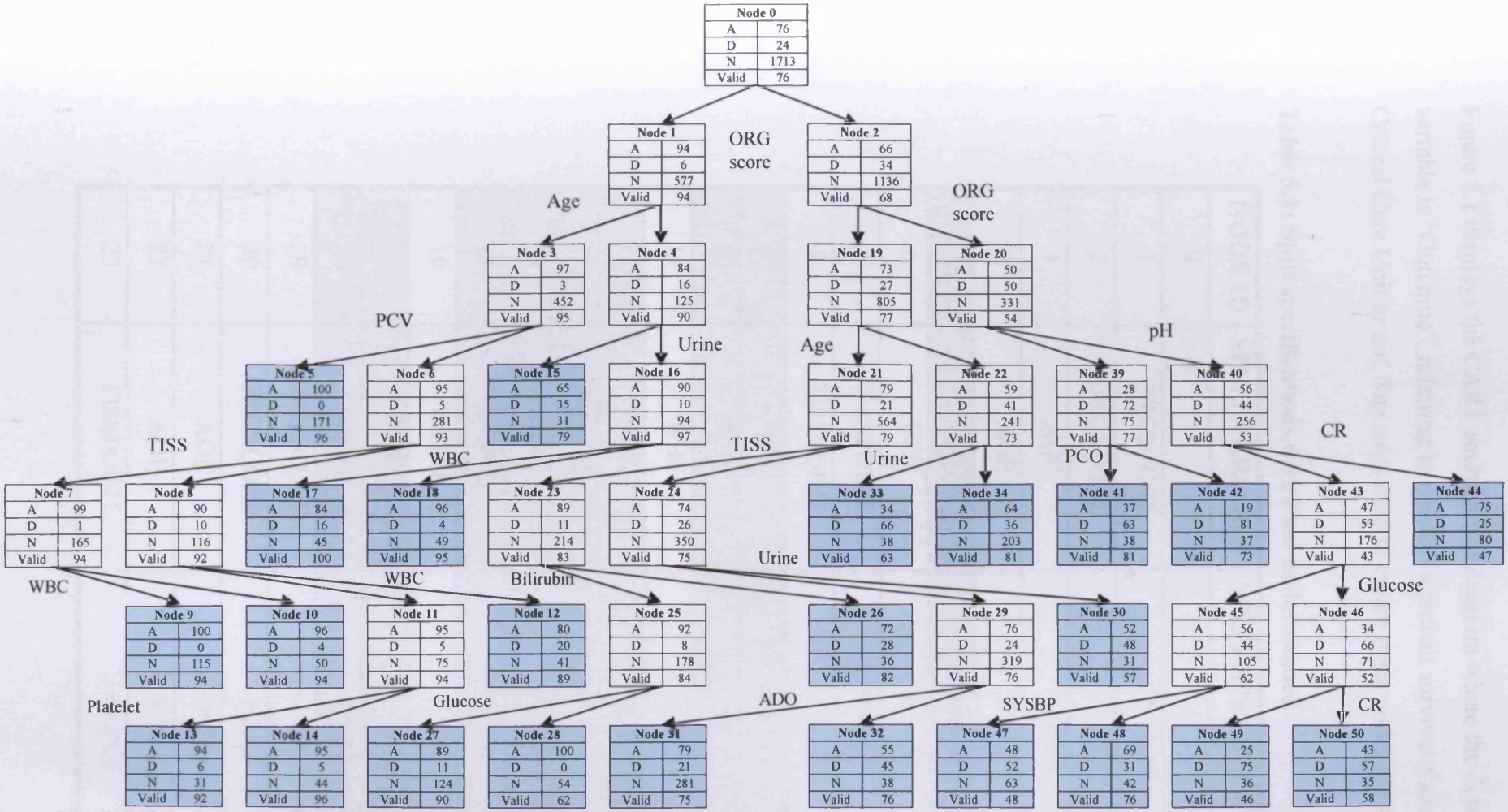


Figure 5.1: CART analysis tree diagram for Mortality – the row Valid refers to the percentage of cases correctly classified in the test sample

Figure 5.1 displays the CART analysis tree diagram where the dependent variable is "Outcome", referring to whether a patient survives their stay in the Critical Care Unit or not. The rules used to create these nodes are in Table 5.1.

Table 5.1: Split specifications with final nodes shaded

NODE ID	SPLIT VARIABLE	SPLIT SPECIFICATION
0		
1	ORGSCORE	<= 12.15
2	ORGSCORE	> 12.15
3	AGE	<= 62.5
4	AGE	> 62.5
5	PCV	<= 30.95
6	PCV	> 30.95
7	TISSPOINT	<= 40.5
8	TISSPOINT	> 40.5
9	WBC	<= 12.5
10	WBC	> 12.5
11	WBC	<= 12.25
12	WBC	> 12.25
13	PLATELET	<= 190.5
14	PLATELET	> 190.5
15	URINE	<= 1095
16	URINE	> 1095
17	WBC	<= 9.25
18	WBC	> 9.25
19	ORGSCORE	<= 23.3
20	ORGSCORE	> 23.3
21	AGE	<= 70.5
22	AGE	> 70.5
23	TISSPOINT	<= 43.5

24	TISSPOINT	> 43.5
25	BILIRUBIN	<= 1.19
26	BILIRUBIN	> 1.19
27	GLUCOSE	<= 176.88
28	GLUCOSE	> 176.88
29	URINE	<= 3735
30	URINE	> 3735
31	ADO	<= 331
32	ADO	> 331
33	URINE	<= 535
34	URINE	> 535
39	PH	<= 7.08
40	PH	> 7.08
41	PCO	<= 49.9
42	PCO	> 49.9
43	CR	<= 2.19
44	CR	> 2.19
45	GLUCOSE	<= 183.92
46	GLUCOSE	> 183.92
47	SYSBP	<= 144.5
48	SYSBP	> 144.5
49	CR	<= 1.12
50	CR	> 1.12

There are 24 final nodes present in the tree. These nodes reduce the overall impurity of the data by over 25%. The "Valid" row of each node represents the percentage from the test sample which were correctly classified as dead or alive using the splitting rules. Sixteen out of the 24 nodes correctly classified over 75% of the test sample. From the remaining 8 nodes, three classified less than 50% of the test sample accurately (nodes 44, 47 and 49).

Nodes 5, 9, 10, 13, 14, 17, 18 and 27 correctly classify at least 90% of the test sample. Node 5, which correctly classifies 96% of the test sample, consists of only those patients who survive their stay. Therefore, with 96% confidence, if a patient presents at the CCU with an ORGSCORE of less than or equal to 12.15, is younger than 62.5 years old and has a PCV level of no more than 30.95, they will survive their stay.

Very few nodes have a higher proportion dying than surviving (6 nodes in total). Also, the percentage of the test sample which are correctly classified in these cases is fairly low in comparison with the other nodes. For example, only 19% of patients in node 42 survive their stay in the CCU, but having said this, only 73% of the test sample were classified correctly. This implies that far more uncertainty is present when considering the factors which contribute to the death of a patient.

The most important variables when attempting to create homogeneous nodes of data are listed below. Note all variables were measured on day 1 of a patients stay.

- ORGSCORE – Knauss's organ failure score
- Age – the age of a patient
- Pcv – packed cell volume (volume of packed red cells in ml per 100ml of blood)
- Tisspoint – Therapeutic intervention scoring system (TISS) score
- Wbc – white blood cell count (healthy range 3,000 – 14,900 cu.mm)
- Platelet – platelet count
- Urine – urine level
- Bilirubin – a pigment produced when the liver processes waste products
- Glucose – the end product of carbohydrate metabolism
- Ado - Arterial-Alveolar oxygen difference

- pH – pH level
- pco – Arterial partial pressure of carbon dioxide (mmHg)
- cr – creatinine – a waste product of protein metabolism, found in urine (can be used as a measure of kidney failure) healthy range 0.6-1.4 mg/dl
- sysbp – systolic blood pressure

5.3. Logistic regression

Logistic regression is a useful tool to use when predicting the outcome of a binary response from a set of explanatory variables. These explanatory variables need not be continuous and many of the variables used in this analysis will be categorical in nature.

The logistic model in its simplest form is

$$\text{logit}(\pi) \equiv \ln\left(\frac{\pi}{1-\pi}\right) = \alpha + \sum_{i=1}^{s} \beta_i x_i \qquad (5.1)$$

where the β_i are the parameter estimates, α is the intercept and x_i are the explanatory variables. In this case, π corresponds to the probability of a patient dying in the CCU. There are no assumptions associated with logistic regression unlike linear regression, other than the explanatory variables should be independent. The logit can then be calculated from a given set of values for the explanatory variables, x_i.

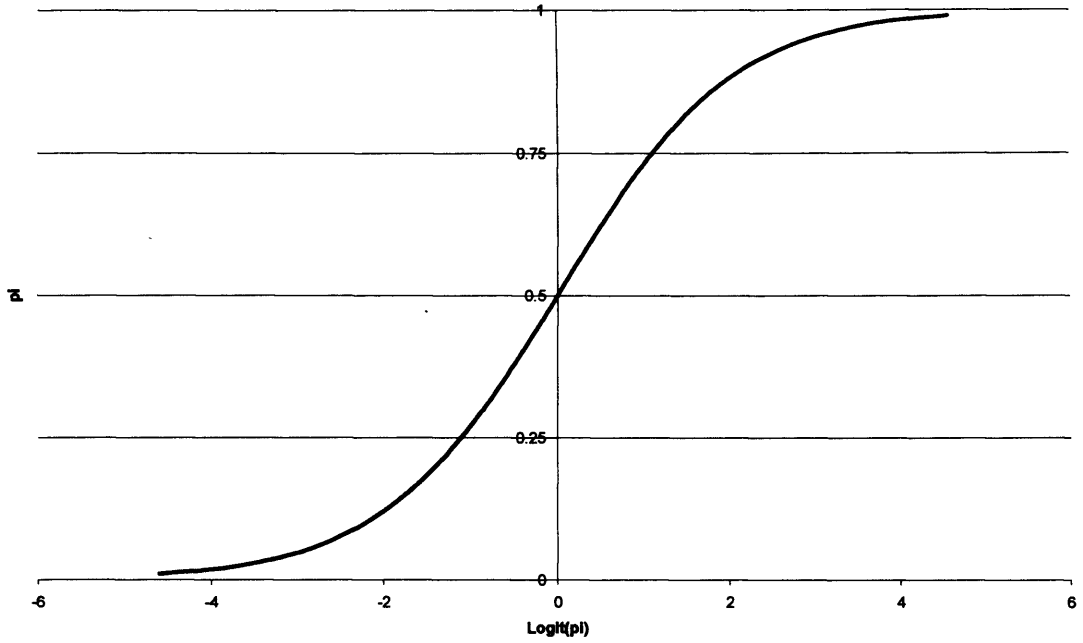


Figure 5.2: Logit(π) against π

Figure 5.2 illustrates that the value of logit(π) is strictly monotonic increasing as π varies between zero and one. It is clear that for values of π less than 0.5, logit is negative and for values of π greater than 0.5, logit is positive. Also $\lim_{\pi \to 1}(\text{logit}(\pi)) = \infty$ and $\lim_{\pi \to 0}(\text{logit}(\pi)) = -\infty$. Note that π = probability of a patient dying in the CCU.

Therefore, it is clear that any parameter which decreases the value of logit(π) will cause π to decrease and when logit(π) = 0 we have $\pi = 0.5$.

It is useful to consider the logistic equation in a slightly different way. The logit model can be modified to calculate the odds of dying by manipulating (5.1).

$$\frac{\pi}{1-\pi} = e^{\alpha + \sum_{i=1}^{s} \beta_i x_i} \tag{5.2}$$

Therefore, if all explanatory variables are kept constant apart from x_i, then for a unit increase in x_i, the odds of dying is β_i times higher.

All continuous explanatory variables have been standardised to have a mean of zero and a standard deviation of 1, and all categorical variables have a value of 0 (negative) or 1 (positive).

5.3.1. Analysis

The first step is to consider the Akaike Information Criterion (AIC), where $AIC = -2\log(L) + 2k(s+1)$. In this equation, L is the likelihood function, k is the number of levels of the dependent variable (2 in this model) and s is the number of predictors in the model. Note a stepwise regression procedure was used to develop the model and therefore s is not fixed at the beginning of the analysis. AIC is useful for comparing different models and, generally, the model with the lowest AIC is considered to be the best model. The logistic regression procedure in SAS calculates the AIC for a model in two ways: firstly using intercept alone and then using the other covariates. In this case the AIC is lower when the covariates are included; thus, the remainder of the output generated by SAS can be considered (2065.187 compared with 2743.778). The likelihood ratio test tests the hypothesis that $\boldsymbol{\beta} = \mathbf{0}$, where $\boldsymbol{\beta}$ is the vector of parameters. This hypothesis is rejected at the 95% level ($p < 0.0001$) and therefore indicates that the model parameters are worth considering.

Table 5.2 highlights the parameter estimates that are significant. The final column indicates the effect on the odds of dying. Note that the numbers in the final column represent the multiplying factor for the odds if all other variables are kept constant and the variable in the row is increased by 1 unit. See Appendix 4.1 for definitions of these variables.

Table 5.2: Parameter estimates for the logistic regression model to predict patient survival

Parameter	Estimate	Standard Error	Pr > ChiSq	Odds Multiplier
INTERCEPT	-2.7554	0.2498	<.0001	
TISS064	1.4285	0.3627	<.0001	4.17
HAEMCAN	1.3005	0.3836	0.0007	3.67
VENT1	1.2026	0.2611	<.0001	3.33
EYESCORE1	0.8782	0.183	<.0001	2.41
PRECIP26	0.712	0.3555	0.0452	2.04
TISS049	0.5905	0.1494	<.0001	1.80
AGE	0.5053	0.069	<.0001	1.66
ADMSOURCE4	0.4835	0.1997	0.0155	1.62
ORGSCORE	0.4577	0.078	<.0001	1.58
ADO	0.3191	0.0566	<.0001	1.38
TISS063	0.2583	0.124	0.0373	1.29
GLUCOSE	0.1506	0.0596	0.0115	1.16
PO	0.1175	0.0521	0.0242	1.12
SYSBP	-0.1309	0.0575	0.0229	0.88
HCO3	-0.2514	0.0564	<.0001	0.78
PRECIP34	-0.3539	0.157	0.0242	0.70
ADMSOURCE1	-0.4693	0.2176	0.031	0.63
SDCODE43	-1.1771	0.3648	0.0013	0.31
SDCODE20	-1.5318	0.5394	0.0045	0.22
PRECIP04	-2.0136	0.4666	<.0001	0.13
PRECIP27	-2.1891	0.9088	0.016	0.11

There are 21 variables included in the model. The odds ratio estimates highlighted in Table 5.2 quantify the effect of the independent variables on the dependent variable. They have been sorted by the value of the odds ratio estimate for ease of interpretation.

The logistic regression equation is therefore

$$\begin{aligned}\text{logit}(\pi) = &-2.8+1.4(TISS064)+1.3(HAEMCAN)+1.2(VENT1)+0.9(EYESCORE1)\\ &+0.7(PRECIP26)+0.6(TISS069)+0.5(AGE)+0.5(ADMSOURCE4)\\ &+0.5(ORGSCORE)+0.3(ADO)+0.3(TISS063)+0.2(GLUCOSE)+0.1(PO)\\ &-0.1(SYSBP)-0.3(HCO3)-0.4(PRECIP34)-0.5(ADMSOURCE1)\\ &-1.2(SDCODE43)-1.6(SDCODE20)-2.0(PRECIP04)-2.2(PRECIP28)\end{aligned}$$

Table 5.3 has the R^2 and the maximum rescaled R^2 value. The maximum rescaled R^2 is the best measure to use with logistic regression. The usual measure of variation explained by a model is the R^2 value. However, the maximum value of R^2 achieved by a discrete model is $1-(L(0))^{\frac{2}{n}}$ where *L(0)* is the likelihood of the intercept-only model and *n* is the sample size. For this model, the maximum R^2 achievable is 0.6671.

The maximum R^2 value takes the maximum achievable R^2 into account and scales the actual R^2 value using the following equation:

$$Max-rescaled\ R^2 = \frac{R^2}{Max(R^2)}, \text{ (Nagelkerke 1991)}$$

Table 5.3: R-Square and Max-Rescaled R-Square for the logistic regression model

R-Square	0.2511	**Max-rescaled R-Square**	0.3764

Another useful measure of the fit of the model is the percentage of concordant observations. More specifically, a pair of observations is concordant if the observation with the lower value has a lower predicted mean logit(π) score than the higher observed value. This value ranges from zero to 100, with 100 as the best possible model. The percentage of concordant pairs for this model is 83.8.

The Somers'D statistic simply determines the direction and strength of a relationship between pairs of observations. A value of 1 indicates that all pairs agree and a value of -1 indicates that all pairs disagree. Somers'D is equal to 0.677 for the above logistic regression model.

The logistic regression procedure also outputs c, the rank correlation of ordinal variables. It has a value between zero (no association) and 1 (perfect association). For the above regression model, c = 0.838.

Finally, the lack of fit statistic, Hosmer and Lemeshow simply identifies how good the fit is to the data. The above model displays no evidence of lack of fit (p = 0.1768).

The following classification table summarises the predictive power of the model.

Prob Level	Correct		Incorrect		Percentages				
	Event	Non-Event	Event	Non-Event	Correct	Sensitivity	Specificity	False POS	False NEG
0.0	596	0	1896	0	23.9	100.0	0.0	76.1	.
0.1	562	903	993	34	58.8	94.3	47.6	63.9	3.6
0.2	496	1287	609	100	71.5	83.2	67.9	55.1	7.2
0.3	404	1520	376	192	77.2	67.8	80.2	48.2	11.2
0.4	307	1658	238	289	78.9	51.5	87.4	43.7	14.8
0.5	229	1765	131	367	80.0	38.4	93.1	36.4	17.2
0.6	157	1824	72	439	79.5	26.3	96.2	31.4	19.4
0.7	105	1861	35	491	78.9	17.6	98.2	25.0	20.9
0.8	61	1879	17	535	77.8	10.2	99.1	21.8	22.2
0.9	28	1889	7	568	76.9	4.7	99.6	20.0	23.1
1.0	0	1896	0	596	76.1	0.0	100.0	.	23.9

For higher probability levels, the models predictive power is far better and correctly classifies a large proportion of the data. With lower probabilities, the predictive power is not as good and incorrectly classifies many patients.

5.3.2. Two example cases.

For the sake of illustration, two patients will be considered and the predicted probability of dying will be calculated. The measurements for the significant variables for the logistic regression are outlined in the Table 5.4 below:

Table 5.4: Two example patients with values for each significant variable

PARAMETER	ESTIMATE	PATIENT 1, x_i	PATIENT 2, x_i	$\beta_i x_i$ FOR PATIENT 1	$\beta_i x_i$ FOR PATIENT 2
ADMSOURCE1	-0.4693	0.00	0.00	0.00	0.00
ADMSOURCE4	0.4835	0.00	0.00	0.00	0.00
ADO	0.3191	-0.64	2.62	-0.21	0.84
AGE	0.5053	1.42	-0.34	0.72	-0.17
EYESCORE1	0.8782	0.00	1.00	0.00	0.88
GLUCOSE	0.1506	0.34	6.68	0.05	1.01
HAEMCAN	1.3005	0.00	0.00	0.00	0.00
HCO3	-0.2514	-0.70	-2.45	0.18	0.62
ORGSCORE	0.4577	-0.67	2.67	-0.31	1.22
PO	0.1175	-0.09	-0.44	-0.01	-0.05
PRECIP04	-2.0136	0.00	0.00	0.00	0.00
PRECIP26	0.712	0.00	0.00	0.00	0.00
PRECIP27	-2.1891	0.00	0.00	0.00	0.00
PRECIP34	-0.3539	0.00	0.00	0.00	0.00
SDCODE20	-1.5318	1.00	0.00	-1.53	0.00
SDCODE43	-1.1771	0.00	0.00	0.00	0.00
SYSBP	-0.1309	0.51	-1.59	-0.07	0.21
TISS049	0.5905	0.00	1.00	0.00	0.59
TISS063	0.2583	0.00	1.00	0.00	0.26
TISS064	1.4285	0.00	1.00	0.00	1.43
VENT1	1.2026	0.00	1.00	0.00	1.20

Addition of the intercept -2.7554 and rearranging, gives the probability of dying for patient 1 to be 0.01 and for patient 2 to be 0.97. Reference to the original data set shows that in fact patient 1 survives and patient 2 does not.

5.4. Comparison between CART analysis and Regression analysis

There are many differences in the variables proved significant from the CART analysis and the logistic regression analysis for the mortality variable. For example, specific diagnostic codes and Precipitating factors prove significant according to the regression analysis, whereas physiological variables such as Bilirubin and platelet counts prove significant in the CART analysis. Five variables appear in both the CART analysis splitting rules and the logistic regression implying that they can be judged as important influencers of mortality. They are as follows:

- ADO - Arterial-alveolar oxygen difference
- AGE – the age of a patient
- GLUCOSE – the end product of carbohydrate metabolism
- ORGSCORE – Knauss's organ failure score
- SYSBP – systolic blood pressure

This next section investigates these variables in more depth to ascertain the relationship between these variables and mortality.

5.4.1. ADO

The arterial-alveloar difference (ADO) is simply the difference between the partial pressure of oxygen in the alveolus and the mean arterial pressure of oxygen. The efficiency of gaseous exchange in the lungs is measured by this quantity. It is known that ADO can effect mortality in patients with acute health

conditions such as pulmonary embolism (Te et al. 2006). The ADO variable is a continuous variable measured on a scale from zero to 650. Table 5.5 below contains summary statistics for the ADO variable by outcome:

Table 5.5: Summary statistics for ADO by CCU outcome

Outcome	N	Mean	Std Dev	Minimum	Median	Maximum
Alive	1897	94.59	167.97	0	0	643
Dead	596	207.80	224.72	0	221	648

The patients who survived their stay in the CCU have a significantly shorter median ADO value compared to those who do not survive their stay. Note that the median value for the patients who survive their time in the CCU is zero, thus indicating the fact that over half the sample have zero values for ADO.

Before considering the CART analysis output, it is sensible to decide which measurements will be useful for analysing discrete data.

Chi-square contingency tables

Chi-square contingency tables are widely used to examine whether the frequency of a certain condition is different between several groups. The assumptions of chi-square tests are that the observations must be independent, each observation should only appear in the table once and the expected value of each cell should be greater than five. The chi-square test statistic is given by the following formula:

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

The software program SAS (version 9.1.3) will calculate the chi-square value for the data and produce the corresponding p-value. This p-value will indicate whether there the variables in question are independent

Odds ratios

Simply, the odds of an event is the ratio of the probability of the event occurring to the probability of the event not occurring. The odds ratio is then calculated by comparing the odds of the exposure group with the odds of the non-exposed group. The assumptions for odds ratios are the same as with the chi-square test. Using the following general table, the odds ratio is calculated as follows:

	Dead	**Alive**	**Total**
Exposure present	A	b	a + b
Exposure absent	C	d	c + d
Total	a + c	b + d	a + b + c + d

$$OR = \frac{\frac{a}{b}}{\frac{c}{d}}$$

To interpret this value, some thought is required. The odds ratio is the odds of a patient not surviving their stay in the Critical Care Unit given that they have been exposed to the disease/treatment.

Relative Risk (also known as mortality risk)

Next, the final categorical data analysis, relative risk, will be considered. Although very similar to odds ratio, relative risk is calculated in a different manner. Relative risk compares the conditional probability of the event occurring in the exposed and non-exposed groups. Again, the assumptions for relative risks are the same as with the chi-square test. The above table (defining events a, b, c and d) will be used for defining relative risk:

$$RR = \frac{\frac{a}{a+b}}{\frac{c}{c+d}}$$

The interpretation of relative risk is much simpler. A relative risk of X suggests that the mortality in the patients who has been exposed to the disease/treatment is X times as high as the mortality in patients who have not been exposed.

Chi-square vs. Odds ratio vs. Relative risk

The chi-square test, as previously mentioned, is limited in its use as it only indicates an association between variables rather than the degree of association present. Odds ratios are always greater than relative risks and therefore often over-estimate the true association between variables. Also the interpretation of relative risk is much simpler than that of odds ratios. Therefore, taking these factors into consideration, the relative risk (or mortality risk) will be considered in this Chapter.

Following the result of the splitting rules of the CART analysis a new variable was constructed, ADO_less_than_331. Table 5.6 below contains information regarding the mortality of patients with an ADO level less than and greater than 331:

Table 5.6: Frequency of ADO ≤ 331 and > 331 by CCU outcome

Frequency	Outcome		Total
	Alive	Dead	
ADO > 331	198	165	363
ADO ≤ 331	1699	431	2130
Total	1897	596	2493

The chi-square contingency test on the above table indicates that the variables are not independent ($p < 0.0001$). The mortality risk calculated for the above data is 2.25 (95% CI: 1.95, 2.59) indicates that there is a significant increase in mortality of patients with an ADO level greater than 331 compared to those with an ADO level less than 331. The odds ratio is 3.29 (95% CI: 2.60, 4.14).

The ADO variable is used to create nodes 31 and 32. Node 31 consists of patients with an ADO level less than 331 and has a higher proportion of surviving patients. This result confirms what was found in the mortality risk above. As can be seen in Table 5.2 the regression coefficient for ADO is 0.3191, giving an odds ratio for mortality of 1.38 times.

5.4.2. Age

The following graph illustrates the age distribution by CCU outcome:

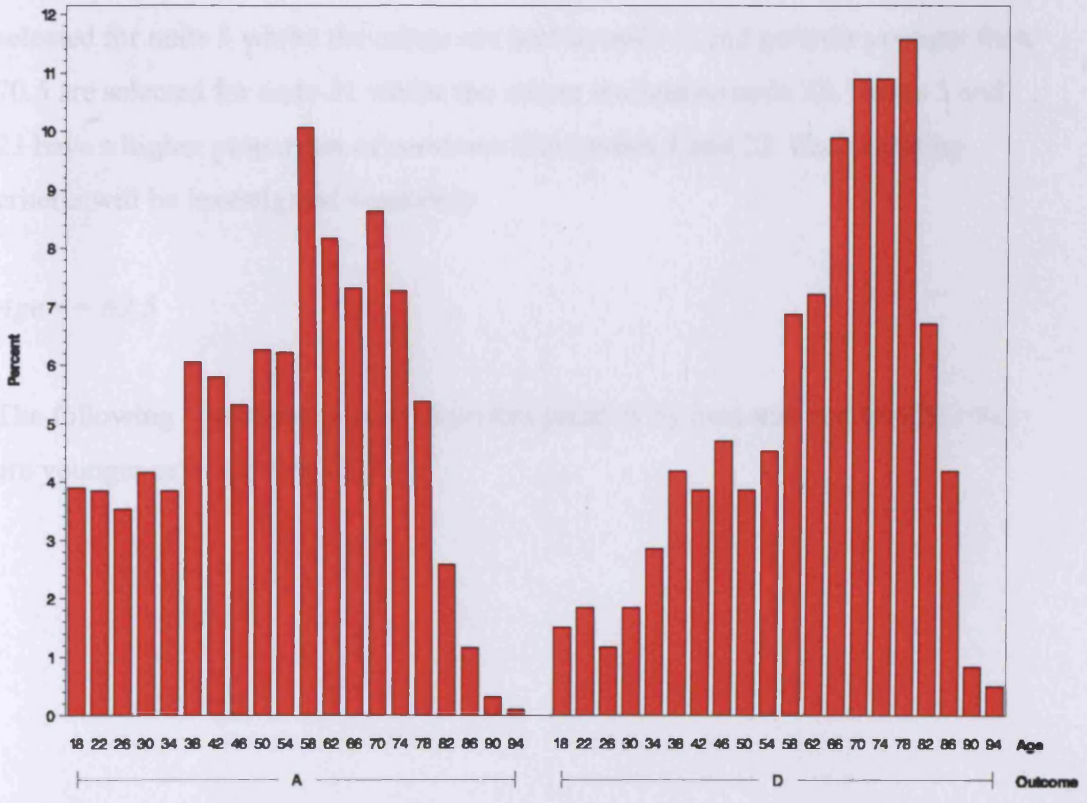


Figure 5.3: Frequency distribution for Age by CCU outcome

Table 5.7 outlines the summary statistics for age by outcome:

Table 5.7: Summary statistics for Age by CCU outcome

Outcome	N	Mean	Std Dev	Minimum	Median	Maximum
Alive	1897	53.08	17.97	16	56	95
Dead	596	61.79	17.08	16	66	93

The Mann-Whitney test establishes that there is a significant difference ($p < 0.0001$) in median for these two groups of patients (the above graph strengthens this point).

The age variable appears twice in the CART analysis tree diagram, once to create nodes 3 and 4 and once to create nodes 21 and 22. Patients younger than 62.5 are selected for node 3 whilst the others are sent to node 4, and patients younger than 70.5 are selected for node 21 whilst the others are sent to node 22. Nodes 3 and 21 have a higher proportion of survivors than nodes 4 and 22. Each splitting criteria will be investigated separately.

Age <= 62.5

The following contingency table classifies patients by outcome and whether they are younger or older than 62.5

Table 5.8: Frequency of Age ≤ 62.5 and > 62.5 by CCU outcome

Frequency	Outcome		Total
	Alive	Dead	
Older than 62.5	655	338	993
Younger than 62.5	1242	258	1500
Total	1897	596	2493

The chi-square contingency test indicates that the variables age and outcome are not independent at the 95% level.

The mortality risk for patients over 62.5 years old is 1.98 times higher than those under 62.5 (95% CI: 1.72, 2.28) and the odds ratio is 2.48 (95% CI: 2.06, 3.00) and both are significant.

Age <= 70.5

The same analysis is presented here for patients less than 70.5 years old compared with those older than 70.5 years old:

Table 5.9: Frequency of Age ≤ 70.5 and > 70.5 by CCU outcome

Frequency	Outcome		Total
	Alive	Dead	
Older than 70.5	361	231	592
Younger than 70.5	1536	365	1901
Total	1897	596	2493

Again, the chi-square test indicates that these variables are independent at the 95% level. In this case, the mortality risk (2.03, 95% CI: 1.77, 2.33) and the odds ratio (2.69, 95% CI: 2.20, 3.29) are higher indicating that the older the patient, the higher the mortality risk.

The above analysis strengthens the argument presented by the CART analysis, that the younger patients have a lower mortality risk than the older patients.

The regression coefficient for age (which has been standardised) is 0.5053. Therefore for a unit increase in age, provided that all other variables are kept constant, π will be 1.66 times higher.

5.4.3. Glucose

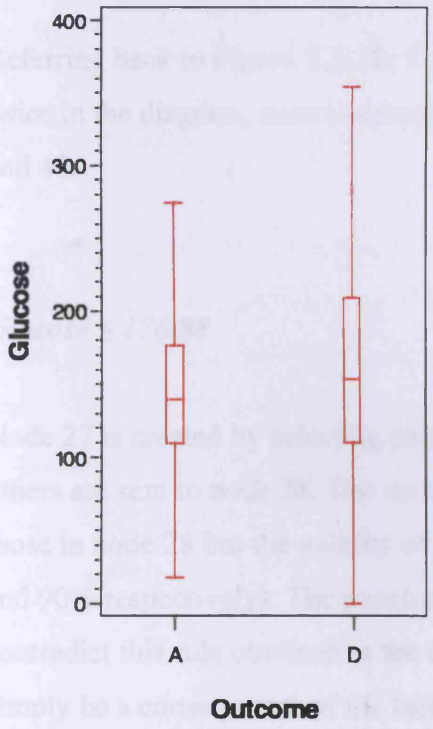


Figure 5.4: Boxplot of Glucose level by CCU outcome

Figure 5.4 demonstrates the distribution of glucose level in patients who survived their stay in the CCU compared with those who do not survive. The median

value of glucose is lower for patients who survive compared with those who do not. The tails of the distribution are longer for the patients who do not live. Table 5.10 below illustrates the summary statistics for glucose for the patients who live compared with those who do not live.

Table 5.10: Summary statistics for Glucose level by CCU outcome

Outcome	N	Mean	Std Dev	Minimum	Median	Maximum
Alive	1897	149.77	74.32	0	140.80	1232
Dead	596	181.65	161.35	0	157.52	2464

A Mann-Whitney test was performed on the above data and the difference in median glucose level was significant ($p < 0.0001$).

Referring back to Figure 5.1, the CART analysis tree diagram, Glucose appears twice in the diagram, once to create nodes 27 and 28, and once to create nodes 45 and 46.

Glucose ≤ 176.88

Node 27 is created by selecting patients with a glucose level less than 176.88, the others are sent to node 28. The survival rate for patients in node 27 is lower than those in node 28 but the validity of node 28 is much lower than node 27 (62% and 90% respectively). The conclusions drawn from Figure 5.4 and Table 5.7 contradict this rule obtained in the CART analysis. This contradiction could simply be a consequence of the lack of validity of node 28.

Table 5.11 summarises the number of patients with a glucose level less than 176.88 and their survival status.

Table 5.11: Frequency of Glucose ≤ 176.88 and > 176.88 by CCU outcome

Frequency	Outcome		Total
	Alive	Dead	
Glucose level greater than 176.88	478	225	703
Glucose level less than 176.88	1419	371	1790
Total	1897	596	2493

The chi-square contingency test affirms that the variables Glucose level and Outcome are not independent ($p < 0.0001$) and the mortality risk for a patient with a glucose level higher than 176.88 is 1.544 times higher (95% CI: 1.3414, 17776) than a patient with a lower glucose level. The odds of death is 1.80 times higher (95% CI: 1.48, 2.19) for patients with a higher glucose level.

Glucose ≤ 183.92

Nodes 45 and 46 are also created using a splitting rule concerning glucose. Patients with a glucose level less than 183.92 are sent to node 45, whereas patients with a glucose level greater than this are sent to node 46. The survival rate for patients in node 45 are higher than those in node 46, but both nodes have questionable validity. Despite this fact, the conclusions drawn from nodes 45 and 46 agree with the overall conclusions drawn from Figure 5.4 and Table 5.7 thus strengthening the argument that lower rates of glucose imply a higher survival rate.

Table 5.12 below indicates the frequencies of patients who survive and those who do not survive classified by their glucose level.

Table 5.12: Frequency of Glucose ≤ 183.92 and > 183.92 by CCU outcome

Frequency	Outcome		Total
	Alive	Dead	
Glucose level greater than 183.92	4199	215	634
Glucose level less than 183.92	1478	381	1859
Total	1897	596	2493

Again, at the 95% level, the variables Glucose < 183.92 and Outcome are not independent. The mortality risk for patients with a glucose level greater than 183.92 is 1.65 times higher (95% CI: 1.44, 1.90) than those with a glucose level less than 183.92, and the odds of death are 1.99 times higher (95% CI: 1.63, 2.43)

Finally, the regression coefficient for glucose is 0.1506 which translates to an odds multiplier of 1.16. This suggests that for a unit increase in glucose level (note that glucose was standardised for the regression analysis), the odds of death increases by 1.16 times thus concurring with the evidence above.

5.4.4. ORGSCORE

The next variable for consideration is ORGSCORE. This is a measure which refers to the organ failure score of a patient. The higher the ORGSCORE, the more severe the organ failure. Figure 5.5 illustrates the distribution of ORGSCORE by the CCU outcome;

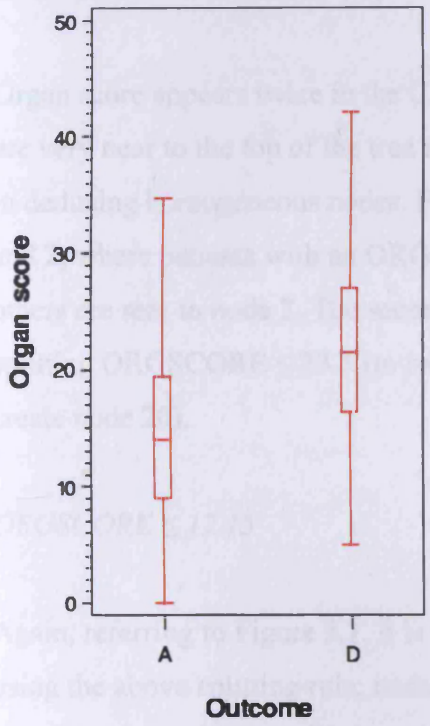


Figure 5.5: Boxplot of Organ Score by CCU outcome

It is clear from Figure 5.5 that the patients who survive their stay at the CCU have a lower organ score compared with those patients who do not survive. The tails of both distributions are fairly similar in length and both distributions are very symmetrical. Table 5.13 contains summary statistics for organ score by outcome

Table 5.13: Summary statistics for Organ score by CCU outcome

Outcome	N	Mean	Std Dev	Minimum	Median	Maximum
Alive	1897	14.61	7.24	0	14	48.6
Dead	596	22.12	7.73	5	21.5	52.9

The Mann-Whitney test indicates that the median value of ORGSCORE for patients who do not survive is significantly higher than those who do survive ($p < 0.0001$). Comparing mean and median for both outcomes, it is clear that the distributions are in fact fairly symmetrical.

Organ score appears twice in the CART diagram (Figure 5.1). Both occurrences are very near to the top of the tree indicating that ORGSCORE is a key variable in deducing homogeneous nodes. Firstly, ORGSCORE is used to create nodes 1 and 2, where patients with an ORGSCORE ≤ 12.15 are sent to node 1, and the others are sent to node 2. The second occurrence is to create nodes 19 and 20 by splitting ORGSCORE ≤ 23.3 (to create node 19) and ORGSCORE > 23.3 (to create node 20).

ORGSCORE $\leq$ *12.15*

Again, referring to Figure 5.1, it is clear that of the two primary nodes created using the above splitting rule; node 1 has the higher survival rate which corresponds to patients with an organ score less than 12.15 units. Table 5.14 indicates the survival distribution of patients with organ score less than and greater than 12.15 units.

Table 5.14: Frequency of Organ score ≤ 12.15 and > 12.15 by CCU outcome

Frequency	Outcome		Total
	Alive	Dead	
Organ score > 12.15	1115	546	1661
Organ score ≤ 12.15	782	50	832
Total	1897	596	2493

At the 95% level, these variables are not independent and the mortality risk for patients with an organ score greater than 12.15 is 5.47 times higher (95% CI: 4.14, 7.22) than those with an organ score less than this. Affirming this, the odds of death are 7.66 times higher (95% CI: 5.65, 10.37) for patients with an organ score greater than 12.15.

ORGSCORE ≤ 23.3

Comparing nodes 19 and 20, node 19 (which has patients with an ORGSCORE ≤ 23.3) has the higher survival rate. Table 5.15 below contains the frequencies of survivors versus non-survivors.

Table 5.15: Frequency of Organ score ≤ 23.3 and > 23.3 by CCU outcome

Frequency	Outcome		Total
	Alive	Dead	
Organ score > 23.3	240	247	487
Organ score ≤ 23.3	1657	349	2006
Total	1897	596	2493

Again, the variables in the above table are not independent ($p < 0.0001$). The mortality risk for patients with an organ score > 23.3 is 2.915 times higher (95%

CI: 2.56, 3.32) than those with a lower organ score, and the odds of death are raised by 4.89 times (95% CI: 3.95, 6.04).

Finally, considering the regression coefficient, it is clear to see that for a unit increase in ORGSCORE (once again, it has been standardised), the odds of dying will increase by 1.58 units, which concurs with the evidence above.

5.4.5. SYSBP

The final variable which appears in both the CART analysis tree diagram and the logistic regression is Systolic blood pressure. In a healthy adult, the systolic blood pressure level should be approximately 120 mm Hg. Figure 5.6 illustrates the systolic blood pressure levels for patients who survive their stay in CCU compared to those who do not.

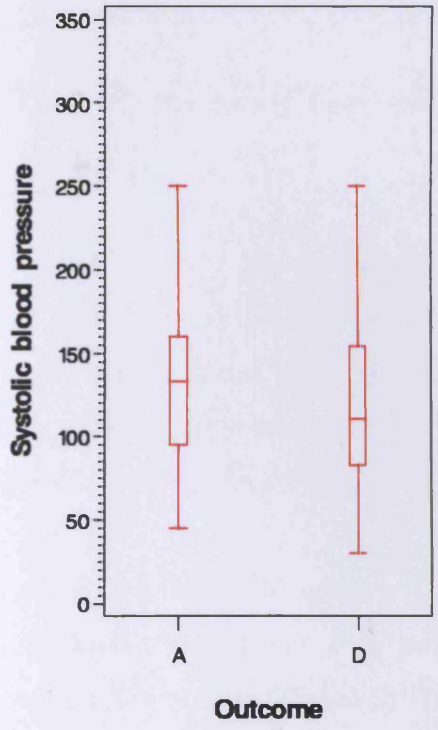


Figure 5.6: Boxplot of Systolic blood pressure by CCU outcome

It is clear from Figure 5.6 that the median value of SYSBP for patients who live is higher than those who do not live. Apart from this, the distributions are fairly similar. However, patients who do not survive their stay have a longer lower tail. Table 5.16 contains summary statistics for systolic blood pressure by outcome:

Table 5.16: Summary statistics for Systolic blood pressure by CCU outcome

Outcome	N	Mean	Std Dev	Minimum	Median	Maximum
Alive	1897	131.74	39.82	45	133	311
Dead	596	119.72	45.51	30	111	263

The Mann-Whitney test found significant differences between the median systolic blood pressure for the patients who survived compared to those who did not survive.

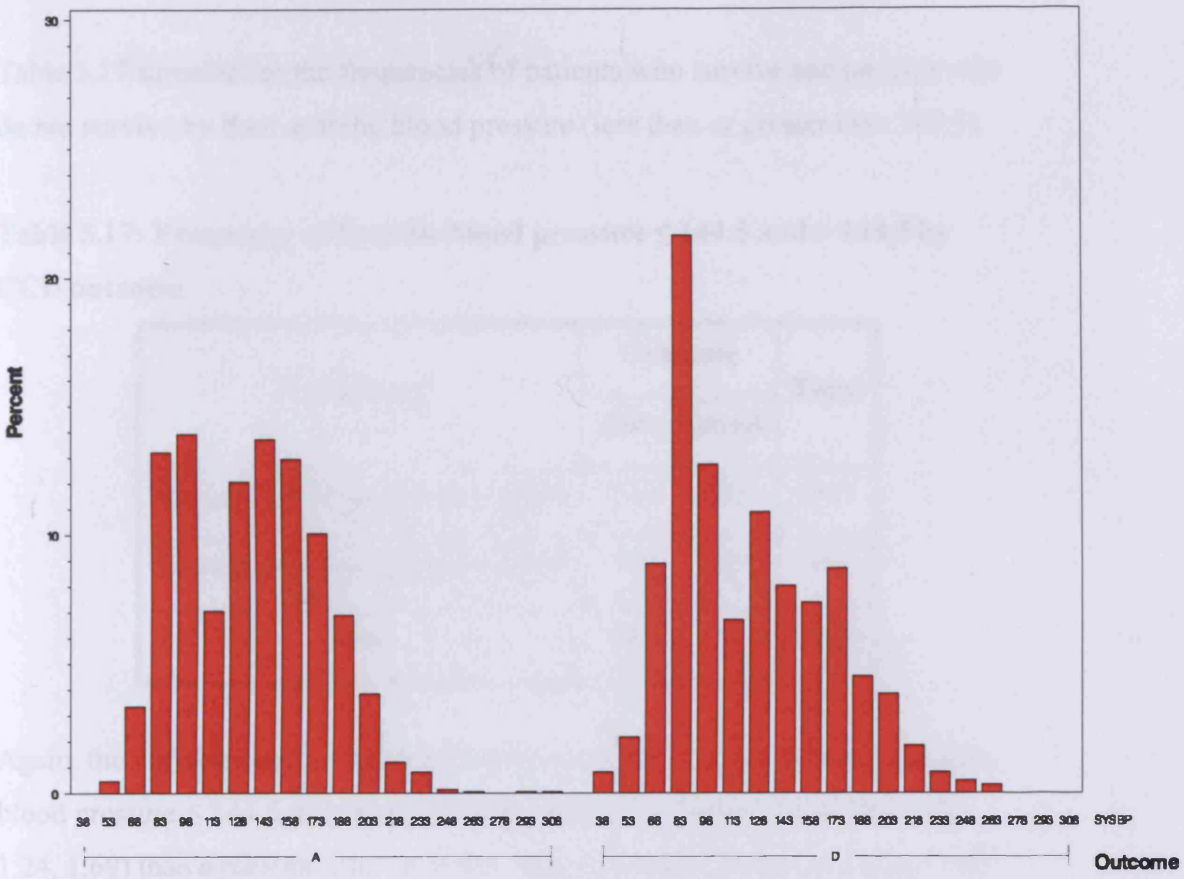


Figure 5.7: Frequency distribution of Systolic blood pressure by CCU outcome

Figure 5.7 displays the distribution of systolic pressure in a different way. Each bar represents a range of 15 points of blood pressure, and the set of bars corresponding to each outcome sum to 100%. It is clear from Figure 5.7 that over 20% of the patients who do not survive their stay at the CCU have a systolic blood pressure of between 76 mm Hg and 90 mm Hg (midpoint = 83 mm Hg).

Considering the CART analysis tree diagram (Figure 5.1), Systolic blood pressure appears once to create nodes 47 and 48. If a patient has a systolic blood pressure of less than 144.5 then they are sent to node 47, otherwise they are sent to node 48. The survival percentages for nodes 47 and 48 are 48% and 69% respectively. This implies that patients with a lower value of systolic blood pressure are less likely to survive their stay in the CCU.

Table 5.17 summarises the frequencies of patients who survive and patients who do not survive by their systolic blood pressure (less than or greater than 144.5).

Table 5.17: Frequency of Systolic blood pressure ≤ 144.5 and > 144.5 by CCU outcome

Frequency	Outcome		Total
	Alive	Dead	
Systolic blood pressure ≤ 144.5	1104	413	1517
Systolic blood pressure > 144.5	793	183	976
Total	1897	596	2493

Again, the variables are not independent ($p < 0.0001$). If a patient has a systolic blood pressure ≤ 144.5 then their mortality risk is 1.45 times higher (95% CI: 1.24, 1.69) than a patients who does not. Also, the odds of death are raised 1.62 times (95% CI: 1.33, 1.97).

The logistic regression coefficient found for systolic blood pressure is -0.1309. This translates to an odds ratio estimate of 0.88 which implies that for a unit increase in standardised systolic blood pressure, the odds of survival will be 0.88 times of the previous value. The relationship between survival and systolic blood pressure is inverse in this case. Therefore, if the value of systolic blood pressure of a patient is less than the average for the entire data set, the survival chances of that person will be lower.

5.5. Other factors which contribute to mortality risk

To complete this Chapter, an investigation into other factors which affect mortality risk follows. Firstly, various chronic diseases are investigated and their affect on mortality is documented. Following this, variables relating to the type of surgery (if any) that a patient undergoes, and whether there are any complications with the surgery, are discussed and the mortality risk is recorded.

5.5.1. Patients with a history of chronic disease

To complete this section of analysis, since the variables relating to a history of chronic disease and also CCU outcome are fairly complete, the data set used consists of 4205 patients. To ascertain the relationship between these chronic diseases and mortality, statistical techniques such as chi-square contingency tables, odds ratios and relative risks were used. Note, that unless stated, the chi-square contingency table test (using the continuity correction) yielded a significant p-value at the 95% level thus rejecting the hypothesis of independence. Discussion will be based on the relative risks calculated in each case rather than the odds ratio to aid ease of interpretation.

Table 5.18: Odds ratios and Mortality risk associated with various chronic diseases

DISEASE	FREQUENCY (%) OF SAMPLE	MORTALITY RISK (95% CI)	ODDS RATIO (95%CI)
Liver disease	36 (0.86%)	2.17 (1.40, 3.37)	2.84 (1.43, 5.62)
Haematological cancer	50 (1.19%)	2.80 (2.06, 3.81)	4.33 (2.47, 7.60)
Lymphoma	66 (1.57%)	1.92 (1.34, 2.75)	2.35 (1.39, 3.98)
Cardiovascular disease	127 (3.02%)	1.73 (1.30, 2.29)	2.01 (1.36, 2.99)
Renal failure	181 (4.3%)	1.77 (1.39, 2.25)	2.08 (1.49, 2.90).
Respiratory failure	190 (4.52%)	1.37 (1.04, 1.80)	1.48 (1.04, 2.10)
Immunological disease	226 (5.37%)	1.76 (1.41, 2.19)	2.05 (1.52, 2.78)

Table 5.18 has been ordered by disease incidence in the CCU sample. Note that information on two chronic diseases have not been included in the above table (HIV and other forms of cancer). The reason for this is the low frequency of HIV sufferers in the sample (only 7 in total) and the chi-square test for independence of Cancer and Mortality was not rejected at the 95% level ($p = 0.2366$).

Less than 1 percent (0.86%) of patients admitted onto the CCU has a history of liver disease. This rare condition significantly affects the survival outcome of a patient, since the mortality risk of 2.17 is significantly higher than 1.

The condition with the highest associated mortality is Haematological cancer which affects 1.19% if the sample. This condition increases the mortality risk of a patient 2.8 times. This variable also appears in the Logistic Regression analysis with a coefficient of 1.30 which further confirms the importance of this variable.

The condition which affects mortality least is Respiratory failure. This condition is not so rare and affects 4.52% of the CCU sample.

5.5.2. Variables relating to surgery

The final set of variables under consideration are those relating to the surgery experienced by a patient. In this section, only patients who have entered the CCU from either the Elective surgery source or the Emergency surgery source are included, reducing the sample size to 1892.

Table 5.19: Odds ratios and Mortality risk associated with surgical variables

VARIABLE	FREQUENCY (%) OF SAMPLE	MORTALITY RISK (95% CI)	ODDS RATIO (95%CI)
Emergency surgery (compared with Elective surgery)	711 (37.58%)	5.02 (3.59, 7.03)	5.90 (4.10, 8.48)
Complications in surgery	222 (11.73%)	2.57 (1.87, 3.52)	2.94 (2.01, 4.30)

Of the patients who experience some form of surgery before entering the CCU, 62% have undergone elective surgery and the remaining 38% have undergone Emergency surgery. Concentrating on the Elective patients alone, 3.6% of these do not survive their stay on the CCU whereas the corresponding percentage of emergency patients is as much as 17.9%. Table 5.19 demonstrates the mortality risk associated with Emergency surgery (compared with Elective surgery). This mortality risk indicates that a patient who undergoes Emergency surgery is 5.02 times more likely to die than an Elective patient.

Another interesting variable relating to surgery is whether the surgeons experienced any complications during the operation. Almost 12% of patients admitted onto the CCU from a surgical source (Elective or Emergency) experience complications in their surgery. The mortality risk of 2.57 implies that patients who experience complications in their operations are 2.57 times more likely to die in the CCU.

As an aside, it is also interesting to determine whether there is a link with surgery type and complications in surgery. Almost 5% of patients who undergo elective surgery experience complications in their surgery, whereas the corresponding percentage for Emergency patients is 18.6%. The relative risk is 2.42 (95%CI: 1.88, 3.10) suggesting that a patient who undergoes Emergency surgery is 2.42 times more likely to experience complications in their surgery compared to the Elective patients.

5.6. Five logistic regression equations

Consultants at the Critical Care Unit desire a tool which will predict whether a patient will stay in the Unit until the next day. The following section of analysis creates a set of predictive equations to estimate the probability that a patient is present in the Unit on day $n + 1$ given that a patient is present at the Unit on day n.

Initially, seven equations were constructed but two equations showed a significant lack of fit according to the Hosmer and Lemeshow statistic. The remaining five equations, correspond to the first five days of a patient's stay at the CCU. Note, only data from day n will be used to construct the probabilistic equation for day $n + 1$.

To construct these equations, a stepwise logistic regression procedure was adopted. The outcome variable is simply, was the patient present in the system on day $n + 1$ given that they were present on day n. Note, no distinction is made between a patient being discharged from the CCU and a patient dying. The results displayed in the section are kept to a minimum since there are 5 equations to analyse. Note that π_n is the probability that a patient will remain in the Unit until day $n + 1$ given that they are present on day n.

Day 1

The first step was to develop an equation for logit(π_1) where π_1 is the probability that a patient will remain in the Unit until day 2 given that they are present on day 1.

Table 5.20 highlights the parameter estimates that are significant. The final column indicates the effect on the odds of remaining in the system until day 2 given that the patient is present in the system on day 1. Note that the numbers in

the final column represent the multiplying factor for the odds if all other variables are kept constant and the variable in the row is increased by 1 unit.

Table 5.20: Parameter estimates for the logistic regression equation for survival until day 2

Parameter	Estimate	Standard Error	Pr > ChiSq	Odds multiplier
INTERCEPT	2.5061	0.2366	<.0001	
DIALYSED1	4.009	1.1121	0.0003	55.09
TPNT	2.5473	1.0313	0.0135	12.77
VENT1	1.6754	0.2378	<.0001	5.34
TISSPOINT	0.3129	0.0792	<.0001	1.37
HCO3	0.3046	0.0731	<.0001	1.36
URINE	0.1991	0.0757	0.0085	1.22
ALBUMIN	-0.1322	0.054	0.0145	0.88
PO	-0.2313	0.0632	0.0003	0.79
ADO	-0.3052	0.0549	<.0001	0.74
RESPSCORE0	-0.5015	0.1506	0.0009	0.61
PRECIP37	-0.5506	0.1745	0.0016	0.58
MBPSCORE0	-0.5854	0.129	<.0001	0.56
PHSCORE0	-0.6507	0.138	<.0001	0.52
TISSCODE63	-0.6851	0.218	0.0017	0.50
PHSCORE4	-0.7216	0.2139	0.0007	0.49
NVVERBAL0	-0.7721	0.1437	<.0001	0.46
ADMISSION5	-0.848	0.3467	0.0144	0.43
MBPSCORE4	-0.8828	0.262	0.0008	0.41
TISSCODE38	-0.8932	0.2166	<.0001	0.41
SDCODE19	-1.1249	0.2877	<.0001	0.32
TISSCODE65	-1.2835	0.354	0.0003	0.28
SDCODE40	-1.3634	0.3348	<.0001	0.26
PRECIP04	-1.3758	0.2162	<.0001	0.25
TISSCODE64	-1.4147	0.385	0.0002	0.24
TISSCODE68	-2.681	1.1432	0.019	0.068

The above table has been ordered by parameter estimate, beginning with the largest.

The logistic regression equation for the probability of a patient remaining in the system until day 2 given that they are present on day 1 is therefore:

$$\begin{aligned}\text{logit}(\pi_1) = 2.51 \; + \; & 4.00(DIALYSED) + 2.55(TPNT) + 1.68(VENT1) \\ & + 0.31(TISSPOINT) + 0.30(HCO3) + 0.20(URINE) \\ & - 0.13(ALBUMIN) - 0.23(PO) - 0.31(ADO) \\ & - 0.50(RESPSCORE0) - 0.55(PRECIP37) - 0.59(MBPSCORE0) \\ & - 0.65(PHSCORE0) - 0.69(TISSCODE63) - 0.72(PHSCORE4) \\ & - 0.77(NVVERBAL0) - 0.85(ADMISSION5) - 0.88(MBPSCORE4) \\ & - 0.89(TISSCODE38) - 1.12(SDCODE19) - 1.28(TISSCODE65) \\ & - 1.36(SDCODE40) - 1.38(PRECIP04) - 1.41(TISSCODE64) \\ & - 2.68(TISSCODE68)\end{aligned}$$

Referring back to Table 5.20, the variables DIALYSED1, TPNT and VENT1 which are most influential on outcome will now be considered in greater detail. If a patient is dialysed on day 1 (thus giving DIALYSED1 a value of 1), π_1 will be 55.09 times higher, which indicates that the odds of remaining in the system until day 2 will be 55.09 times higher compared to a patient who is not dialysed. Also, if a patient is given Total Parenteral nutrition (TPNT), π_1 will be 12.77 times higher and if a patient is ventilated, their odds of remaining in the system until day 2 will be 5.34 times higher. The remaining variables, although they are significant in terms of the model do not affect π_1 in such a dramatic way.

The most noteworthy variable in decreasing the value of π_1 is TISSCODE68 (which simply indicates that a patient has had the CAVHD/CVVHD intervention on day 1). When a patient has received this intervention, their odds of survival is 0.068 times that of a patient who has not received this intervention. Considering the odds in a slightly different manner, if a patient has not received the intervention CAVHD/CVVHD, they are have an odds of remaining in the system until day 2 14.60 times higher than patients who have received this intervention.

The variables TISSCODE64 (Arrest/defibrillation (24h)), PRECIP04 (Self-intoxication), SDCODE40 (Post-op Laminectomy and other Spinal cord surgery) and TISSCODE65 (Pulmonary artery (SG) catheter) all cause a large decrease in the value of π_1. For example, a patient who has arrested or been defibrillated within the previous 24 hours has a value of π_1 which is 0.24 times that of a patient who has not undergone such an episode (which corresponds to π_1 being 3.96 times higher for a patient who has not had an arrest or been defibrillated within the past 24 hours).

One final interesting point is the inclusion of the variable PRECIP37 in the regression model. This precipitating factor represents patients who receive Planned post operative monitoring, and are thus usually Elective surgery patients (83% are Elective surgery patients, the remaining 17% are Emergency surgery patients). Patients who enter the CCU from Elective surgery are very often short stay patients, thus the regression parameter of -0.55 (or the odds multiplier of 0.576604) suggests that a patient who has a precipitating factor of 37 has a lower value of π (0.58 times) than those patients who do not have this precipitating factor.

The maximum-rescaled R^2 value, which is 0.3542 for this model.

The percentage of concordant pairs of observed responses and predicted probabilities, is 82.2% for this model. More specifically, a pair of observations is concordant if the observation with the lower value has a lower predicted mean logit(π_1) score than the higher observed value. This value ranges from zero to 1, with 1 as the best possible model. The percentage discordant and the percentage tied which are 17.6 and 0.2 respectively.

Recalling that the Somers'D statistic simply determines the direction and strength of a relationship between pairs of observations. Somers'D is equal to 0.646 for the above logistic regression model.

For the above regression model, c, the rank correlation of ordinal variables is 0.823.

The above model displays no evidence of lack of fit ($p = 0.2710$).

Two example patients:

Once the logistic equation has been derived, it is possible to calculate the value of π_1 for a given patient. Table 5.21 contains two example patients for which the value of π_1 has been calculated. Note that patient 1 remained in the system until day 2, and patient 2 did not.

Table 5.21: Parameter estimates for the logistic regression equation for survival until day 2

PARAMETER	ESTIMATE	PATIENT 1, x_i	PATIENT 2, x_i	$\beta_i x_i$ FOR PATIENT 1	$\beta_i x_i$ FOR PATIENT 2
DIALYSED1	4.009	1	0	4.009	0
TPNT	2.5473	1	0	2.5473	0
VENT1	1.6754	1	0	1.6754	0
TISSPOINT	0.3129	0	-2	0	-0.6258
HCO3	0.3046	-1.4	-0.2	-0.42644	-0.06092
URINE	0.1991	-2	0	-0.3982	0
ALBUMIN	-0.1322	-2	1.8	0.2644	-0.23796
PO	-0.2313	0.2	0.4	-0.04626	-0.09252
ADO	-0.3052	-1	1	0.3052	-0.3052
RESPSCORE0	-0.5015	0	1	0	-0.5015
PRECIP37	-0.5506	0	0	0	0
MBPSCORE0	-0.5854	0	1	0	-0.5854
PHSCORE0	-0.6507	0	1	0	-0.6507
TISSCODE63	-0.6851	0	0	0	0
PHSCORE4	-0.7216	0	0	0	0
NVVERBAL0	-0.7721	0	0	0	0
ADMISSION5	-0.848	0	0	0	0
MBPSCORE4	-0.8828	0	0	0	0
TISSCODE38	-0.8932	1	0	-0.8932	0
SDCODE19	-1.1249	0	0	0	0
TISSCODE65	-1.2835	0	0	0	0
SDCODE40	-1.3634	0	0	0	0
PRECIP04	-1.3758	0	0	0	0
TISSCODE64	-1.4147	0	0	0	0
TISSCODE68	-2.681	0	0	0	0
INDICATH	-3.8643	0	1	0	-3.8643
Intercept				2.5061	2.5061
Logit(π_1)				9.5433	-4.4182
π_1				0.9999	0.0119

Here it is clear to see that the probability of patient 1 remaining in the system until day 2 is 0.9999. As was previously mentioned, this patient does in fact remain in the system until day 2. However, patient 2 has a probability of remaining in the system until day 2 of 0.0119 and this patient is discharged at the end of day 1.

Day 2

Note the day 2 equation seeks to estimate the probability that a person is still in the Unit on day 3 given that they are there on day 2, denoted by π_2. Another way of expressing this is the probability that a person has a length of stay greater than or equal to three days given that they have a length of stay of at least 2 days. The model parameters are listed in Table 5.22:

Table 5.22: Parameter estimates for the logistic regression equation for survival until day 3

Parameters	Estimate	Error	Pr > ChiSq	Odds multiplier
INTERCEPT	3.0599	0.2948	<.0001	
DIALYSED1	2.483	0.485	<.0001	11.98
SDCODE20	1.1109	0.4214	0.0084	3.04
PRECIP03	1.068	0.2957	0.0003	2.91
VENT1	0.6724	0.2071	0.0012	1.96
TISSPOINT	0.433	0.0734	<.0001	1.54
GLUCOSE	0.3923	0.1487	0.0083	1.48
SYSBP	0.2963	0.0754	<.0001	1.34
URINE	0.2738	0.0827	0.0009	1.31
ADO	-0.1502	0.0654	0.0217	0.86
CR	-0.1606	0.0687	0.0194	0.85
PCV	-0.208	0.0653	0.0015	0.81
MBPSCORE0	-0.3116	0.1366	0.0225	0.73
PO	-0.4533	0.1176	0.0001	0.64
EYESCORE1	-0.6681	0.2379	0.005	0.51
TEMPSCORE0	-0.6761	0.1523	<.0001	0.51
SPEC3	-0.7917	0.3529	0.0249	0.45
RESPSCORE0	-0.8404	0.1485	<.0001	0.43
SPEC1	-0.9142	0.3482	0.0086	0.40
RENALT	-1.023	0.3389	0.0025	0.36
SPEC7	-1.0765	0.4335	0.013	0.34
TPNN	-1.2789	0.1429	<.0001	0.28
SDCODE48	-1.3181	0.4273	0.002	0.27
INDICATN	-1.5479	0.1983	<.0001	0.21
PHSCORE4	-1.592	0.2607	<.0001	0.20
SDCODE43	-1.6484	0.2155	<.0001	0.19
TEMPSCORE2	-1.7589	0.6563	0.0074	0.17
TISSCODE68	-1.8047	0.4965	0.0003	0.16
SDCODE40	-2.6956	0.4554	<.0001	0.07

Observing Table 5.22, it is clear that whether a patient receives kidney dialysis on day 2, whether the patient's specific diagnostic code (SDCODE20) is Non-

Operative Neurologic Seizure disorder, whether the factor precipitating admission is (PRECIP03) Trauma or whether a patient is ventilated on day 2, affect logit(π_2) considerably. For example, once again, if a patient is dialysed on day 2, their odds of remaining in the system until day 3 is increased by almost 12 times.

Many variables cause a decrease in the value of π_2 as Table 5.22 clearly shows. For example, a patient with SDCODE40 (which represents people who have the specific diagnostic code Post-op Laminectomy and other Spinal cord surgery) have a much decreased chance of remaining in the system until day 3 compared with those who do not receive this diagnosis (in fact patients who are not diagnosed with this condition have a 14.29 times higher odds of remaining in the system until day 3).

Three medical specialty variables now appear in the regression equation: Cardiac (SPEC1), Intensivist (SPEC3) and Obstetrics and gynaecology (SPEC7). Each of these medical specialties decrease the odds of remaining in the system until day 3. Also, the intervention CAVHD/CVVHD (TISSCODE68) causes the value of π_2 to decrease.

One final variable is worthy of comment. TPNN (which indicates patients who have not received any intravenous feeding) has a regression parameter of -1.28 which indicates that a patient who has not received intravenous feeding on day 2 is more likely to leave the system on day 2 than to remain in the system. This is intuitively sensible since patients who are likely to remain in the system for more than a few days are fed intravenously (TPNT) or enterally (TPNE).

The Maximum rescaled R^2 value is 0.4884 for this model and the percentage of concordant pairs is 87% with 12.8% discordant (0.2% tied). The Somers'D statistic is 0.742 and c = 0.871. These all indicate that there fit is adequate. Finally, the Hosmer and Lemeshow lack of fit p-value is 0.1389.

Day 3

Table 5.23 contains the parameter estimates for the logistic regression model. Note π_3 is the probability that a patient will be present in the CCU on day 4 given that they are present on day 3:

Table 5.23: Parameter estimates for the logistic regression equation for survival until day 4

Parameter	Estimate	Standard error	Pr > ChiSq	Odds multiplier
INTERCEPT	1.2077	0.6831	0.0771	
DIALYSED1	2.5474	0.5522	<.0001	12.77
TISSCODE39	1.3294	0.3602	0.0002	3.78
SDCODE20	1.2068	0.5605	0.0313	3.34
TPNE	1.1829	0.1696	<.0001	3.26
GLUCOSE	0.9001	0.247	0.0003	2.46
CHPOINTS0	0.6178	0.2195	0.0049	1.85
TISSCODE38	0.573	0.2078	0.0058	1.77
TISSPOINT	0.389	0.0887	<.0001	1.48
SYSBP	0.3402	0.09	0.0002	1.41
URINE	0.2371	0.0925	0.0104	1.27
PHSCORE0	-0.4186	0.1851	0.0238	0.66
MBPSCORE0	-0.4513	0.1627	0.0055	0.64
ADMISSION1	-0.5823	0.1861	0.0018	0.56
PO	-0.6408	0.1441	<.0001	0.53
SPEC3	-0.9406	0.4084	0.0213	0.39
SDCODE15	-1.1805	0.3614	0.0011	0.31
SDCODE43	-1.2262	0.2868	<.0001	0.29
INDICATN	-1.2525	0.2286	<.0001	0.29
PHSCORE4	-1.34	0.3294	<.0001	0.26
SDCODE46	-1.4474	0.6464	0.0251	0.24
INDICATK	-1.7298	0.6149	0.0049	0.18
PRECIP13	-2.0334	1.0327	0.0489	0.13
TISSCODE64	-2.0459	0.8992	0.0229	0.13
TISSCODE68	-2.2186	0.5722	0.0001	0.11
TEMPSCORE4	-2.4523	0.8532	0.004	0.09
PRECIP10	-3.0095	1.3652	0.0275	0.05

Once again, kidney dialysis is the most significant factor in determining the probability that a patient will remain in the system until day 4 given that they are present on day 3. This factor will cause π_3 to increase by a factor of 12.77. Other variables also cause an increase in π_3, such as whether a patient has received the intervention CPAP (TISS039) on day 3, whether the specific diagnostic code for a patient is Non-Op Neurologic Seizure disorder (SDCODE20) or whether the patient has been fed enterally (TPNE).

The most significant factor in reducing π_3 is PRECIP10 (which refers to a patient who has had a Myocardial infarction). An identical patient who has not had a myocardial infarction is 20 times more likely to remain in the system until day 4. Again, TISSCODE68 proves an important factor which causes π_3 to decrease in this case, a patient who has not received this intervention has odds of remaining in the system increased by a factor of approximately 9.

The Maximum rescaled R^2 value is 0.4132 for this model which is slightly lower than the model for day 2. The percentage of concordant pairs is 85.2% with 14.6% discordant (0.2% tied). The Somers'D statistic is 0.706 and c = 0.853. Finally, the Hosmer and Lemeshow lack of fit p-value is 0.5558.

Day 4

This next piece of analysis seeks to model the probability of patient remaining in the CCU until day 5 given that they are present in the Unit on day 4 (π_4). Table 5.24 outlines the parameter estimates for this model:

Table 5.24: Parameter estimates for the logistic regression equation for survival until day 5

Parameter	Estimate	Standard error	Pr > ChiSq	Odds multiplier
INTERCEPT	1.6185	0.236	<.0001	
DIALYSED1	2.7015	0.6362	<.0001	14.90
TPNE	1.2144	0.2149	<.0001	3.37
URINE	0.4668	0.1261	0.0002	1.59
HCO3	0.4196	0.1505	0.0053	1.52
TISSPOINT	0.3273	0.1195	0.0062	1.39
ADO	-0.318	0.1017	0.0018	0.73
INDICATN	-0.6364	0.2868	0.0265	0.53
PO	-0.7088	0.2149	0.001	0.49
EYESCORE1	-1.1153	0.359	0.0019	0.33
PHSCORE4	-1.1244	0.526	0.0325	0.32
CHPOINTS2	-1.4173	0.5641	0.012	0.24
VPCO	-1.4408	0.7313	0.0488	0.24
RESPD0	-1.4854	0.2662	<.0001	0.23
NVVERBAL5	-1.5296	0.7554	0.0429	0.22
PRECIP24	-1.5864	0.3629	<.0001	0.20
TISSCODE68	-1.839	0.6743	0.0064	0.16
SDCODE49	-2.2683	0.8726	0.0093	0.11
SDCODE16	-2.4157	0.759	0.0015	0.09
TEMPSCORE4	-3.9226	1.3232	0.003	0.02

Nineteen parameters are deemed significant in this model at the 95% level. Two parameters, DIALYSED1 and TPNE, contribute a large amount to π_4. If all other parameters are kept constant then the odds of remaining in the Unit until day 5 given that a patient is present on day 4 is increased by 14.9 times if the patient is

dialysed on day 4, whereas the odds is increased by 3.37 times if a patient receives Enteral feeding (TPNE).

Conversely, specific diagnostic codes 49 (Post-op Gastrointestinal) and 16 (Non-Op CVS F Cardiogenic shock) contribute negatively to π_4. This implies that if all other parameters are kept constant, then the odds of remaining in the Unit until day 5 given that a patient is present on day 4 is decreased by 9.09 times if the patient is given SDCODE49, whereas the odds is decreased by 11.11 times if a patient is given SDCODE16.

Also, here the factor TEMPSCORE4 has a large affect on logit(π_4). In fact, if a patient does not have TEMPSCORE4, their odds of remaining in the system until day 5 is approximately 50 times higher than a patient who has a TEMPSCORE of 4.

The Maximum rescaled R^2 value of 0.4456 indicates that this model is fairly useful at predicting the outcome probability from the variables in the model. Over 87 percent of pairs are concordant with 12.6% discordant (0.2% tied). Somers'D and c have values greater than 0.5 (0.745 and 0.873 respectively) thus indicating a reasonable fit. Finally, the Hosmer and Lemeshow goodness of fit test indicates that at the 95% level there is no evidence for lack of fit ($p = 0.2389$).

Day 5

The final model for consideration is logit(π_5). Here, π_5 denotes the probability of remaining in the Unit until day 6 given that the patient is present in the Unit until day 5. Table 5.25 holds information concerning the parameter estimates:

Table 5.25: Parameter estimates for the logistic regression equation for survival until day 6

Parameter	Estimate	Standard error	Pr > ChiSq	Odds multiplier
INTERCEPT	2.1025	0.1853	<.0001	
DIALYSED1	1.1839	0.3157	0.0002	3.27
HCO3	0.5805	0.1478	<.0001	1.79
TISSPOINT	0.3607	0.1239	0.0036	1.43
SYSBP	0.324	0.1168	0.0055	1.38
BILIRUBIN	-0.1056	0.0474	0.0258	0.90
ADO	-0.253	0.1033	0.0143	0.78
SDCODE18	-1.0081	0.4423	0.0226	0.36
SDCODE39	-1.1152	0.5043	0.027	0.33
PRECIP06	-1.2627	0.4621	0.0063	0.28
RESPD0	-1.341	0.2678	<.0001	0.26
PRECIP15	-1.4054	0.5631	0.0126	0.25
CHPOINTS2	-1.7449	0.7126	0.0143	0.17
VPCO	-3.2294	1.3032	0.0132	0.04

There are 13 variables present in the model. Once again, DIALYSED1 is the variable which contributes the most to logit(π_5).Three other variables cause an increase in π_5; HCO3, TISSPOINT and SYSBP.

The remaining variables contribute negatively to π_5, the most influential being VPCO. This variable has a dramatic influence on π_5, causing it to decrease by a factor of 25 for each unit increase in standardised VPCO.

The diagnostic tests all provide satisfactory findings, with 80% concordant, 19.7% discordant and values of Somers'D and c of 0.603 and 0.801 respectively.

The Hosmer and Lemeshow test has a p-value of 0.5678 indicating no evidence for lack of fit and the maximum rescaled R^2 value is 0.2842.

Summary

To summarise, two variables appeared in all models, DIALYSED1 and TISSPOINT and on each occasion both had a positive valued parameter. Since both variables have positive valued parameters, if a patient is dialysed on days 1 to 5 or has a higher than average TISSPOINT on these days, they are very likely to remain in the CUU until day 6. Several variables appeared in four of the models, the most interesting being TISSCODE68, and each appearance yielded a negative value for the parameter. Four variables appeared in three regression equations, including HCO3 which appeared on days 1, 4 and 5, and SYSBP which appeared on days 2, 3 and 5. Both variables had positive regression coefficients each time.

5.7. Conclusion

The second objective of this thesis, to identify the factors which affect length of stay and mortality in the CCU has been addressed in this Chapter. The evidence in this Chapter suggests that mortality risk is very complicated factor which depends upon many variables. Fourteen variables were included in the splitting rules for the CART and this reduced the impurity of the data by over 25%. Twenty one variables appeared in the logistic regression analysis and yielded a maximum rescaled R^2 value of 0.3764. Only five variables were concordant between the two pieces of analysis, namely ADO, Age, Glucose, ORGSCORE and SYSBP. Both pieces of analysis gave similar results for ADO, Age and ORGSCORE but some discrepancies were found for Glucose and SYSBP. A reason for this discrepancy may be that both Glucose and SYSBP appear at the bottom of the CART analysis tree.

The following graph summarises the Mortality risks associated with all variables in this section:

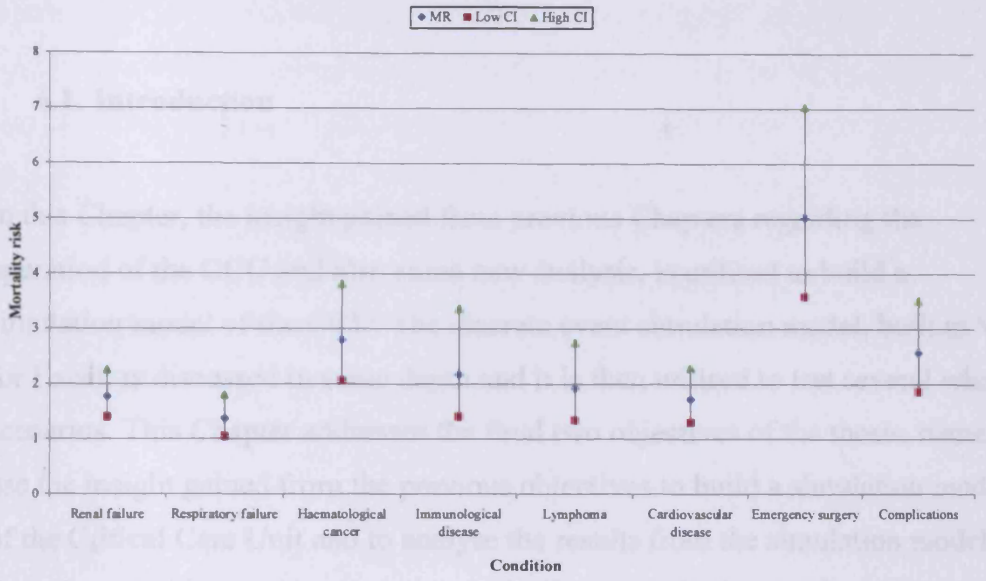


Figure 5.8: Mortality risk confidence intervals for chronic diseases and surgical variables.

It is clear that the variable which affects mortality risk the most is Emergency surgery. Respiratory failure is the variable which affects mortality risk least but it still has a significant effect.

When considering the five logistic regression equations, two variables appeared in all models, DIALYSED1 and TISSPOINT and on each occasion both had a positive valued parameter. Several variables appeared in four of the models, the most interesting being TISSCODE68, and each appearance yielded a negative value for the parameter.

An alternative regression model could have been utilised here, namely ordinal regression. This regression technique is similar to logistic regression but it allows the possibility having more than two outcomes. In this application, the outcomes could be; survive until the next day, discharge or death. This is an interesting area for future research.

Chapter 6: Bed occupancy modelling of a Critical Care Unit.

6.1. Introduction

In this Chapter, the insight gained from previous Chapters regarding the operation of the CCU and also some new analysis, is utilised to build a simulation model of the CCU. The discrete event simulation model, built in VBA for Excel, is discussed in some depth and it is then utilised to test several what-if scenarios. This Chapter addresses the final two objectives of the thesis, namely to use the insight gained from the previous objectives to build a simulation model of the Critical Care Unit and to analyse the results from the simulation model to show how varying some parameters will affect cancellations and cost.

As has been previously mentioned, a Critical Care Unit with insufficient beds to facilitate demand is not satisfactory as Elective surgeries may require cancellation which results in the extension of some waiting times.

This Chapter will focus on ensuring that there are an optimal number of beds available in CCU to minimise cancellations of Elective surgery.

6.1.1. The Critical Care Unit

The CCU is the sector of the hospital where, as the name suggests, critically ill patients receive treatment. More importantly, a significant proportion of patients who have an operation will be admitted to the CCU for post-operative care. For example, all patients who are ventilated need to be cared for in the CCU. This study is based on patients admitted to the CCU in the University Hospital of Wales (UHW) which was formed in 2003. Previous to this there were two Units, the Intensive Care Unit and the High Dependency Unit.

One of the main factors that differentiates the CCU from other high dependency wards is the level of nursing care needed – the majority of patients in the CCU require one-to-one nursing care. The nurses who care for these patients are specially trained, in short supply and expensive. Another important factor is the cost of beds in CCU. In 2005-2006 it was estimated by the Department of Health (DOH 2006) that each CCU bed costs the NHS £1,716 per day. However, in 2006-2007 the Department of Health (DOH 2007) changed their policy on costing and now calculate the cost per patient in CCU according to the number of organ failures they have rather than the average cost of a bed.

Since the hospital has no control over the number of emergency patients admitted onto the CCU, the focus of this study is Elective patients only. Preliminary data analysis shows that the mean number of organ failures an Elective patient has is 0.36 (SD = 0.73). In order to capture as much of the variation as possible (and to have a "worst case" cost), for this study the cost of an Elective patient in the CCU is taken to be £990 per day which corresponds to the cost of one organ failure.

6.1.2. Data

The data set, which is routinely collected, has complete records between April 2004 and May 2007. As has been previously noted, it contains a vast amount of information about each patient who is admitted into the CCU. For the sake of this Chapter, the important variables which are considered are arrival time and date, discharge time and date, and source of arrival.

The CCU at UHW has 24 beds with 5 additional beds available for use at times of peak demand. Initial analysis revealed that on 9% of occasions there were more than 24 beds occupied; that is, one or more of the extra five beds stored for

peak demand use were utilised. Taking account of the length of stay of patients, this amounts to over 7% of the available CCU time.

6.2. Methods

To construct a simulation model, the first factor to ascertain is the arrival sources. The CART analysis nodes developed in Chapter 4 could be utilised as arrival sources for a simulation model (as was mentioned in Chapter 4), however, it would be difficult for staff to pick which node a patient belonged to in a pressure situation due to the many rules required to form the nodes. Therefore, since arrival source appeared to be a significant factor in length of stay (again, see Chapter 4 and subsequent analysis), the arrival sources would simply be A&E, Elective surgery, Emergency surgery, other hospitals, the wards and X-Ray.

Several analytical techniques are presented in this Chapter to investigate the length of stay distribution of patients from different arrival sources. The techniques include survival analysis, optimisation and simulation.

6.2.1. Survival analysis

In Chapter 4, there is a thorough investigation of variables which affect length of stay. Before beginning with the simulation model, it is important to ensure that the source of admission is in fact a good indicator of length of stay. This preliminary section seeks to do this by utilising the statistical technique survival analysis. Simply, survival analysis is a method which seeks to investigate the nature and duration of survival. It can be used in many different fields in real life, but commonly it is used in the medical field. Very often it is used to investigate the effectiveness of a drug on survival. For example, earlier this year, Wang et al (Wang et al. 2008) considered the efficacy of treatment schedules for hepatocellular carcinoma. They compared Kaplan-Meier survival curves for

different treatments during different stages in the cancer. Also Vasilakis and Marshall (Vasilakis and Marshall 2005) used survival analysis to model hospital length of stay of stoke patients.

In this section, rather than considering treatments, comparison will be made according to which source the patient arrived from.

The analysis was undertaken in SAS 9.1.3. There are several assumptions associated with survival analysis and these need to be met before commencement of the analysis.

Firstly, a clear event must be defined in order for us to measure the time to that event. In this case, the event will simply be discharge from the Critical Care Unit. For the sake of this work, the departure destination is not considered (for example, we will consider patients who died, patients who were discharged onto a different ward, and patients who were sent home in the same manner).

Secondly, the time to event must be measured in a precise way, avoiding errors. All erroneous data was excluded prior to this analysis, and therefore we consider the remaining data to be valid and precise.

Thirdly, each subject within the study must only appear once in the dataset. This is true for the CCU data; any repeat admissions during the study period were omitted for the sake of independence.

The survival function is defined as $S(t) = \text{Probability}(T > t)$ where T is the life time of a randomly selected observation. Other factors which are noteworthy are the CDF, defined as $F(t) = 1 - S(t)$ and the PDF $f(t) = \frac{dF(t)}{dt}$. The hazard function is then defined as $h(t) = \frac{f(t)}{S(t)}$. The simplest (and most common)

survival distribution is the single Exponential distribution, which is characterised by a constant hazard function λ. To test whether this is an appropriate distribution function, the plot of the negative log of the survival function must be inspected. If the distribution function is appropriate, the plot will be close to a straight line through the origin. Figure 6.1 demonstrates this plot for each different arrival source:

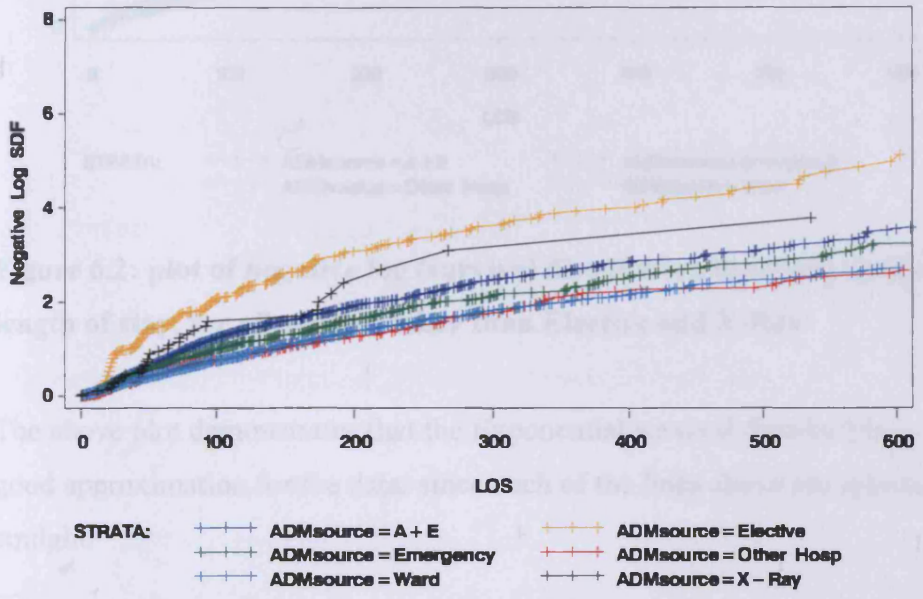


Figure 6.1: plot of negative log (survival distribution function) against length of stay

It is clear from the Figure 6.1 that the sources Elective and X-Ray do not produce straight line plots. Figure 6.2, is simply the same as above but this time excluding the Elective and X-Ray patients:

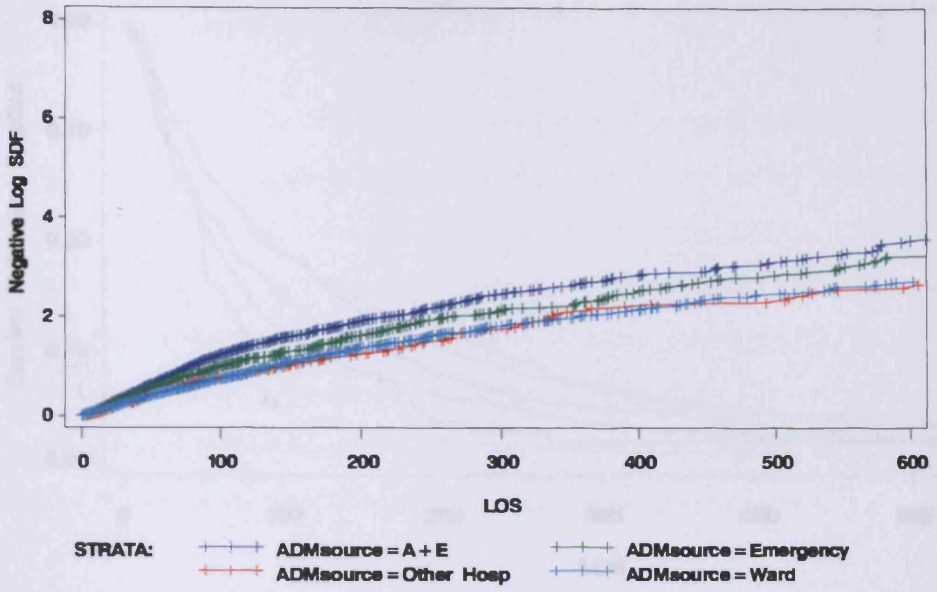


Figure 6.2: plot of negative log (survival distribution function) against length of stay, for all sources other than Elective and X-Ray

The above plot demonstrates that the Exponential survival distribution is a fairly good approximation for the data, since each of the lines above are approximately straight.

As was previously mentioned, the event of note is simply discharge from the CCU. Figure 6.3 demonstrates the survival function for patients from different arrival sources:

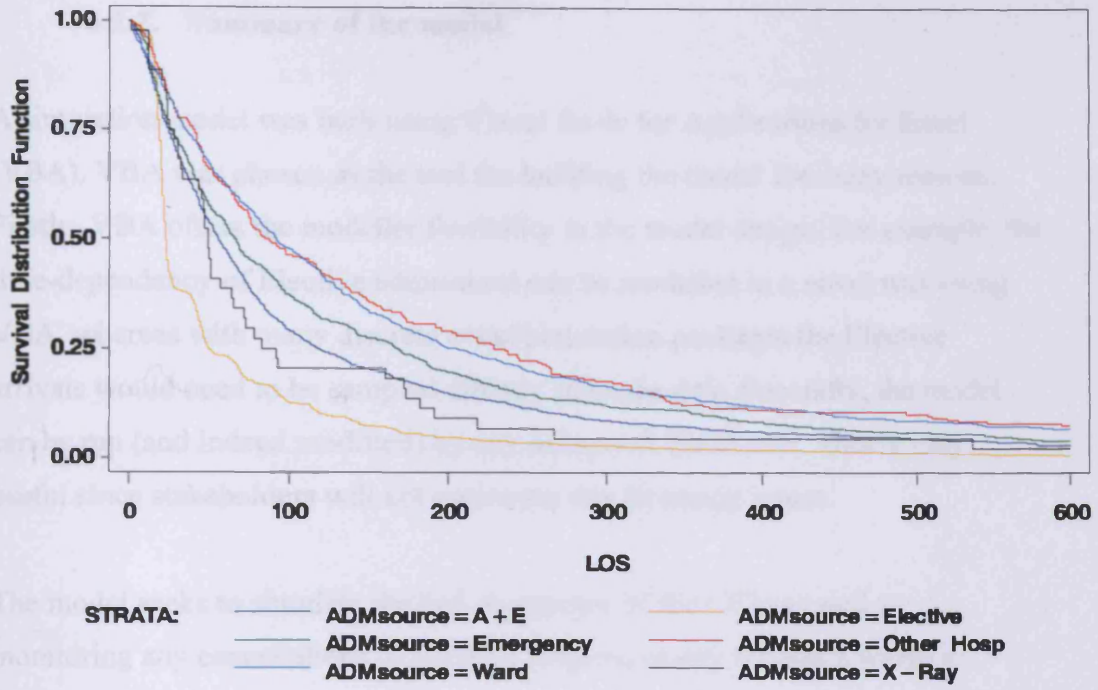


Figure 6.3: Survival curve for patients from different sources

Note that patients admitted from the ward and Other Hospitals have the highest probability for staying in the Critical Care Unit (surviving) for a long time. Patients admitted from Elective surgery have the highest probability of a quick discharge, and a far smaller probability of staying within the unit for a long time. This conclusion is intuitive, considering that the mean length of stay for patients from Elective Surgery is significantly lower than the length of stay of patients admitted from other sources. Even though patients from Elective Surgery and X-Ray do not conform to the assumptions of the survival model fitted, the survival curves above do accord with the expected pattern for length of stay for patients from these groups.

Thus, it is clear that source of admission is a factor which affects length of stay in the Critical Care Unit and will thus be used to distinguish between groups of patients in the simulation model.

6.2.2. Summary of the model

A simulation model was built using Visual Basic for Applications for Excel (VBA). VBA was chosen as the tool for building the model for many reasons. Firstly, VBA offers the modeller flexibility in the model design. For example, the time-dependency of Elective admissions can be modelled in a novel way using VBA, whereas with many discrete event simulation packages the Elective arrivals would need to be sampled directly from the data. Secondly, the model can be run (and indeed modified) by any Microsoft Excel user. This is very useful since stakeholders will not encounter any licensing issues.

The model seeks to simulate the bed-occupancy of the CCU as well as monitoring any cancellations of Elective surgery, or any instances where a patient experienced a delay before admission to the CCU.

The model was constructed to allow patients to arrive at the CCU from 6 different sources, namely Accident and Emergency, Wards, Elective surgery, Emergency surgery, other hospitals, or the X-Ray department. Each source had a different inter-arrival time distribution, which was ascertained using the statistical distribution fitting software Stat::Fit utilising the Chi-square goodness of fit test.

In the model it is assumed that once a patient is booked for admission to the CCU, (s)he is classified as either a 'planned admission' (Elective patients only) or as an 'unplanned admission' (all other patients) If an arriving patient finds that all beds are occupied, they are sent to a queue. There are two queues built into the model, the "Unplanned Admissions" queue and the "Planned Admissions" queue. The patients in the "Planned Admissions" queue - that is the Elective surgery patients - have their surgery cancelled and are then sent home. The patients in the "Unplanned Admissions" queue wait until a bed becomes available. They are then served and leave the system once their service is complete.

If a patient arrives at a time where there are unoccupied beds, they are admitted into the CCU and are then treated (i.e. they are served). The service time distributions were linked with the source of arrival.

The model has a user friendly front-end whereby parameters can be changed readily. For example, the mean and standard deviation for each arrival distribution can be changed corresponding to the arrival source. The results from the model are then output to an Excel spreadsheet where they can be analysed further.

6.3. Results

As was previously mentioned, the model was constructed so that patients would arrive from six different arrival sources and would be served according to a statistical distribution based upon their arrival source. The following sections describe the inter-arrival time and service time distributions for each of the six arrival sources. Verification of the model and a brief outline of results are also presented.

6.3.1. Inter-arrival time distributions

Each arrival source will be considered independently and the results will be collated in a table to finish this section.

Accident and Emergency

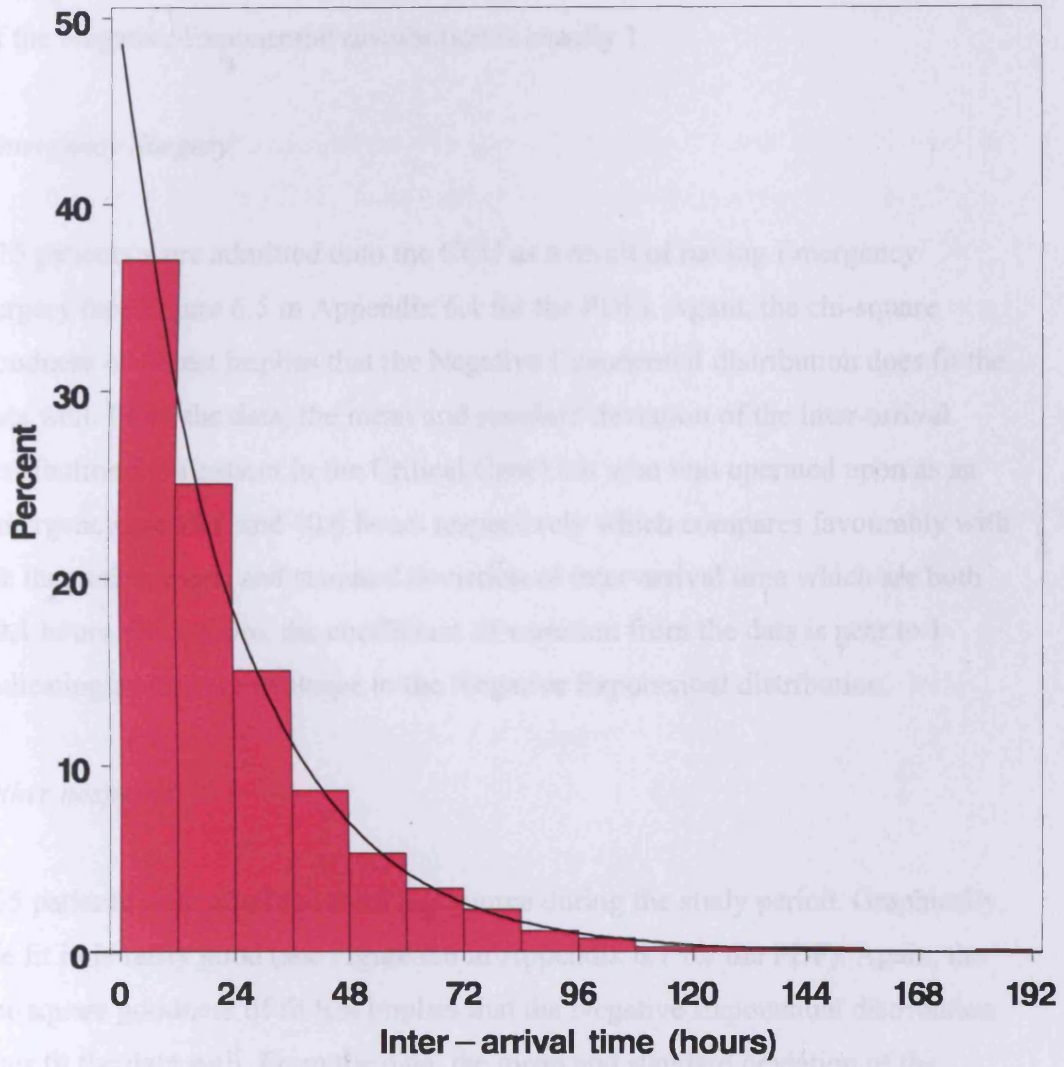


Figure 6.4: Inter-arrival time for A&E patients with Negative Exponential fit (mean = 24.7 hours)

Figure 6.4 encapsulates the inter-arrival distribution for patients arriving onto the Critical Care Unit from the A&E department. Visually, the fit in Figure 6.4 is

very good. 1,134 were in this group. The chi-square goodness of fit test suggests that the Negative Exponential distribution does fit the data well. From the data, the mean and standard deviation of the inter-arrival distribution for a patient in the Critical Care Unit who was admitted from the A&E department are 24.7 and 25.1 hours respectively which compares favourably with the theoretical mean and standard deviation of inter-arrival times which are both 24.7 hours. Also, having a coefficient of variation near to 1 is also indicative since the coefficient of the Negative Exponential distribution is exactly 1.

Emergency Surgery

715 patients were admitted onto the CCU as a result of having Emergency surgery (see Figure 6.5 in Appendix 6.1 for the PDF). Again, the chi-square goodness of fit test implies that the Negative Exponential distribution does fit the data well. From the data, the mean and standard deviation of the inter-arrival distribution for a patient in the Critical Care Unit who was operated upon as an Emergency are 39.1 and 40.6 hours respectively which compares favourably with the theoretical mean and standard deviation of inter-arrival time which are both 39.1 hours. Once more, the coefficient of variation from the data is near to 1 indicating a close resemblance to the Negative Exponential distribution.

Other hospitals

235 patients were admitted from this source during the study period. Graphically, the fit in is fairly good (see Figure 6.6 in Appendix 6.1 for the PDF). Again, the chi-square goodness of fit test implies that the Negative Exponential distribution does fit the data well. From the data, the mean and standard deviation of the inter-arrival distribution of this group of patients are 118.0 and 108.2 hours respectively which compares favourably with the theoretical mean and standard deviation of length of stay which are both 118 hours. Once more, the coefficient of variation from the data is near to 1 indicating a close resemblance to the Negative Exponential distribution.

Ward

There were 916 patients fitting into this category (see Figure 6.7 in Appendix 6.1 for the PDF). Again, visually, the fit is good and the null Hypothesis for the chi-square test cannot be rejected. The mean inter-arrival time from the data is 30.5 hours and the standard deviation is 31.4 hours. The Negative Exponential distribution gives both mean and standard deviation to be 30.5 hours.

X-Ray

The final unplanned admission source is the X-Ray department. Only 44 patients were admitted onto the CCU from this source during the study period; thus the graph does not demonstrate such a good fit (see Figure 6.8 in Appendix 6.1 for the PDF). However, the data does fit the distribution according to the Chi-square test and the mean and standard deviation from the data (604.2 and 638.0 hours respectively) are not vastly different from the theoretical mean and standard deviation (604 hours).

Arrival distributions for unplanned patients are often found to be Negative Exponential (Coats and Michalis 2001, Moore 2003) due to the random nature of emergency events. Next, consideration will be given to the only source where admissions are planned in nature, Elective surgery.

Elective Surgery

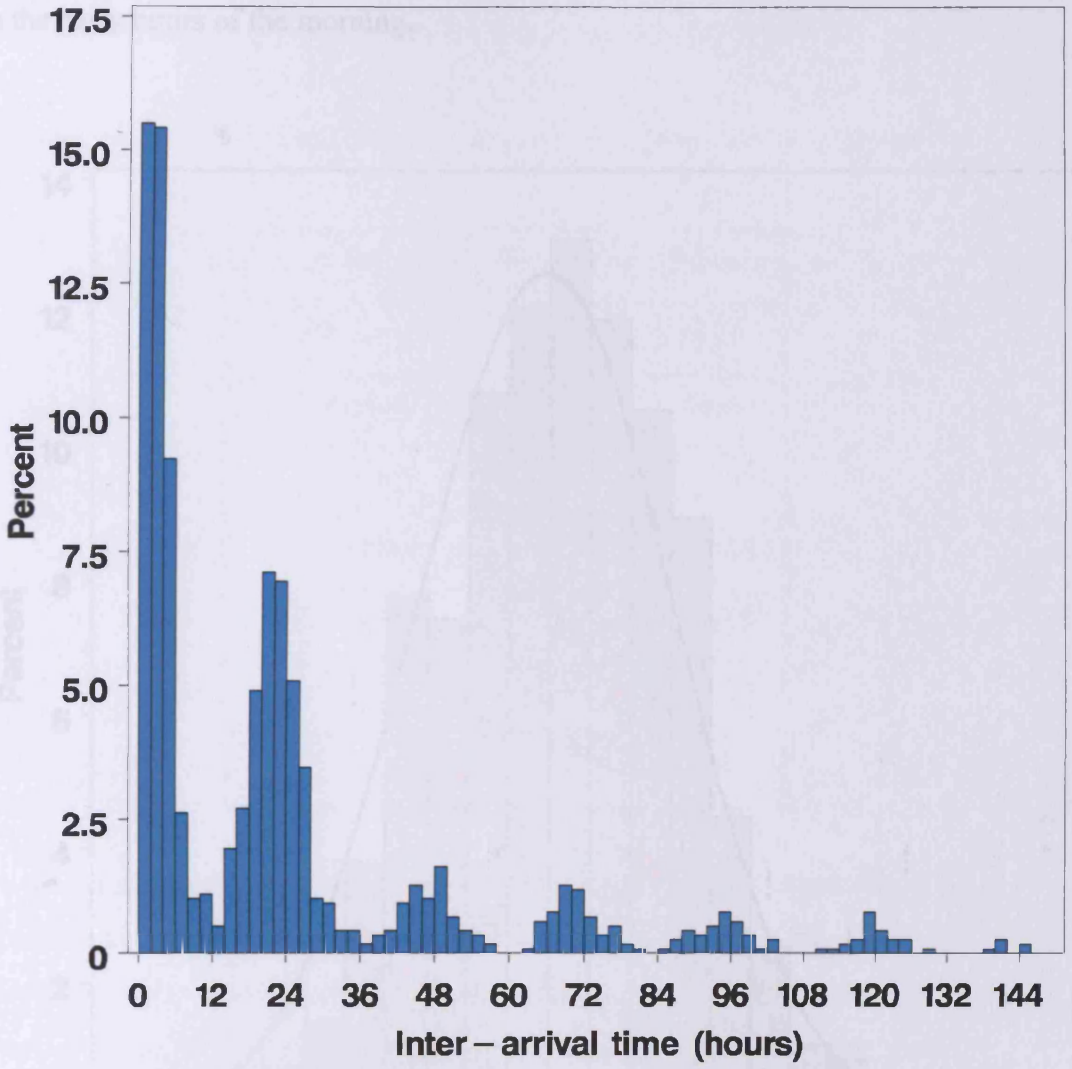


Figure 6.9: Inter-arrival time for Elective Surgery patients

The inter-arrival time of Elective patients is time-dependent in nature (see Figure 6.9). This graph shows that for many arrivals the inter-arrival time is between 0 and 4 hours. Another peak occurs between 22 and 26 hours, and then other smaller peaks at approximately 24 hour intervals. These observations are consistent with the practice of Elective surgery being performed at specific sessions spaced at 24 hour intervals.

Figure 6.10 highlights the precise time of day of these elective arrivals. The graph shows that the majority of arrivals occur in the early evening with very few in the early hours of the morning.

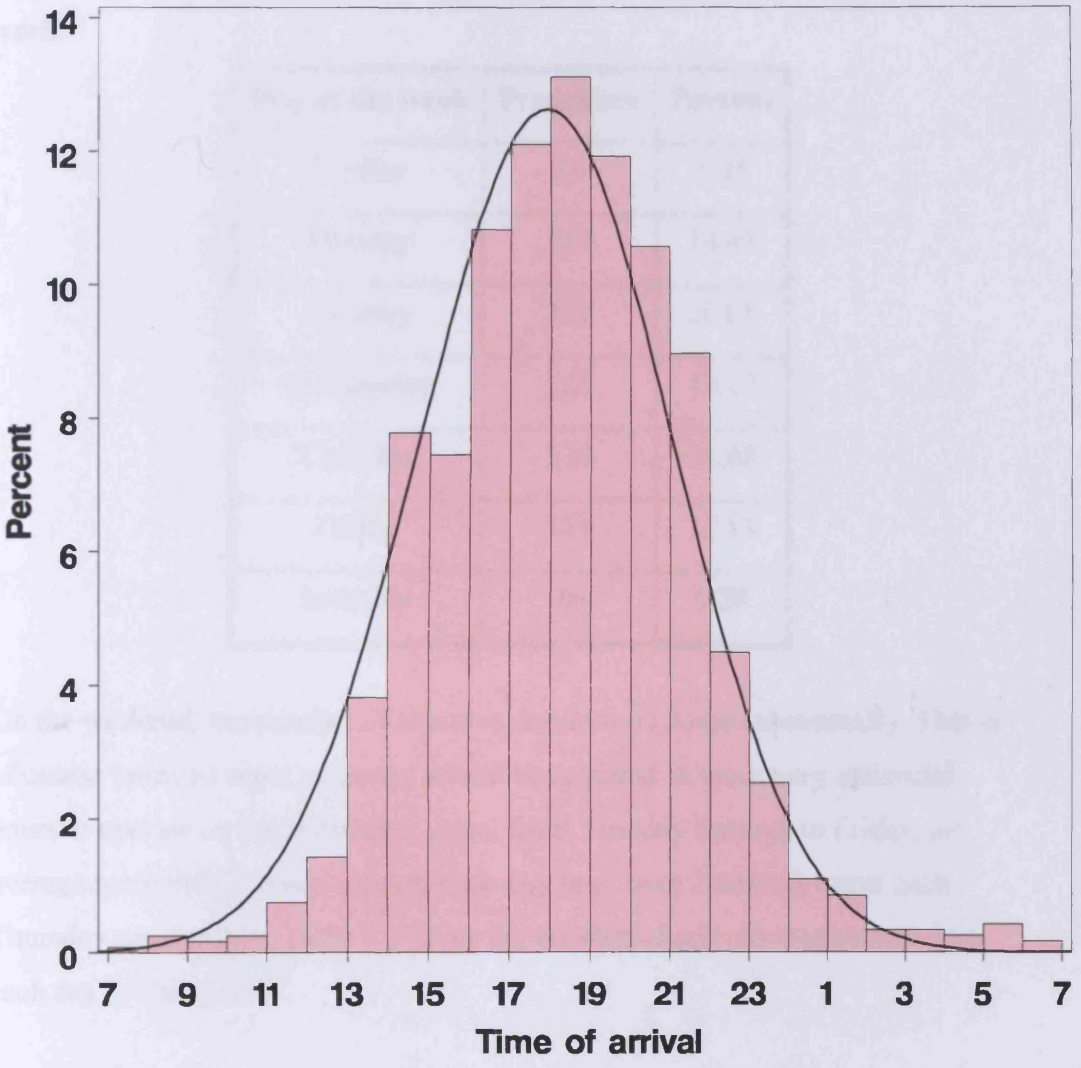


Figure 6.10: Arrival times of Elective surgery patients

The fitted normal distribution has a mean value of 17.91 and a standard deviation of 3.16. Unfortunately, the chi-square goodness of fit test yields a p-value which is significant, but gives a good indication of the pattern of arrival times.

Another interesting factor when considering Elective surgery admissions is the day of the week. Table 6.1 indicates the number of arrivals on each day of the week during the study period (note the study period is 165 weeks):

Table 6.1: the number of Elective Surgery admissions on each day of the week

Day of the week	Frequency	Percent
Sunday	16	1.35
Monday	159	13.45
Tuesday	237	20.05
Wednesday	226	19.12
Thursday	339	28.68
Friday	179	15.14
Saturday	26	2.20

On the weekend, the number of Elective admissions drops substantially. This is of course intuitive since surgeons would be required to work very antisocial hours to operate on these patients. Also, from Tuesday through to Friday, on average more than 1 arrival occurs each day (e.g. over 2 arrivals occur each Thursday on average). Table 6.2 gives the number of arrivals that occurred on each day of the week:

Table 6.2: The number of admissions on each day of the week

	Number of admissions						
	1	**2**	**3**	**4**	**5**	**6**	**Total**
Sunday	16	0	0	0	0	0	16
Monday	56	26	10	4	1	0	97
Tuesday	57	41	22	8	0	0	128
Wednesday	64	45	20	3	0	0	132
Thursday	38	48	30	16	9	1	142
Friday	63	35	11	2	1	0	112
Saturday	26	0	0	0	0	0	26
Total	320	195	93	33	11	1	653

The simulation model gathers information from the above table to generate inter-arrival times for the elective surgery patients. For each day of the week, the simulation model samples a random number between 0 and 1. It then uses this random number to decide how many arrivals should occur, if any. For example, each Sunday, if the random number is less than 16/165 then an arrival occurs, otherwise, no patients arrive that day. Each Monday, if the random number is less than 159/165 then at least one arrival occurs. If the random number is less than 159/165 but greater than 103/165, one arrival will occur. If the random number is less than 103/165 but greater than 51/165, two arrivals will occur. If the random number is less than 51/165 and greater than 21/165, three arrivals will occur. If the random number is less than 21/165and greater than 5/165, four arrivals will occur. Finally, if the random number is less than 5/165, five arrivals will occur.

Once the number of arrivals has been calculated, the arrival times are generated by sampling from the normal distribution with $\mu = 17.91$ and $\sigma = 3.16$ (see

Figure 6.10), then simply converting the given number into the hour of the day represented.

Table 6.3 summarises the inter-arrival times for each of the arrival sources that are used to populate the model.

Table 6.3: Summary of Inter-arrival distributions in hours

Source	N	Theoretical Mean and Standard Deviation	Actual Mean	Actual Standard Deviation
A&E	1134	24.7	24.7	25.1
Ward	916	30.5	30.5	31.4
Emergency	715	39.1	39.1	40.6
Other hospital	235	118.0	118.0	108.2
X-Ray	44	604.0	604.2	638.0
Elective	1182	-----	23.7	30.4

6.3.2. Service time distributions

There are six different service time distributions, each of which relates to a specific arrival source. This may not be intuitive but the data indicates significant differences between the length of stay distributions (or service time) recorded in hours for each of the arrival sources ($p < 0.0001$).

Chi-square goodness of fit tests were performed using Stat::Fit and it was found that the lognormal distribution was a good fit for each of the groups of patients, other than Elective surgery. The lognormal distribution is commonly fitted to length of stay data (Litvak et al. 2008, Lowery 1993) due to its long tails. Also, Hyperexponential distributions were fitted to the data to ensure that a phase-type distribution would not improve results. Up to five phases were used. Phase-type distributions have been utilised in the field of healthcare since, very often, they

can adequately model length of stay data. In 2008, Fackrell (Fackrell 2008) conducted a literature review of the use of phase-type distributions in health care modelling.

Accident and Emergency

Using Stat::Fit, the best fit for the A&E patients was the lognormal distribution with a shape parameter of 1.25 and a scale parameter of 4.

The fitted lognormal distribution yields a theoretical mean of 119.25 hours (compared with the actual mean of 123.59 hours) and a theoretical standard deviation of 231.57 hours (compared with 304.40 hours). Therefore the mean compares favourably with the theoretical approximation but the theoretical standard deviation is rather higher than the actual standard deviation.

A different set of distributions was then considered, namely the Hyperexponential distribution with a various number of phases.

Stat::Fit will not fit the Hyperexponential distribution; therefore a different method was sought. Many different fitting procedures were used, beginning with Least Squares Estimators and finishing with Maximum Likelihood Estimators.

Recalling that the probability density function of the Hyperexponential distribution with four phases is:

$$f(t) = \sigma\left(4\mu\sigma e^{-4\mu\sigma t}\right) + \omega\left(4\mu\omega e^{-4\mu\omega t}\right) + \nu\left(4\mu\nu e^{-4\mu\nu t}\right) + (1-\sigma-\omega-\nu)\left(4\mu(1-\sigma-\omega-\nu)e^{-4\mu(1-\sigma-\omega-\nu)t}\right)$$

The method of least squares was used to estimate the values of the parameters μ, σ, ω and ν. For the Accident and Emergency

arrivals, $\mu = 0.007764874$ per hour, $\sigma = 0.052587058$, $\omega = 0.447428003$, $\nu = 0.052541675$ and $1 - \sigma - \omega - \nu = 0.4474433$.

Looking at these parameters closely, it is clear to see that $\sigma \approx \nu$ and $\omega \approx 1 - \sigma - \omega - \nu$. If these were in fact identically equal, the 4 phase Hyperexponential distribution would collapse to the two stage Hyperexponential distribution. To ensure that these values do in fact yield the minimum value, the value of μ and ν were fixed ($\mu = 0.007764874$ and $\nu = 0.052541675$) and the objective function was plotted for different values of σ and omega.

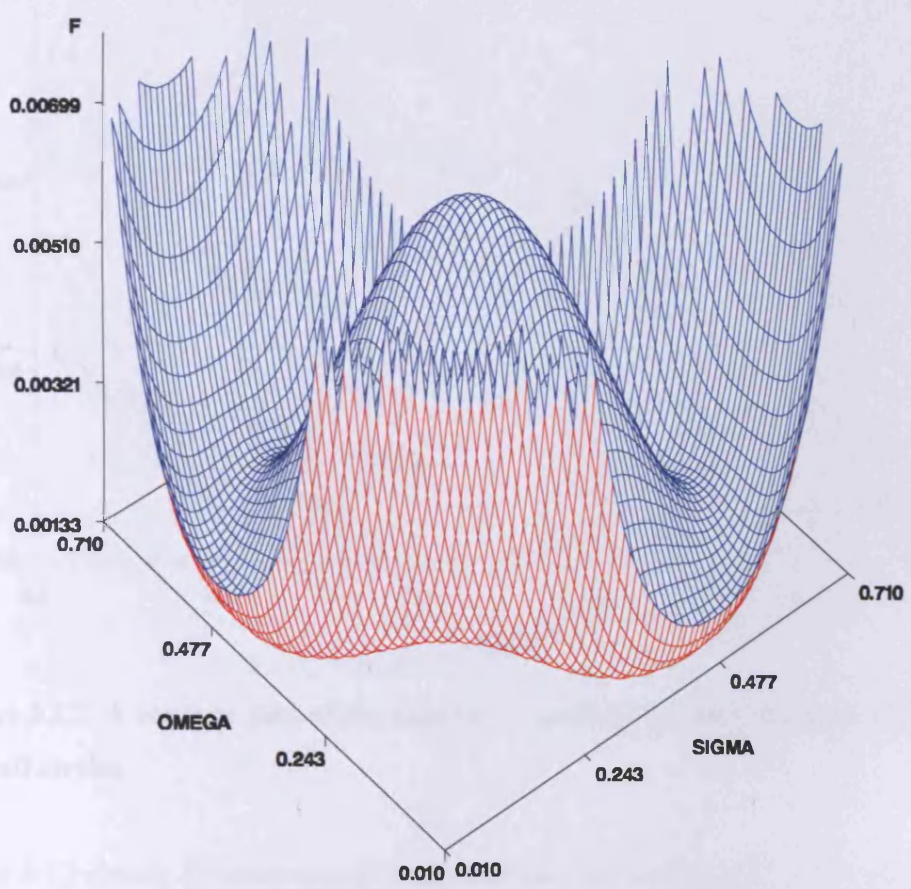


Figure 6.11: 3D plot of objective function, μ and ν fixed

Figure 6.11 above shows that we are indeed finding minimum values of our function and they are the global minima in the feasible region. Having said this, the graph does not clearly show the values of σ and omega which yield the minimum, thus, Figure 6.12, a contour plot is also included:

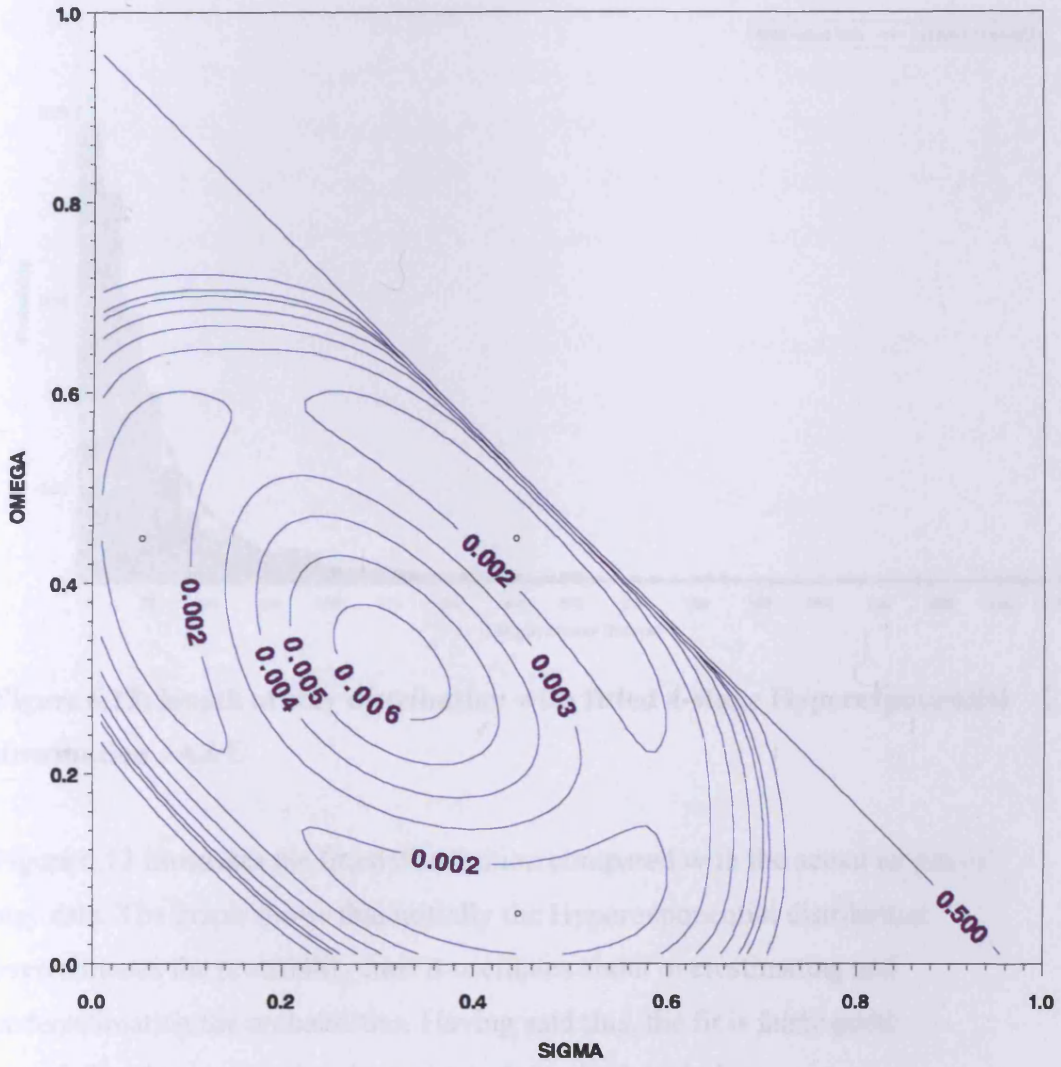


Figure 6.12: A contour plot of the objective function, μ and ν fixed, minima = small circles

Figure 6.12 clearly demonstrates that the minima occur at (0.44725, 0.052574), (0.052574, 0.44725) and (0.44725, 0.44725). For each minimum, the parameters of the four phase Hyperexponential distribution will be as was stated above.

Using these parameters, the mean length of stay is 128.79 hours with a standard deviation of 267.41 hours compared with a mean and standard deviation of 123.59 and 304.40 hours respectively from the data.

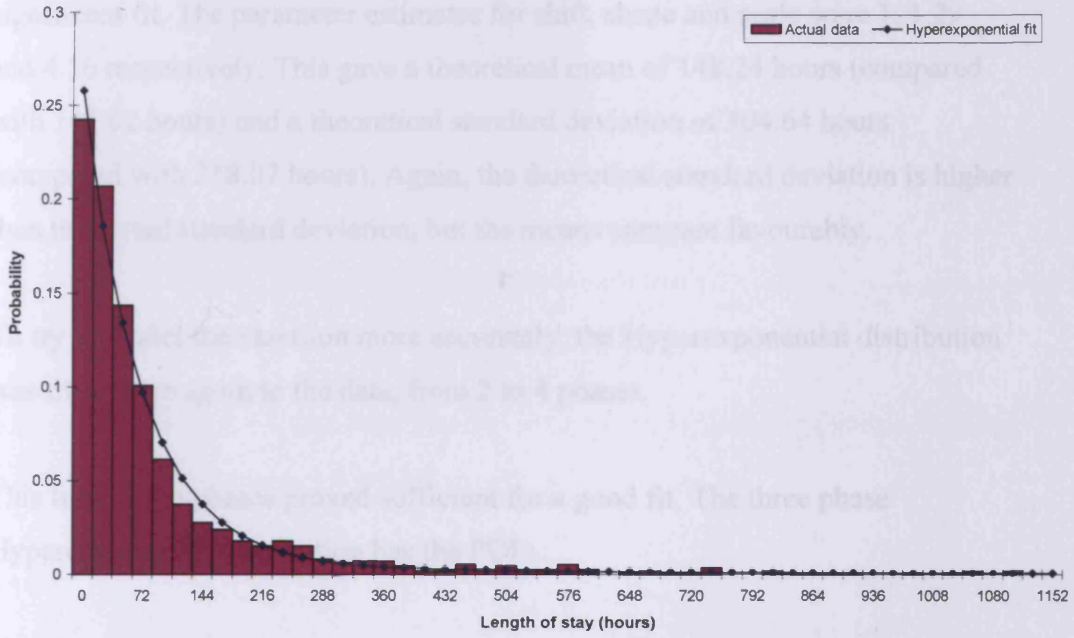


Figure 6.13: length of stay distribution with fitted 4-stage Hyperexponential distribution - A&E

Figure 6.13 illustrates the fitted distribution compared with the actual length of stay data. The graph shows that initially the Hyperexponential distribution overestimates the probability then it oscillates about overestimating and underestimating the probabilities. Having said this, the fit is fairly good, especially when comparing the mean and standard deviation.

Emergency Surgery

On closer inspection, an outlier was found at 8973 hours, this outlier was excluded from the statistical analysis. Again, the lognormal distribution gave a significant fit. The parameter estimates for shift, shape and scale were 1, 1.29 and 4.16 respectively. This gave a theoretical mean of 148.24 hours (compared with 141.02 hours) and a theoretical standard deviation of 304.64 hours (compared with 218.07 hours). Again, the theoretical standard deviation is higher than the actual standard deviation, but the means compare favourably.

To try to model the variation more accurately, the Hyperexponential distribution was fitted once again to the data, from 2 to 4 phases.

This time, three phases proved sufficient for a good fit. The three phase Hyperexponential distribution has the PDF:

$$f(t)=\sigma\left(3\mu\sigma e^{-3\mu\sigma t}\right)+\omega\left(3\mu\omega e^{-3\mu\omega t}\right)+(1-\sigma-\omega)\left(3\mu(1-\sigma-\omega)e^{-3\mu(1-\sigma-\omega)t}\right)$$

The method of least squares was used to estimate the values of the parameters μ, σ and ω. For the Emergency Surgery arrivals, $\mu = 0.007135078$, $\sigma = 0.177996263$ and $\omega = 0.702212231$.

Figure 6.14 shows the data with the corresponding Hyperexponential fit. Once again, the fit is fairly good with a few deviant points.

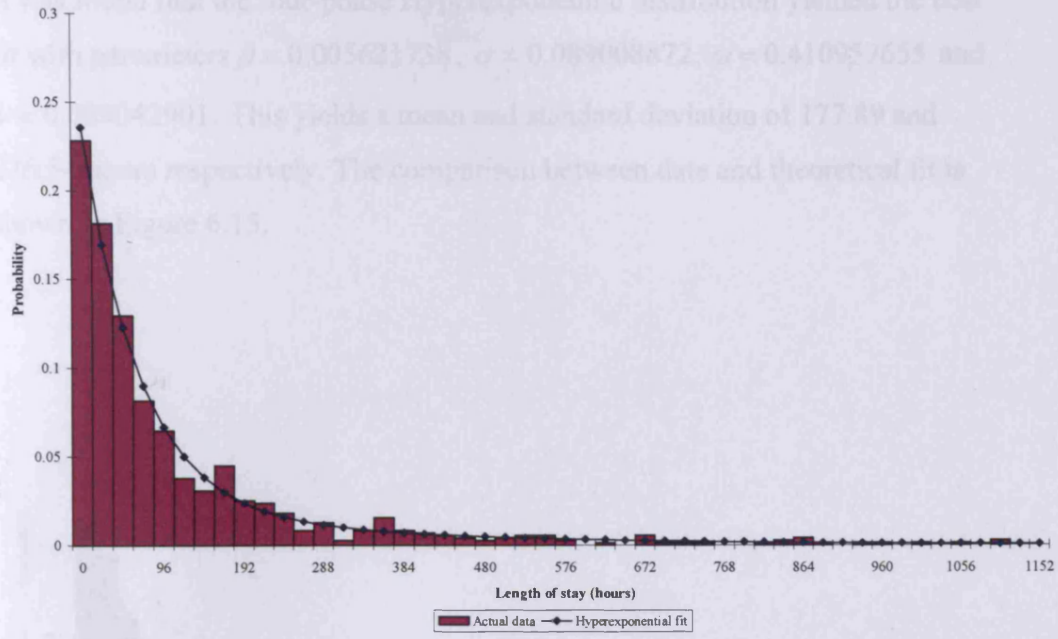


Figure 6.14: length of stay distribution with fitted 3-phase Hyperexponential distribution - Emergency Surgery

Ward

916 patients arrived at the Critical Care Unit from the Wards. Of these, four had missing values for their length of stay and four had an arrival time later than their discharge time. One patient had a length of stay of 7378 hours and was thus excluded from the analysis. The remaining 907 patients had a mean length of stay of 181.73 hours with a standard deviation of 273.61 hours. Using Stat::Fit, a lognormal distribution, with shift, shape and scale parameters of 1, 1.39 and 4.41 respectively, gave a significant fit. This distribution has a mean value of 217.17 hours and a standard deviation of 525.24. Therefore it is clearly not a very good fit, even though the null hypothesis of chi-square goodness of fit test cannot be rejected.

Since the lognormal proves to be an inadequate fit, it is certainly worth looking at the Hyperexponential class of distributions.

It was found that the four-phase Hyperexponential distribution yielded the best fit with parameters $\mu = 0.005621738$, $\sigma = 0.089008872$, $\omega = 0.410957655$ and $\nu = 0.089042901$. This yields a mean and standard deviation of 177.89 and 276.54 hours respectively. The comparison between data and theoretical fit is shown in Figure 6.15.

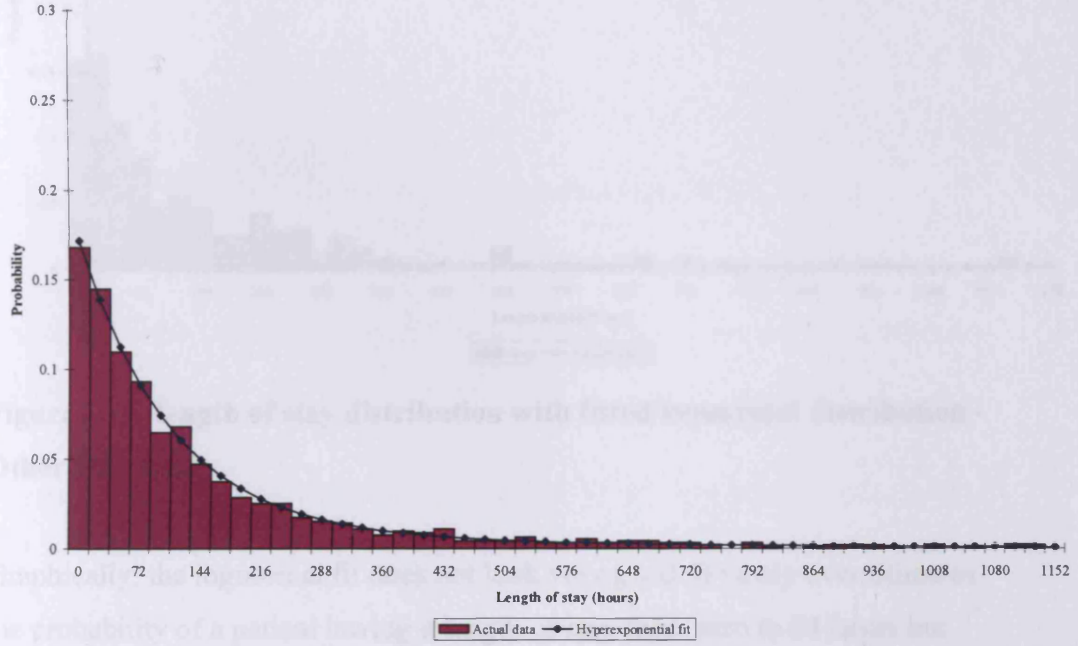


Figure 6.15: length of stay distribution with fitted 4-phase Hyperexponential distribution - Ward

Other Hospital

The Other Hospital data proved to be very clean. There were 233 observations in the data set and all were valid. No outliers were found. Once again, the appropriate distribution according to Stat::Fit was the lognormal distribution with shift, shape and scale parameters of 2, 1.32 and 4.48 respectively. This gave the theoretical mean to be 212.86 hours and the theoretical standard deviation to be 457.67 hours. The actual mean and standard deviation were 210.13 and 431.48 hours respectively. The lognormal distribution fits the data well in this instance. The fit is shown in Figure 6.16.

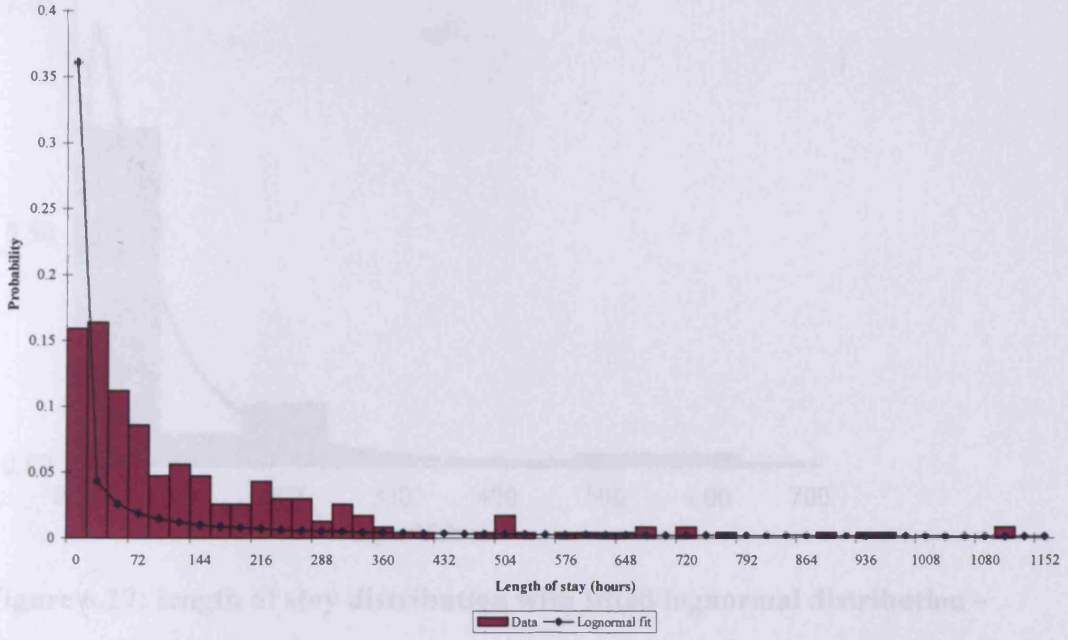


Figure 6.16: length of stay distribution with fitted lognormal distribution - Other Hospitals

Graphically, the lognormal fit does not look very good. It vastly overestimates the probability of a patient having a length of stay from zero to 24 hours but underestimates virtually every other category.

The Hyperexponential class of distributions did not effectively fit the data.

X-Ray

Arrivals from X-Ray are very rare in the CCU. Only 44 occurred between the 1st of April 2004 and the 31st of May 2007. Thus distribution fitting is not entirely sensible with such a small sample. However, the data was entered into Stat::Fit and again, the lognormal distribution proved to fit the data well with shape, scale and threshold parameters of 0.97, 3.99 and 1 respectively. This gave a theoretical mean and standard deviation of 87.53 and 108.15 hours respectively comparing with the actual mean of 91.27 hours and the actual standard deviation of 124.33 hours. The fit is shown in Figure 6.17.

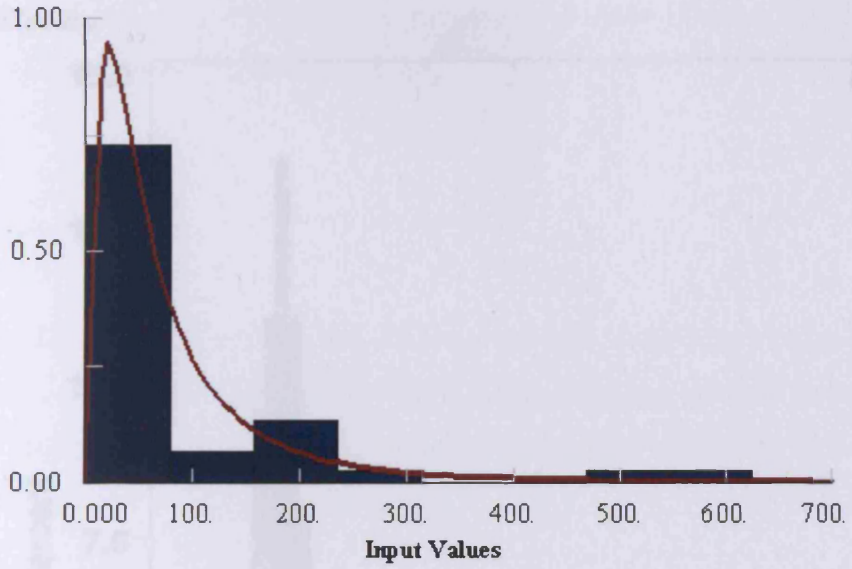


Figure 6.17: length of stay distribution with fitted lognormal distribution – X-Ray

It does not seem sensible to fit the two or three phase Hyperexponential to the X-Ray data due to the small sample size. Thus the lognormal distribution, as it's a fairly adequate fit, will be used.

Elective

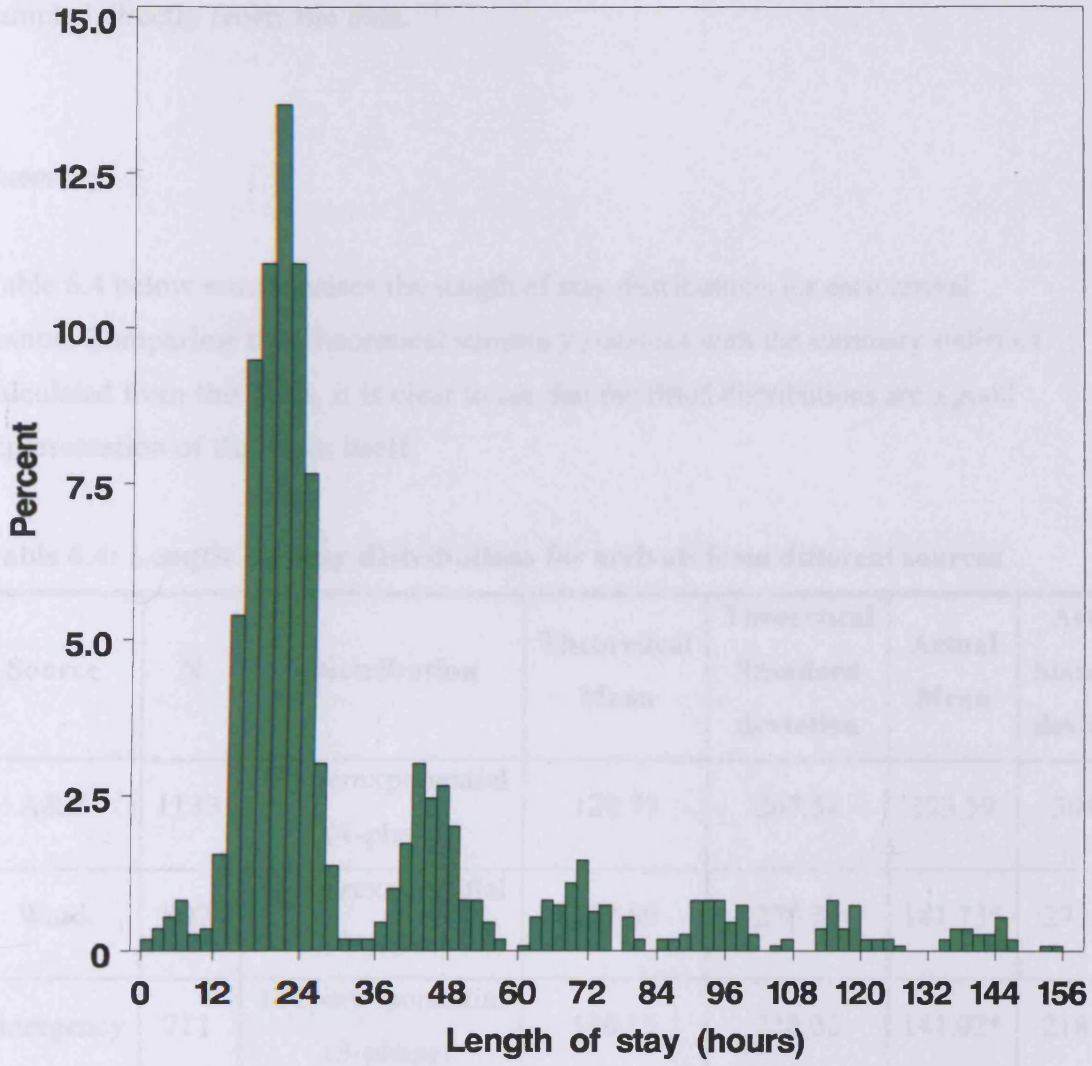


Figure 6.18: length of stay distribution with fitted lognormal distribution – Elective surgery

Once again, the Elective patients have a time-dependent length of stay distribution. The majority of patients stay in the CCU between 16 and 26 hours. The next group of patients have a length of stay roughly 24 hours longer than the first group. This is a result of the time of day at which the clinicians make their ward rounds and patients cannot be referred for discharge until the ward round has been completed. Rather than trying to find a distribution which adequately

models this data, the lengths of stay of arrivals from Elective surgery were sampled directly from the data.

Summary

Table 6.4 below summarises the length of stay distributions for each arrival source. Comparing the theoretical summary statistics with the summary statistics calculated from the data, it is clear to see that the fitted distributions are a good representation of the data itself.

Table 6.4: Length of stay distributions for arrivals from different sources

Source	N	Distribution	Theoretical Mean	Theoretical Standard deviation	Actual Mean	Actual Standard deviation
A&E	1133	Hyperexponential (4-phase)	128.79	267.51	123.59	304.40
Ward	907	Hyperexponential (4-phase)	177.89	276.54	181.73*	273.61*
Emergency	711	Hyperexponential (3-phase)	140.15	218.02	141.02*	218.07*
Other hospital	233	Lognormal	212.86	457.67	210.13	431.48
X-Ray	44	Lognormal	87.53	108.15	91.27	124.33

* Outliers have been removed from these groups, one from Ward (length of stay = 7,378 hours) and one from Emergency surgery (length of stay = 8,973 hours).

After consultation with hospital managers and consultants at UHW, it was clear that the actual length of stay of a patient was not a true reflection of the amount of time a bed was used. After each discharge, the bed and the surrounding area require intensive cleaning. This can take anything between two and eight hours.

Since no data is collected on this issue, it was assumed that a changeover time of 5 hours was required after discharge.

6.3.3. Verification and Validation

To allow the model to enter steady state conditions, a warm up period equivalent to 1 month was included. The model was subsequently run to represent one year and then replicated 100,000 times.

Validation of the model consisted of comparing the actual data with the output from the model. The model was run with 24 beds available – i.e. the number of beds currently present in the CCU. This resulted in 57 cancellations per year and an 82% bed-occupancy rate, compared with observed rates of 57 cancellations and 87% bed occupancy. However, in reality, no cancellations can occur here as the data only includes patients who have actually entered the CCU and hence had not had their operations cancelled. By looking at the data, it was clear that on occasions more than 24 beds were occupied in the CCU at any one time. In fact, twenty nine beds were occupied at one stage. Initial analysis revealed that on 9% of occasions there were more than 24 beds occupied; that is, one or more of the extra five beds stored for peak demand use were utilised. Taking account of the length of stay of patients, this amounts to over 7% of the available CCU time. The reason for this could be that patients are allowed to queue on trolleys if the hospital staff know there will be a bed available shortly. After consultation with staff, it was suggested that when a patient dies it takes time for the nurses to prepare that bed for the next patient which means that patients would feasibly be waiting on a trolley for a period of time. In order to avoid the situation where the model predicted cancellations occurring, it was run with 50 beds available.

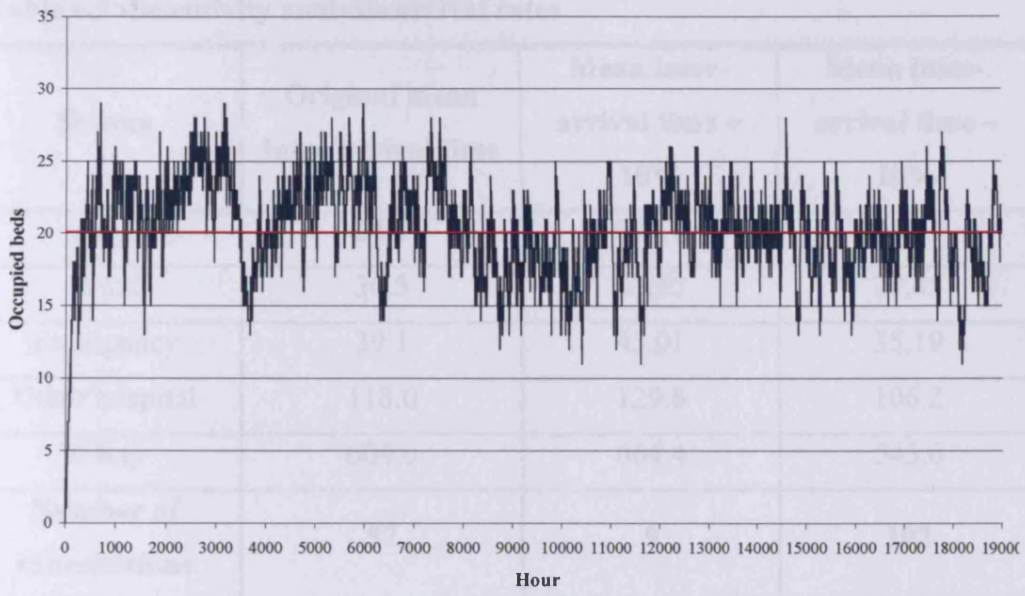


Figure 6.19: Actual bed occupancy profile for study period

The number of admissions in an average year, according to the data, was 1359, and the simulation recorded 1341 admissions. The mean number of beds occupied at any time during the year, as shown in Figure 6.19, was 20.10 according to the data, compared with 20.23 according to the simulation. This different was found not to be significant at the 95% level

Sensitivity analysis

A sensitivity analysis was performed on the inter-arrival times of the simulation model to ascertain how sensitive the results were to the mean inter-arrival times. The mean inter-arrival times were increased by 10% and then decreased by 10%. Note, this was only done for the unplanned admissions. The new mean inter-arrival times are given in Table 6.5 below:

Table 6.5: Sensitivity analysis arrival rates

Source	Original mean inter-arrival time	Mean inter-arrival time + 10%	Mean inter-arrival time – 10%
A&E	24.7	27.17	22.23
Ward	30.5	33.55	27.45
Emergency	39.1	43.01	35.19
Other hospital	118.0	129.8	106.2
X-Ray	604.0	664.4	543.6
Number of cancellations	**57**	**9**	**102**

The affect of the increase and decrease by 10% in parameter estimates is highly influential. For example, the number of cancellations given 24 beds and the original arrival rates is 57 (see Table 6.6), whereas when the mean inter-arrival time is increased by 10% the number of cancellations decreases to 9, and if the mean inter-arrival time is decreased by 10%, the number of cancellations increases to 102. But, since the original mean inter-arrival times above have been derived from a relatively large data set, there is no need to be overly concerned by the substantial effect of the sensitivity analysis.

6.3.4. "What-if" scenarios

The model was run for several "what-if" scenarios. Firstly, bed numbers were increased incrementally from 22 (which are fewer than the current number in the CCU) to 29 (this is the maximum number which can be accommodated), and the results produced are recorded in Table 6.6. As the number of beds increases, the mean number of occupied beds also increases. Although the mean number of occupied beds increases, the percentage of occupied beds decreases. Unsurprisingly, the number of cancellations in a year decreases as the number of beds increases.

Table 6.6 Summary measures using simulation model

Number of beds	Mean number of occupied beds	% occupied	Number of cancellations
22	19.15	87%	146
23	19.40	84%	101
24	19.76	82%	57
25	20.05	80%	20
26	20.18	78%	3
27	20.20	75%	0
28	20.20	72%	0
29	20.20	70%	0

There are two main interesting outcomes to this model and the various scenarios tested, namely the number of cancellations of Elective surgery annually, and bed-occupancy levels.

Firstly, the number of cancellations of Elective surgery which occur is of particular interest to the hospital. The cancellations recorded in this model occur on the day of admission. Such cancellations can be most traumatic for patients and will also result in lengthening of the hospital waiting lists, because cancellations often imply re-admissions at a later date.

The current formation of the Unit, with 24 beds, was taken to be the baseline. This formation resulted in 57 Elective cancellations with a bed-occupancy level of 82%. This cancellation level translates to roughly 5 cancellations per month which is very high considering that there are only approximately 30 Elective admissions scheduled each month. Also, the bed-occupancy level of 82% is considerable higher than the 60-70% occupancy guidelines given by the Intensive Care Society (ICS 1997).

With 25 beds in the CCU, the model showed that the number of cancellations would drop to approximately 20 per year with 80% bed-occupancy. The cost of the extra bed is around £361,350 per annum (assuming only one organ failure per Elective patient); therefore the resulting reduction of 37 cancellations implies that the cost per avoided cancellation is £9,747.

The main aim of this study was to investigate the number of Elective cancellations that could be avoided by having more bed resources, but if a new bed is to be added, emergency patients will also be able to use it. This is also a worthwhile use of NHS funding as the emergency patients could be seen more quickly, and trolley waits could be reduced. To calculate this additional benefit, the number of bed days generated by the addition of a bed was calculated and then appropriately scaled by the bed occupancy rate of the Unit. The number of bed days for elective and non-elective patients generated from the additional bed was calculated as follows. Using the mean length of stay of Elective patients and the number of additional Elective patients that could be seen given the additional bed, the number of bed days generated by an additional bed was decremented by the number of bed days the Elective patients would use. The average length of stay of emergency patients was then used to determine the average number of emergency patients that could utilise the remaining bed days so that this new bed would be full.

Table 6.6 shows that the number of cancellations and bed-occupancy levels continues to drop as additional beds are provided. Table 6.7 summarises the incremental cost of each cancellation avoided and also the "free" unplanned patients that would be seen as a result of adding a bed. As the number of additional beds increases, the cost of each avoided cancellation increases dramatically. Twenty nine beds, which is the maximum number which could be physically accommodated, results in no cancellations and a bed-occupancy level of 70%. Clearly, the bed-occupancy level of 70% is desirable and meets the ICS guidelines. The important question to consider is whether adherence to the ICS guidelines warrants the extra investment required for the additional CCU beds.

This is a matter for stakeholders to discuss. Turning to the issue of cancellations, with only 27 beds no Elective cancellations occur, implying that the investment in these extra beds, in terms of Elective cancellations avoided, is wasted.

Table 6.7: Costing for each additional bed and number of additional Emergency patients admitted to the new bed

Number of beds	Number of cancellations avoided	Cost per avoided cancellation	Number of additional emergency patients admitted
22	0	-----	0
23	45	£8,030	2
24	89	£8,213	6
25	126	£9,747	14
26	143	£21,750	26
27	146	£115,525	38
28	146	---------	50
29	146	---------	62

Another scenario tested was the creation of two separate CCUs, one for Elective (or planned) patients and one for "Unplanned" patients. Focussing on the ring-fenced bed situation, to achieve the 60-70% target recommended by the Intensive Care Society, 25 Emergency beds (Table 6.8) would be required and 1 Elective bed (Table 6.9). This would result in 295 cancellations each year. Clearly this is not an acceptable number of cancellations. If four Elective beds were funded, no Elective cancellations would occur, but these beds would only be utilised 65% of the time. Comparison of this "ring-fenced beds" scenario with the CCUs current formation, indicates that to achieve the ICS bed-occupancy target and to achieve a low number of cancellations, 29 beds are required, which suggests that ring-fencing is not beneficial.

Table 6.8 Summary measures using simulation model - Emergency patients only

Number of beds	Mean number of occupied beds	% occupied
22	17.58	80%
23	17.58	76%
24	17.58	73%
25	17.58	70%
26	17.58	68%
27	17.58	65%
28	17.58	63%
29	17.58	61%

Table 6.9 Summary measures using simulation model - Elective patients only

Number of beds	Mean number of occupied beds	% occupied	Number of cancellations
1	0.68	68%	295
2	1.73	87%	133
3	2.47	82%	19
4	2.60	65%	0
5	2.60	52%	0
6	2.39	40%	0

Additional data referring to the delay experienced by a patient prior to discharge was available but only for the period of time from the 1st of April 2004 to the 31st

of December 2005. From this data it was evident that 57% of patients who were admitted onto the CCU during this time period experienced a delayed discharge, and the mean delay experienced by a patient who was delayed was 33.64 hours. Investigation into the delay data found considerable time dependencies; therefore to account for this delay, a delay value was sampled from the original delayed discharge data and this value was subtracted from the calculated length of stay. Again, the model with this decreased length of stay was run for several different bed numbers.

Table 6.10 Summary measures using simulation model - removing delayed discharge

Number of beds	Mean number of occupied beds	% occupied	Number of cancellations
20	16.64	83%	110
21	16.94	81%	56
22	17.15	78%	16
23	17.23	75%	2
24	17.24	72%	0
25	17.25	69%	0

Table 6.10 addresses the issue of delayed discharge. As was previously mentioned, if the length of stay of each patient was decreased by the delay before discharge many of the cancellation issues are avoided. With 24 beds, no Elective cancellations would occur. In fact, the current formation of 24 beds yielded a 72% bed-occupancy level also. The addition of 1 bed causes the bed-occupancy rate to drop to 69% thus complying with the ICS guideline.

Finally, it is clear from the CCU admissions data that the majority of Elective surgery (around 83%) is performed from Tuesday until Friday. The remaining 17% is mostly on Monday, with 3% spread over the weekend. It is of interest to

discover whether a more uniform distribution of Elective admissions is beneficial to the overall running of the Unit given that the average number of Elective surgeries is approximately 365 per year. Two final scenarios, relating to the scheduling of Elective surgery were considered. The first being that Elective surgery would take place daily, with one operation each day (for seven days). The second being that either one or two surgeries are performed each day during the five days, Monday to Friday. For this scenario to be implemented, a significant change to working practice of surgeons would need to be agreed. The results follow in Figure 6.20.

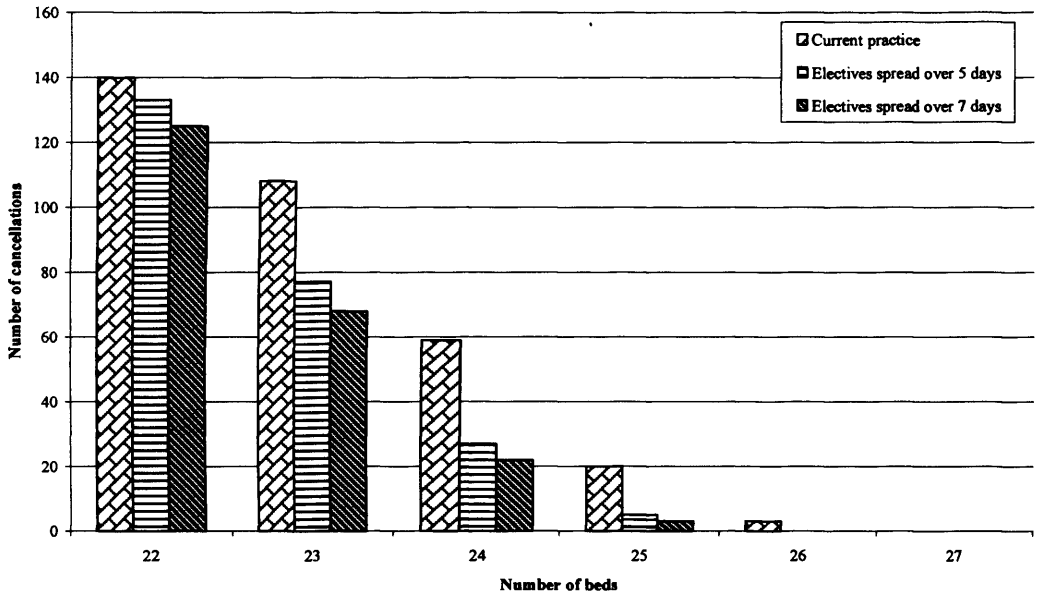


Figure 6.20 Number of Elective cancellations per year, for current practice, Electives spread over 5 days and Electives spread over 7 days

Figure 6.20 gives the number of cancellations which occur in one year when certain changes in practice occur. For example, when Elective surgery is performed over 5 days with either one or two surgeries performed daily, it is clear that the number of annual cancellations drops accordingly. If Elective surgery took place over 7 days, a larger affect is seen. The bed-occupancy levels for each of these scenarios are similar to those of the current formation and are therefore not included graphically. It is therefore clearly beneficial to schedule

the Elective surgery over at least 5 days and up to 7 days but the latter would require major changes in surgeons' working conditions which may result in clinician opposition.

It seems clear that the CCU under study is only able to achieve the high level of bed-occupancy by utilisation of the 5 additional beds when circumstances demanded this extra capacity. If the main 24 beds alone were used, the bed occupancy would have been 82%. But there are ways to use fewer beds if the circumstances outlined in the what-if scenarios exist.

6.4. Conclusion

The CCU, under its current formation, does not adhere to ICS bed-occupancy guidelines. Also, the waiting lists associated with Elective surgery patients who require a period of care in the CCU will also increase due to the number of cancellations which occur at the Unit. This study has highlighted several different configurations of the management of the CCU including resolving the issue of bed-blocking, ring-fencing beds for Elective and unplanned admissions and changing the working practice of theatre staff.

The most beneficial action for reducing bed-occupancy levels and cancellation rates is to resolve the issue of bed-blocking. If this were possible, the CCU would only require one additional bed to meet the ICS bed-occupancy guidelines.

If this were not possible, the next best option is to change the working practice of theatre staff. If Elective surgery was performed daily, the number of annual cancellations would drop considerably. This option would require the backing of clinicians and may be difficult to administer. However, if Elective surgery was performed on 5 days per week, in a more regular fashion, cancellation rates would also drop. This second option is much simpler to implement, with minimal disruption to theatre staff.

Chapter 7: Conclusions and further work

7.1. Introduction

This study was carried out to analyse data provided by the Critical Care Unit at the University Hospital of Wales. The theoretical work developed and extended in Chapters 2 and 3 was chosen due to its obvious application in medical research. The statistical analysis and Operational Research techniques utilised in Chapters 4 and 5 highlighted the important factors which influence bed occupancy and patient flow, such as length of stay and mortality. The simulation model developed in Chapter 6 can be used as an aid to decision making in the CCU regarding bed-numbers for example. This concluding Chapter addresses the initial objectives of the thesis and how these have been met. Also, it summarises some of the interesting and important findings of this study, including an indication of further research where appropriate.

7.2. Objective 1

To identify and investigate appropriate queueing systems which can model activities at the Critical Care Unit

In Chapter 2, batch service queueing theory was investigated. Once the steady state probability equations were derived, solving these required use of Laplace transforms and generating functions. The final solution required a transcendental equation to be solved which involved the use of iterative procedures.

As ρ and s (the batch size) increase, the mean number of customers in the system, L, will also increase. Several different service time distributions were considered including the Erlangian family and Hyperexponential with k phases.

When considering the distributions from the Erlangian family (the Negative Exponential, E_2, E_{10} and the Constant) we see that as k increases, L decreases. The values of z_i, when plotted on an Argand diagram, are nearer to the unit circle when ρ is small. Also, when s is large and ρ is small, these roots tend quickly towards the unit circle. Although, for larger values of k, ρ must be very small to have the roots approach the unit circle. When k is large, the roots are much nearer to the origin than when k is small.

Considering the Hyperexponential distribution, it is clear to see that when σ is small, the value of L is larger. The values of z_i are further away from the origin and closer together when σ is small.

It would be very interesting to apply this theoretical work to a practical real-life situation. For example, in a healthcare context, it would be interesting to investigate the benefit of parallel operating theatres. Specifically, if two operating theatres were set up in parallel to one another, and if two patients arrive singly, both requiring use of the operating theatre, would it be beneficial to use both operating theatres in parallel at the same time, using one surgical team, or would it be more beneficial to treat one patient at a time? Questions of this nature can be answered by applying the theory of batch service.

Also, queueing theory is not utilised as much as other Operational Research techniques, such as simulation, in real-life applications. One of the reasons for this may be the complicated nature of some of the solution methods which are used (for example the iterative technique to find the z_i in this case). The statistical analysis package SAS has an Operational Research module which has a simulation branch and an optimisation branch, but has no queueing theory. One possible avenue for future work would be to develop a queueing theory add-on for SAS which was user- friendly in nature. The user would select a queue of their choice and the program would then generate summary measures and probabilities automatically.

Chapter 3 began with the derivation of the steady state equations and summary measures for the $M^{(b)}/M/1$ queue. Then, the work was extended to investigate the Hyperexponential service time distribution, with one server. Finally, the steady state queueing equations of the $M^{(b)}/He/c$ queue.

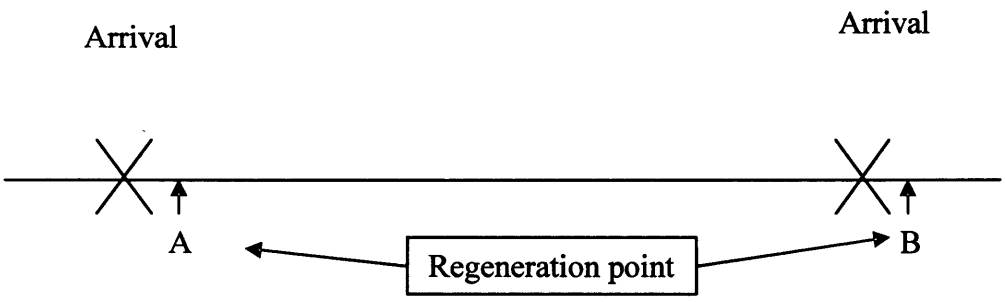
The steady state for the $M^{(b)}/He/c$ queueing system were solved using an iterative method. Parameters derived from the CCU data were then utilised to populate this queueing system and to obtain results for the mean number of customers in the system. The queueing system proved to be a good approximation to the CCU but it is possible that further improvements can be made in future.

Several improvements could be made to this queueing system. Some of these will be discussed here. Firstly, the service time distribution could be extended from the two-phase Hyperexponential distribution to the four-phase Hyperexponential distribution. The length of stay data follows the four-phase distribution more closely thus making the queueing system more reflective of reality. To do this, a new solution method would be required. The method utilised in Chapter 3 to find the steady-state equations would be difficult to use in this extension since the notation would be complicated and the number of equations requiring consideration would be very large and cumbersome. Rather, the generating function method would need to be applied.

Secondly, priority queueing could be utilised here. As has been mentioned in this thesis several times, unplanned admissions have priority over planned admissions and will cause planned admissions to be cancelled if there are no available beds.

The final area for further work considered would be to change the time between arrivals to be Deterministic rather than Markovian.

D/M/c



Let $P_n(R)$ be the probability of n customers in the system at regeneration point R, $n = 0, 1, 2, ...$

Let $k_j(i)$ be the probability that j customers will depart in the inter-arrival time when there are i customers present in the system at the beginning of the interval, $j = 0, 1, 2, ..., i = 0, 1, ...$ $j \le i$

Then

$$P_1(B) = P_1(A)k_1(1) + P_2(A)k_2(2) + P_3(A)k_3(3) + ...$$
$$P_2(B) = P_1(A)k_0(1) + P_2(A)k_1(2) + P_3(A)k_2(3) + ...$$
$$P_3(B) = \quad P_2(A)k_0(2) + P_3(A)k_1(3) + ...$$

$$P_r(B) = P_{r-1}(A)k_0(r-1) + P_r(A)k_1(2) + P_{r+1}(A)k_2(r+1) + ...$$
$$= \sum_{j=r-1}^{\infty} P_j(A)k_{i+1-r}(i)$$

Now, $k_j(i)$ takes on two different forms, depending on if $i \le c$ or $i > c$.

$i \le c$

$k_j(i) = {}^iC_j p^j q^{i-j}$ where $p = 1 - e^{-\mu}$

$i > c$

Now, there are some customers waiting for service, so if a server finishes with one of the initial customers, another would enter service immediately:

$$k_j(i) = \frac{(c\mu)^j e^{-c\mu}}{j!} \text{ if } j \leq i-c-1$$

$$k_j(i) = \int_0^1 \frac{c\mu(c\mu t)^{i-c-1} e^{-c\mu t}}{(i-c-1)!} {}^cC_{j-i+c} p(1-t)^{j-i+c} q(1-t)^{i-j} dt \text{ where } p(t) = 1-e^{-\mu t}$$

It is clear from the above description that this queueing system is difficult to solve and that the solution is not trivial, thus providing an obvious avenue for future research.

7.3. Objective 2

Understand the factors which affect length of stay and outcome at the Critical Care Unit

Chapter 4 used both CART analysis and linear regression analysis to attempt to model CCU length of stay. It is clear from the analysis undertaken in Chapter 4 that length of stay is a difficult and complex entity to model.

The CART analysis demonstrated that a total of 12 distinct variables were required to create 20 groups of patients with a similar length of stay profile. This caused a reduction in variance of 12%. The linear regression analysis indicated that a total of 39 variables were required to model length of stay which yielded an R^2 value of 0.3614. However, length of stay could be modelled with an R^2 value of 0.2955 with only 12 variables.

Many of the variables which appeared in the CART analysis also appeared in both regression models. In the majority of cases the conclusions drawn about the variables in both pieces of analysis were in agreement thus strengthening the data analysis.

The nodes constructed from the CART analysis could be utilised as arrival sources for a simulation model. This model would have 20 arrival sources (corresponding to the final nodes) with different length of stay distributions. Also, the regression model could be utilised in a simulation to determine a patient's length of stay in the Critical Care Unit.

In Chapter 5, an investigation into the factors which affected mortality was presented. Clearly the likelihood of survival is of fundamental importance to the patients and their relatives, so this work sought to summarise the most important factors influencing mortality risk. The analysis undertaken was similar to that of Chapter 4, beginning with CART analysis and then moving onto Logistic regression.

The evidence in Chapter 5 suggests that mortality risk is again very complex and fourteen variables were included in the splitting rules for the CART analysis thereby reducing the impurity of the data by over 25%. Twenty one variables appeared in the logistic regression analysis and yielded a maximum rescaled R^2 value of 0.3764. Only five variables were concordant between the two pieces of analysis, namely ADO, Age, Glucose, ORGSCORE and SYSBP. Both pieces of analysis gave similar results for ADO, Age and ORGSCORE but some discrepancies were found for Glucose and SYSBP. A reason for this discrepancy may be that both Glucose and SYSBP appear at the bottom of the CART analysis tree.

Other significant factors which contribute to mortality risk, such as chronic disease and operative complications, were also considered. Haematological

cancer and emergency surgery were found to have a large effect on mortality risk.

When considering the five logistic regression equations, two variables appeared in all models, DIALYSED1 and TISSPOINT and on each occasion both had a positive valued parameter. Several variables appeared in four of the models, the most interesting being TISSCODE68, and each appearance yielded a negative value for the parameter.

Apart from using the CART analysis nodes to classify patients in terms of their mortality risk in real life, and using the Logistic regression equation to give family and friends an indication of the risk of mortality of their loved one, the analysis could be used, once more, to populate a simulation model. The research undertaken in Chapters 4 and 5 could be combined and a model could be built which would incorporate both length of stay and mortality.

Also, it would be very interesting to investigate which variables affect both length of stay and mortality. For example, the vairable ORGSCORE appeared in the CART analysis and regression analysis for both Chapters 4 and 5. This indicates that this variable is crucial in determining both of these factors and thus warrants further investigation.

7.4. Objectives 3 and 4

Use the insight gained from the previous objectives to build a simulation model of the Critical Care Unit.

Analyse the results from the simulation model to show how varying some parameters will affect cancellations and cost.

Finally, in Chapter 6, a discrete event simulation model was built in VBA for Excel, to model bed-occupancy in the CCU. This model was used to test various what-if scenarios and the results were presented and discussed.

Currently, the CCU does not adhere to Intensive Care Society bed-occupancy guidelines (maximum of 60-70% occupancy). Since the bed-occupancy level is high, elective cancellations inevitably follow resulting in the extension of waiting lists.

Bed-occupancy levels can be controlled in a number of ways. Firstly, additional beds can be funded. For each additional bed, the number of cancellations will drop and the bed-occupancy percentage also drops. Unfortunately, the cost of each additional bed is approximately £361,350 and it is therefore not always cost effective to add beds.

The most sensible option for reducing bed occupancy is to deal with the problem of delayed discharge. If delayed discharge were eradicated, the CCU would require only one additional bed to meet the ICS bed-occupancy guidelines.

Less dramatic improvements can be made by changing the working practices of the surgeons. If elective surgery were scheduled in a uniform way each day, the number of cancellations could be greatly reduced.

An interesting avenue for future research would be to incorporate the research undertaken in Chapters 4, 5 and 6 to create a new simulation model which would act as diary planner. Each morning, before any elective surgery is undertaken, information regarding elective patients who are near the top of the waiting list and are due to undergo surgery within the next few days, could be input into the model to obtain an estimated length of stay and mortality risk. This information could be utilised to schedule the appropriate elective patients who are near the top of the list given the current bed-occupancy level of the CCU. Of course, the model would also indicate the probability of an unplanned admission and if one did occur then an estimate of length of stay and mortality risk could also be calculated for this patient.

7.5. Final conclusion

This study has considered many aspects of theoretical and practical application of mathematical modelling in the CCU setting. The analyses have shown that by developing and applying novel queuing theory, improvements can be made in the management and efficiency of the CCU at UHW.

References

Abolniko L M (1967) A multichannel queueing system with group arrival of demands. *Engineering Cybernetics* (4) 39-48.

Adan I and Resing J A C (2000) Multi-server batch-service systems. *Statistica Neerlandica* 54 (2) 202-220.

Alfa A S (2005) Analysis of A Gl/G(Y)/1 system in discrete-time. *Stochastic Models* 21 (1) 185-213.

Alfa A S and Yannopoulos E (1993) Bulk Server Queues with Markov Renewal Type of Service. *Journal of the Operational Research Society* 44 (3) 271-278.

Armero C and Conesa D (2000) Prediction in Markovian bulk arrival queues. *Queueing Systems* 34 (1-4) 327-350.

Armero C and Conesa D (2006) Bayesian hierarchical models in manufacturing bulk service queues. *Journal of Statistical Planning and Inference* 136 (2) 335-354.

Arora K L (1964) Two-Server Bulk-Service Queuing Process. *Operations Research* 12 (2) 286-294.

Artalejo J R, Atencia I and Moreno P (2005) A discrete-time Geo([X])/G/1 retrial queue with control of admission. *Applied Mathematical Modelling* 29 (11) 1100-1120.

Ashton R, Hague L, Brandreth M, Worthington D and Cropper S (2005) A simulation-based study of a NHS Walk-in centre. *Journal of the Operational Research Society* 56 (2) 153-161.

Baba Y (1996) A bulk service GI/M/1 queue with service rates depending on service batch size. *Journal of the Operations Research Society of Japan* 39 (1) 25-35.

Baek J W and Lee H W (2007) Waiting time analysis of a batch arrival queue under a workload control. *International Journal of Industrial Engineering-Theory Applications and Practice* 14 (3) 298-306.

Bahary E and Kolesar P (1972) Multilevel Bulk Service Queues. *Operations Research* 20 (2) 406-420.

Bailey N T J (1954) On Queueing Processes with Bulk Service. *Journal of the Royal Statistical Society Series B-Statistical Methodology* 16 (1) 80-87.

Baily D E and Neuts M F (1981) Algorithmic methods for multi-server queues with group arrivals and Exponential services. *European Journal of Operational Research* 8 (2) 184-196.

Banik A D and Gupta U C (2007) Analyzing the finite buffer batch arrival queue under Markovian service process: GI(X)/MSP/1/N. *Top* 15 (1) 146-160.

Bertsimas D and Papaconstantinou X (1988) On the Steady-State Solution of the M/C2(a,B)/S Queuing System. *Transportation Science* 22 (2) 125-138.

Bocharov P P and Yakoutina S V (1994) Stationary queue-length distribution in a queueing system with limited capacity, batch arrival, and phase-type service time. *Automation and Remote Control* 55 (9) 1324-1333.

Brahimi M and Worthington DJ (1991) The finite capacity multi-server queue with inhomogeneous arrival rate and discrete service time distribution and its

application to continuous service time problem. *European Journal of Operational Research* 50 310-324.

Bratiychuk M S and Kempa W (2003) Application of the superposition of renewal processes to the study of batch arrival queues. *Queueing Systems* 44 (1) 51-67.

Briere G and Chaudhry M L (1987) Computational analysis of single-server bulk-arrival queues: GI(X)/M/1. *Queueing systems* 2 (2) 173-185.

Briere G and Chaudhry M L (1988) Computational analysis of single-server bulk-arrival queues M(X)/G/1. *Computers & Operations Research* 15 (3) 283-292.

Briere G and Chaudhry M L (1989) Computational Analysis of Single-Server Bulk-Service Queues, M/Gy/1. *Advances in Applied Probability* 21 (1) 207-225.

Chaudhry M L and Chang S H (2004) Analysis of the discrete-time bulk-service queue Geo/G(Y)/1/N+B. *Operations Research Letters* 32 (4) 355-363.

Chaudhry M L and Gupta U C (1992) Exact computational anlaysis of waiting-time distribitions of single-server bulk arrival queues - MX/G/1. *European Journal of Operational Research* 63 (3) 445-462.

Chaudhry M L and Gupta U C (1997) Queue-length and waiting-time distributions of discrete-time GI(X)/Geom/1 queueing systems with early and late arrivals. *Queueing Systems* 25 (1-4) 307-324.

Chaudhry M L and Gupta U C (1998) Performance analysis of discrete-time finite-buffer batch-arrival GI(X)/Geom/1/N queues. *Discrete Event Dynamic Systems-Theory and Applications* 8 (1) 55-70.

Chaudhry M L and Gupta U C (1999) Modelling and analysis of M/G(a,b)/1/N queue – A simple alternative approach. *Queueing Systems* 31 (1) 95-100.

Chaudhry M L and Gupta U C (2001) Computing waiting-time probabilities in the discrete-time queue: GI(X)/G/1. *Performance Evaluation* 43 (2-3) 123-131.

Chaudhry M L, Gupta U C and Agarwal M (1992a) Computational analysis of the distribution of number in the system M(X)/G/1 - an alternative approach. *Infor* 30 (1) 30-43.

Chaudhry M L, Gupta U C and Goswami V (2001) Modeling and analysis of discrete-time multiserver queues with batch arrivals: GI(X)/Geom/*m*. *Informs Journal on Computing* 13 (3) 172-180.

Chaudhry M L, Gupta U C and Madill B R (1991) Computational Aspects of Bulk-Service Queuing System with Variable Capacity and Finite Waiting Space - M/Gy/1/N+B. *Journal of the Operations Research Society of Japan* 34 (4) 404-421.

Chaudhry M L and Kim J D (2004) Complete analytic and computational analyses of the discrete-time bulk-arrival infinite-server system: GI(X)/GEOM/infinity. *Computers & Operations Research* 31 (13) 2119-2135.

Chaudhry M L and Kim N K (2003) A complete and simple solution for a discrete-time multi-server queue with bulk arrivals and deterministic service times. *Operations Research Letters* 31 (2) 101-107.

Chaudhry M L and Lee A M (1972) Single channel constant capacity bulk service queueing process with an intermittently available server. *INFOR* 10 (3) 284-291.

Chaudhry M L, Madill B R and Brière G (1987) Computational analysis of steady-state probabilities of M/G(a,b)/1 and related nonbulk queues. *Queueing Systems* 2 (2) 93-114.

Chaudhry M L and Templeton J G C (1981) The queueing system M/G(B)/1 and its ramifications. *European Journal of Operational Research* 6 (1) 56-60.

Chaudhry M L and Templeton J G C (1983) *A first course in bulk queues*. Wiley.

Chaudhry M L, Templeton J G C and Medhi J (1992b) Computational results of multiserver bulk-arrival queues with constant service time M(X)/D/C. *Operations Research* 40 S229-S238.

Chen S P (2004) A mathematical programming approach to fuzzy queues with batch arrivals. *Engineering Optimization* 36 (6) 635-644.

Chen S P (2005) Parametric nonlinear programming approach to fuzzy queues with bulk service. *European Journal of Operational Research* 163 (2) 434-444.

Choi B D and Han D H (1994) G/M(a,B)/1 Queues with Server Vacations. *Journal of the Operations Research Society of Japan* 37 (3) 171-181.

Choudhury G (1998) On a batch arrival Poisson queue with a random setup time and vacation period. *Computers & Operations Research* 25 (12) 1013-1026.

Choudhury G (2000) An M-X/G/1 queueing system with a setup period and a vacation period. *Queueing Systems* 36 (1-3) 23-38.

Choudhury G (2002) A batch arrival queue with a vacation time under single vacation policy. *Computers & Operations Research* 29 (14) 1941-1955.

Choudhury G (2007) A two phase batch arrival retrial queueing system with Bernoulli vacation schedule. *Applied Mathematics and Computation* 188 (2) 1455-1466.

Choudhury G and Krishnamoorthy A (2004) Analysis of the M-x/G/1 queue with a random setup time. *Stochastic Analysis and Applications* 22 (3) 739-753.

Choudhury G and Madan K C (2004) A two phase batch arrival queueing system with a vacation time under Bernoulli schedule. *Applied Mathematics and Computation* 149 (2) 337-349.

Choudhury G and Madan K C (2005) A two-stage batch arrival queueing system with a modified Bernoulli schedule vacation under N-policy. *Mathematical and Computer Modelling* 42 (1-2) 71-85.

Choudhury G and Paul M (2004) A batch arrival queue with an additional service channel under N-policy. *Applied Mathematics and Computation* 156 (1) 115-130.

Choudhury G and Paul M (2006) A batch arrival queue with a second optional service channel under N-policy. *Stochastic Analysis and Applications* 24 (1) 1-21.

Choudhury G, Tadj L and Paul M (2007) Steady state analysis of an M-x/G/1 queue with two phase service and Bernoulli vacation schedule under multiple vacation policy. *Applied Mathematical Modelling* 31 (6) 1079-1091.

Chydzinski A (2006) Duration of the buffer overflow period in a batch arrival queue. *Performance Evaluation* 63 (4-5) 493-508.

Chydzinski A and Winiarczyk R (2008) On the blocking probability in batch Markovian arrival queues. *Microprocessors and Microsystems* 32 (1) 45-52.

Claeys D, Walraevens J, Laevens K and Bruneel H (2007) Discrete-time queueing model with a batch server operating under the minimum batch size rule. *Next Generation Teletraffic and Wired/Wireless Advanced Networking, Proceedings* 4712 248-259.

Clermont G, Angus D C, DiRusso S M, Griffin M and Linde-Zwirble W T (2001) Predicting hospital mortality for patients in the intensive care unit: A comparison of artificial neural networks with logistic regression models. *Critical Care Medicine* 29 (2) 291-296.

Coats T J and Michalis S (2001) Mathematical modelling of patient flow through an accident and emergency department. *Emergency Medicine Journal* 18 (3) 190-192.

Cohen J W (1976) Single-server queue with group arrivals. *Journal of Applied Probability* 13 (3) 619-622.

Cong T D (1994) On the MX/G/Infinity queue with heterogeneous customers in a batch. *Journal of Applied Probability* 31 (1) 280-286.

Conolly B W (1960) Queueing at a Single Serving Point with Group Arrival. *Journal of the Royal Statistical Society. Series B (Methodological)* 22 (2) 285-298.

Cooke C R, Kahn J M, Caldwell E, Okamoto V N, Heckbert S R, Hudson L D and Rubenfeld G D (2008) Predictors of hospital mortality in a population-based cohort of patients with acute lung injury. *Critical Care Medicine* 36 (5) 1412-1420.

Coombs M and Lattimer V (2007) Safety, effectiveness and costs of different models of organising care for critically ill patients: Literature review. *International Journal of Nursing Studies* 44 (1) 115-129.

Cooper J K and Corcoran T M (1974) Estimating Bed Needs by Means of Queuing Theory. *New England Journal of Medicine* 291 (8) 404-405.

Costa A X, Ridley S A, Shahani A K, Harper P R, Senna V D and Nielsen M S (2003) Mathematical modelling and simulation for planning critical care capacity. *Anaesthesia* 58 320-327.

Cullen D J, Civetta J M, Briggs B A and Ferrara L C (1974) Therapeutic intervention scoring system: a method for quantitative comparison of patient care. *Critical Care Medicine* 2 (2) 57-60.

Dagsvik J (1975a) The general bulk queue as a matrix factorisation problem of the Wiener-Hopf type I. *Advances in Applied Probability* 7 (3) 636-646.

Dagsvik J (1975b) The general bulk queue as a matrix factorisation problem of the Wiener-Hopf type II. *Advances in Applied Probability* 7 (3) 647-655.

Denteneer D and van Leeuwaarden J (2005) The delayed bulk service queue: a model for a reservation process. *Performance Challenges for Efficient Next Generation Networks, Vols 6a-6c* 6A-6C 909-918.

DOH 2008 *18 weeks Referral to Treatment statistics.* http://www.dh.gov.uk/en/Publicationsandstatistics/Statistics/Performancedataandstatistics/18WeeksReferraltoTreatmentstatistics/index.htm Accessed: 01/12/2008

DOH (2006) *NHS Trust and PCT Combined Reference Cost Schedules.* http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_062884 Accessed: 18/03/2008

DOH (2007) *NHS Trust and PCT Combined Reference Cost Schedules.* http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_082571 Accessed: 18/03/2008

Downton F (1955) Waiting Time in Bulk Service Queues. *Journal of the Royal Statistical Society Series B-Statistical Methodology* 17 (2) 256-261.

Downton F (1956) On Limiting Distributions Arising in Bulk Service Queues. *Journal of the Royal Statistical Society Series B-Statistical Methodology* 18 (2) 265-274.

Duke G J, Green J V and Briedis J H (2004) Night-shift discharge from Intensive Care unit increases the mortality-risk of ICU survivors. *Anaesthesia and Intensive Care* 32 (5) 697-701.

Fabens A J (1961) The Solution of Queuing and Inventory Models by Semi-Markov Processes. *Journal of the Royal Statistical Society Series B-Statistical Methodology* 23 (1) 113-127.

Fabens A J and Perera A (1963) A correction to "The Solution of Queuing and Inventory Models by Semi-Markov Processes". *Journal of the Royal Statistical Society Series B-Statistical Methodology* 25 (2) 455-456.

Fackrell M (2008) Modelling healthcare systems with phase-type distributions *Health Care Management Science.*

Fakinos D (1990) On the M/G/K group-arrival loss system. *European Journal of Operational Research* 44 (1) 75-83.

Fakinos D (1991) The relation between limiting queue size distributions at arrival and departure epochs in a bulk queue. *Stochastic Processes and Their Applications* 37 (2) 327-329.

Falin G (1994) The M(K)/G/Infinity batch arrival queue with hterogeneous dependent demands. *Journal of Applied Probability* 31 (3) 841-846.

Falin G I (1988) On a multiclass batch arrival retrial queue. *Advances in Applied Probability* 20 (2) 483-487.

Ferrandiz J M (1993) The BMAP/GI/1 queue with server set-up times and server vacations. *Advances in Applied Probability* 25 (1) 235-254.

Franx G J (2005) The M-x/D/c batch arrival queue. *Probability in the Engineering and Informational Sciences* 19 (3) 345-349.

Friedrich J O, Wilson G and Chant C (2006) Long-term outcomes and clinical predictors of hospital mortality in very long stay intensive care unit patients: a cohort study. *Critical Care* 10 (2).

Friere A X, Umpierrez G E, Afessa B, Latif K A, Bridges L and Kitabchi A E (2002) Predictors of Intensive Care Unit and Hospital Length of Stay in Diabetic Ketoacidosis. *Journal of Critical Care* 17 (4) 207-211.

Fuhrmann S W (1991) Second moment relationships for waiting times in queueing systems with Poisson input. *Queueing Systems* 8 397-406.

Gaver D P (1959) Markov Chain Analysis of a Waiting-Line Process in Continuous Time. *The Annals of Mathematical Statistics* 30 (3) 698-720.

Ghare P M (1968) Multichannel queuing system with bulk service. *Operations Research* 16 (1) 189-192.

Goldfrad C and Rowan K (2000) Consequences of discharges from intensive care at night. *The Lancet* 355 1138-1142.

Gomes R V, Tura B, Mendonca Filho H T F d, Almeida Campos L A d, Rouge A, Matos Nogueira P M d, Oliveira Fernandes M A d, Rocha Dohmann H F d and Cunha A B (2007) A first postoperative day predictive score of mortality for cardiac surgery. *Ann Thorac Cardiovasc Surg* 13 (3) 159-164.

Gornescu F, McClean S I and Millard P H (2002) A queueing model for bed-occupancy management and planning of hospitals. *Journal of the Operational Research Society* 53 (1) 19-24.

Goswami V, Mohanty J R and Samanta S K (2006) Discrete-time bulk-service queues with accessible and non-accessible batches. *Applied Mathematics and Computation* 182 (1) 898-906.

Griffiths J D (1981) A Mathematical-Model of a Non-Signalized Pedestrian Crossing. *Transportation Science* 15 (3) 222-232.

Griffiths J D (1995) Queueing at teh Suez Canal. *Journal of the Operational Research Society* 46 1299-1309.

Griffiths J D and Cresswell C (1976) A mathematical model of a pelican crossing. *Journal of the Institute of Mathematics and Applications* 18 381-394.

Griffiths J D, Holland W and Williams J E (1991) Estimation of queues at the Channel Tunnel. *Journal of the Operational Research Society* 42 (5) 365-373.

Griffiths J D, Price-Lloyd N, Smithies M and Williams J E (2005a) Modelling the requirement for supplementary nurses in an intensive care unit. *Journal of the Operational Research Society* 56 126-133.

Griffiths J D, Price-Lloyd N, Smithies M and Williams J E (2005b) Modelling the requirement for supplementary nurses in an intensive care unit. *Journal of the Operational Research Society* 56 (2) 126-133.

Griffiths J D, Price-Lloyd N, Smithies M and Williams J E (2006) A queueing model of activities in an intensive care unit. *IMA Journal of Management Mathematics* 17 277-288.

Gupta S K and Goyal J K (1965) Queues with Batch Poisson Arrivals and Hyper-Exponential Service. *Naval Research Logistics Quarterly* 12 (3-4) 323-329.

Gupta U C and Sikdar K (2004) The finite-buffer M/G/1 queue with general bulk-service rule and single. *Performance Evaluation* 57 (2) 199-219.

Halfin S and Segal M (1972) Priority queuing model for a mixture of 2 types of customers. *Siam Journal on Applied Mathematics* 23 (3) 369-379.

Harper P and Leite E (2008) TreeWorks: Advances in Scalable Decision Trees. *International Journal of Information Technology and Web Engineering* 3 (4) 53-68.

Harper P R (2002) A framework for operational modelling of hospital resources. *Health Care Management Science* 5 (3) 165-173.

Harper P R and Shahani A K (2002) Modelling for the planning and management of bed capacities in hospitals. *Journal of the Operational Research Society* 53 11-18.

Harris C M (1970) Some results for bulk-arrival queues with state-dependent service times. *Management Science Series a-Theory* 16 (5) 313-326.

Hawkes A G (1965) Time-dependent solution of a priority queue with bulk arrival. *Operations Research* 13 (4) 586-595.

Hebuterne G and Rosenberg C (1999) Arrival and departure state distributions in the general bulk-service queue. *Naval Research Logistics* 46 (1) 107-118.

Holland W (1991) *A study of bulk queueing systems.* Cardiff University.

Holland W and Griffiths J D (1999) A time-dependent approximation for the queue M/M(1,s)/c. *IMA Journal of Mathematics Applied to Business and Industry* 10 213-223.

Hur S and Ahn S (2005) Batch arrival queues with vacations and server setup. *Applied Mathematical Modelling* 29 (12) 1164-1181.

ICS (1997) *Standards for Intensive Care Units.* http://www.ics.ac.uk/icmprof/standards.asp?menuid=7 Accessed: 18/03/2008

Ishizaki F, Takine T, Takahashi Y and Hasegawa T (1994) A generalized SBBP/G/1 queue and its applications. *Performance Evaluation* 21 (1-2) 163-181.

Jacob M J, Krishnamoorthy A and Madhusoodanan T P (1988) Transient Solution of a Finite-Capacity M/Ga,B/1 Queuing System. *Naval Research Logistics* 35 (3) 437-441.

Jaiswal N K (1960a) Bulk-Service Queuing Problem. *Operations Research* 8 (1) 139-143.

Jaiswal N K (1960b) Time-Dependent Solution of the Bulk-Service Queuing Problem. *Operations Research* 8 (6) 773-781.

Jun J B, Jacobson S H and Swisher J R (1999) Application of discrete-event simulation in health care clinics: A survey. *Journal of the Operational Research Society* 50 (2) 109-123.

Kabak I W (1968) Blocking and delays in M(*n*)/M/C bulk queuing systems. *Operations Research* 16 (4) 830-840.

Kabak I W (1970) Blocking and delays in M(X)/M/C bulk arrival queueing systems. *Management Science Series a-Theory* 17 (1) 112-115.

Kapadia A S, Chan W, Sachdeva R, Moye L A and Jefferson L S (2000) Predicting duration of stay in a pediatric intensive care unit: A Markovian approach. *European Journal of Operational Research* 124 353-359.

Kawasaki N, Takagi H, Takahashi Y, Hong S J and Hasegawa T (2000) Waiting time analysis of M-x/G/1 queues with/without vacations under random order of service discipline. *Journal of the Operations Research Society of Japan* 43 (4) 455-468.

Ke J C (2007a) Batch arrival queues under vacation policies with server breakdowns and startup/closedown times. *Applied Mathematical Modelling* 31 (7) 1282-1292.

Ke J C (2007b) Operating characteristic analysis on the M-[x]/G/1 system with a variant vacation policy and balking. *Applied Mathematical Modelling* 31 (7) 1321-1337.

Ke J C and Chu Y K (2006) A modified vacation model M-[X]/G/1 system. *Applied Stochastic Models in Business and Industry* 22 (1) 1-16.

Ke J C and Wang K H (2003) Analysis of operating characteristics for the heterogeneous batch arrival queue with server startup and breakdowns. *Rairo-Operations Research* 37 (3) 157-177.

Keenan S P, Dodek P, Martin C, Priestap F, Norena M and Wong H (2007) Variation in length of intensive care unit stay after cardiac arrest: Where you are is as important as who you are. *Critical Care Medicine* 35 (3) 836-841.

Keene A R and Cullen D J (1983) Therapeutic Intervention Scoring System: Update 1983. *Critical Care Medicine* 11 (1) 1-3.

Kendall D G (1951) Some Problems in the Theory of Queues. *Journal of the Royal Statistical Society Series B-Statistical Methodology* 13 (2) 151-185.

Kendall D G (1953) Stochastic Processes Occurring in the Theory of Queues and Their Analysis by the Method of the Imbedded Markov Chain. *Annals of Mathematical Statistics* 24 (3) 338-354.

Kim S C, Horowitz I, Young K K and Buckley T A (1999) Analysis of capacity management of the Intensive Care Unit in a hospital. *European Journal of Operational Research* 115 36-46.

Kimura T and Ohsone T (1984) A diffusion-approximation for an M/G/M queue with group arrivals. *Management Science* 30 (3) 381-388.

Krishnamoorthy A and Ushakumari P V (2000) A queueing system with single arrival bulk service and single departure. *Mathematical and Computer Modelling* 31 (2-3) 99-108.

Kulkarni V G (1986) Expected waiting-times in a multiclass batch arrival retrial queue. *Journal of Applied Probability* 23 (1) 144-154.

Kumar B K, Arivudainambi D and Vijayakumar A (2002) On the M-x/G/1 retrial queue with Bernoulli schedules and general retrial times. *Asia-Pacific Journal of Operational Research* 19 (2) 177-194.

Langaris C and Moutzoukis E (1995) A retrial queue with structured batch arrivals, priorities and server vacations. *Queueing Systems* 20 (3-4) 341-368.

Laupland K B, Shahpori R, Kirkpatrick A W and Stelfox H T (2008) Hospital mortality among adults admitted to and discharged from intensive care on weekends and evenings. *Journal of Critical Care* 23 (3) 317-324.

Laxmi P V and Gupta U C (1999) On the finite-buffer bulk-service queue with general independent arrivals: GI/M([b]/)1/N. *Operations Research Letters* 25 (5) 241-245.

Laxmi P V and Gupta U C (2000) Analysis of finite-buffer multi-server queues with group arrivals: GI(X)/M/c/N. *Queueing Systems* 36 (1-3) 125-140.

Lee H W (1989) Bulk arrival queues with server vacations. *Applied Mathematical Modelling* 13 (6) 374-377.

Lee H W and Lee S S (1991) A batch arrival queue with different vacations. *Computers & Operations Research* 18 (1) 51-58.

Lee H W, Lee S S, Chae K C and Nadarajan R (1992) On a Batch Service Queue with Single Vacation. *Applied Mathematical Modelling* 16 (1) 36-42.

Lee H W, Park J G, Kim B K, Yoon S H, Ahn B Y and Park N I (1998) Queue length and waiting time analysis of a batch arrival queue with bilevel control. *Computers & Operations Research* 25 (3) 191-205.

Lee S S, Lee H W, Yoon S H and Chae K C (1995) Batch arrival queue with N-policy and single vacation. *Computers & Operations Research* 22 (2) 173-189.

Lin S-P, Lee C-H, Lu Y-S and Hsu L-N (2006) A comparison of MICU survival prediction using the logistic regression model and artificial neural network model. *J Nurs Res* 14 (4) 306-314.

Litvak N, van Rijsbergen M, Boucherie R J and van Houdenhoven M (2008) Managing the overflow of intensive care patients. *European Journal of Operational Research* 185 (3) 998-1010.

Liu F H and Tseng J W (1999) Bilevel hysteretic service rate control for bulk arrival queue. *Computers & Industrial Engineering* 37 (1-2) 269-272.

Liu L, Kashyap B R K and Templeton J G C (1987) The Service System M/M(R)/infinity with Impatient Customers. *Queueing Systems* 2 (4) 363-372.

Lopezsoriano E M, Matthews M E and Norback J P (1981) Improving the flow of customers in a hospital cafeteria. *Journal of the American Dietetic Association* 79 (6) 683-688.

Love R F (1970) Steady-State Solution of Queuing System Ew/M/S with Batch Service. *Operations Research* 18 (1) 160-171.

Lowery J C (1993) Multi-hospital validation of critical care simulation model. In Evans, G W, Mollaghasemi, M, Russell, E C and Biles, W E (Eds.), *Proceedings of the 1993 Winter Simulation Conference.*

Luchak G (1956) The Solution of the Single-Channel Queuing Equations Characterized by a Time-Dependent Poisson-Distributed Arrival Rate and a General-Class of Holding Times. *Operations Research* 4 (6) 711-732.

Luyt C E, Combes A, Aegerter P, Guidet B, Trouillet J L, Gibert C, Chastre J and Network C U-R (2007) Mortality among patients admitted to intensive care units during weekday day shifts compared with "off" hours. *Critical Care Medicine* 35 (1) 3-11.

Madill B R, Chaudhry M L and Buckholtz P G (1985) On the Diophantine Queuing System, D/Da,B/1. *Journal of the Operational Research Society* 36 (6) 531-535.

Masuyama H and Takine T (2002) Analysis of an infinite-server queue with batch Markovian arrival streams. *Queueing Systems* 42 (3) 269-296.

McManus M L, Long M C, Cooper A and Litvak E (2004) Queuing theory accurately models the need for ctitical care resources. *Anesthesiology* 100 (5) 1271-1276.

Medhi J (1975) Waiting Time Distribution in a Poisson Queue with a General Bulk Service Rule. *Management Science Series a-Theory* 21 (7) 777-782.

Miller R G, Jr. (1959) A Contribution to the Theory of Bulk Queues. *Journal of the Royal Statistical Society. Series B (Methodological)* 21 (2) 320-337.

Moore S (2003) Capacity planning – modelling unplanned admissions in the UK NHS. *International Journal of Health Care Quality Assurance* 16 (4).

Moran J L, Bristow P, Solomon P J, George C, Hatt G K, Australian and New Zealand Intens C (2008) Mortality and length-of-stay outcomes, 1993-2003, in the binational Australian and New Zealand intensive care adult patient database. *Critical Care Medicine* 36 (1) 46-61.

Moustafa M S and Elsayed K M F (1992) Matrix-geometric solution of a multiserer queue with Markovian group arrivals and coxian servers. *Applied Mathematics and Computation* 49 (2-3) 177-196.

Moustafa M S and Elsayed K M F (1994) Iterative solution of the M(X)/C(2)/S/N queues. *Applied Mathematics and Computation* 66 (2-3) 313-323.

Murari K (1972) Time dependent solution of a queueing problem with correlated batch arrivals and general service time distribution. *Metrika* 19 (2-3) 201-208.

Nagelkerke N J D (1991) A note on a general definition of the coefficient of determination. *Biometrika* 78 (3) 691-692.

Neuts M F (1965) Busy Period of a Queue with Batch Service. *Operations Research* 13 (5) 815-819.

Neuts M F and Nadarajan R (1982) A multiserver queue with thresholds for the acceptance of customers into service. *Operations Research* 30 (5) 948-960.

Nierman D M, Schechter C B, Cannon L M and Meier D E (2001) Outcome prediction model for very elderly critically ill patients. *Critical Care Medicine* 29 (10) 1853-1859.

Nobel R D and Tijms H C (1999) Optimal control for an M-X/G/1 queue with two service modes. *European Journal of Operational Research* 113 (3) 610-619.

OECD (2006) *Health Expenditure.* http://www.oecd.org/dataoecd/20/51/37622205.xls Accessed: 15 October 2008

Ohwaki K, Yano E, Nagashima H, Nakagomi T and Tamura A (2008) Impact of infection on length of intensive care unit stay after intracerebral hemorrhage. *Neurocritical Care* 8 271-275.

Omosigho SE and Worthington DJ (1988) An approximation of known accuracy for single server queues with inhomogeneous arrival rate and continuous service time distribution. *European Journal of Operational Research.* 33 304-313.

Ozekici S (1987) Average Waiting Time in Queues with Scheduled Batch Services. *Transportation Science* 21 (1) 55-61.

Pechinkin A V (1990) Queuing system with absolute continuous priorities and group arrivals. *Automation and Remote Control* 51 (10) 1400-1408.

Pilgrim H and Chilcott J (2008) Assessment of a 7-day turn-around for the reporting of cervical smear results using discrete event simulation. *Journal of the Operational Research Society* 59 (7) 902-910.

Ranhoff A H, Rozzini R, Sabatini T, Cassinadri A, Boffelli S, Ferri M, Travaglini N, Ricci A, Morandi A and Trabucchi M (2006) Subintensive care unit for the elderly: a new model of care for critically ill frail elderly medical patients. *Intern Emerg Med* 1 (3) 197-203.

Reddy G V K and Anitha R (1998) Markovian bulk service queue with delayed vacations. *Computers & Operations Research* 25 (12) 1159-1166.

Reddy G V K, Nadarajan R and Kandasamy P R (1991) Markovian General Bulk Service Queuing-Systems with Vacation and Additional Server. *Asia-Pacific Journal of Operational Research* 8 (1) 80-89.

Reynolds J F (1968) Some results for bulk-arrival infinite-server Poisson queue. *Operations Research* 16 (1) 186-189.

Ridge J C, Jones S K, Nielsen M S and Shahani A K (1998) Capacity planning for intensive care units. *European Journal of Operational Research* 105 346-355.

Roes P B M (1966) A Many Server Bulk Queue. *Operations Research* 14 (6) 1037-1044.

Rosenfeld R, Smith J M, Woods S E and Engel A M (2006) Predictors and outcomes of extended intensive care unit length of stay in patients undergoing coronary artery bypass graft surgery. *Journal of Cardiac Surgery* 21 (2) 146-150.

Ryden T (1993) Waiting Time Distributions in Buffers with Batch Service. *Ieee Transactions on Communications* 41 (7) 1027-1030.

Schellhaas H (1994) Single-server queues with a batch Markovian arrival process and server vacations. *Or Spektrum* 15 (4) 189-196.

Selim S Z (1997) Time-dependent solution and optimal control of a bulk service queue. *Journal of Applied Probability* 34 (1) 258-266.

Shahani A K, Ridley S A and Nielsen M S (2008) Modelling patient flows as an aid to decision making for critical care capacities and organisation. *Anaesthesia* 63 (10) 1074-1080.

Shanthikumar J G (1994) Convexity of single-stage queueing-systems with bulk arrival. *Queueing Systems* 16 (3-4) 287-299.

Sharma R and Sharma G C (1996) A queueing model with the bulk arrival rate depending upon the nature of service. *Applied Mathematical Modelling* 20 (6) 470-475.

Shin Y W and Pearce C E M (1998) The BMAP/G/1 vacation queue with queue-length dependent vacation schedule. *Journal of the Australian Mathematical Society Series B-Applied Mathematics* 40 207-221.

Sivasamy R and Senthamaraikannan K (1994) Embedded Processes of a Markov Renewal Bulk Service Queue. *Asia-Pacific Journal of Operational Research* 11 (1) 51-65.

Stadje W (1994) On the behavior of the batch arrival queue M(X)/M/1 during a busy period. *Naval Research Logistics* 41 (2) 153-169.

Stanford D A and Pagurek B (1992) On the serial-correlation for number in systems in the stationary GI/M/1 bulk arrival and GI/EM/1 queues. *Journal of Applied Probability* 29 (2) 404-417.

Su S and Shih C L (2003) Modeling an emergency medical services system using computer simulation. *International Journal of Medical Informatics* 72 (1-3) 57-72.

Tadj L (2003) A quorum queueing system under D-policy. *Applied Mathematics and Computation* 144 (2-3) 325-336.

Tadj L and Sarhan A (2003) Effect of the server capacity distribution on the optimal control of a bulk service queueing system. *Chaos Solitons & Fractals* 18 (5) 1101-1110.

Tadj L and Tadj C (2002) Recursive solution to a quorum queueing system. *Mathematical and Computer Modelling* 35 (3-4) 283-293.

Takacs, L (1963) Delay distributions for one line with Poisson input, general holding times, and various orders of service, *The Bell System Technical Journal* 42 487-503.

Takagi H and Takahashi Y (1991) Priority-queues with batch Poisson arrivals. *Operations Research Letters* 10 (4) 225-232.

Takagi H and Wu D A (2004) Multiserver queue with semi-Markovian batch arrivals. *Computer Communications* 27 (6) 549-556.

Takahashi M, Osawa H and Fujisawa T (1999) Geo([x])/G/1 retrial queue with non-preemptive priority. *Asia-Pacific Journal of Operational Research* 16 (2) 215-234.

Takahashi Y and Takagi H (1990) Structured priority queue with batch arrivals. *Journal of the Operations Research Society of Japan* 33 (3) 242-261.

Tang Y H (2007) On the transient departure process of M-x/G/1 queueing system with single server vacation. *Journal of Systems Science & Complexity* 20 (4) 562-571.

Tang Y H and Tang X W (2000) The queue-length distribution for M-x/G/1 queue with single server vacation. *Acta Mathematica Scientia* 20 (3) 397-408.

Te H J, Ming C C, Tai C S, Wen C H, Jan C N, Wan-Ching H and Min C C (2006) Prognostic Role of Alveolar-Arterial Oxygen Pressure Difference in Acute Pulmonary Embolism. *Circulation Journal* 70 (12) 1611-1616.

Towsley D and Tripathi S K (1991) A single-server priority queue with server failures and queue flushing. *Operations Research Letters* 10 (6) 353-362.

Van Hoorn M H (1981) Algorithms for the State Probabilities in a General-Class of Single-Server Queuing-Systems with Group Arrivals. *Management Science* 27 (10) 1178-1187.

Van Houdenhoven M, Nguyen D T, Eijkemans M J, Steyerberg E W, Tilanus H W, Gommers D, Wullink G, Bakker J and Kazemier G (2007) Optimizing intensive care capacity using individual length-of-stay prediction models. *Critical Care* 11 (2).

Vanommeren J C W (1990) Simple approximations for the batch-arrival M(x)/G/1 queue. *Operations Research* 38 (4) 678-685.

Vasilakis C and Marshall A H (2005) Modelling nationwide hospital length of stay: opening the black box. *Journal of the Operational Research Society* 56 862–869.

Vinck B and Bruneel H (1994) Analyzing the discrete-time G(G)/GEO/1 queue using complex contour integration. *Queueing Systems* 18 (1-2) 47-67.

WAG (2007a) *NHS Beds by specialty (NHS Region, Trust, Hospital, Specialty)* http://www.statswales.wales.gov.uk/index.htm Accessed: 18/03/2008

WAG (2007b) *NHS hospital waiting times latest month by patient type (Speciality, Trust, LHB).* http://www.statswales.wales.gov.uk/index.htm Accessed: 18/03/2008

Wang J H, Changchien C S, Hu T H, Lee C M, Kee K M, Lin C Y, Chen C L, Chen T Y, Huang Y J and Lu S N (2008) The efficacy of treatment schedules according to Barcelona Clinic Liver Cancer staging for hepatocellular carcinoma - Survival analysis of 3892 patients. *European Journal of Cancer* 44 (7) 1000-1006.

Wang K H, Chan M C and Ke J C (2007) Maximum entropy analysis of the M-[x]/M/1 queueing system with multiple vacations and server breakdowns. *Computers & Industrial Engineering* 52 (2) 192-202.

Wang Y C and Lu C C (2007) Loss behavior in space priority queue with batch Markovian arrival process - continuous-time case. *Performance Evaluation* 64 (1) 93-101.

Worthington D J (1987) Queuing Models for Hospital Waiting Lists. *Journal of the Operational Research Society* 38 (5) 413-422.

Worthington D J and Wall A (1999) Using the discrete time modelling approach to evaluate the time-dependent behaviour of queueing systems. *European Journal Operational Research* 50 777-788.

Wunsch H, Mapstone J, Brady T, Hanks R and Rowan K (2004) Hospital mortality associated with day and time of admission to intensive care units. *Intensive Care Medicine* 30 (5) 895-901.

Xia C H, Michailidis G, Bambos N and Glynn P W (2002) Optimal control of parallel queues with batch service. *Probability in the Engineering and Informational Sciences* 16 (3) 289-307.

Yao D D W, Chaudhry M L and Templeton J G C (1984) On bounds for bulk arrival queues. *European Journal of Operational Research* 15 (2) 237-243.

Yi X W, Kim N K, Yoon B K and Chae K C (2007) Analysis of the queue-length distribution for the discrete-time batch-service Geo/G(a,Y)/1/K queue. *European Journal of Operational Research* 181 (2) 787-792.

Zare M M, Itani K M F, Schifftner T L, Henderson W G and Khuri S F (2007) Mortality after nonemergent major surgery performed on Friday versus Monday through Wednesday. *Annals of Surgery* 246 866-874.

Zhu Y X (1994) Tandem queue with group arrivals and no intermediate buffer. *Queueing Systems* 17 (3-4) 403-412.

Zili M, Wang S M and Li Q L (1991) A Class of Cyclic Queue with Priority Batch Service. *Applied Mathematical Modelling* 15 (9) 450-458.

Appendix

Appendix 3.1

Rearrangement of steady state queueing equations for the $M^{(b)}$/He/c queueing system.

$$\begin{aligned}\left(\lambda+(j)2\mu\sigma\right)P_{new}(j,j) &= \lambda\sum_{i=1}^{j}\sigma^{i}b_{i}P_{new}(j-i,j-i)\\ &+(j+1)2\mu\sigma P_{old}(j+1,j+1)\\ &+2\mu(1-\sigma)P_{old}(j+1,j)\end{aligned}$$

$1\le j\le c-1$

$$\begin{aligned}\left(\lambda+(j)2\mu(1-\sigma)\right)P_{new}(j,0) &= \lambda\sum_{i=1}^{j}(1-\sigma)^{i}b_{i}P_{new}(j-i,0)\\ &+2\mu\sigma P_{old}(j+1,1)\\ &+(j+1)2\mu(1-\sigma)P_{old}(j+1,0)\end{aligned}$$

$1\le j\le c-1$

$$\begin{aligned}\left(\lambda+(j-m)2\mu(1-\sigma)+(m)2\mu\sigma\right)P_{new}(j,m) &= \lambda\sum_{i=1}^{j-m}(1-\sigma)^{i}b_{i}P_{new}(j-i,m)\\ &+\lambda\sum_{i=1}^{m}\sigma^{i}b_{i}P_{new}(j-i,m-i)\\ &+\lambda\sum_{n=1}^{m}\sum_{i=1}^{j-m}(1-\sigma)^{i}\sigma^{n}b_{i+n}P_{new}(j-n-i,m-n)\\ &+(j+1-m)2\mu(1-\sigma)P_{old}(j+1,m)\\ &+(m+1)2\mu\sigma P_{old}(j+1,m+1)\end{aligned}$$

$1\le j\le c-1$ and $1\le m\le j-1$

$$\lambda P_{new}(0)=2\mu\sigma P_{old}(1,1)+2\mu(1-\sigma)P_{old}(1,0)$$

$$\left(\lambda+(c)2\mu\sigma\right)P_{new}(c,c)=\lambda\sum_{i=1}^{c}\sigma^{i}b_{i}P_{new}(c-i,c-i)$$
$$+(c)2\mu\sigma^{2}P_{old}(c+1,c)$$
$$+2\mu(1-\sigma)\sigma P_{old}(c+1,c-1)$$

$$\left(\lambda+(c)2\mu(1-\sigma)\right)P_{new}(c,0)=\lambda\sum_{i=1}^{c}(1-\sigma)^{i}b_{i}P_{new}(c-i,0)$$
$$+2\mu\sigma(1-\sigma)P_{old}(c+1,1)$$
$$+(c)2\mu(1-\sigma)^{2}P_{old}(c+1,0)$$

$$\left(\lambda+(m)2\mu\sigma+(c-m)2\mu(1-\sigma)\right)P_{new}(c,m)=\lambda\sum_{i=1}^{c-m}b_{i}(1-\sigma)^{i}P_{new}(c-i,m)$$
$$+\lambda\sum_{i=1}^{m}b_{i}\sigma^{i}P_{new}(c-i,m-i)$$
$$+\lambda\sum_{n=1}^{m}\sum_{i=1}^{c-m}b_{i+n}\sigma^{n}(1-\sigma)^{i}P_{new}(c-n-i,m-n)$$
$$+\left((m)2\mu\sigma^{2}+(c-m)2\mu(1-\sigma)^{2}\right)P_{old}(c+1,m)$$
$$+(m+1)2\mu\sigma(1-\sigma)P_{old}(c+1,m+1)$$
$$+(c-m+1)2\mu\sigma(1-\sigma)P_{old}(c+1,m-1)$$

$1\le m\le c-1$

$$\left(\lambda+(c)2\mu\sigma\right)P_{new}(j,c)=\lambda\sum_{i=1}^{j-c}b_{i}P_{new}(j-i,c)$$
$$+\lambda\sum_{i=1}^{c}b_{i+j-c}\sigma^{i}P_{new}(c-i,c-i)$$
$$+(c)2\mu\sigma^{2}P_{old}(j+1,c)$$
$$+2\mu(1-\sigma)\sigma P_{old}(j+1,c-1)$$

$j\ge c+1$

$$\left(\lambda+(c)2\mu(1-\sigma)\right)P_{new}(j,0)=\lambda\sum_{i=1}^{j-c}b_{i}P_{new}(j-i,0)$$
$$+\lambda\sum_{i=1}^{c}b_{i+j-c}(1-\sigma)^{i}P_{new}(c-i,0)$$
$$+2\mu\sigma(1-\sigma)P_{old}(j+1,1)$$
$$+(c)2\mu(1-\sigma)^{2}P_{old}(j+1,0)$$

$j\ge c+1$

$$\left(\lambda+(m)2\mu\sigma+(c-m)2\mu(1-\sigma)\right)P_{new}(j,m)=\lambda\sum_{i=1}^{j-c}b_i P_{new}(j-i,m)$$
$$+\lambda\sum_{i=1}^{c-m}b_{i+j-c}(1-\sigma)^i P_{new}(c-i,m)$$
$$+\lambda\sum_{i=1}^{m}b_{i+j-c}\sigma^i P_{new}(c-i,m-i)$$
$$+\lambda\sum_{n=1}^{m}\sum_{i=1}^{c-m}b_{j-c+i+n}\sigma^n(1-\sigma)^i P_{new}(c-n-i,m-n)$$
$$+(c-m+1)2\mu(1-\sigma)\sigma P_{old}(j+1,m-1)$$
$$+\left((m)2\mu\sigma^2+(c-m)2\mu(1-\sigma)^2\right)P_{old}(j+1,m)$$
$$+(m+1)2\mu\sigma(1-\sigma)P_{old}(j+1,m+1)$$

$j\geq c+1$ and $1\leq m\leq c-1$

Appendix 3.2

Pseudo-code for the M(b)/He/c queueing system

```
sumPP = 0
sumPP is the sum of the newly calculated probabilities before they are
normalised

Sum = 0
Sum us the sum of the probabilities that are output on the last iteration

For i = 0 To 2000
        batchprob(i) = Cells(i + 2, 2)
Next i
Reads in the batch distribution from the Excel sheet

For m1 = 0 To probno
For m2 = 0 To Application.WorksheetFunction.Min(m1, c)
        Pnew(m1, m2) = 0
        Pold(m1, m2) = 0
Next m2
Next m1

Set initial probabilities
startingval = 0.5
Pold(0, 0) = startingval

For m1 = 1 To probno
For m2 = 1 To Min(m1, c)
        Pold(m1, 0) = startingval * Pold(m1 - 1, Min(m1 - 1, c))
        Pold(m1, m2) = startingval * Pold(m1, m2 - 1)
Next m2
```

```
Next m1

Pnew(0, 0) = startingval

For m1 = 1 To probno
For m2 = 1 To Min(m1, c)
        Pnew(m1, 0) = startingval * Pnew(m1 - 1, Min(m1 - 1, c))
        Pnew(m1, m2) = startingval * Pnew(m1, m2 - 1)
Next m2
Next m1

x = 0
Do x = x + 1

For j = 1 To c - 1

Equation 3.18
Arrivals1 = 0
For i = 1 To j
        Arrivals1 = (lambda * batchprob(i) * (1 - sigma) ^ i * Pnew(j - i, 0)) +
        Arrivals1
Next i

        Pnew(j, 0) = (Arrivals1 + (2 * mu * sigma * Pold (j + 1, 1)) + (2 * mu *
        (1 - sigma) * (j + 1) * Pold(j + 1, 0))) / (lambda + (2 * mu * (1 - sigma) *
        j))

Equation 3.19

For m = 1 To (j - 1)

Arrivals2 = 0
```

```
For i = 1 To (j - m)
        Arrivals2 = (lambda * (1 - sigma) ^ i * batchprob(i) * Pnew(j - i, m)) +
        Arrivals2
Next i

Arrivals3 = 0
For i = 1 To m
        Arrivals3 = (lambda * (sigma) ^ i * batchprob(i) * Pnew(j - i, m - i)) +
        Arrivals3
Next i

Arrivals4 = 0
For n = 1 To m
For i = 1 To (j - m)
        Arrivals4 = (lambda * (1 - sigma) ^ i * (sigma) ^ n * batchprob(i + n) *
        Pnew(j - n - i, m - n)) + Arrivals4
Next i
Next n

        Pnew(j, m) = (Arrivals2 + Arrivals3 + Arrivals4 + ((j + 1 - m) * 2 * mu *
        (1 - sigma) * Pold(j + 1, m)) + ((m + 1) * 2 * mu * sigma * Pold(j + 1, m
        + 1))) / (lambda + ((j - m) * 2 * mu * (1 - sigma)) + (2 * m * mu *
        sigma))

Next m

Equation 3.16
        Pnew(0, 0) = ((2 * mu * sigma * Pnew(1, 1)) + (2 * mu * (1 - sigma) *
        Pnew(1, 0))) / (lambda)

Equation 3.17
Arrivals5 = 0
```

```
For i = 1 To j
        Arrivals5 = (lambda * batchprob(i) * sigma ^ i * Pnew(j - i, j - i)) +
        Arrivals5
Next i

        Pnew(j, j) = (Arrivals5 + ((j + 1) * 2 * mu * sigma * Pold(j + 1, j + 1)) +
        (2 * mu * (1 - sigma) * Pold(j + 1, j))) / (lambda + (2 * mu * sigma * j))
Next j

Equation 3.21
Arrivals6 = 0
For i = 1 To c
        Arrivals6 = (lambda * (1 - sigma) ^ i * batchprob(i) * Pnew(c - i, 0)) +
        Arrivals6
Next i

        Pnew(c, 0) = (Arrivals6 + (2 * mu * sigma * (1 - sigma) * Pold(c + 1, 1))
        + (c * 2 * mu * (1 - sigma) ^ 2 * Pold(c + 1, 0))) / (lambda + (2 * mu * (1
        - sigma) * c))

Equation 3.22
For m = 1 To (c - 1)
Arrivals7 = 0

For i = 1 To (c - m)
        Arrivals7 = (lambda * batchprob(i) * (1 - sigma) ^ i * Pnew(c - i, m)) +
        Arrivals7
Next i

Arrivals8 = 0
For i = 1 To m
```

```
        Arrivals8 = (lambda * batchprob(i) * sigma ^ i * Pnew(c - i, m - i)) +
        Arrivals8
Next i

Arrivals9 = 0
For n = 1 To m
For i = 1 To (c - m)
        Arrivals9 = (lambda * batchprob(i + n) * sigma ^ n * (1 - sigma) ^ i *
        Pnew(c - n - i, m - n)) + Arrivals9
Next i
Next n

        Pnew(c, m) = (Arrivals7 + Arrivals8 + Arrivals9 + (2 * (c - m) * mu * (1
        - sigma) ^ 2 + (2 * m * mu * sigma ^ 2)) * Pold(c + 1, m) + ((m + 1) * 2
        * mu * sigma * (1 - sigma) * Pold(c + 1, m + 1)) + ((c - m + 1) * 2 * mu
        * sigma * (1 - sigma) * Pold(c + 1, m - 1))) / (lambda + (2 * mu * sigma
        * m) + (2 * mu * (1 - sigma) * (c - m)))

Next m

Equation 3.20
Arrivals10 = 0

For i = 1 To c
        Arrivals10 = (lambda * sigma ^ i * batchprob(i) * Pnew(c - i, c - i)) +
        Arrivals10
Next i

        Pnew(c, c) = (Arrivals10 + (2 * mu * sigma ^ 2 * c * Pold(c + 1, c)) + (2
        * mu * sigma * (1 - sigma) * Pold(c + 1, c - 1))) / (lambda + (2 * mu *
        sigma * c))
```

```
Equation 3.24
For j = (c + 1) To probno
Arrivals11 = 0

For i = 1 To (j - c)
        Arrivals11 = (lambda * batchprob(i) * Pnew(j - i, 0)) + Arrivals11
Next i

Arrivals12 = 0

For i = 1 To c
        Arrivals12 = (lambda * batchprob(i + j - c) * (1 - sigma) ^ i * Pnew(c - i,
        0)) + Arrivals12
Next i

        Pnew(j, 0) = (Arrivals11 + Arrivals12 + (2 * mu * sigma * (1 - sigma) *
        Pold(j + 1, 1)) + (c * 2 * mu * (1 - sigma) ^ 2 * Pold(j + 1, 0))) / (lambda
        + (2 * c * mu * (1 - sigma)))

Equation 3.25
For m = 1 To (c - 1)
Arrivals13 = 0

For i = 1 To (j - c)
        Arrivals13 = (lambda * batchprob(i) * Pnew(j - i, m)) + Arrivals13
Next i

Arrivals14 = 0
For i = 1 To (c - m)
        Arrivals14 = (lambda * batchprob(i + j - c) * (1 - sigma) ^ i * Pnew(c - i,
        m)) + Arrivals14
Next i
```

```
Arrivals15 = 0
For i = 1 To m
        Arrivals15 = (lambda * batchprob(i + j - c) * sigma ^ i * Pnew(c - i, m -
        i)) + Arrivals15
Next i

Arrivals16 = 0
For n = 1 To m
For i = 1 To (c - m)
        Arrivals16 = (lambda * batchprob(j - c + i + n) * sigma ^ n * (1 - sigma)
        ^ i * Pnew(c - n - i, m - n)) + Arrivals16
Next i
Next n

        Pnew(j, m) = (Arrivals13 + Arrivals14 + Arrivals15 + Arrivals16 + ((c -
        m + 1) * 2 * mu * sigma * (1 - sigma) * Pold(j + 1, m - 1)) + ((2 * mu *
        sigma ^ 2 * m) + (2 * mu * (1 - sigma) ^ 2) * (c - m)) * Pold(j + 1, m) +
        ((m + 1) * 2 * mu * sigma * (1 - sigma) * Pold(j + 1, m + 1))) / (lambda +
        (2 * mu * sigma * m) + (2 * mu * (1 - sigma) * (c - m)))

Next m

Equation 3.23
Arrivals17 = 0

For i = 1 To (j - c)
        Arrivals17 = (lambda * batchprob(i) * Pnew(j - i, c)) + Arrivals17
Next i

Arrivals18 = 0
```

```
For i = 1 To c
        Arrivals18 = (lambda * batchprob(i + j - c) * sigma ^ i * Pnew(c - i, c - i))
        + Arrivals18
Next i

        Pnew(j, c) = (Arrivals17 + Arrivals18 + (2 * mu * sigma ^ 2 * c * Pold(j
        + 1, c)) + (2 * mu * sigma * (1 - sigma) * Pold(j + 1, c - 1))) / (lambda +
        (2 * mu * sigma * c))

Next j

sumPP = 0

For m1 = 0 To probno - 1
For m2 = 0 To Min(m1, c)
        sumPP = Pnew(m1, m2) + sumPP
Next m2
Next m1

For m1 = 0 To probno
For m2 = 0 To Min(m1, c)
        Pnew(m1, m2) = Pnew(m1, m2) / sumPP
Next m2
Next m1

For m1 = 0 To probno
For m2 = 0 To Min(m1, c)
        Pold(m1, m2) = Pnew(m1, m2)
Next m2
Next m1

Loop Until x > limit
```

Appendix 4.1

The table below contains a list of all variables used in the analysis.

Variable	Description
TIMEIN	Time into ICU
TIMEOUT	ICU discharge time
TPN	Did the patient have TPN? (Y/N)
OUTCOME	ICU outcome Alive/Dead (A/D)
HOUTCOME	Hospital outcome Alive/Dead (A/D)
AGE	(adate - dob)/365
SEX	(M/F)
RACE	(W)hite (B)lack (A)sian
ADSOURCE	(1)theatre (2)A&E (3)Ward (4)other hosp (5)Recovery (6)HDU (7)X-Ray
SURTYPE	(1)elective (2)emergency
OPCOM	Operative complications present (Y/N)
WARD	Name of Ward admitted from.
INDICAT	Apache admission indication: Resp Cardiovasc Neuro Gastro Kidney Metabolic Haem
PRECIP	Apache Factor precipitating the primary system failure
CHE	Chronic (ie long-standing) ill-health present (Y/N)
LIVER	If CHE is true, is there liver failure? (Y/N)
CVS	If CHE is true, is there cardiac failure? (Y/N)
RESP	If CHE is true, is there respiratory failure? (Y/N)
RENAL	If CHE is true, is there renal failure? (Y/N)
IMMUNE	If CHE is true, is there immune deficiency? (Y/N)
HIV	Does the patient have AIDS (Y/N)
LYMPHOMA	Does the patient have Lymphoma (Y/N)
CANCER	Does the patient have cancer with metastases (Y/N)
HAEMCAN	Does the patient have a haematological malignancy (Y/N)
PREDICT	RIP prediction. (D)ie/(U)nknown. Is D if apasco.predict is D in any line with this icuid
SDCODE	Specific Diagnostic Code
SPEC	Specialty patient admitted to ICU under
DOCTOR	Consultant patient admitted to ICU under
TOTTISS	Total TISS point accumulated for patient during ICU stay
VENT	Was the patient ventilated during ICU stay (Y/N)
DIALYSED	Did the patient have haemodialysis or haemofiltration during ICU stay (Y/N)
AGEPTS	APACHE points awarded for age
CHPOINTS	APACHE points awarded for Chronic Health probs
TEMP	Core body temp Deg Centigrade

TEMPSCORE	APACHE points awarded for temp
VMBP	mean BP
MBP	Most abnormal Mean (averaged over one beat) Arterial Blood Pressure
MBPSCORE	APACHE points awarded for MBP
VSYSBP	systolic BP
SYSBP	Most abnormal Systolic Blood Pressure
HR	Heart Rate (beats per minute)
HRSCORE	APACHE points awarded for HR
RESPR	Most abnormal Respiratory Rate (breaths per minute)
RESPSCORE	APACHE points awarded for RESPR
VFIO	inspired oxygen (modern units)
FIO	Inspired oxygen concentration (%)
VPO	Arterial po2 (modern units)
PO	Arterial partial pressure of oxygen (mmHg)
VPCO	carbon dioxide (modern units)
PCO	Arterial partial pressure of carbon dioxide (mmHg)
ADO	Arterial-Alveolar oxygen difference ((FIO*760)-PO)
OXYSCORE	APACHE points awarded for ADO
PH	Arterial pH
PHSCORE	APACHE points awarded for pH
HCO3	Plasma Bicarbonate in mmol
SOD	Plasma Sodium (mmol)
SODSCORE	APACHE points awarded for SOD
POT	Plasma Sodium (mmol)
POTSCORE	APACHE points awarded for POT
VCR	creatinine (modern units)
CR	Plasma Creatinine (old units; multiply by 90.9 for micromol/L)
CRSCORE	APACHE points awarded for CR
UREA	Plasma Urea Concentration (old units; divide by 6 to get mmol/L)
VUREA	urea (modern units)
URINE	Urine Output (ml/day)
VPCV	Haematocrit (modern units)
PCV	Packed Cell Volume (% of blood which is red cells)
PCVSCORE	APACHE points awarded for Packed Cell Volume
WBC	No. of White Blood Cells per cubic mm in thousands
WBCSCORE	APACHE points awarded for WBC
PLATELET	No. of Platelets in Blood per cubic mm in thousands
PROTIME	Ratio of Prothrombin Time to Control
SGOT	Aspartate Transaminase Concentration
VBILIRUBIN	bilibubin (modern units)
BILIRUBIN	Plasma Bilirubin (old units; mutiply by 17.24 for mmol/L)
VALBUMIN	albumin (modern units)
ALBUMIN	Plasma Albumin concentration (mg/100ml)

VGLUCOSE	glucose (modern units)
GLUCOSE	Plasma glucose (mmol)
TPN	Is the patient having TPN today?
EYES	Menu choice for eyes movements 1 = good, 4 = bad
EYESCORE	Glasgow Coma Score (GCS) points for EYES (?5-EYES)
MOTOR	Menu choice for motor (ie muscle) function 1 = good, 6 = bad
MOTORSCORE	Glasgow Coma Score (GCS) points for MOTOR (?7-MOTOR)
NVVERBAL	Menu choice for "verbal" responses in intubated patients
NVVERBSCO	Glasgow Coma Score (GCS) points for NVVERBAL
VVERBAL	Menu choice for verbal responses in non-intubated patients
VVERBSCO	Glasgow Coma Score (GCS) points for VVERBAL
SEDATION	Is the patient sedated today (Y/N)
GCS	Glasgow Coma Score SUM(EYESCORE,MOTORSCORE, NorVVERBSCO
GCSSCORE	APACHE points awarded for GCS
APSCORE	APACHE score
APS	APACHE score before Age CHE and Diagnostic Category points added
PRODEATH	Percentage Risk Of Death in Hospital (%risk that HOUTCOME = D)
ARF	Acute Renal Failure ie Urine < 400 ml/day and Creatinine > 200 (Y/N)
CARF	Clinical Diagnosis of Acute Renal Failure ie. doctor says there is.
NEUF	Neurological Failure today? (Y/N)
CVSF	Cardiovascular Failure today? (Y/N)
RESPF	Respiratory Failure today? (Y/N)
HAEMF	Haematological Failure today? (Y/N)
HEPF	HepaticFailure today? (Y/N)
GIF	GastroIntestinal Failure today? (Y/N)
FIB	Ventricular Fibrillation today? (Y/N)
ENCEP	Encephalopathy today? (Y/N)
PREDICT	Today's prediction to survive the illness (U)nknown/(D)ie
ORGF	No. of failed organs today
ORGFD	No. of days that ORGF has been at that score
ORGSCORE	Knauss's Organ failure coefficient for (ORGF*ORGFD)
DVENT	Ventilated today? Y/N
VENTD	Days continuously on ventilator so far
CVSD	No of days of continuous CVS failure so far
RESPD	No of days of continuous Resp failure so far
NEUD	No of days of continuous Neurological failure so far
HEPD	No of days of continuous Hepaticfailure so far
GID	No of days of continuous GastroIntestinal failure so far
ARFD	No of days of continuous Acute Renal failure so far
HAEMD	No of days of continuous Haematological failure so far

SUR	Surgery today? Y/N
COMP	Complications occur with today's surgery? Y/N
TISSPOINT	TISS points for today
TISSCODE	List of tisscodes which were present today. Lookup table = aptiss.dbf
TOTSAPS	SAPS total for today
SAPSRISK	% Risk Of Death for todays SAPS Score
SAPSAGE	SAPS points for age (?duplicated each day)
SAPSHR	SAPS points for heart rate
SAPSSYS	SAPS points for Systolic BP
SAPSTEMP	SAPS points for Temperature
SAPSPF	SAPS points for ??
SAPSURINE	SAPS points for Urine ?volume per 24 hrs
SAPSUREA	SAPS points for plasma urea
SAPSWBC	SAPS points for White Blood Cell count
SAPSPOT	SAPS points for Plasma Potassium
SAPSSOD	SAPS points for Plasma Sodium
SAPSHCO3	SAPS points for Plasma Bicarbonate
SAPSBIL	SAPS points for Plasma Bilirubin
SAPSGCS	SAPS points for Glasgow Coma Score
SAPSCHPT	SAPS points for Chronic Health (?duplicated each day)
AP3AGE	Apache 3 points for age (?duplicated each day)
AP3CH	Apache 3 points for Chronic health (?duplicated each day)
AP3HR	Apache 3 points for HR
AP3MBP	Apache 3 points for Mean BP
AP3TEMP	Apache 3 points for Temp
AP3RESPR	Apache 3 points for Resp rate
AP3PO	Apache 3 points for Arterial Oxygen Tension
AP3PCV	Apache 3 points for Packed cell volume (Haematocrit)
AP3WBC	Apache 3 points for White Blood Cell count
AP3CR	Apache 3 points for Plasma Creatinine
AP3URINE	Apache 3 points for Urine ?volume per 24 hrs
AP3BUN	Apache 3 points for Blood Urea Nitrogen
AP3SOD	Apache 3 points for Plasma Sodium
AP3ALB	Apache 3 points for Plasma Albumin
AP3BIL	Apache 3 points for Plasma Bilirubin
AP3GLU	Apache 3 points for Plasma Glucose
AP3PH	Apache 3 points for Plasma pH
AP3NEURO	Apache 3 points for Neurological Status
AP3SCORE	Total Apache 3 points for today

Appendix 4.2

The table below contains each SDCODE as well as a description of the condition experienced by the patient.

SDCODE	Description
01	Non-Op Asthma/Allergy
02	Non-Op Chronic Obstructive Pulmonary Disease
03	Non-Op Pulmonary oedema (non cardiogenic)
04	Non-Op Post respiratory arrest
05	Non-Op Aspiration/poisoning/toxic
06	Non-Op Pulmonary embolus
07	Non-Op Respiratory failure from Infection
08	Non-Op Respiratory failure from Neoplasm
09	Non-Op CVS failure from Hypertension
10	Non-Op CVS F from Rhythm disturbance
11	Non-Op Congestive heart failure
12	Non-Op CVS F Haemorrhagic/hypovolaemic shock
13	Non-Op CVS F from Coronary artery disease
14	Non-Op CVS F from Sepsis
15	Non-Op CVS F Post cardiac arrest
16	Non-Op CVS F Cardiogenic shock
17	Non-Op CVS F Dissecting Abdo/Thoracic Aneurysm
18	Non-Op Multiple Trauma
19	Non-Op Head Trauma
20	Non-Op Neurologic Seizure disorder
21	Non-Op Neurologic ICH/SDH/SAH
22	Non-Op Drug Overdose
23	Non-Op Diabetic ketoacidosis
24	Non-Op Gastrointestinal bleeding
25	Non-Op Metabolic/Renal
26	Non-Op Respiratory
27	Non-Op Neurologic
28	Non-Op Cardiovascular
29	Non-Op Gastrointestinal
30	Post-op Multiple Trauma
31	Post-op Chronic cardiovascular disease
32	Post-op Peripheral vascular surgery
33	Post-op Heart valve surgery
34	Post-op Craniotomy for Neoplasm
35	Post-op Renal surgery for Neoplasm
36	Post-op Renal transplant
37	Post-op Head Trauma
38	Post-op Thoracic surgery for Neoplasm

39	Post-op Craniotomy for ICH/SDH/SAH
40	Post-op Laminectomy and other Spinal cord surgery
41	Post-op Haemorrhagic shock
42	Post-op Gastrointestinal bleeding
43	Post-op GI surgery for Neoplasm
44	Post-op Respiratory insufficiency
45	Post-op GI Perforation/Obstruction
46	Post-op Neurologic
47	Post-op Cardiovascular
48	Post-op Respiratory
49	Post-op Gastrointestinal
50	Post-op Metabolic/Renal
51	Post-op Sepsis
52	Post-op Cardiac Arrest

Appendix 4.3

The table below contains each Precipitating factor as well as a description of the condition experienced by the patient.

Precip	Description
01	Infection
02	Neoplasm
03	Trauma
04	Self-intoxication (overdose)
05	Intracerebral haemorrhage
06	Extracerebral (Subdural/arachnoid haemorrhage)
07	Seizures
08	Neuromuscular failure
09	Coronary artery disease
10	Myocardial infarction (documented)
11	Valvular heart disease
12	Peripheral vascular disease
13	Embolus (localised)
14	Congenital anomaly/anatomic defect
15	Congestive heart failure/pulmonary edema
16	Hypertension
17	Rhythm disturbance
18	Pericardial disease
19	Cardiogenic shock/myocardiopathy
20	Septic shock/sepsis
21	Anaphylactic/drug induced shock
22	Haemorrhagic/hypovolaemic shock
23	Bleeding (significant but not shocked)
24	Post arrest (cardiac and/or respiratory)
25	Allergic reaction
26	Obstruction/perforation
27	Coma/mental derangement (metabolic)
28	Electrolyte/Acid-base disturbance
29	Diabetic ketoacidosis
30	Endocrine emergency
31	Hypo/hyperthermia
32	Haematologic insufficiency/crisis
33	Post-transplant surgery
34	Unplanned post-op ventilation
35	Acute-on-chronic end stage disease
36	Toxic/chemical poisoning
37	Planned post operative monitoring

Appendix 6.1

Inter-arrival time graphs for the simulation model

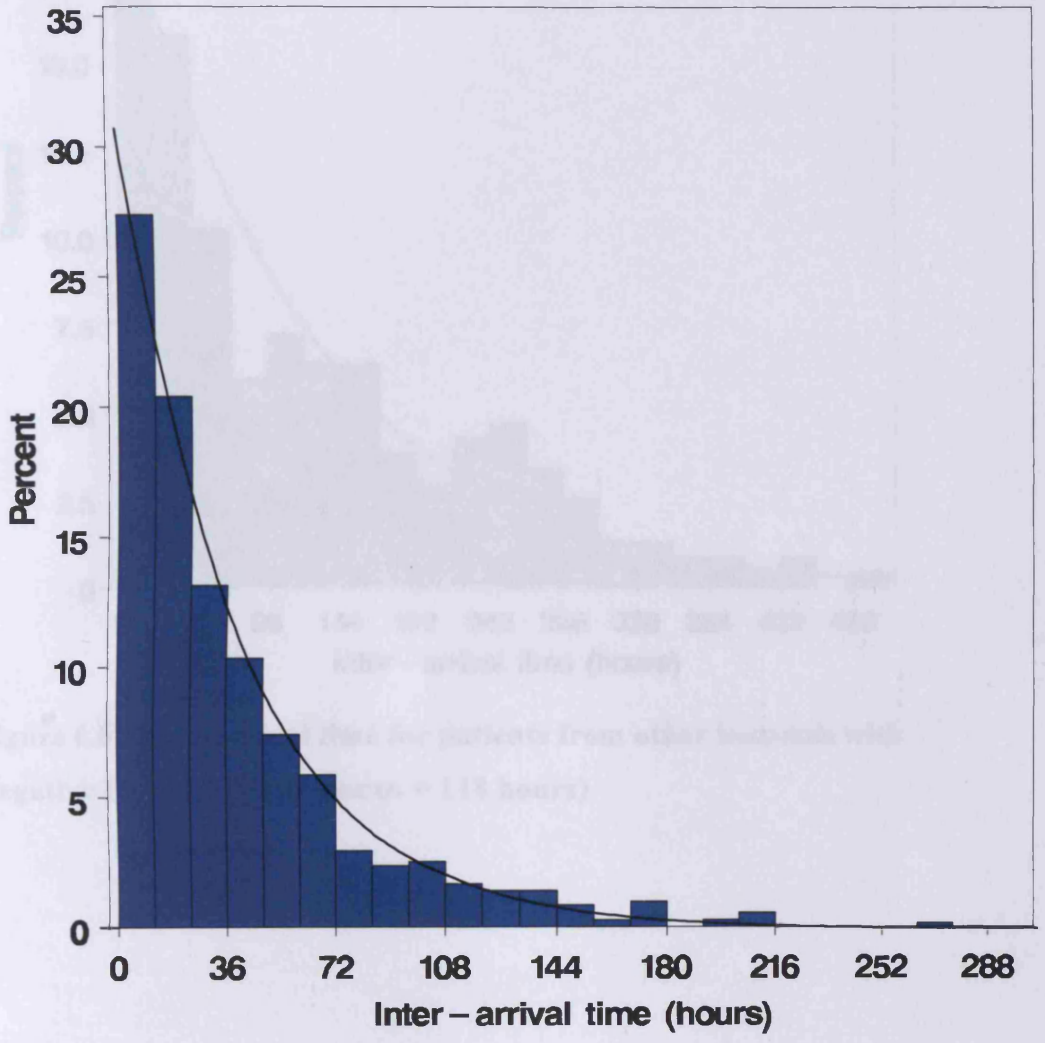


Figure 6.5: Inter-arrival time for Emergency Surgery patients with Negative Exponential fit (mean = 39.1 hours)

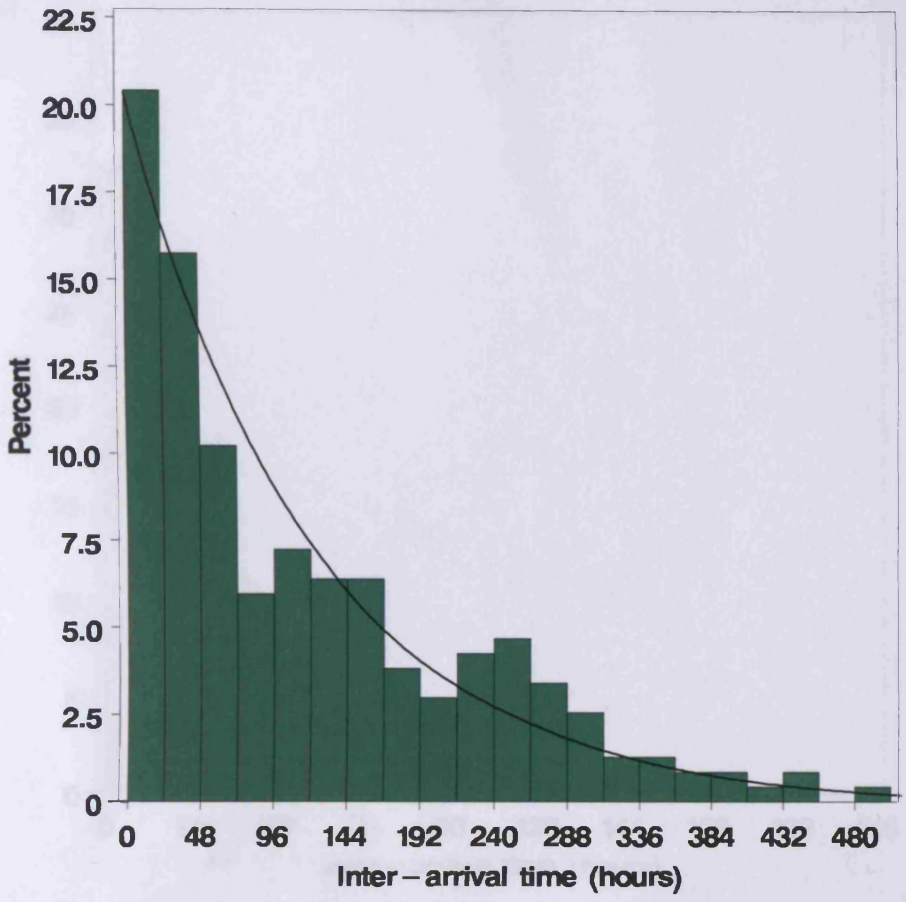


Figure 6.6: Inter-arrival time for patients from other hospitals with Negative Exponential fit (mean = 118 hours)

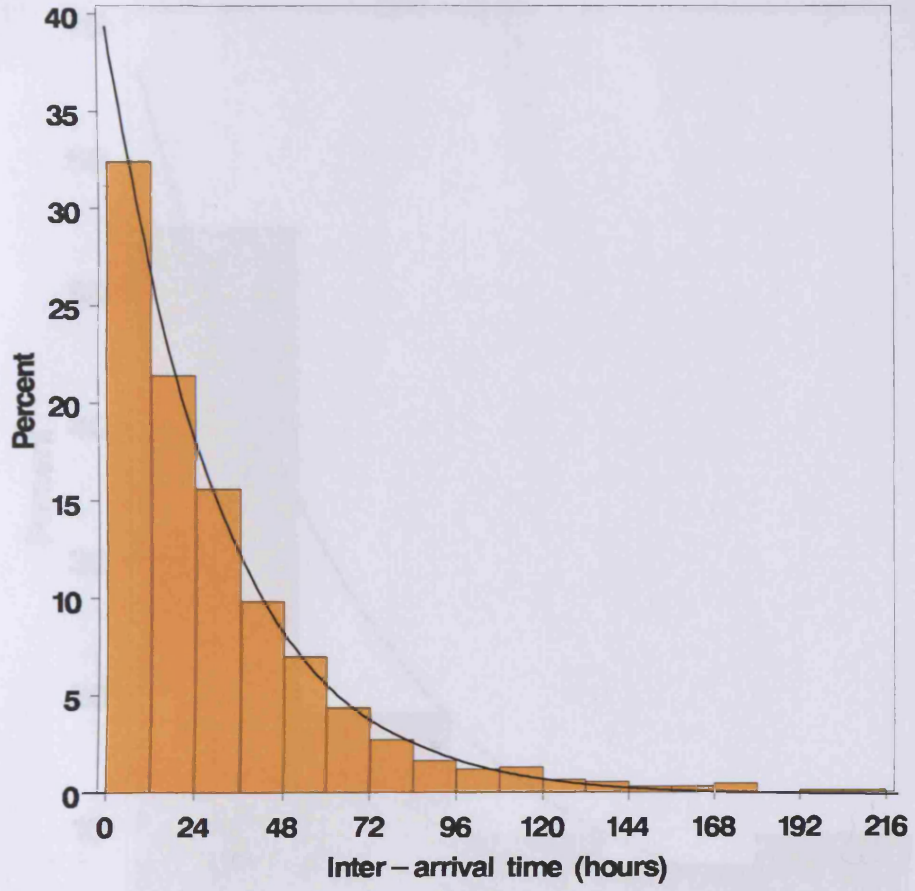


Figure 6.7: Inter-arrival time for Ward patients with Negative Exponential fit (mean = 30.5 hours)

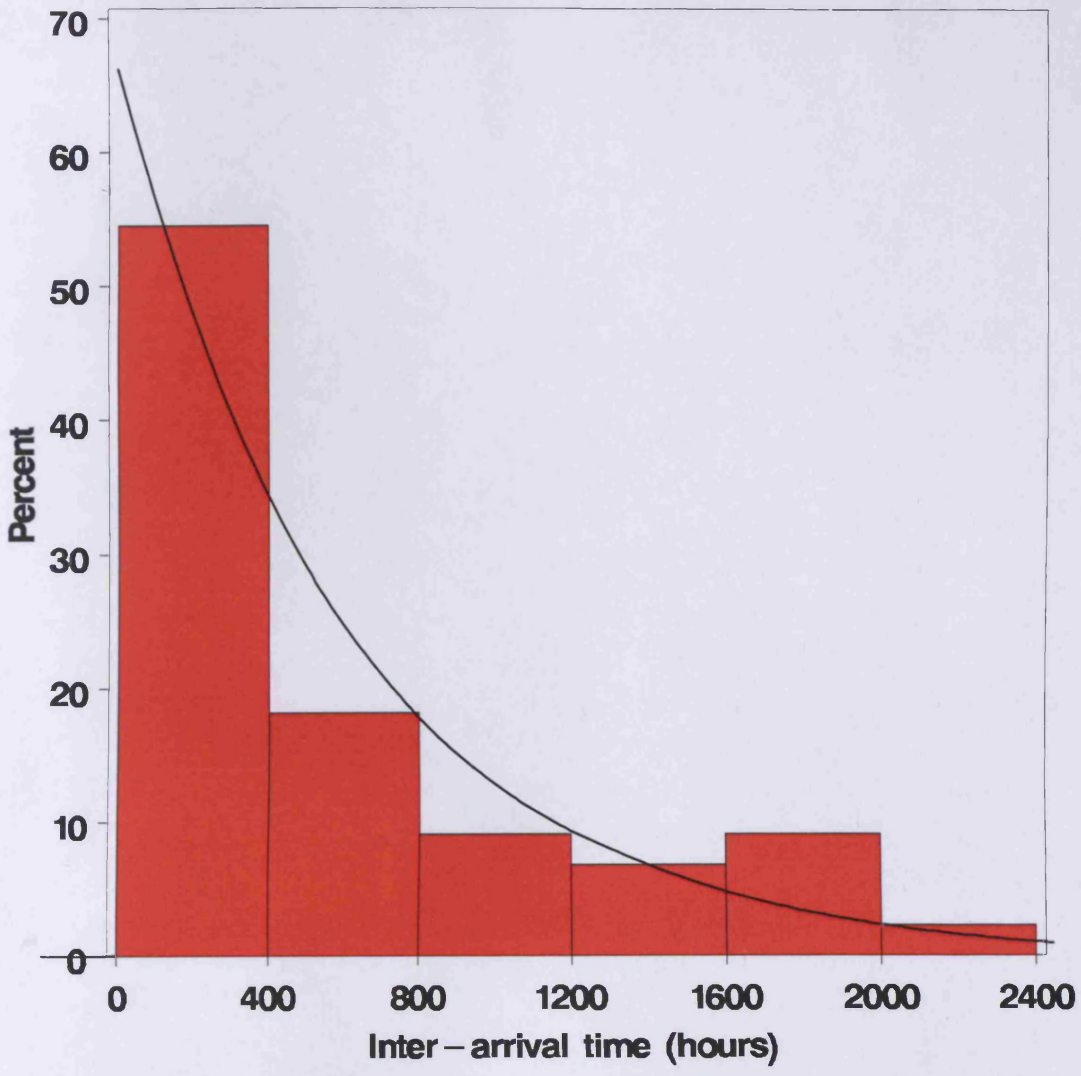


Figure 6.8: Inter-arrival time for X-Ray patients with Negative Exponential fit (mean = 604.0 hours)