



Faculty of Computing Sciences and Engineering

# Dealing with variability in the design, planning and evaluation of Healthcare inpatient units:

---

a modelling methodology for patient dependency variations.

**Matías Urenda Moris**

A dissertation submitted in partial fulfilment of the requirements for the degree of:

**Doctor of Philosophy**

**March 2010**

Sponsored by the University of Skövde, Sweden



# Abstract

This research addresses the fluctuating demand and high variability in healthcare systems. These system's variations need to be considered whilst at the same time making efficient use of the systems' resources. Patient dependency fluctuation, which makes determining the level of adequate staffing highly complex, is among the variations addressed. *Dealing with variability is found to be a key feature in the design, planning and evaluation of healthcare systems.*

Healthcare providers are facing increasing challenges resulting from an aging population, higher patient expectancies, a shortage of healthcare professionals, as well as increasing costs and reduced funding. Despite the accentuated need for effective healthcare systems and efficient use of resources, many healthcare organisations are inadequately designed and, moreover, poorly managed. Hospital systems consist of complex interrelations between relatively small units, each of which is sensitive to stochastic variations in demand. In addition to this aspect of the system view, a critical resource for the patients' wellbeing and survival is the staffing level of nurses. This puts the planning and scheduling of human resources as one of the system's foremost aims. Current tools for staffing and personnel planning in healthcare organisations do not take into consideration the workload variations that result from the variable nature of patient dependency levels.

The work presents the empirical findings of a number of case studies conducted at a regional hospital in Sweden. Principles and practical suggestions for the robust system design of *inpatient wards* using Discrete Event Simulation (DES) have been identified. Although DES techniques have, in principle, all the features for modelling the variation and stochastic nature of systems, DES has not been previously used for workload studies of inpatient wards. The main contribution of this work is therefore how a combination of DES and the data of Patient Classification Systems (PCSs) can be used to model workload variations and, subsequently, plan the nurse staffing requirements in systems with high variability. The work presented gives step by step guidance in how the analysis and subsequent modelling of an inpatient ward should be carried out. It defines a novel *modelling methodology for patient dependency variations* and length of stay modelling of a patient's dependency progression, including an adaptation to the ward's discharge figures. The modelling approach opens a novel way of analysing and evaluating the system design of inpatient wards.

# Acknowledgements

I would like to express my sincere gratitude to my supervisors Prof. Philip R. Moore, Prof. Patric Eriksson and Prof. Leo de Vin for their guidance, support and help in completing this thesis. They did not lose hope with me and gave me constructive criticism along the way. I would also like to thank the University of Skövde for its financial support during my research, in particular the Vice-Chancellor Leif Larsson aswell as the current and past Heads of the School of Technology and Society, Dr. Nomie Eriksson and Prof. Anne Persson.

I am very thankful to many individuals who provided me with help and support of different kinds. They include all my colleagues at the Centre for Intelligent Automation, the people at Skaraborg's Hospital who have been involved in this work, among others Gerald Wallsterson, Sonja Kvint and Per Sjöli. I would additionally like to express my gratitude to Dr. Mats Jägstam and Dr. Amos Ng who I have always been able to discuss ideas and problems with and who have always been supportive and encouraging, Tehseen Aslam and Marcus Frantzén who I have shared the anxiety of doing research with, Jacob Svensson, Ingemar Karlsson and Martin Andersson, who have certainly made my work easier, Dr. Thomas Lezama, Dr. Jan Oscarsson and Dr. Felipe Baesler for their advice and help, and to Vera Lindroos who has made a huge contribution in making this thesis enjoyable to read.

I am deeply grateful to my parents for their support whenever I needed it, and above all to God for giving me the possibility and sustain in life to fulfil it.

I would like to dedicate this work to my beloved wife Angélica for her whole hearted support, to my dear son Samuel for occupying my mind with things other than my thesis and to my lovely daughter Linnéa for being a ray of sunshine in my life.

Matías Urenda Moris

Skövde 2010

# Table of contents

<b>Abstract</b>	<b><i>i</i></b>
<b>Acknowledgements</b>	<b><i>ii</i></b>
<b>Table of contents</b>	<b><i>iii</i></b>
<b>List of figures</b>	<b><i>vi</i></b>
<b>List of abbreviations</b>	<b><i>vii</i></b>
<b>1 Chapter one: Introduction</b>	<b>1</b>
<b>1.1 Introduction</b>	<b>1</b>
<b>1.2 Research context</b>	<b>2</b>
1.2.1 System variability and system design considerations	3
1.2.2 Discrete Event Simulation and Modelling	4
1.2.3 Swedish healthcare system and DES	5
1.2.4 Nurse staffing and patient dependency level	6
1.2.5 Conclusions identified following literature review	6
<b>1.3 Research Aim and Objectives</b>	<b>7</b>
<b>1.4 Research Methodology</b>	<b>8</b>
<b>1.5 Thesis organisation</b>	<b>12</b>
<b>1.6 Summary</b>	<b>13</b>
<b>2 Chapter two: System and theoretical background</b>	<b>15</b>
<b>2.1 Introduction</b>	<b>15</b>
<b>2.2 The Swedish Healthcare System</b>	<b>15</b>
2.2.1 Management structure and organisation of health services	16
2.2.2 Financing and fees	18
2.2.3 Market reforms and challenges	19
2.2.4 Inpatient care	21
<b>2.3 Understanding the effects of demand variation</b>	<b>22</b>
2.3.1 Consequences of high variability	23
2.3.2 Robust system design	26
<b>2.4 Nurse staffing</b>	<b>29</b>
2.4.1 Nurse demand methods	31
<b>2.5 Discrete Event Simulation</b>	<b>35</b>
2.5.1 DES - modelling problems for Inpatient clinics	37
2.5.2 DES – First stumbling steps in the Swedish Healthcare domain	39
<b>2.6 Communication problems and barriers</b>	<b>40</b>
<b>2.7 Summary</b>	<b>42</b>

<b>3</b>	<b><i>Chapter three: Patient Classification Systems</i></b>	<b>45</b>
3.1	Introduction	45
3.2	PCS background and development	45
3.3	Identified critics and further development	47
3.4	PCSs in a Swedish case-costing context	48
3.5	Beakta® – PCS	50
3.5.1	Factor based patient classification	50
3.5.2	Activity study	54
3.5.3	Staffing	57
3.5.4	Beakta® – software tool	57
3.6	Summary	57
<b>4</b>	<b><i>Chapter four: The orthopaedic case study</i></b>	<b>59</b>
4.1	Introduction	59
4.2	Background	59
4.3	The Orthopaedic Unit – Issues to Consider	60
4.4	Case study questions	63
4.5	Model development	63
4.5.1	Software Package Used	63
4.5.2	Data Acquisition and Input Analysis	64
4.5.3	Modelling Logic and Using Graphics	65
4.6	Validation of the model	67
4.7	Experimentation and Analysis	68
4.7.1	First set of Experiments	68
4.7.2	Second set of Experiments	68
4.7.3	Analysis	70
4.8	Discussions and contribution to framework	74
4.9	Summary	76
<b>5</b>	<b><i>Chapter five: Robust design of a maternity ward</i></b>	<b>78</b>
5.1	Introduction	78
5.2	Contribution to framework	78
5.3	Background	79
5.4	Project aims	81
5.5	Project development	85
5.5.1	Data acquisition and input analysis	85
5.5.2	Model building, Verification and Validation	86
5.6	Experimentation and Analysis	88
5.6.1	First set of experiments - DoE	88
5.6.2	Second set of experiments	91
5.7	Discussions and contributions to framework	93
5.8	Summary	96

<b>6</b>	<b>Chapter six: Rehab case study – definition of modelling methodology</b>	<b>98</b>
6.1	Introduction	98
6.2	Development of framework	98
6.3	Modelling Approach	100
6.4	The rehabilitation ward at SkaS	103
6.4.1	Organisational and juridical obstacles	105
6.4.2	Mapping and modelling difficulties	105
6.4.3	Dependency variance modelling today	106
6.4.4	Stochastic modelling of dependency and LoS	109
6.4.5	Generation of transition probability matrixes	113
6.4.6	Point estimate data analysis	116
6.4.7	Adaptation of transitions to fit LoS distribution	121
6.4.8	Discharge modelling adaptation	125
6.4.9	Indicator and determinant estimation	127
6.4.10	Calculation of workload requirements	129
6.5	Proper modelling abstraction level for implementation of modelling procedure	132
6.5.1	Detailed model	133
6.5.2	Aggregated model	138
6.6	Modelling Methodology Summary	144
<b>7</b>	<b>Chapter seven: Verification, validation and future system modelling</b>	<b>147</b>
7.1	Introduction	147
7.2	Validation procedure	147
7.2.1	Patient groups	149
7.2.2	Key parameter measurement	150
7.2.3	Validation conclusions	155
7.3	Experimental case study	156
7.3.1	Experimental case system setup	157
7.3.2	Experiments and analysis	159
7.3.3	Experiments conclusions	164
7.4	Conclusions	164
7.5	Summary	165
<b>8</b>	<b>Chapter Eight: Conclusions</b>	<b>167</b>
8.1	Introduction	167
8.2	Dissertation results in summary	167
8.3	Meeting the research aim and objectives	168
8.4	Contributions of this research	170
8.5	Further work	171
8.5.1	Improvement in patient grouping	172
8.5.2	National dependency matrix libraries	172
8.5.3	IT solutions and operational scheduling	172
<b>9</b>	<b>References:</b>	<b>174</b>
	<b>Earlier Publications</b>	<b>189</b>

# List of figures

Figure 1. Relevant situations for different research strategies. (Yin 2007)	10
Figure 2. The operational procedure during the research work	11
Figure 3. Ranking of countries by type of health index (SALAR 2005a, pp.6)	16
Figure 4. The organisation of Swedish health services (Swedish Institute 2007)	17
Figure 5. Internal arrival rate of patients into an orthopaedic trauma ward	24
Figure 6: Research areas in managing variation (McLaughlin 1996)	26
Figure 7. The structural model of the Beakta system	51
Figure 8 Description of work procedure for patient classification	53
Figure 9. The general flow of patients between the different units	61
Figure 10. Wards 83 and 84, modelled in Quest®, patient colour visualises gender	66
Figure 11. Results from the first set of experiments where contusion and fracture patients are moved on the 4 <sup>th</sup> day, contrasted with 2002 values.	70
Figure 12. Results from second set of experiments – relation between number of refused, number of beds and occupancy level of wards 83 and 84 under the different scenarios	72
Figure 13. Description of the Perinatal centre and the patient flow between the different units. Patient categories are identified by their DRG groups.	80
Figure 14. Arrival rate figures of patient's p/d for 2004-2005 and LoS figures for the same period	81
Figure 15. Total number of births: in Sweden 1973 to 2006 and at SkaS 1990 to 2005	82
Figure 16. Original ward configuration	87
Figure 17. Main effects of factors and their SN ratios	91
Figure 18. Results from model 3 analysis presenting both occupancy and service level	92
Figure 19 Suggestion for future Perinatal centre with an 18 room maternity ward.	94
Figure 20. The flows among units related to ward 75 and their patient contribution in percentage	104
Figure 21. Ward 75's room configuration and layout.	104
Figure 22. Modelling steps of suggested methodology leading to the analysis and design of a new improved system.	102
Figure 23. Dependency profile of a specific patient, dependency H represents leaving the system	106
Figure 24. Four patients' dependency variance and LoS	108
Figure 25: Illustrates the states a patient can be in and the probability of transition between these states	109
Figure 26. Proportion of patients, in the different dependency groups from arrival up to LoS of 27 days	116
Figure 27. CI length vs. number of observations for two point estimations	118
Figure 28 Points estimates of transition probability $\hat{p}_{CB}^{(n)}$ , CI Low (CBL) and Upper limit (CBU), Moving average smoothing (CB-MA) and Polynomial regression (Poly. CB)	120
Figure 29. Frequency Comparison plot between the historical data and the estimated distribution.	122
Figure 30. Absorbing probabilities to state H for patients in the sample group	123
Figure 31. Weekday's variation of discharge index	126
Figure 32. Determinant definition flow for the 7 indicators	129
Figure 33. Correlation between Time for Hygiene activity and the morning routines	135
Figure 34. Snapshot of male patients' morning activities	136
Figure 35: Model abstraction in the simulation process (Frantz 1995)	138
Figure 36. Aggregated model view of the 14-bedplaces rehabilitation ward	140
Figure 37. Weekdays' Base work time distribution	142
Figure 38. Distribution of patient's total workload requirement over periods of the day and staff categories for weekdays.	143
Figure 39. Summary of workload variations in minutes from 164 days of ward 75 where the number of patients is equivalent to 14, historical data.	154
Figure 40. Workload variations in ward 75 for different numbers of patients and days	155
Figure 41 System configuration and flows, where ward 85 represents the new ward	158
Figure 42. The day to day record of the three wards' workload variation, presented in minutes.	160
Figure 43. Aggregated workload variation for the three wards, presented in minutes	161
Figure 44. The relation between the number of patients and the system's workload in ward 83.	162
Figure 45. The relation between the number of patients and the system's workload. System size 160 beds	163

# List of abbreviations

AI	Artificial Intelligence
AN	Auxiliary Nurse
CHSA	Centre for Healthcare Analysis
DES	Discrete Event Simulation
DoE	Design of Experiments
DRG	Diagnosis Related Groups
ED	Emergency Department
GA	Genetic Algorithm
ICU	Intensive Care Unit
KAVA	Surgical emergency ward
KKP	Cost per Patient
KKS	Kärnsjukhuset (main hospital of SkaS)
LoS	Length of Stay
MCMC	Markov Chain Monte Carlo
MHSA	Ministry of Health and Social Affairs, Sweden
MLE	Maximum Likelihood Estimation
NICU	Neonatal Intensive Care Unit
NNs	Neural Networks
NPOB	Nurses per occupied bed
OEC	Overall Evaluation Criteria
PCS	Patient Classification System
QC	Quality Characteristics
REHAB	Rehabilitation ward
RN	Registered Nurse
RSM	Response Surface Methodology
SALAR	Swedish Association of Local Authorities and Regions
SBO	Simulation Based Optimization
SEK	Swedish Krona, currency
SFS	The Health and Medical Services Act, Sweden
SkaS	Skaraborg Regional Hospital
S/N	Signal to Noise
SoS	The National Board of Health and Welfare, Sweden
TQM	Total Quality Management
V&V	Verification and Validation
WNM	Ward's Nurse Manager
WHO	World Health Organization



# 1 Chapter one: Introduction

## 1.1 Introduction

According to the World Health Report of 2006, the cost of healthcare services accounts for an average of more than 11 per cent of the OECD nations' gross domestic product (GDP) and prognoses point to increasing costs in the coming decades (WHO 2006). A newly presented research study reveals that Sweden's healthcare costs are expected to increase by 270% before the year 2040 (Klevmarken and Lindgren 2008). Despite increasing costs arising from the use of new and more expensive pharmaceuticals and innovative treatment options, the major cost of healthcare services is personnel related, accounting for approximately 70 per cent of the total cost (Hallin and Siverbo 2003; SALAR 2005a). Proficient personnel are a scarce resource in many countries and there are strong indications that this situation may deteriorate (Buerhaus et al. 2000; WHO 2006). The current situation and future development emphasise the need to plan, schedule and use resources, especially human ones, intelligently. The need for suitable tools and techniques to design and manage appropriate personnel levels in healthcare systems is increasing as the current circumstances become more urgent.

Unfortunately, most of the tools that are available for managers to plan staffing levels do not take into consideration the dynamic and stochastic nature of the system. Staffing levels are based on heuristics or, in the best of cases, static patient to nurse ratio. This may work well in an "average" world, but reality is not the average of things. Reality fluctuates; it contains multiple variations of system variables that dictate the system's performance over time (McLuaghlin 1996; Noon et al. 2003). If the system configuration is not robust and/or staffing does not take into consideration the changing resources needed, it will be extremely difficult to make efficient use of scarce resources. In order to improve system design and staff planning, tools that adequately take into consideration the system's variability need to be used (McLuaghlin 1996).

This research presents a new modelling methodology which provides the opportunity to build realistic simulation computer models of inpatient units with an existing workload measurement system, "Patient Classification System" (PCS). The benefits and outcomes

from the merger of Discrete Event Simulation (DES) and PCS data may represent the most promising combination for the extension of simulation solutions to inpatient clinics and, in its prolongation, the possibility of designing robust healthcare systems. These systems would take variability in the fluctuating resource demand into consideration and use not only room and bed places as design parameters but also nurse staffing levels.

## ***1.2 Research context***

There are several terms and relations that need to be defined in order to grasp the settings and background of this work. These include both techniques and concept descriptions, which are interrelated and consequently presented in four subsections. The description starts by explaining the importance of being aware of how variance and randomness affect healthcare systems' efficiency and the need of having inherent variability in mind when these systems are designed. This section is followed by a discussion on the adequateness of DES as a tool for healthcare system design and introduces the importance of system modelling being both art and science. The subsequent section contains a brief description of the Swedish healthcare system, its pros and cons, but, above all, it discusses the lack of tools used in system design, planning and evaluation. The section that follows focuses on the importance of nurse staffing and its relation with patient dependency, while the final section represents a gap analysis which highlights the current situation needs. This serves as a foundation for the definition of the research's main objectives.

Finally, a few words on the nature of this research which, from the researcher's point of view, represents a clear domain change from the strictly industrial applications of production engineering and management to the healthcare area. This has resulted in a cultural and social change as well as an important mindset adaptation. The following sections present the current scene that will hopefully provide the reader with the necessary background to understand the future work.

### 1.2.1 System variability and system design considerations

The Centre for Healthcare Analysis (CHSA) in Sweden describes the Swedish healthcare system to be:

*“complex ... a business which consists of many varying activities, with different mutual dependencies and interested parties and which often is ruled by events and therefore difficult to define and unpredictable” (CHSA 2002, pp.54).*

This description is probably adequate for healthcare systems in general. Several authors describe the complex and variable nature of healthcare systems and the difficulty of efficiently managing healthcare delivery (McLaughlin 1996; Noon et al. 2003; Walley et al. 2006). One of the difficulties lies in process owners not understanding the impact that variations have on the system's performance and what can be done to mitigate their effects. It is commonly believed that all the variations are generated from the fluctuations of patient demand, but this is only part of the truth. McLaughlin (1996) defined the variations of healthcare systems to be of two types. The first one is called inherent (common cause) variations that is generated naturally from, for example, the random arrival of patients or differences in the time it might take to complete a manual task. The second type, unnecessary (or special cause) variations, is unnecessary because it can be omitted with better planning and/or scheduling, or by implementing standardized working procedures. This is why Noon et al. (2003) call this variation type preventable. Litvak et al. (2005), on the other hand, recognize the importance of reducing the unnecessary variations, but make a different classification of sources of variability. They identify three sources that put stress on healthcare delivery systems. The first is called *flow stress* and refers to the variability arising from patients' rate of arrival for hospital care. The second, *clinical stress*, refers to the stress arising from the variability in type and severity of disease (dependency level of the patient). The third stress factor comes from the variability that originates in the *competing responsibilities and abilities* of the healthcare providers.

This discussion reveals that reducing unnecessary variability only solves part of the problem. In order to avoid inefficiencies and queues, a system designer must bear in mind the inherent variations. The general approach is therefore 1) to work with reducing the variance that is *preventable* and 2) design the system in a robust manner so that it can handle the *inherent* variations. Several papers present lessons learned and considerations

on system design (McLaughlin 1996; Noon et al. 2003; Litvak et al. 2005; Walley et al. 2006). Both McLaughlin (1996) and Noon et al. (2003) state the importance of using simulation or other analytical tools to be able to define the most adequate system configuration. Walley et al. (2006), on the other hand, use a vast number of “improvement programs” to gain insight. In addition, both Walley et al (2006) and Litvak et al. (2005) worked with the real system when experimenting on new approaches for the purpose of improving healthcare delivery. The value of working with the “real system” is obviously significant, but these improvement programs were made on systems already inadequately designed. Even if this is a feasible way of testing new approaches, it is not the optimal solution for a number of reasons. Firstly, even with a well based heuristic solution, it is a trial and error approach that puts a lot of stress on the organisation. The numbers of possible solutions evaluated are limited for obvious practical, economical and quality reasons. Finally, it is difficult to fine tune the system parameters to obtain a robust and more optimal setup. A better approach is presented in the following section.

### **1.2.2 Discrete Event Simulation and Modelling**

Discrete Event Simulation is a relatively well-established technique in the healthcare domain. Its benefits are well documented and it is gaining more and more recognition among healthcare professionals (Jun et al. 1999; Jacobson et al. 2006; Kopach-Konrad et al. 2007). Its dynamic and stochastic modelling features make it an ideal tool for healthcare system design, understanding and evaluation (Eldabi 2000). DES gives the simulation modeller the possibilities to design and model systems, whether it is a manufacturing, healthcare or other (stochastic) system, in order to study and understand its current or future behaviour. The objective is to obtain a better understanding of the system and be able to test different configurations and control options under the changing conditions the system is facing or will face, in order to evaluate the design and control applied.

System modelling has been described as a process requiring both art and science. The boundaries, with regard to system modelling, are not only related to the modeller’s skills and knowledge but also to the accessible data on the system of interest. This research focuses on how DES and data from PCS can be used to attain a patient dependency modelling that reflects the real system’s variations. And thereby give the modeller the tools for better system design, planning and evaluation.

### **1.2.3 Swedish healthcare system and DES**

The Swedish medical service is among the best healthcare systems in the world (SALAR 2005b). Every registered citizen in Sweden has access to healthcare regardless of income, sex, age or geographic residence (Klevmarken and Lindgren 2008). However, there is still room for improvement. The waiting lists for some operations and treatments are still long and uncertain, and there are flaws in how patients are being treated. In addition, some patient categories have difficulties getting the treatment they need because of poor coordination among the healthcare service providers. Furthermore, Sweden's healthcare expenditure per capita for 2006 was 9.1 per cent of its GDP per citizen and year (SoS 2008a), which is a considerable expense and a 24 per cent increase since 1993, and expected to rise considerably in the coming decades (SALAR 2005a; Klevmarken and Lindgren 2008).

Currently, almost all research in the Swedish healthcare system is dedicated to clinical research, medical devices and equipment, and IT system development, but very little has been done with regard to the use of engineering tools and approaches for the design and efficient management of the system (there are exceptions see, Persson 2007). This is an international problem, even though other countries have a higher awareness of the predicament. Proctor et al. (2005) highlight several reasons for this situation (supported by the CHSA (2002) report) of which two include (1) few incentives for the system's improvement in the current organisation, management and regulation of the healthcare system, and (2) lack of awareness of its importance and a negligent attitude towards production control ideas. The outcome is that healthcare managers seldom have the knowledge or tools for an analysis of their systems. Consequently, despite the obvious benefits of using DES in the healthcare domain, there has been no research study of DES in the Swedish healthcare system. This work thus constitutes one of the first major studies. This means that the work has a distinct exploratory approach, and many of the findings consist of confirming (or rejecting) international conclusions in this field of research. The Swedish context of these issues does not obscure the value that the international healthcare community can obtain from this work. The problems and issues targeted are of common interest in the design, planning and evaluation of healthcare systems.

#### **1.2.4 Nurse staffing and patient dependency level**

Nurse staffing is an area in which considerable effort and research have been made (Hurst 2002). The aim is to have a suitable staffing level to match the fluctuating patients' needs. Managers need to consider healthcare quality, costs and the accessibility of human resources when planning for the right staffing level. There is also an important relationship between work overload and the increase of illness and absenteeism among nurses, as well as between work overload and increased mortality risk among patients, which implies that the aim is not only about achieving high figures on personnel utilization (Rauhala et al. 2007; Litvak et al. 2005). Healthcare managers need to adapt the personnel and its mix to their unit's variable workload requirements, taking into consideration the personnel's well being and work satisfaction (Shuldham 2004). This is a difficult and complex task especially in units with highly variable numbers of patients and/or patient dependency.

A frequently used term for nursing workload systems is "Patient Classification System" (PCS) (Edwardson and Giovannetti 1994). This term might be misinterpreted in the healthcare domain because patient classification systems for other aims are common. In this context, PCS refers to "the identification and classification of patients into groups or categories, and to the quantification of these categories as a measure of the nursing effort required" (Giovannetti 1979). This means that there is a close relation between the patients' dependency level and the nursing effort or workload that this dependency level generates. The sum of all the patients' generated workload with the addition of the administrative tasks would represent the total, day to day workload variation of the ward. When this figure is translated to number of staff and staff categories, it would represent the best possible staffing level.

#### **1.2.5 Conclusions identified following literature review**

The above subsections have highlighted several interdependent topics that set the scene for the research work. They can be summarised as follows:

- The importance of an adequate design that considers the high level of demand variation healthcare systems experience. Robust design is introduced as a design approach that takes into account inherent system variations.

- The use of an appropriate modelling and simulation technique, where Discrete Event Simulation (DES) is introduced as a tool for achieving the aim of designing robust healthcare systems.
- The lack of use of appropriate tools and lack of awareness of the importance of considering the systems' stochastic behaviour in the design and planning of the systems' operations.
- The importance of considering patient dependency levels and their inherent variation when determining appropriate staff levels. One tool that is introduced for measuring patient dependency levels and thereby calculating the appropriate staff levels is Patient Classification Systems (PCS).

Even though some of these topics can be addressed independently, the sum of them is higher than their individual parts. This work, both the literature review as well as the empirical work, identifies that the design of efficient healthcare systems needs to address the different sources of variation affecting the system, of which patient dependency variations and their effect on staffing requirements have previously been neglected. We advocate that the design of robust systems can be achieved through the use of DES. *Moreover, stochastic patient dependency variation and its corresponding effect on staffing levels have not previously been considered in the design of inpatient healthcare systems.*

### **1.3 Research Aim and Objectives**

Demand variations affecting healthcare systems put stress on their resources and make running them efficiently difficult. This has been known for decades, but very little has been done with regard to taking these variations into consideration when systems are designed and controlled. This is especially true when it comes to personnel dependency variations and their relation to staffing levels.

*The primary research Aim is to identify how and why DES can be effectively utilized to design, plan and evaluate inpatient healthcare systems and their nurse staffing requirements.*

Nevertheless, Kopach-Konrad et al. (2007) define a system and its behaviour as a “set of possibly diverse entities (patients, nurses, physicians, etc.) each performing some set of functions. The interaction of these entities as they perform their various functions gives rise to a global system behavior”. It is thus therefore obvious that adequate staffing is a part, a brick, in a bigger context. How big is the context? How much consideration needs to be taken to other system parts? The answer obviously depends on the question being targeted, but it is impossible to define the scope of the system of interest without a proper understanding of the system. Eldabi (2000) states that one of the main objectives in simulation modelling is achieving problem understanding. It is through problem understanding that real success in healthcare management is attained. With this in mind, the work presented has a broader system approach, looking at different aspects of inpatient, ward system design and variation reduction, for the purpose of achieving a clearer picture of the problem and proposed solution. The research is therefore divided into the following sub-objectives:

- To carry out an exploratory pilot study and a complementary literature survey:
  - To gain insight into the specific features of the Swedish healthcare system, as well as its opportunities and limitations as an aid in the research design.
  - Review different approaches to nurse staffing and patient dependency modelling.
  - Identify sources of demand variations and how they affect the healthcare system units.
- Identify the principles and suggestions of the best robust system design practices based on the results of the experimentation and analysis of the case studies.
- Develop an appropriate modelling methodology for inpatient dependency variation, specifying the data requirements and data sources.
- Use a system approach to highlight the benefits of an improved modelling methodology.

## **1.4 Research Methodology**

There are some fundamental differences between research in the engineering disciplines and research in natural science (such as biology and physics). According to Braha and



Maimon (1998), three distinguishing differences are: (1) Engineering is concerned with synthesis while natural science is concerned with analysis. (2) Engineers are concerned with how things ought to be, while natural science is exclusively concerned with how things are. (3) Engineering is creative, intuitive and spontaneous, while natural science is rational and analytic. Viewing engineering research as a problem solving discipline is therefore well in line with its ambitions.

Bearing this in mind, what then is the most suitable research approach for the research objectives presented? Procter et al. (2001) give suitable recommendations when they call the systems engineering community into action to help healthcare professionals solve the problems of healthcare delivery systems. One of the recommendations (recommendation 5-1b) defines the following three-fold mission for multidisciplinary research centres, in summary: (1) to conduct basic and applied research; (2) to demonstrate and diffuse the use of these tools, technologies and knowledge (technology transfer); (3) to educate and train future professionals. Therefore, although not presented as objectives, extensive work in diffusing, training, and education has been performed as project related activities. However, more importantly, the nature of the problems obliges a research strategy that is based on empirical studies and has a clear problem solving ambition.

Yin (2007) defines relevant conditions for the following five different research strategies: *experiment*, *survey*, *archival analysis*, *history* and *case study* (see Figure 1). All strategies have their advantages and disadvantages. In order to choose the most suitable strategy three conditions are presented: (a) the type of research question (b) if control over behavioural events is required and (c) whether the focus is on contemporary events or historical phenomena. Depending on the answers to these conditions and using the matrix in Figure 1, the researcher can identify the most suitable main strategy. There are obviously some overlaps between the different strategies and more than one strategy can be used in the study.

Due to the nature of the research question and its context, no control being required over the system's behavioural events, and a focus on contemporary events, a case study based

research strategy was chosen. In addition, the research work was complemented by regular statistical analysis and experiments.

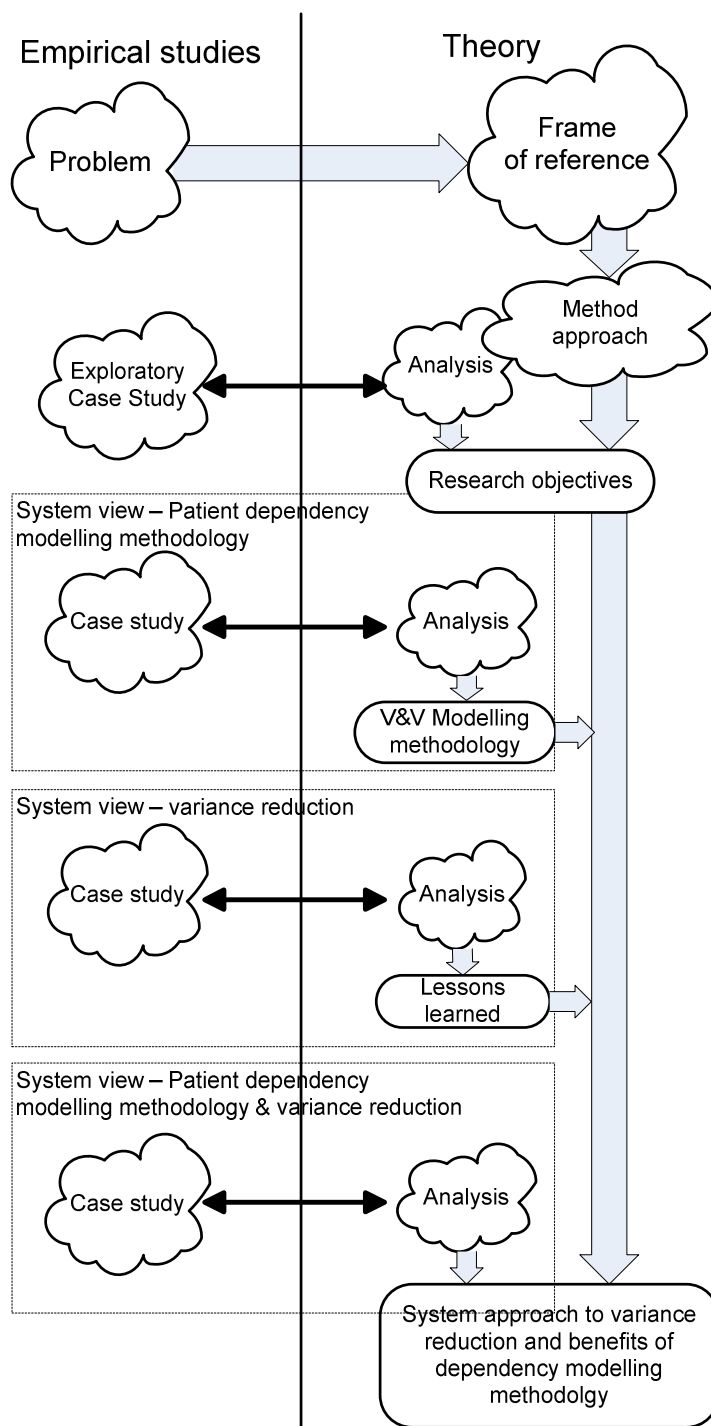
<b>Strategy</b>	<b>Form of research question</b>	<b>Requires control over behavioural events?</b>	<b>Focus on contemporary events?</b>
<b>Experiment</b>	How, why	Yes	Yes
<b>Survey</b>	Who, what, where, how many, how much	No	Yes
<b>Archival analysis</b>	Who, what, where, how many, how much	No	Yes/No
<b>History</b>	How, why	No	No
<b>Case study</b>	How, why	No	Yes

**Figure 1. Relevant situations for different research strategies. (Yin 2007)**

The work involved four case studies, where the first one served as an exploratory pilot study for the purpose of gaining insight and aiding in the definition of the future research design. The first case study is therefore regarded as the opener for the transition from the manufacturing to the healthcare arena. Moreover it contributes to the foundations of the proposed modelling methodology by providing empirical evidence and an understanding of how system variances affect inpatient systems. However, it is also enlightening because it helped adapt the research aim towards the healthcare provider's main questions. The second and third case studies were identified with the help of the initial study and the literature survey. Both these later studies follow a single study approach (Yin 2007) addressing complementary questions:

- The second study centres on how and why a system with high variability needs to use robust system design in order to be efficiently run. The case study corroborates the need to model the different sources of variation in order to adequately address system design issues. It thereby contributes to the framework development by identifying the need and context in which the proposed modelling methodology fills a role.
- The third case study focuses on how DES can be used to define the need of staff resources based on adequately modelled, patient dependency variations. This study serves as an arena for the presentation of the most important contribution of this work, namely a modelling methodology for patient dependency variations.

The fourth and final case study is used to set the proposed methodology for dependency modelling in a wider system context. The objective is to describe its possibilities and benefits with regard to designing inpatient systems for variance reduction. This fourth study is more educational/explanatory in the sense that it is based on a fictional yet realistic scenario. The operational procedure during the research is illustrated in Figure 2.



**Figure 2. The operational procedure during the research work**

## **1.5 Thesis organisation**

This section presents the outline of the dissertation giving a brief summary of the chapters. The dissertation is organised as follows:

**Chapter two** gives the reader the background and the system information that forms the scenario of this thesis. Considering the interdisciplinary nature of the work, one of the objectives of this chapter is to fill the possible gaps for the reader, regardless of whether they have an industrial or healthcare domain background. Among the chapter's contributions is the identification of a nursing demand method that supports the aims of this work. It also defines an appropriate simulation technique and a system design philosophy. Finally, the chapter touches upon some of the organisational and cultural differences between industry and healthcare systems.

**Chapter three** describes Patient Classification Systems, their use, different types, pros and cons, and especially why they are interesting for the modelling of staff requirements in inpatient units. The chapter also puts the use of PCSs in a Swedish context. The later part of the chapter describes a PCS software that is commonly used by the Swedish healthcare providers and how it is used to calculate patient dependency levels and corresponding staff levels. The main objective is to give the reader the knowledge to understand the modelling methodology presented in chapter six.

**Chapter four** provides an account of the first of four case studies. This first study had an exploratory aim and served as a first contact with the healthcare system's modelling and design. It took place in two adjacent orthopaedic wards. The chapter highlights how ward interdependency and demand variation cause delays and difficulties in system planning and scheduling. Moreover, the study identified some of the pending questions that are dealt with in this work. How do we design inpatient systems in a robust manner, and how can DES be used for deciding the appropriate staff level?

**Chapter five** contains the description of the second case study and presents how DES can be used for a robust system design of a maternity ward. This case study emphasises and verifies lessons learned from the first study and gives valuable insight into both simulation modelling and system design. The difficulty which confronted the obstetric unit's

managers was a lack of understanding how a high utilization level affected the aims of a higher system service level. Could a good compromise be found through the use of DES and robust system design? The study takes into consideration demand variation in both the short and long term.

**Chapter six** includes a summary of current approaches for dependency modelling and Markov processes in particular. More important, it describes a modelling methodology that uses PCS data in combination with DES to model patient dependency variations and the resulting staffing implications. In the chapter, this methodology is exemplified by its application in the third case study, which was carried out at an orthopaedic rehabilitation ward. The chapter also presents a new way of modelling patients' LoS that more truthfully considers the weekday's discharge correlations. It concludes with a discussion of two abstraction levels for the modelling of staff requirements.

**Chapter seven** continues from the conclusion of the previous chapter, describing how the proposed dependency modelling methodology is verified and validated using historical data from the rehabilitation case study. It summarizes and concludes what has been learned and exemplifies the findings by applying the modelling methodology and robust system design lessons in a final hypothesised scenario, which was based on both real and fictional data. However, the scenario and the data used represent a realistic suggestion that was agreed upon together with the healthcare professionals at Skaraborg's Hospital.

**Chapter eight** contains the final summary and conclusions. It highlights the contributions to knowledge of the work and discusses the future challenges of the healthcare sector in the area of inpatient ward modelling. It also presents further work in this field in order to achieve the full benefits of the proposed modelling approach.

## **1.6 Summary**

This chapter describes the rising cost level of the healthcare services in general and the Swedish healthcare system in particular. Available prognoses indicate that the ageing world population and its expectations will, if nothing radical is done, increase healthcare costs to a precarious level. Bearing in mind that healthcare is a service enterprise and, as such, mainly relies on personnel resources, most of the variable cost is staff related.

Trained and motivated healthcare staff is one of the key parameters for saving lives and improving the health of the population. Unfortunately, both efficient staffing and design of healthcare systems are, in general, extremely difficult as a result of the high demand variation that healthcare systems experience. In view of these factors, there is an expectation that tools and techniques to efficiently plan, improve and design healthcare systems and their operation would be numerous and diligently used by healthcare professionals. The reality is unfortunately much different. The need for awareness and use of appropriate tools for robust system design is essential.

Several interrelated topics are introduced to clarify the problem and present the approach of this thesis. These topics have, when possible, a contrastive narration between healthcare systems in general and the Swedish healthcare system in particular. The discussion centres on the need to address the different sources of variation affecting healthcare systems, of which patient dependency variations and their effect on staffing requirements have previously been neglected. It advocates the design of robust healthcare systems through the use of DES as a suitable tool. Moreover, it identifies that stochastic patient dependency variation and its corresponding effect on staffing levels, have not previously been considered in the design of inpatient healthcare systems.

The objectives of this research can be summarised as follows: *identify how DES can be effectively utilized to design, plan and evaluate inpatient healthcare systems and their nurse staffing requirements, in a holistic system context.* This means that staff levels are not an isolated modelling parameter. If robust system design is to be achieved, then a more complex scenario including additional parameters and system understanding needs to be part of the solution.

The research methodology is based on four case studies. The ambition has been to gain insight into the systems' true needs and support the technology transfer from the engineering to the healthcare domain. Some of the expected added benefits include facilitating the education of healthcare professionals, providing more efficient healthcare systems and lowering the society's healthcare costs. Moreover, considering that we are all users of the healthcare system, this seems a desirable aim.

## **2 Chapter two: System and theoretical background**

### ***2.1 Introduction***

Words like interdisciplinary and cross-cultural research have been widely used during the past decade. However, true interdisciplinary research is not easy to establish and demands, among other things, a different academic incentive and reward structure to be successfully adopted (Rhoten 2004). The approach presented does not attempt to be a successful example of interdisciplinary research, but can be said to be the fruit of an institutional aim to achieve its benefits, combining knowledge from different domains into something new. For the researcher, it has meant making a domain change and adaptation, not only in his conceptual world, changing from strictly engineering applications to the healthcare domain, but also in terms of a social and cultural change.

The following chapters (two and three) provide a theoretical background, presenting the scene of today and hopefully giving the reader the context with which to understand the final solution. For those with an engineering background, the healthcare related information will be crucial to understanding the particular difficulties that lie in simulation modelling and healthcare system improvement. For healthcare professionals, on the other hand, the more technically related sections will be of more importance. Regardless of the readers' background, the subject that concerns understanding the effects of demand variation is indispensable for future comprehension.

### ***2.2 The Swedish Healthcare System***

The healthcare system is one of the corner stones of the Swedish welfare system. It is something that strongly concerns public opinion and politicians. During recent decades, it has been discussed, criticized and submitted to a considerable number of reforms (Anell 2005). Nevertheless, despite what might seem to be the general public's discontent, it accounts for relatively high public trust (Bergmark 2008). The population and patient survey of 2006 shows that 74 per cent agreed wholly or in part with the statement that "I have access to the healthcare I need", which represents an increase compared to earlier years (SALAR 2008, pp.103). Moreover, despite the problems the Swedish healthcare system is facing, it has been consistently performing well in an international context (SALAR 2005a). In fact, it has performed so well that it is ranked among the top 4 in

several different healthcare indexes, see Figure 3. This positive judgement does not suggest that healthcare management or the general public is content and satisfied with the system's performance. There are still many issues to improve, especially with regard to accessibility of care services and overall cost levels.

	Health status index 2001	SCP composite index 2001	WHO composite index 1997	Highest health-expenditure 2001
Sweden	1	2	1	11
France	10	1	3	6
Austria	6	3	7	9
Luxembourg	17	4	2	3
Germany	14	5	11	2
Belgium	11	6	10	10
Netherlands	12	7	5	7
Australia	2	12	9	8
Finland	3	9	15	15
Spain	4	11	13	19
Canada	8	13	4	4
Denmark	5	8	14	5
Ireland	15	10	17	13
Italy	9	17	8	12
United Kingdom	16	16	6	14
Czech Republic	18	14	19	20
New Zealand	7	15	18	16
United States	19	18	12	1
Greece	13	19	16	18
Slovak Republic	21	20	22	22
Portugal	20	21	20	17
Poland	22	22	21	23
Hungary	23	23	23	21

Figure 3. Ranking of countries by type of health index (SALAR 2005a, pp.6)

### 2.2.1 Management structure and organisation of health services

The responsibility for the healthcare system is divided between the state, county councils and municipalities (see Figure 4). The Health and Medical Services Act (SFS) sets out the respective responsibilities of county councils (county governments) and municipalities for health and medical care. The central government is represented by the Ministry of Health and Social Affairs. This governmental office supervises and evaluates the healthcare delivery of local governments, as well as determines the system's overall objectives with the help of several health and medical care agencies, of which the most important is The National Board of Health and Welfare (SoS).

Sweden is divided into 20 county councils. One municipality, the island of Gotland, carries the same responsibilities for healthcare as the county councils. Approximately 90 per cent



of the county councils' responsibilities are related to healthcare. The other 10 per cent involves areas such as culture and infrastructure. Each county council is responsible for providing services in its geographical area, but collaboration among councils is very common, especially with regard to sharing the resources of regional hospitals (so-called university hospitals). The county councils are run by democratically elected bodies.

The healthcare system is heavily decentralized and each of the county councils has considerable freedom in how to organize and manage the delivery of its healthcare. This freedom is regulated by the new Local Government Act which came into force on the 1st of January 1992. This means that there are differences in how the system components are organized, the level of taxes and patient fees, as well as the percentage of healthcare delivery provided by private healthcare facilities. The dominant employers of healthcare professionals are the different county councils. Private healthcare actors account for only a few per cent of the healthcare provided, even though this level has been increasing during the last decade (Anell 2005; Bergmark 2008). Almost all the hospitals and most of the primary healthcare centres are owned and operated by the county councils.

<b>The organisation of Swedish health services</b>			
<b>Central government</b>		<b>Local government</b>	
Ministry of Health and Social Affairs	Swedish Association of Local Authorities and Regions	20 county councils	8 regional hospitals 65 county/district hospitals 1,000 health centres
National Board of Health and Welfare		290 municipalities	Housing, care and social support services for the elderly and disabled
Responsibilities: <i>legislation</i> <i>supervision</i> <i>evaluation</i>		Responsibilities: <i>finance</i> <i>organization</i> <i>follow-up</i>	

**Figure 4. The organisation of Swedish health services (Swedish Institute 2007)**

The municipalities, on the other hand, are responsible for the care of the elderly and people with disabilities living in special accommodation. The municipalities were given this responsibility in 1992 and 1996 respectively for these two groups. The aim was to better integrate health and medical care with the social services that the municipality already provided. Although this reform did improve part of the service and healthcare processes, it

also causes problems with regard to the discharge process of elderly patients from inpatient clinics back to their nursing homes. These discharge problems create a so-called bed-blocker. The term refers to a patient, whose treatment at the hospital is completed but who remains there because the municipality does not have the resources to accept the patient into a nursing home and/or the patient is delayed at the hospital as a result of poor communication between the healthcare providers. Bed-blockers use resources that were intended for other patients. The problem primarily affects planned elective operations, bouncing scheduled patients and having devastating consequences on the effective use of hospital resources. Even though municipalities must reimburse the hospitals for patients that extend their stay beyond the expected time, the compensation level is low and does not cover the costs the bed blockers generate. The National Board of Health and Welfare addressed these issues with regulations for better collaboration and routines for the enrolment and discharge of inpatients between the municipalities and hospitals (SoS 2005). Nevertheless, there are still problems. The municipalities have, for instance, accused the hospitals of discharging patients too early, when they are still in need of medical attention, and complained about lack of funds to maintain the right level of nursing home beds (Anell 2005).

### **2.2.2 Financing and fees**

The cost of Swedish health and medical care, expressed in GDP, is approximately 9 per cent, a figure that has remained fairly stable over the years (SALAR 2005a). However, when expressed per inhabitant and year, and adjusted to buying power (purchasing power parity, or PPP), the cost was 2517 USD in 2002, compared to 1650 USD in 1994, showing a considerable cost increase. In a comparison with 17 western countries (including the US and Australia), seven had higher costs than Sweden and nine had lower costs (SALAR 2005a).

Seventy one per cent of healthcare is funded through local taxation. The county councils and municipalities have the right to levy taxes. The county council tax is an average 11 per cent of the income. Contributions from the state are another source of funding and represent 16 percent, while patient fees (out-of-pocket) only account for 3 per cent. The remaining 10 per cent comes from other national subsidies (Swedish Institute 2007). The contribution from the state differs depending on the need of the county councils and

represents a levelling arrangement where rich councils give to the poor councils (Anell 2005). Patient fees can also vary between councils; they range from 100 – 150 Swedish Krona (SEK) in primary care and are somewhat higher for attending a hospital unit without a referral.

### **2.2.3 Market reforms and challenges**

The positive judgement presented in Figure 3 does not suggest that healthcare management or the general public is content and satisfied with the system's performance. There are still many issues to improve, especially with regard to accessibility of care services, and cost level. Even more urgent is planning for the challenges of the future. Currently, Sweden has proportionally the largest elderly population in Europe and the proportion of the population over the age of 80 years is expected to increase in the coming decades. This demographic trend creates two main drawbacks. One is that the economic burden for the working part of the population increases. Secondly, the older patients are big consumers of healthcare, putting considerable stress on scarce resources and creating even higher costs for the healthcare system. Consequently, on the one side is the demographic trend affecting the financial base of the system negatively and on the other side it is increasing its cost. The consequences of this have recently been presented by Klevmarken and Lindgren (2008), who estimate that the healthcare cost for Sweden will increase by 270 per cent before the year 2040. This estimation is partly based on a prognosis that the number of inpatient days (bed-days) will increase with 30 per cent. This future cost increase will doubtless put pressure on the whole healthcare system, but will specially increase the demands on hospital beds and nursing staff.

Cost containment and higher efficiency have always been a part of the divergent goals of healthcare systems. The management of this complex system demands a constant balance between cost containment, accessibility of care, medical quality and the best possible outcome for the individual patient. The principles governing how nations choose to balance all these possible matters as well as control and manage the health and medical care system might be quite different. The Swedish healthcare system is “based on the principles that care should be provided on equal terms and according to need, that it should be under democratic control and financed on the basis of solidarity” (Sweden 2007). According to Bergmark (2008), the Swedish or Scandinavian model is based on such

values as, *universalism, solidarity* and *decommodification*. This last statement is reflected in the low level of private healthcare providers, as well as private healthcare insurance among the population (Bergmark 2008; Anell 2005). A consequence of the decommodification of healthcare services and the solidarity principle is that accessibility of care might be experienced as poor, especially for that part of the population that has a high income and is willing to pay for more prompt service. Problems with long waiting lists for treatment are common among publicly funded healthcare systems with limited resources and are partially related to the low level of a parallel private system (Vissers et al. 2001a).

The importance of this ideological question, for the Social Democratic Party and for people in general, was significantly exemplified when Sweden's former Prime Minister, Göran Persson, was diagnosed with hip osteoarthritis, in September 2003, and surgery was recommended. In the true meaning of equality, he chose to go through Sweden's public healthcare system instead of seeking private treatment. As a result, he was put on a waiting list, and finally received a hip replacement operation in June 2004. During this period, his work was greatly affected and several official trips were cancelled as a consequence of the pain he was in. It is not unusual that the waiting time for a hip replacement operation extends beyond 9 months.

A well timed agreement, considering the above mentioned episode, between the Government and the Federation of Swedish County Councils took place in 2004. The agreement consisted of implementing radical measures to improve the accessibility problem. The following year, a maximum, waiting time guarantee programme was launched. The programme's targets are usually expressed in the series of figures 0 – 7 – 90 – 90, which mean that: "(0) The primary care system should offer contact by phone or on site on the same day... (7)... and a doctor's appointment – if required – within no more than seven days. (90) After a decision on a referral has been made, an appointment with a specialist in the field concerned should be offered within no more than 90 days after the date of decision... (90) ...and any treatment decided on should be offered within a further 90 days of this decision, at most" (SALAR 2005b). This type of regulation and incentives was tried in previous decades and only led to temporary improvements (Anell 2005). However, this time the programme is more ambitious, with economical incentives it

contains better management tools and a platform for collaboration among the county councils for sharing resources, defining best practices, and so on. Despite these efforts, newly presented figures show that only two county councils are “close” to fulfilling the guarantee. The new government has therefore promised a bonus of 1 billion SEK (120 million USD) to be shared among the county councils that can, to an 80 per cent level, comply with the maximum, waiting time requirements for appointments with a specialist and obtain treatment on time.

It is fair to say, however, that despite the accessibility problems shown, they do not adequately represent the system’s productivity. Figures show that only three countries (Belgium, Luxembourg and Austria) performed a higher number of hip replacements per 100,000 population during 2002, and no other country had a better quality level in those operations than Swedish healthcare (SALAR 2005a).

Although the right wing government of 2009 does not have the same ideological commitment, it still maintains the status quo of the Swedish healthcare system. And despite a number of market reforms during the last two decades and differences in the rhetoric from the right respectively left party blocks, the output has resulted in cautious steps and minor changes regardless of the government in power (Anell 2005; Bergmark 2008).

#### **2.2.4 Inpatient care**

Inpatient care, as opposed to outpatient care, requires at least one bed-day in a hospital. With the exception of patients enrolled as a result of severe accidents or childbirth, the utilization of inpatient care is determined in a rather complex process. The discussion about the figures 0 – 7 – 90 – 90 described previously, indicates that this process can be quite long. Without a referral from a primary care physician and a subsequent favourable opinion from a specialist, the patient does not get access to these services. The physicians’ role in this process is both as assessor of the patient and regulator of the system. The aim is to satisfy a patient’s real need and at the same time ration the supply of health.

The number of available beds in Swedish hospitals has decreased by 60 percent between 1992 and 2005. “Also hospital stays and the number of bed-days decreased the latter more

than the former, due to a 39 percent reduction in mean length of stay, during the same time period.” (Klevmarken and Lindgren 2008). This development has been part of a continuous process, which, to some extent, exchanged hospital care days for nursing home care days.

During recent years, the observance of an increasing trend and studies show that a 30 per cent increase in bed-days is expected in the next two decades. The prime reason is the ageing population; patients aged 65 and above account for 45 per cent of all hospital stay. Furthermore, inpatient healthcare accounts for approximately 30 per cent of the total healthcare expenditure (Klevmarken and Lindgren 2008).

### **2.3 Understanding the effects of demand variation**

The extensive efficiency problems experienced by the healthcare production system are from a manufacturing management point of view at first intricate and difficult to understand. Why are they so difficult to solve? Wherein lies the difficulty? It seems, at least according to public opinion, that anyone with common sense could do a better job than the healthcare managers and make better use of the hard earned tax payers’ money. Healthcare systems seem to affect people in the same way as national football teams, with everybody having a better idea of how they should play than the “incompetent” managers. Healthcare systems are, however, extremely complex and difficult to manage. The complexity lies in different areas; the organisational structure with two chains of command, the managerial and healthcare professionals and their complex interdependencies, the task in itself (saving people), the many different stakeholders’ views and interests that must find ways of reaching a common solution or compromise and, finally, the unpredictable nature of the events that take place within a changing environment.

The Centre for Healthcare Analysis in Sweden described Swedish healthcare to be:

*“complex ... a business which consists of many varying activities, with different mutual dependencies and interested parties and which often is ruled by events and therefore difficult to define and unpredictable”* (CHSA 2002, pp.54).

This description is probably adequate for healthcare systems in general. Several authors describe the complex and variable nature of healthcare systems and the difficulty of efficiently managing healthcare delivery (McLaughlin 1996; Vissers et al. 2001a; Noon et

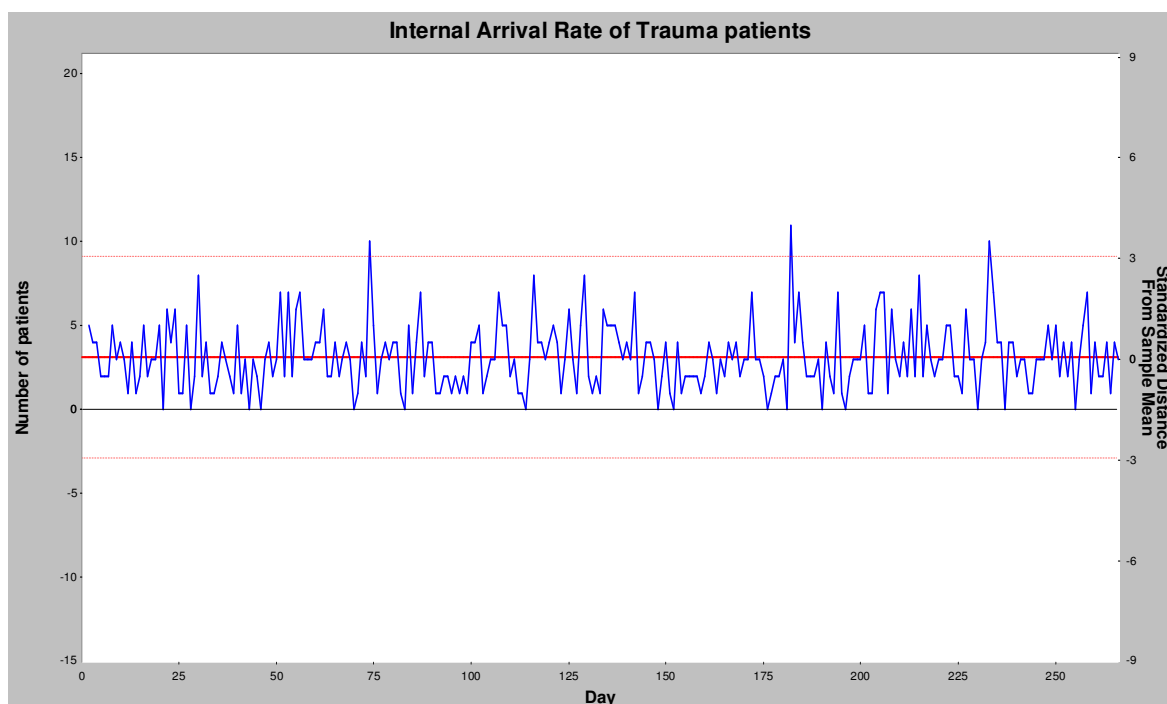
al. 2003; Walley et al. 2006). Jack and Powers (2004) describe some of the reasons for the different emphases in operation strategies between the manufacturing domain and the healthcare domain. The manufacturing operation strategy focuses on cost efficiency, speed and flexibility, but resources and planning are based on a comparatively predictable demand rate (with the exception of the 2008/09 crisis), with the consequences of long delays resulting in spoilage and higher inventory costs, at the most. The healthcare industry, on the other hand, is a “challenge to meet a highly variable rate of demand and a constant rate of high quality service where the consequences of poor service results in patient death” (Jack and Powers 2004). This does not mean that standard production approaches cannot be used or are applicable. Nevertheless, they need to be adapted and take into account the different conditions of the domain (Vissers et al 2001b). Moreover, the complexity of the healthcare system does not justify a settlement with current management practices and certainly does not suggest that there is no room for improvements. On the contrary, there is a significant lack of operational research methods being used for process improvements and control, and this has been so for some considerable time (McLaughlin 1995; Noon et al. 2003; Litvak et al. 2005; Walley et al. 2006; Kopach-Konrad et al. 2007). There are exceptions, obviously, but in general the lack of engineers and engineering knowledge in the healthcare domain results in temporary improvements related to single projects or governmental incentives. However, lessons learned fade away with time and the organisation falls back on its old sins.

### **2.3.1 Consequences of high variability**

One of the most delicate problems when managing healthcare systems is how to handle the system’s variability or demand variation. In contrast to manufacturing systems, healthcare processes are affected by a much higher variation in demand, while in the manufacturing arena, production systems are protected from variations through the use of in-between buffers. Production is either scheduled or planned based on sales prognoses, or, if waiting time is acceptable to the customer, on a production to order basis. The production systems are designed according to the type of products they are intended to produce and each business unit is optimized for its product-market combination (Vissers et al. 2001b).

This is hard to achieve in a healthcare system. The units are seldom independent islands or disconnected through the use of buffers, which means that variations multiply through the system and make it difficult to control. Prognoses might be more accurate at an aggregate

level and show seasonal variations or weekly trends, but it is impossible to forecast the day to day randomness, which is considerable. We can consider, for example, an orthopaedic trauma ward which receives patients from the Emergency Department (Urenda Moris et al. 2004). The average number of patients arriving per day is 3.14, while the data range varies between 0 to 11 patients per day (see Figure 5). Add to this uncertainty the fact that this is not the only variation affecting the ward, the patients' LoS varies significantly between 1 to 51 days with an average of 8.11 days. We must not forget to also take into consideration that the patient's dependency level fluctuates from day to day and dictates, together with the number of patients the number of nurses, the ward needs. Balance into this equation that low levels of utilization are banned by management and that the only buffer your system has consists of taking bed places from elective (scheduled) patients in a nearby ward, resulting in bouncing patients that have waited for at least 9 months for an operation to alleviate their constant pain, and quite an unpleasant reality is revealed. It might seem that the system's bottleneck consists of insufficient bed places, but unfortunately the situation is even more complex. However, what is obvious is that variations put a lot of stress into a system.



**Figure 5. Internal arrival rate of patients into an orthopaedic trauma ward**

These kinds of relationships and demand variation rates are common in healthcare systems and one of the reasons for queues and long waiting lists when system utilization is put into the balance. Noon et al (2003) exemplify the impact of random arrivals and random service

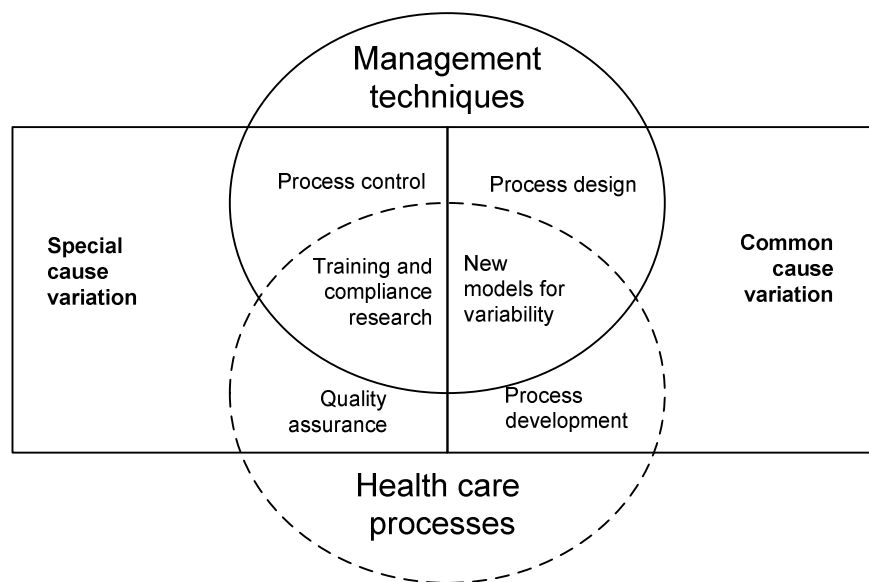


times at a simple walk-in clinic. The basic scenario is the following: on average, a new patient arrives every 6 minutes and it takes an average of 5 minutes to treat the patient. The system might seem in balance, but adding randomness changes the picture completely. Two common distribution families for these types of events are the Poisson distribution to describe the randomness of the arrival time and Exponential for the service time. When their random behaviour is added to the system it leads to a 30 minute average waiting time for the patients, where 10 per cent of them would spend more than 65 minutes waiting for a simple 5 minute treatment. In most of the cases, waiting is a common experience for us as customers, even though nobody appreciates it. However, with regard to healthcare services, waiting or lack of personnel is sometimes deadly (Litvak et al. 2005). The simple but expensive solution to these problems is to over dimension capacity, similar to the way fire department services are planned. However, considering the economical reality of most countries and citizens, this is not an option nowadays, especially not for public funded systems. What can then be done to deal with system variations? How can we achieve both the sufficient utilization of our resources and moderate waiting times? How can healthcare systems be efficiently designed for demand variation?

There is no simple answer to these questions. Different approaches can be useful depending on the planning level of production control (Vissers et al. 2001b). In addition, different strategies are more adequate than others depending on the nature of the healthcare service provided (Jack and Powers 2004). There is, after all, a big difference between the level of healthcare emergency in primary care services and emergency care services, or in antenatal clinic services and maternity ward services. A number of authors address these questions and foremost accentuate the need for healthcare professionals to be aware of the effects of variability and the importance of taking this into consideration in the design and control of the healthcare system, whether we look at ward, department, or hospital levels (McLaughlin 1996; Green and Nguyen 2001; Noon et al. 2003; Litvak et al. 2005; Walley et al. 2006).

Of particular interest is the analysis that McLaughlin (1996) presents with the help of Figure 6 below. He identifies several areas of research and defines as a first step “to see variability as a fact to be analyzed and managed rather than something to be eliminated entirely”. This statement is based on the verity that there are inherent (common cause)

variations in a healthcare system that cannot be eradicated. Moreover, he identifies a second cause of variation (special cause variation) as being unnecessary because it can be omitted with better planning and/or scheduling, or by implementing standardized working procedures. The same distinction between causes of variation is supported by Noon et al. (2003) and Litvak et al (2005), although their name labels are somewhat different. It must be remembered that special cause variation is artificially created by inadequate process control, such as poor patient scheduling or discharge policies.



**Figure 6: Research areas in managing variation (McLaughlin 1996)**

The right side of Figure 6 indicates that management needs to focus on process design in order to handle inherent variation and that healthcare professionals need to focus on developing better processes based on available science. According to McLaughlin (1996), “both, however, need to rely on better guidance than we currently offer on how to understand and model variation”. This guidance would be provided by lessons learned, tools and techniques (“new models for variability”) that help managers model and understand variability, with the goal of designing robust systems. McLaughlin’s call for guidance and research into new models for variability is well in line with the aspirations of this work, which are to increase awareness, provide lessons learned and, foremost, to developed a new methodology for modelling patient dependency variability.

### **2.3.2 Robust system design**

The word *robust*, in engineering design terms, means insensitive or immune to variability from a performance point of view. The driving philosophy behind robust design was

introduced by Dr. Genichi Taguchi (Taguchi 1986), and although it was initially concerned with robust product design or redesign, extensions of this approach make it possible to examine robust design of complex processes and systems (Wild and Pignatiello 1991; Sanchez et al. 1996). The principle behind Taguchi's robust design is simple; instead of trying to eliminate or reduce the causes of product performance variability, adjust the design of the product so that it is insensitive to the effects of uncontrolled (noise) variations. Uncontrolled or noise variations are equivalent to the inherent (or common cause) variations that McLaughlin (1996) refers to.

Benjamin et al. (1995) define system robustness as follows: *“a system is considered robust if its performance is insensitive to the variability of the system's operating conditions”*. One general statement is that systems should not be evaluated on the basis of mean performance; a good system must be relatively insensitive to uncontrolled sources of variation. A healthcare system designer should consequently identify the system's configuration that has the most “optimum” performance under the high demand variation it operates in. The general approach is therefore 1) to work with reducing the variance that is preventable and 2) design the system in a robust manner so that it can handle the inherent variations.

Several papers present lessons learned and considerations on healthcare system design (McLaughlin 1996; Green and Nguyen 2001; Noon et al. 2003; Litvak et al. 2005; Walley et al. 2006). However, the approach to reach a robust system differs between the writers. McLaughlin (1996), Green and Nguyen (2001) and Noon et al. (2003) state the importance of using simulation or other analytical tools to be able to define the most adequate system configuration. Walley et al. (2006), on the other hand, use a vast number of “improvement programs” to gain insight. Furthermore, both Walley et al (2006) and Litvak et al. (2005) worked with the real system when they experimented on new approaches for improving healthcare delivery. The value of working with the “real system” is obviously important, but these improvement programs were made on systems that were already inadequately designed. Even if this is (sometimes) a feasible way of testing new approaches, it is not the optimal solution for a number of reasons. Firstly, even with a well based heuristic solution, it is a trial and error approach that puts a lot of stress on the organisation. The numbers of possible solutions evaluated are limited for obvious reasons (practical, economical, and

quality). Finally, it is difficult to fine tune the system parameters to obtain a robust and more optimal setup.

Discrete Event Simulation (for a more detailed description see section 2.4), on the other hand, does not have these limitations (Benjamin et al. 1995; Sanchez et al. 1996; Gaury and Kleijnen 1998). The analyst has the possibility to literally control all inputs to the model and at the “same” time simulate the stochastic and random demand variation that affects the real or proposed system. Even though they all describe the use of DES as an experimental platform for different system configurations, their working methodologies and analysis techniques are different from one another (Sanchez et al. 1996; Gaury and Kleijnen 1998). Greatly simplified, the differences consist in 1) how the search space of possible solutions, which include both experimentation with controllable variables, so-called design factors, and uncontrollable (noise factors), is covered, and 2) how robustness is measured. The Taguchi approach suggests Design of Experiments (DoE) and the quadratic loss functions or Signal to Noise (S/N) ratio to evaluate robustness. Others have more sophisticated solutions (Sanchez et al. 1996; Al-Aomar 2002; Kleijnen and Gaury 2003) using combinations of many complementary techniques, such as Genetic Algorithm (GA), Neural Networks (NNs), Response Surface Methodology (RSM), Risk analysis (combining Monte Carlo simulation with Latin Hypercube Sampling), Bootstrapping, and so on. Occasionally, personal preferences dictate which technical solution is to be used for the task at hand. But more often, it is the complexity of the project that forces the simulation analyst to choose more sophisticated approaches. The complexity depends, inter alia, on the number of possible solutions (search space), the simulation model execution time (computer power) and the number of optimization objectives (multi or single objective).

This work has no aim in presenting a new robust design methodology. It does, on the other hand, make use of the available approaches, choosing among the different methodologies according to the necessity of the case study. Considering its aim of using DES to *design, plan and evaluate inpatient healthcare systems and their nurse staffing requirements*, it supports robust design by contributing with a modelling methodology for inpatient dependency variability. Without a modelling possibility there is no sense in addressing staff levels as a design factor for a robust system. Moreover, with regard to demand

variation, Green and Nguyen (2001) do present a simple but crucial truth; small systems are more vulnerable to demand variation. Inherent variations such as inter arrival rates and LoS are hard to deal with for a small ward manager and, as mentioned earlier, the healthcare system has many divergent goals, for example, utilization vs. service level. The trade offs and system configuration decisions are many and understanding the effects of demand variation is certainly just a first step.

## **2.4 Nurse staffing**

In 2006, The World Health Organization (WHO) presented the World Health Report on the theme “Working together for health” (WHO 2006). The report addresses the increasing and, in some nations, urgent need for healthcare workers. In November 2005, Lee Jong-wook, the 6<sup>th</sup> Director General of WHO, stated:

*“We have to work together to ensure access to a motivated skilled, and supported health worker by every person in every village everywhere.”* (WHO 2006, pp. 3)

The report highlights the significant challenges the world is facing in recruiting, maintaining and financing healthcare workforces. It specially puts stress on the difficult situation of 57 countries that are experiencing a critical shortage of health service providers, which includes physicians, nurses and midwives. This current crisis is expected to deepen in the coming years when the demand for service providers will escalate in all countries. “Richer countries face a future of low fertility and large populations of elderly people, which will cause a shift towards chronic and degenerative diseases with high care demands.” (WHO 2006, pp. 7). Among the strategies for addressing this issue is “enhancing worker performance” through improvements in the availability, competence, responsiveness and productivity of the workforce. Considering the labour-intensiveness of the healthcare sector and the close relation between health workers’ density and the probability of patient survival (WHO 2006), it is not surprising that nurse staffing is an area in which considerable effort and research have been made (Hurst 2002).

Enhancing workforce productivity by cutting waste and improving performance is not a trivial task and not an indiscriminate aim either. Firstly, it is extremely difficult to have a suitable staff level that constantly matches the fluctuating patients’ needs. The workload

generated by patients comes not only from the number of patients currently admitted, but also from the dependency level they are in. Among the many different tradeoffs managers need to take into consideration when planning for the right staffing level are healthcare quality, costs, and the accessibility of human resources. There is also an important relationship between work overload and the increase of sickness and absenteeism among nurses, as well as between work overload and increased mortality risk among patients, which implies that the aim is not only to achieve high figures with regard to personnel utilization (Litvak 2005; Rauhala et al. 2007). Therefore, healthcare managers are quick to adapt the personnel and personnel mix to the unit's variable workload requirement, taking into consideration the personnel's well being and work satisfaction (Shuldham 2004). However, the goal is simple, "to get the right worker with the right skills in the right place doing the right things!" (WHO 2006).

Nurse planning is considered to be part of a wider field which implies requirement planning at different levels and for different time perspectives (Burke et al. 2004). There is no consensus in how these activities should be divided and named (Burke et al. 2004; Punna 2006), but *nurse staffing* is generally used for a long term strategic management decision, while *nurse scheduling or rostering* describes a shorter term (several weeks) planning of the workforce requirement. Warner (2006) uses the term *shift staffing* to describe the planning and scheduling activities that take place just a day or hours before the shift. This last minute planning takes into consideration the current status in the ward, absences among nurses, higher patient demands, and so on. It simply deals with how to adjust supply to meet demand. It might require calling in extra nurses, moving a nurse from one unit to another or arranging for a nurse not to come in.

The current work focuses on how staffing requirements at a strategic level for robust strategies can be modelled for DES inpatient ward models. It discusses the benefits of taking into consideration the system's variability, concentrating on patient dependency variability when modelling and simulating inpatient systems. Furthermore, it presents a modelling methodology that uses available patient dependency data and combines this data with simulation in order to give the user the possibility to design more robust healthcare systems.

## 2.4.1 Nurse demand methods

Nurse demand methods have been defined as “any system of determining the number and/or mix of nursing staff” (Arthur and James 1994). There is an enormous range of options in determining nurse staffing requirements. If aggregated and listed, a taxonomy identifying five groups of methods can be distinguished as commonly used for estimating size and mix of nursing teams (Arthur and James (1994) present a different taxonomy, for more or less the same nurse demand methods). An extensive review of them is presented by Hurst (2002). These methods, by themselves or in combination, are used to calculate adequate staffing levels.

They are:

- Professional judgement approach
- Nurses per occupied bed (NPOB) method
- Acuity-quality method
- Timed-task/activity method
- Regression analysis

The methods are listed from the simplest to the most complex. The review covers more than 500 articles, books and reports on the subject of demand-size planning of the nursing workforce, and related issues are listed and briefly commented on. A brief discussion of their pros and cons can serve as a background to identify the need for a more sophisticated solution.

The *professional judgement approach* uses the judgement of experienced, nurse managers in determining the right staff and mix levels per day and shift. The method is based on intuitiveness and subjectivity, but can, to some degree, be improved if a team of professionals consults with each other before setting the levels. Arthur and James (1994) consider that the Telford Method (defined by Hurst (2002) as a type of professional judgement approach) is less intuitive and has a more elaborate methodology, but it is still subjective in its nature and not always trusted by management. In addition, Arthur and James (1994) claim that professional judgement approaches are part of a so-called Consensus approach group and state that “Consensus approaches are considered to be overtly subjective, and attempt to take a critical and reflective view on nursing workload.”

Among their strengths is that they are easy to use and encourage a critical view of staffing and practice.

The *NPOB* method is, on the other hand, considered to be a top-down norm method that specifies the required number of staff per certain number of patients. Professional bodies set these staffing norms as recommendations and may suggest a nurse:bed ratio to determine shift or establishment staffing levels. For example, Litvak et al. (2005) point out that the estimated ratio for a surgical unit in the U.S. is 1:4, while (Hurst 2002) presents a matrix which, depending on the type of ward, suggests specific ratios. The *NPOB* method is presumed to be used as a guide or to initially set minimum requirement staffing levels, but since it does not take local variations into consideration it needs to be complemented with other nurse demand methods (Arthur and James 1994).

An important weakness of both the *Professional judgement method* and the *NPOB* method is that they do not take the system's workload variation into consideration. They are based on average calculations and, if they consider patient dependency, they view it as a static value that does not change over time. These types of demand methods determine staff levels according to an average, and therefore lead to problems related to both over and under staffing.

The third method is better described as *Dependency-activity-quality method*. As its full name suggests, the method is based on several steps and activities. The first step consists of determining the dependency levels of the patients and putting them into categories, normally ranked 1 to 4, where four is the category for patients with the most dependency. There are different approaches used to assess the dependency level of the patients, of which the most common are 1) the prototype method and 2) the factor method. These methods are described in more detail in chapter three: Patient Classification Systems. The second step consists of determining the average amount of direct (and indirect) care time for each patient category. This is normally done by an activity study at the implementing ward. These two steps give answers to the total minutes of nursing care per patient stay, per diagnosis or Diagnosis Related Groups (DRG) and, consequently, staffing requirements per day and average per month. The method is classified by Arthur and James (1994) as a bottom-up management approach. The bottom-up nature of this



approach has the appeal of focusing on patients' nursing needs and thereby influences nurses into reflecting on their practice in a more critical way.

Hurst (2002) also presents data for calculating staffing requirements based on quality assessed medical wards in the UK. This data can serve as acceptable estimates for staffing levels at wards for different patient care groups, without the wards having to make their own activity study. Hurst's study presents information on the recommended staff mix of the nurse category and the total amount of staff required. These estimations are, nevertheless, based on UK best practice and may not necessarily be accurate estimations for Swedish healthcare. One reason for this is presented by Levenstam and Bergbom Engberg (1993). They point out that an organisational change in the nursing care methodology leads to increased staff demand in a ward. In their case, the ward changed from a more traditional method of nursing care to "a kind of team nursing". These kinds of organisational differences are common even within the same country, and more so between different healthcare systems.

Several commercial software products are based on this method, but there are differences in how the methodology is implemented and also on the additional features of these products (Levenstam and Bergbom Engberg 1997). Many of these products are called "Patient Classification System" (PCS) or patient acuity system and are being used in numerous hospitals (Adams-Wendling 2003; Galan Perroca and Ek 2007; Rauhala et al. 2007). This system is of special interest for two main reasons. Firstly, the software solution contains a database containing all the relevant historical data needed to make a stochastic and dynamic model of the ward of interest. Data in combination with process and logistic information gives the opportunity to design robust systems of whole units, simulation models that can answer fundamental questions about design and of an operational nature. Secondly, these are software systems that are already in use, the data is available and therefore easy to retrieve and use without any additional cost to the proposed solution. The identified features of PCS, as well as its accepted use and diffusion among healthcare providers, give credence to regarding PCS as an interesting tool in achieving the aim of this work.

*A thorough description of a factor based PCS and its functionality is presented in chapter three: Patient Classification Systems.*

The fourth method mentioned in the review is called *Timed-task/Activity Method*, and concerns using a time-task/activity study to measure the nursing care needed by the patients. Each patient's direct care needs for the coming day are recorded using a locally developed check list of nursing interventions, which might have several hundred different interventions to select from. Furthermore, the time required for each intervention has been measured. The patient intervention list, comprising a selected number of interventions, is therefore the basis for a sum of timed activities representing the total amount of direct care time that the patient will need the following day. The sum of all patients' lists becomes the basis for calculating the total need of work staff. This method does not classify patients into dependency groups, but instead gives each patient a unique list of interventions and, therefore, a unique nursing time requirement. The most recognized commercial system utilizing this methodology is known as GRASP®.

One of the main disadvantages of this method is that it is time consuming and therefore adds considerable "overhead" to the ward. From a simulation analyst perspective, the detailed data stored in the software product could be both a benefit and a curse. The data is a curse if it cannot be grouped in a meaningful way or if the time measurement does not represent how the actual work is performed. For example, the experienced nurse will most certainly perform several tasks simultaneously, while the timed-task method will measure the time of each task separately. This could negatively bias the analysis of the ward's situation. Moreover, it is not clear whether the software solution stores historical data of the direct care time of patients' dependencies. This data is crucial in order to model the stochastic dependency variability of patient groups.

The fifth method, *Regression analysis*, is broadly described as a method to predict the required number of nurses for a given level of activity. The aim of the analysis is to find a statistical relationship between predictors (independent variables) and outcomes (dependent variables). Once this relationship has been found, the user is able to estimate the number of nurses needed based on the predictors. Among the weaknesses of the method is that it is unsafe to make predictions outside the regression model's observed

range. There are other weaknesses, such as that the statistical relationship between predictors and outcomes is not easy to set and that the relationship will show an average for a certain predictor combination and not the variance. The consequence is that the model, with its relationship, is useful only if the main setup is maintained; bigger ward design changes and patient mix changes may render the model useless. The ability to forecast is limited to “what if ” analysis, answering strategic matters and not forecasting near future developments. It does not monitor variability in workload, patient admission, and so on, and therefore does not simulate the dynamic and stochastic behaviour of the ward, unit, et cetera.

Arthur and James (1994) state that there is no perfect system for workload measurement and that most of the methods in use include certain levels of subjectivity. This is certainly true, and it would be presumptuous to claim that the optimal method is presented in this work. However, there are some important disadvantages with the current solutions that might be overcome. All the above mentioned methods are based on mathematical calculations and they give the user answers in form, figures, or relations between variables. Many of them are very simple and easy to use, and have a significant value for managers in their planning activities both for nurse staffing and as information for scheduling activities. Complex systems, such as healthcare units, are difficult to control, manage and fully understand, and while simple tools and heuristics are valuable, in the long term they cannot cope with this complex task. These methods undoubtedly fill a need and purpose, but they lack the possibility of providing a more thorough understanding of the whole system’s current and future behaviour. Some of them do not even allow you to try different ideas and see the outcomes. In addition, they do not help you to see complex relations between patient mix, shifts, wards, and so on. Neither do they give you a stochastically and dynamic representation of the system. Simulation, on the other hand, is more versatile and gives you the possibility to address these issues.

## ***2.5 Discrete Event Simulation***

Simulation has been described as “the activity of producing conditions which are similar to real ones, especially in order to test something...” (Law 2000). These conditions could obviously be produced in a real environment using physical models, but from our point of view we are going to limit the use of simulation to mathematical computer models with or

without animation. The term simulation is very broad and it is therefore useful to define the type of simulation models being used. For instance, they can be classified along three different dimensions (Law 2000).

- Static vs. Dynamic simulation models. If time does not play any role, that is, the system variables do not change over time, then the model is static. A dynamic model on the other hand evolves over time.
- Deterministic vs. Stochastic simulation models. If the simulation model does not contain any randomness, it is said to be deterministic. This means that you will always get the same answer, given that you use the same variable setups. A simulation model that is stochastic, on the other hand, will give you different (random) outputs every time, despite the same setup. This means that you get an estimated answer, but a realistic estimation.
- Continuous vs. Discrete simulation models. This dimension is more concerned with how variables are updated and time is controlled during the execution of the model.

Different simulation techniques and tools address different combinations of the above dimensions. The simulation technique that is the focus of this work, named “Discrete Event Simulation” (DES), is classified as a dynamic, stochastic and discrete simulation technique. DES is superior to other modelling techniques, with regard to modelling complex systems with many stochastic variables and dynamic behaviour.

There are other ways of modelling and analysing stochastic systems besides DES. One widely used technique is Markov processes. A description of how Markov processes are used for modelling the patient’s transition between different dependency states is presented in chapter six. In order to compare Markov process modelling and DES, it is of value to read Le Lay et al. (2006). The authors state that the great benefits of DES are “that it allows the analyst to model more complex and dynamic systems compared to other types of modelling and that it permits experiments that might not otherwise be feasible”. They also maintain that the modelling flexibility “enables the model to capture more details about the uncertainty in the system being modelled”. Similar statements about the modelling flexibility of DES compared to Markov process models are presented by

Simpson et al. (2009) and Karnon (2003), although the latter considers DES to be a more complicated decision modelling technique. Nevertheless, part of the solution presented contains an absorbing Markov process with a non-homogenous transition matrix. Furthermore, and in support of the above discussion, Lowery (1996) raises some of the differences between analytical models and DES. She claims that analytical models are a better choice if the system being modelled is simple, has low variability, and few (single) performance objectives being studied.

Consequently, all simulation or analysis techniques have their specific pros and cons. However, what happens if your system is too complex or you have a system that displays high variability? What if you want to understand the nature of the trade-offs between competing objectives? Considering the nature and complexity of healthcare systems, the most flexible and appropriate choice for the purposes of this work is Discrete Event Simulation.

### **2.5.1 DES - modelling problems for Inpatient clinics**

DES has been used for healthcare system modelling and analysis for more than two decades (Jun et al.1999; Jacobson et al. 2006). Nevertheless, many of the models and simulation studies have not taken staffing requirements into consideration. The focus has instead been on bed/room sizing and planning, patient scheduling and admission, patient routing, appointment scheduling, availability of resources, et cetera. When staff sizing and staff scheduling have been part of the simulation project's objectives, it appears to have been limited to, for example, care centres, walk-in clinics, pharmacies, and Emergency Departments (EDs) (McGuire 1998), in other words mainly outpatient services. These studies represent systems with well defined activities which are carried out in a relatively clear process flow. The most complex models which include staffing are those that simulate EDs. Many of them utilise simple PCS in order to categorise incoming patients, as well as test different strategies and resource configurations to handle the patient's acuity illness properly (Kumar 1989).

The author has not found a single inpatient ward study that has addressed the issues of staffing. There are references to an Intensive Care Unit (ICU) study addressing staffing (Masterson et al. 2004), however, in this case, staffing seems to have been modelled only

as a discrete variable that controlled the number of available ICU beds. The study did not address modelling methodology or patient dependency levels vs. staffing levels. Groothuis et al. (2001) present a study based on an imaginary ward. This ‘typical’ ward was designed with data from different studies. Their study illustrates modelling issues related to the evaluation on how nurses organise their common workload, whether it should be done in a functional manner (each nurse category is in charge of specific tasks) or with a team approach. Although the model does not cover all the issues related to nursing activities in an inpatient ward, it does raise interesting modelling and programming comments.

There are several intuitive reasons for the lack of DES studies in this field. Firstly, there is always considerable modelling complexity when human resources are modelled. This is particularly true within the healthcare domain (McGuire 1998), but especially true when it comes to modelling inpatient wards, where the activities are more stochastically distributed, the process is not very well defined and many activities are not exclusive for one personnel category (Groothuis et al. 2001). Secondly, a common problem is the lack of input data. If accurate input data is not available, it might take months or years to gather enough to make reasonable estimations in order to address patient dependency variability or time activities. Finally, it is always difficult to map and study personnel in any category of work, because they are distrustful and feel threatened by the observations. For all these reasons concerning the effort, the time and the additional cost involved, one could easily get the impression that addressing the staffing issue as one of this study’s purposes is not worth the effort. It must also be remembered that the healthcare administrators’ and the nurses’ busy working situation, with all its demanding tasks, can make them feel that they do not have time for a prolonged study.

An often repeated mantra within the simulation community, when it comes to model complexity, is “keeping it as simple as possible” (Lowery 1996, McGuire 1998; Chwif et al. 2000; Sánchez 2006). This advice is obviously important and valid, but it may have caused simulation projects to focus more on defining “easier” modelling objectives, limiting the scope of the project and avoiding the difficult modelling areas. This mindset could have been influenced by the need to start (and continue) with success stories (Barnes et al. 1997), in order to overcome the barriers associated with implementing simulation in healthcare service providers (Lowery 1996). Timothy Ward, on Sanchez et al. (2000),

comments that simulation studies do not focus on the right context and pinpoints staffing as one of the areas where simulation modelling belongs. It must be remembered that the performance of healthcare systems is not only dependent on inanimate resources, but indeed more on human ones. Properly modelled or quantified human resources are central in order to model entire systems as well as their resource dependency and interaction, and in so doing being able to improve system design including, for example, staffing flexibility approaches. Jacobson et al. (2006) indicate one of the future directions of DES research and state that estimations of patient demand, utilisation of staff and overall cost “may not be possible in a microscopic, single level model...” The future aim is “models that capture the interaction of major service departments and support services in a hospital...analysing the system as a whole...can be invaluable for hospital planners and administrators”. An efficient modelling approach of staff requirements will most certainly contribute to achieving this aim.

### **2.5.2 DES – First stumbling steps in the Swedish Healthcare domain**

In October 2002, a first stumbling step was taken in introducing DES as a tool for the improvement and analysis of the Swedish Healthcare’s systems (Urenda Moris et al. 2004). At that time, DES was extensively used by the main manufacturing industries in Sweden, particularly the automotive industry (Jägstam 2004). However, no other record of research in the area of DES in Swedish hospitals was documented or presented at that time. Discussions with an early DES consultant confirmed that some studies had been conducted, but that these singularities had not raised any awareness of the potential benefits of DES. Furthermore, they had not targeted questions about working or modelling methodology and implementation, or any possible adaptation needed for the Swedish healthcare sector. These are areas in which there is a legitimate interest from the academic community in order to add and obtain insight and understanding. The goal at that time was to introduce DES for decision support in the Swedish healthcare system. Looking back to the efforts undertaken in propagating knowledge and awareness, it can be said that the work presented in this thesis certainly contributes to the progress in achieving that goal. Furthermore, it has been achieved by targeting knowledge transfer. The author has actively participated on innumerable occasions at various events promoting the importance of using DES to model healthcare systems for better system design. These lectures/presentations have been conducted for future IT nursing students, simulation networks, healthcare managers, those

responsible for healthcare processes, hospital CEOs, and politicians. The presentations have been made at course lectures, simulation symposiums, project meetings, research proposals, board presentations, and so on.

This effort, together with others, has without doubt increased the awareness and understanding of the need for deeper analysis and simulation in the healthcare area. Today, important contributions have been presented by Persson (2007) and Elf (2003) although the latter did not apply DES. Elf (2003) discusses the use of simulation in order to evaluate and plan the healthcare environment from an architectonic perspective, as well as how the patient and work environment could be improved using simulation tools. Her work addresses more intangible parameters such as patient well-being, professional behaviour, organisational outcomes, and so on. A new network called SimSIC – Simulation for Swedish Innovative Care - has recently (2008) been launched. Its aim is to support the Swedish healthcare system in its task of system and process improvement. The network involves important partners both governmental, academic (universities) and business (companies) under the same umbrella.

## ***2.6 Communication problems and barriers***

Although many of the production control strategies in healthcare systems have their correspondence in the manufacturing domain, which can be observed in the language and concept description of Vissers et al. (2001b), there are important differences in how fluctuating demand is approached and volume flexibility is acquired (Jack and Powers 2004). These differences are not only found in a comparison between healthcare, and manufacturing industry settings, but also in important cultural and organisational settings, which affect the way in which successful simulation projects in the healthcare service sector are run.

Lowery (1996) describes several barriers associated with implementing simulation in healthcare. Some are based on traditions, long held beliefs, and the fear of engineering solutions that are regarded as dehumanizing. According to McGuire (1998), three of these beliefs are:

- Practices designed for manufacturing are not transferable to healthcare.
- Efforts to increase efficiency will shortcut patient care



- The public will interpret efforts to increase efficiency as a reduction in the quality of medical care provided to patients.

It is interesting, though, how new cost control incentives, as well as quality improvement programmes, have changed the attitude towards engineering solutions. McGuire (1998) describes how the acceptance of Total Quality Management (TQM) in the mid 1990s has reduced this resistance. The 1990's economic depression in Sweden forced the providers of healthcare services to improve efficiency and become aware of new solutions (CSHA 2002). Today, one example of this awareness is the work being carried out at Skaraborg Regional Hospital (SkaS). This major healthcare facility works actively with quality and process improvement, and has introduced Six Sigma as a working methodology for managers, head physicians and nurses, in their ongoing process improvement work (Urenda Moris et al. 2007).

One of the barriers presented in Lowery (1996) is the “number and variety of customers with competing priorities for solutions suggested by simulation”. The difficulty of conflicting objectives between hospital managers and medical personnel is also addressed by Sanchez et al. (2000). The Swedish healthcare system is considered to be divided into three domains, the political, the management and the service (Hallin and Siverbo 2003). Each of these domains is, generally speaking, formed by different professions: politicians (owners of the system), health organisation managers (those in charge of implementing the political decisions), and physicians (who generally feel that politicians interfere in the running of the system about which they have little knowledge). Physicians do not want to renounce their autonomy and hospital management struggles to implement unpopular decisions under the pressure of the decision makers in the other two domains. This makes the Swedish healthcare organisation complex and its administration resemble an ongoing struggle for power and influence. Unfortunately, there are also several conflicts of interest between different departments and units, as well as between professions (physicians, nurses etc.), which make healthcare organisations extremely complex environments. According to Eldabi et al. (2002), this reality, of multiple decision makers, should cause healthcare simulation projects to be run differently to the way they are normally run in a manufacturing environment, for example. They state that because many problems in the healthcare area are not well defined, a result of interactions that are poorly understood and

not easy to capture as a consequence of the different interests and background of the stakeholders, the simulation process should therefore not follow the typical waterfall structure more common in other domains. Eldabi et al. (2002) propose a new approach called MAPIU which will help to enhance stakeholders' understanding and communication. Lowery (1996) points out the need of understanding and communication with the different stakeholders as a prerequisite for promoting simulation in the healthcare domain. However, Eldabi et al. (2002) take this aspect even further, giving the stakeholders much more authority over the modelling development. This is an important difference compared to the way simulation engineers traditionally work, but something that experience has shown to be a decisive factor between success or not.

A discussion with Six Sigma project leaders at Skaraborg Hospital revealed that consensus and well defined goals are matters achieved during the early phases of the Six Sigma methodology, that is, the Define, Measure, and Analyse phases. This means that when simulation is used to confirm whether the proposed solutions have the stipulated effect or not, the project can proceed in a similar manner as in an industrial project. There was a fear among some of the members that if too many stakeholders were included in the simulation modelling work it might lead to irrational requests from the stakeholders, which could slow the project down with detailed, out of proportion discussions. These project leaders have more experience in leading improvement projects without the use of simulation. Their experience might be different if they use DES as a discussion platform in order to avoid endless discussions and reach a consensus. Nevertheless, these concerns both confirm the MAIPU2 approach and point out that the most important issues are to clarify the project goals and achieve consensus among the different stakeholders. The process of how this is achieved seems to be of secondary importance.

## **2.7 Summary**

Chapter two discusses a number of subjects that form the theoretical background and scenario of this work. The main objectives of this chapter are to describe the healthcare domain and its peculiarities from an engineering perspective, and give the reader the basis with which to understand the work and approach.

It describes the Swedish healthcare system as a national health organisation, which consists of a complex and decentralized structure involving a central government, local governments and municipalities. Despite the decentralised structure, the number of private healthcare providers is very low, accounting for just a few per cent in some county councils (local governments). When the system is viewed as a whole, it performs well in an international context, but there are clear problems with access to care services that are not of an acute nature. Problems with system demand variation and process control lead to long waiting times for orthopaedic surgery, for example. Moreover, as a result of an ageing population, healthcare costs are expected to increase by some 270 per cent over the coming decades, and inpatient bed-days are expected to increase with 30 per cent.

The demand variations on healthcare services are extremely high in many system's components. This aspect, together with the complex relations and dependencies among units, illustrate a clear difference between manufacturing systems and healthcare systems. It is more difficult to "protect" healthcare systems from demand variations and, considering the implications, they cannot afford long waiting times for emergency patients, for example. The most rational approach is to reduce the preventable variations and design the healthcare systems in a robust manner to be able to handle the inherent variations. Among the variations that the system faces is the patients' fluctuating dependency levels which, together with the random arrival of new patients, make maintaining the right staff level difficult. A staff level with not enough nurses puts a lot of stress on the system, which leads to both work dissatisfaction among nurses and jeopardises the patients' health. Today there are many different approaches to nurse staffing. Most of them are used for average staffing, which simply means that the staff level is calculated on the average need of the ward. Average staffing is used as a trade-off that leads to both over and under staffing depending on the ward situation, neither of which is of benefit. Over staffing leads to economic waste and under staffing leads to the stress related consequences already mentioned. One of the approaches used for staffing is called Patient Classification Systems (PCSs). These have a number of interesting features and are based on a bottom-up approach that daily measures the patients' dependency levels, which are stored in a database. This classification system monitors the patients' fluctuating dependency needs and converts them into work hours for staff categories.

Among the different simulation and analytical tools used for system analysis, design and redesign, DES is the most adequate tool for the robust design of systems with high variability in the healthcare domain. The main reasons for this conclusion are 1) the stochastic nature of DES, 2) the modelling flexibility and versatility it offers, and 3) its strong visual attributes which make it an excellent choice for solving communication problems and creating consensus among healthcare stakeholders. Furthermore, it is the opinion of the researcher that a combination of PCS and DES provides important benefits for the purpose of designing healthcare systems that are robust to inherent demand variation. And in contrast to earlier approaches, this combination does provide the possibility to evaluate more robust staffing approaches and address healthcare systems in a more holistic way. This last aim is of major importance considering the complex interactions between the healthcare system's components.

## **3 Chapter three: Patient Classification Systems**

### ***3.1 Introduction***

A common feeling when a large number of people enter a lift/elevator is to check its maximum load capacity. This will normally be given in two different ways; the maximum number of persons allowed in the lift and the maximum weight in kilograms this number of people might weigh in total. Intuitively, people check the number of persons first and then their “size” to estimate whether they should squeeze in an extra person or not. However, this intuitive thinking often seems to be lacking when calculating personnel requirements in a ward. People only reflect on the number of beds available when determining whether to admit a new patient and then adapt, in the best of cases, the number of human resources according to that specific number of patients. However, the number of personnel needed to provide adequate nursing also depends on the “size” or weight of the patients, or more specifically, on the amount of nursing assistance they need. This is one of the primary aims of Patient Classification Systems (PCSs). They are used, among other things, to determine the nursing weight of all the patients individually and, by doing so, measure the unit’s total need of human resources.

This section does not focus on how to evaluate, implement or design PCSs. Instead, it focuses on describing the basic building blocks of a PCS, provides insight into the measuring methodologies, illustrates some of the differences between systems and gives a background to PCSs used in Sweden. What is more important is that it provides the reader with sound information about how the data of a PCS is valuable from a DES point of view.

### ***3.2 PCS background and development***

The classification of the nursing care requirements of patients can be traced to the days of Florence Nightingale in the middle of the 19<sup>th</sup> century. Based on the professional judgement of the ward sister, the most seriously ill patients were placed closest to her office to facilitate observation, whereas patients who could fend for themselves tended to be placed at the far end of the open wards (Edwardson and Giovannetti 1994). Its modern application started in the USA in the 1950s, and in Sweden the first attempt was in the late 1960s at the Vasa hospital in Gothenburg, (The Federation of the Swedish County

Councils 2000). Today, 23 of a total of 93 Swedish hospitals use PCSs in at least one unit (Perroca and Ek 2007). Patient classification systems have gone through many refinements and improvements since those first attempts. They have become more reliable and versatile, and their use more automated (Malloch and Conovaloff 1999; Soliman 1998; Perroca and Ek 2007). In addition, both their scope in terms of how to use the information stored in them and their actual utilisation in different units have increased during the last few decades (Botter 2000; Perroca and Ek 2007).

Two common types of PCSs, differentiated by the method of evaluation they use, are the “prototype” and the “factor” evaluation types (Giovannetti 1979). Prototype evaluation is characterised by a relatively general description of different patient types, so called “prototypes”. Patients in a unit are then compared to these prototypes and the staff requirement is calculated according to the standard times for nursing care per prototype. Factor evaluation, on the other hand, is based on the judgement of several critical indicators or descriptors of direct care requirements, for example, feeding, bathing and ambulation. A patient is evaluated according to the number of “points” received from these different indicators. The total sum from each indicator determines the patient’s dependency level. Every indicator rate is translated to a nursing care time, and this average time has been determined through an activity study.

Acceptance of and improvements to PCSs have not been established without criticism and considerable research in the area. De Groot (1989a, b) presented two papers that both identified essential system components and provided valuable information about system selection and implementation. The essential system components comprise the following (De Groot 1989a):

- A tool to predict nursing care requirements for individual patients.
- A sound method of validating the amount of care given to each category or type of patient on each unit and shift.
- A sound method of evaluating the patterns of care delivery of each unit, shift and staff level.
- A mechanism to revalidate the amount of care by patient category and patterns of care delivery on a periodic basis.

- A method of relating nursing care requirements to the allocation of staff resources on a shift-by-shift and unit-by-unit basis.
- A method of monitoring the reliability of the patient classification system over time.

These system components form together with De Groot's suggested PCS selection criteria and the keys to successful PCS implementation (De Groot 1989b), an evaluation framework to make a knowledgeable evaluation, implementation and maintenance of a PCS.

### ***3.3 Identified critics and further development***

There are some critics who oppose the use of PCSs and question their true value in either helping to measure staffing requirements or patient needs. These critics not only focus on the methodology itself, but also on the difficulty of maintaining system reliability (accuracy of the measurements) and validity (that the method measures what should be measured over time) (Edwardson and Giovannetti 1994). These two measurements are key selection criteria and fully operational PCS are expected to have a methodology that gives the user evidence of both reliability and validity. Consequently, a lot of effort has been put into confirming these features (Adams-Wendling 2003; Bergqvist and Edberg 2005). If validity or reliability is an issue, it seems rather to be related to the implementation and operational management of the tool.

In response to the critics, several researchers defend the use of PCSs and strongly support their value and benefit. Giovannetti (1979) states that even though PCS are imperfect in "determining the true need of patients, they do, when used appropriately, provide a rational approach to the problems of nurse staffing" and he continues stating that "a well developed patient classification system is better than no system at all". Van Slyck (2000) argues about the need for a holistic view of the use of PCSs and does not mention its staffing features too much, stating only that "the PCS is not about the staff. It is about the patient". Botter (2000) confirms Van Slick's view on the multiple uses of PCS and points to several areas where the information stored in a PCS is used and valuable. It is interesting to read two of her statements; "Findings from this research suggest that uses for PCS information will continue to evolve...use the information creatively in combination with other available information". Finally she points out that if PCS information is to be used as a component

of decision making regarding staffing the users “might consider having patients classified more frequently ...at least three times a day.”

The reason for this last statement is that the evaluation of the patients' dependency level is normally done in retrospect, establishing what the needs of the patient were during the last 24 hours. Obviously, this kind of evaluation does not reveal what the patient's needs will be in the coming 24 hours. There are systems where the users estimate the patients' future development, but these estimations are not done in a reliable way. Instead, they are just hunches and not statistically based. By making more frequent evaluations, the time horizon goes from 24 to 8 hours, which obviously increases the reliability of the estimation. However, this is normally not done for the basic reason that it is unpractical. It takes time and effort and, even if it is done, what are the options/consequences with regard to finding additional staff or sending people home at such short notice. This means that PCSs are not used to evaluate staffing needs on a short time horizon or for daily decision making (Botter 2000). Instead, they are utilised to solve strategic staff management issues, for example, they are able to monitor seasonal changes, as well as staffing requirement variations, shift-by-shift or unit-by-unit over a longer period. (Rainio and Ohinmaa 2004)

Finally, Edwardson and Giovannetti (1994) conclude their evaluation of the development of the nursing workload measurement systems and future directions by stating “that staffing predictions would seem less important than that which focuses on the cost and outcomes of care.” The use of PCS for calculating patients' real cost to the healthcare system is an essential question for the healthcare sector. The following section describes a Swedish approach to this aim and how it affects the selection of a suitable PCS tool.

### ***3.4 PCSs in a Swedish case-costing context***

The Federation of Swedish County Councils<sup>1</sup> presented a report (2000) of PCSs used in Sweden and their possible use as an intermediate tool in a case-costing system called Cost per Patient (KPP). Case-costing refers to the cost calculation of an individual patient's stay or visit to a health or medical care facility. The aims of the KPP project are to better identify the real cost of patients for the healthcare system, improve the possibilities of

---

<sup>1</sup> The Federation of Swedish County Councils merged with the organisation for Swedish municipalities in 2005 and formed the Swedish Association of Local Authorities and Regions (SALAR). The merge of the two organisations was judicially settled in 2007.



making comparisons and analyses, and use the KPP findings to improve DRG grouping and weights (Nilsson 2002). Staff related labour costs are the main contributor to the total cost of inpatient care. The previous cost estimation approach was to use a standard cost per diem. In order to achieve a more reliable distribution of nursing costs, PCSs have been examined to determine whether they can become this intermediate product when allocating labour costs to the individual patient (Heurgren 2000). The report (The Federation of Swedish County Councils 2000) stated that the general objective for using a PCS was:

*“To create balance between patients’ nursing care requirements  
and the nursing care unit’s personnel resources.”*

They also identified four areas where patient classification can be of considerable help:

- Monitoring the need of care
- Unit staffing and planning
- Monitoring and calculating nursing costs
- Monitoring and analysing a unit’s performance.

The report confirmed that these were the PCSs main aims, but stated that “PCS for nursing can be used in a case costing system when the classification is based on what has been performed” (Heurgren 2000). The report defined several criteria of PCSs in order to support the case costing system.

The criteria are:

- The intermediate products must be based on real events (not what should have happened or normally happens).
- The intermediate products must offer a way to describe what the patients receive (not what they are expected to receive).
- The PCS must have a method for continuously monitoring and if needed recalculating the amount of resources used for each product.

The report studied the four most common PCSs in Sweden: Zebra, Beakta, Rush and RiL. Two of the systems, Zebra and Beakta, were considered to fulfill the three criteria and the

quality requirements defined by De Grooth (1989a), Levenstam and Bergbom Engberg (1993), and Heurgren (2000). Both systems are so-called factor evaluation systems.

These criteria are important from a simulation point of view. If the criteria are in line with simulation criteria, the recommendations from the Federation of Swedish County Councils (currently SALAR) would support a methodology and system that can also be used for the modelling and simulation of the personnel requirements of inpatient units. If the opposite is true, the results of the report would be an impediment. Before we analyse these criteria and their correspondence to the DES modelling requirements, let us take a closer look at the structure and features of PCS Beakta®, which is one of the PCSs that complies with the above mentioned criteria.

### **3.5 Beakta® – PCS**

Beakta® is one of the largest PCSs in Sweden, and over 200 hospitals wards and units for the elderly have implemented this software based system. The methodology, based on a Canadian study, was developed in Norway and Sweden at the end of the 1980s. It consists of three parts: *patient classification/assessment*, *activity study*, and *staffing* (see Figure 7). The description of Beakta® is based on the information and the teaching material provided together with the installation of the software (Beakta 1995) and the Federation of Swedish County Council's report on PCSs (2000). Both of these documents are in Swedish<sup>2</sup>.

#### **3.5.1 Factor based patient classification**

There are two ways of measuring the patient's dependency level or nursing care workload in Beakta. One way is determining the care needs of the patient in advance. This is referred to as a normative approach. The second way is by measuring retrospectively the care the patient has actually received. This is called the empiric approach. One of the requirements of the KPP project was that PCSs would measure the nursing care that the patients really received, in other words, an empiric approach would be used. This is the usual way Beakta's patient classification work is done. Nevertheless, having the ability to use both ways of measuring is beneficial for quality reasons. The combination of methods is one way of scrutinising whether the patients actually receive the nursing care they require.

---

<sup>2</sup> The English translation of the Swedish indicator names and other concept descriptions are a free translation to English by the author and not necessarily a direct translation from Swedish or a translation supported by SYSTeam AB, the company behind Beakta®.

Additionally, the Beakta patient classification methodology assesses patients and their relatives as one whole. This means that if relatives help in the care of a patient, the patient's dependency score would be lower. However, the opposite is also true, that is, if relatives need a lot of support and information from the nurses, the patient's dependency score would increase. In describing how the patient's dependency level is determined, the focus is on the empiric approach.

### The Beakta system

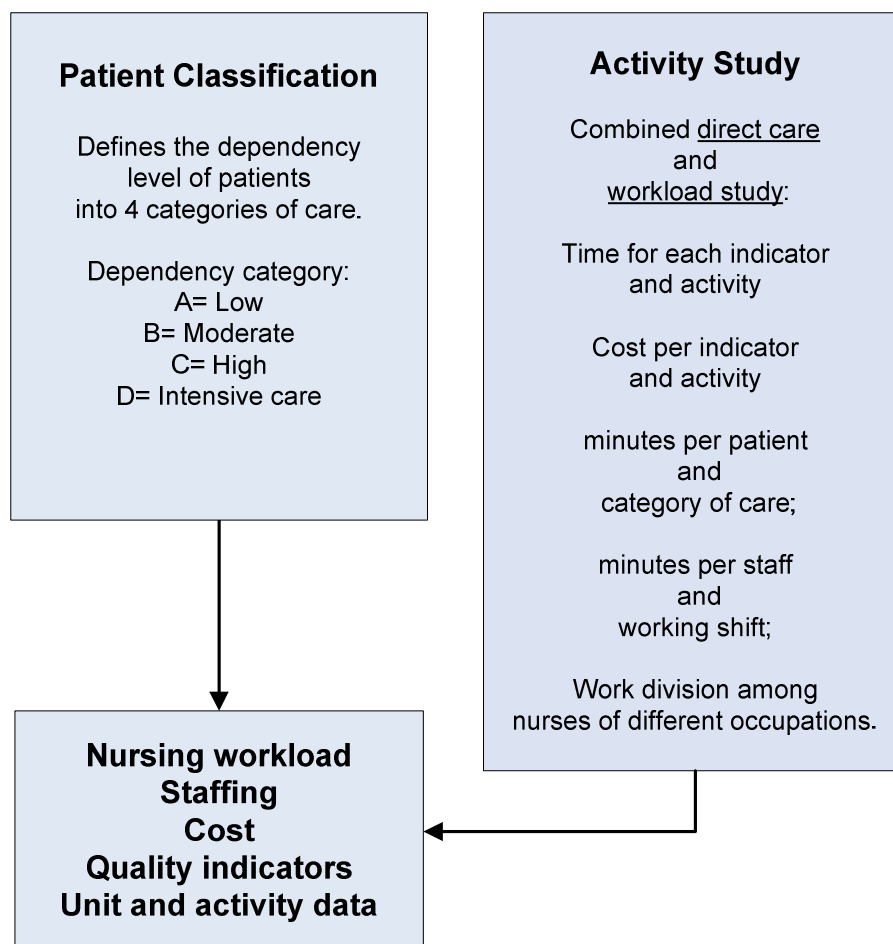


Figure 7. The structural model of the Beakta system

Patient assessment is a daily procedure in which staff members give a collective picture of the nursing care that a patient has received in the last twenty-four hours. In order to help them Beakta has a number of factors/indicators (seven for somatic care) representing the different types of direct care that the patient receives. Every indicator has three determinants which represent the amount of help and care that the patient needed according to the particular indicator area. The determinants are: small need of help (1), medium need of help (2) or large need of help (3). The total sum of the determinant points

(in brackets 1 – 3) for each of the indicators gives a final definition of the patient's category. Patients are categorised into four groups A to D, where D is the patient category for the highest dependency.

The Beakta tool has indicators for different nursing areas. There are indicators for somatic, psychiatric, and paediatric care, as well as maternity, gynaecology, and delivery care. Somatic care has seven indicators. These indicators are used in a number of units, which include surgery, orthopaedics, and rehabilitation wards. The indicators are:

1. Hygiene/elimination
2. Nutrition
3. Observation/examination
4. Treatment
5. Ambulation/training
6. Psychological and social support
7. Communication and education

Every indicator has a precise definition of the types of activities included in its area. This also applies to the sub-definitions of every indicator. The sub-definition is quantified by the determinants of each indicator. The determinants define the level of dependency in a specific area, for example, nutrition. Beakta provides a standard definition, which can be adapted to what constitutes a patient with small, medium, or large need of nursing care. As previously mentioned, the points that a patient receives from every indicator of the determinants' definition are from 1-3. These points, (defined by the determinators) from the seven indicators, are added up to a total score which gives the patient's dependency category A, B, C or D for that particular day (see Figure 8).

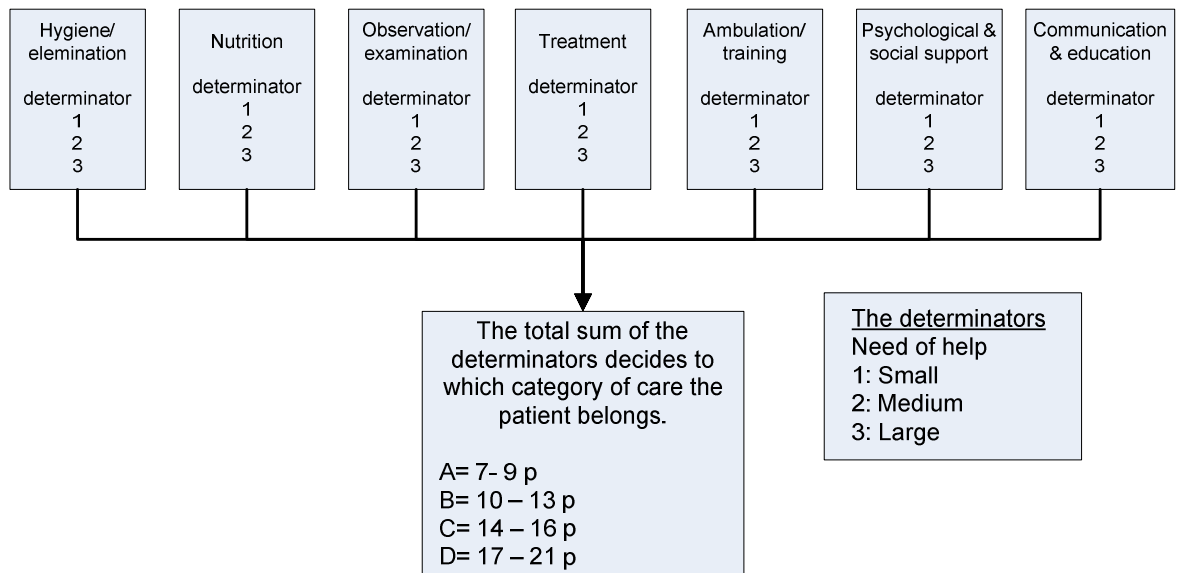
Dependency level A, B, C or D stand for:

A= Low dependency

B= Moderate dependency

C= High dependency

D= Intensive care dependency



**Figure 8 Description of work procedure for patient classification**

The dependency category is primarily defined from the total score of the different indicator values according to the following:

A= 7 – 9 points

B= 10- 13 points

C= 14- 16 points

D= 17- 21 points

When the activity study has been carried out, a standard time is calculated for every determinator. Some indicators might represent a greater effort (expressed in more time) than others, therefore the sum of determinator points and corresponding dependency determination is not always representative.

There are three complementary ways of expressing the total dependency of a patient:

- Unweighted points
- Weighted points
- Number of minutes per patient

An unweighted point does not consider the time devoted to a specific patient. It shows only the registered dependency level. For example, a category C patient in unweighted points might be considered a category B or D if weighted points are used. Weighted points take the total time devoted to a patient into consideration. As previously mentioned, some indicators represent activities that are more time consuming than others, and consequently there could be a clear difference between these two ways of representing dependency levels.

A standard time is calculated for every determinant and a distribution of the different dependency levels can be expressed in the minutes of nursing care required per day according to the following:

A= 0- 75 minutes

B= 76 – 135 minutes

C= 136 – 180 minutes

D= 181 – above minutes

Beakta has a complete mapping process between these different ways of expressing dependency levels and the information is easily accessible. An advantage of unweighted points and number of minutes instead of weighted points is that the data from one unit is easily compared to other similar units. The work in the two last case studies used unweighted points and number of minutes to represent the total patient dependency level.

### **3.5.2 Activity study**

The activity study is performed as a self –observation work sampling study. The aim of the work sampling technique is to investigate the proportions of total time devoted to the various activities that constitute the work of the different personnel categories (Niebel and Freivalds 2008). However, in contrast to normal, work sampling measurements, the activity study in Beakta has several aims. Firstly, it is used to map how the personnel, in their different categories, use their time and skills. Secondly, it shows how the work time is distributed between different work areas, tasks, indicators/determinators and dependency groups. Thirdly, it gives the basis for calculating the time requirement per indicator/determinator in the nursing workload measurement. Finally, it can be used to map

working conditions, ward routines and bottlenecks. After completing the activity study it would be a natural assumption to calculate if it is possible to use more time for direct care activities and less for administrative work.

Beakta has, in contrast to other PCS tools used in Sweden (Levenstam and Bergbom Engberg 1993), a combined patient and workload study. The workload study captures how the staff's time is distributed among different work tasks and areas. The patient study shows how much time is spent on different indicators/determinators (direct care activities), that is, on the different needs/areas of direct nursing care.

The activity study is performed over a period of 10 -14 days. The entire nursing staff take part in measuring and documenting all their activities during this period. The study activities are grouped in four main areas:

1. direct care activities
2. indirect care activities
3. unit-related work
4. personal time

Direct care activities are always patient related. Indirect care can be both patient related and non-patient related activities. Unit-related activities and personal time are non-patient related activities. A number of specified activities is under each of these activity areas (for a more detailed description of activities, see Table 14). The activities under direct care correspond to the indicators which are used during the patient classification process.

In the procedure of the activity study, the staff members document the activity they are performing. The members use a form to help them, which lists the different activities on separated rows and the time periods (10 minutes) in different columns. Every ten minutes the nurses document their activities by making a cross or, when a patient is involved in the activity, writing the patient's bed number in the square which intersects the activity row and the time column.

Having a high number of registered activities provides the possibility of statistically securing the data on how the nursing time is used and in what way the different activities are distributed during the day. The connection between the activities and the assessed patients gives the possibility of describing how much direct care time each indicator and determinant represents. The Beakta methodology suggests that the activity study should be repeated at least once a year. An example of indicator and determinant values is illustrated in Table 1.

Indicator	determinator 1	determinator 2	determinator 3
Hygiene/elimination	10	46	69
Nutrition	3	7	22
Observation/examination	4	10	64
Treatment	1	6	19
Ambulation/training	4	12	38
Psychological and social support	8	14	18
Communication and education	3	5	10

**Table 1. Calculation of time consumption in Beakta, figures expressed in minutes.**

The values presented in Table 1 are referred to as *direct patient care time*. The values of the indirect care activities that are patient related are referred to as *remaining patient care time*. The relationship between the *direct patient care time* and the *remaining patient care time* is calculated and represented as a fraction of a whole according to Table 2, below.

Dependency group	A	B	C	D
Direct patient care time/ Remaining patient care time	43/57	58/42	68/32	68/32

**Table 2. Relation between direct and remaining patient care time per dependency group**

The remainder of the indirect care time (not patient related), the unit-related care work and the personal time is added together and designated *base work time*. The number of care days is calculated for the study period (a care day is equivalent to the time a patient is in the ward, which is somewhat different to the time the patient is registered in the ward). The base work time is related to the total number of care days and this gives an average *base work time* per care day.



### **3.5.3 Staffing**

The staffing tool in Beakta is based on information from the activity study and additional data that needs to be entered manually. Information about theoretical staffing is accessible by its software implementation, while information about how the different work shifts are divided, the grade mix of nurses and the salary cost per occupational group needs to be registered separately. From the assessment/patient classification part, data about the direct patient care time of the current patients is transferred and the base work time is added to this data. The result is a staffing that is adapted to the dependency level, which can be compared to the actual staffing level.

### **3.5.4 Beakta® – software tool**

The software implementation of the methodology is developed on a standard platform suitable for both small and large organisations. The tool is able to collect and send data to both nursing care, and personnel administration data systems. This gives the advantage of a single registration that automatically updates other systems, minimising the administration time. The tool provides a user friendly interface which facilitates the documentation of the daily patient classification, giving the users access to the determinant definition for consultation. Furthermore, it contains a powerful report interface giving users the possibility to see all the key figures of their wards' activities.

Moreover, Beakta stores the ward's historical data in a SQL database, which contains data from every patient's dependency development and its variation during their stay in the ward. This information, together with the activity study data, gives the possibility to model stochastic workload variance for an inpatient ward.

## **3.6 Summary**

This chapter describes Patient Classification Systems and shows their use in an international and Swedish context. It emphasises that a patient's dependency level has an important impact on staff requirements and demonstrates how PCSs classify and quantify these requirements. Moreover, it indicates some of the essential components that a PCS should have in order to confirm its reliability and validity. Critics declare that it is difficult to maintain reliability and validity, but with regard to the PCS tools it does not seem to be a methodology problem. If validity or reliability is an issue, it is related rather to the

implementation and operational management of the tool. The value of PCS is not only what is of interest to us, that is, the work requirement quantification and the historical data of patients' dependency trajectory, but also its ability to take into consideration human factors that are more difficult to quantify.

The chapter also puts the use of PCSs into a Swedish context and describes the requirements that a Swedish case-costing strategy puts on PCSs as an intermediate product for case-costing calculations. The requirements or criteria made by SALAR, which is the most influential stakeholder in the Swedish healthcare system, are obviously important with regard to choosing an appropriate PCS. An inpatient modelling methodology that is built on a PCS that is outcast would definitely not improve its chances of being used and applied. The choice of PCS fell on Beakta, which represents one of the intermediate products that fully fields the criteria prescribed by SALAR. Beakta is described in detail to give the readers the background information needed to understand how it is used in the inpatient modelling solution presented in chapter six.

## **4 Chapter four: The orthopaedic case study**

### ***4.1 Introduction***

This is the first of three case studies conducted at a regional hospital in Skövde. The project began in October 2002 and ended in November 2003. The study was initially presented by Urenda Moris et al. (2004), but a more thorough presentation follows. The overall aim of this first study was to explore a new domain. More precisely, this required obtaining a better understanding of the questions and problems the healthcare services are facing and of the particular difficulties in running a simulation project in this new area. Difficulties related to communication, organisational barriers, data access and modelling challenges were unknown at this initial stage. Consequently, it was with an open mind and stumbling steps that this first study took place, well aware that first impressions and success are of huge importance when a new technique is introduced to an organisation.

From the thesis perspective, this first case study and the results it presents are the stepping stones for the contents of the following chapters. Consequently, the objective of chapter four is to present the initial work that led to the identification of two related questions which identify the core of this work; how can robust system design be applied when systems are small and vulnerable to inherent variations, and how can DES be used to model inpatient dependency variations and thereby variable staff requirements? Furthermore, it contributes to the empirical understanding of the importance of dealing with variability in the design and evaluation of inpatient wards.

### ***4.2 Background***

One of the most important regions in Sweden, both in terms of population, size and economic strength is Västra Götaland with its main centre in Gothenburg. The Västra Götaland region was formed in January 1998 when two former counties, Älvsborg and Skaraborg Counties, were merged. The County of Skaraborg is geographically situated in the eastern part of Västra Götaland, between Sweden's two largest lakes. Although Skaraborg County no longer exists, the region still maintains its old name (from the year 1634). This region, which has an area of 7,400 square kilometres and a population of only 250,000 inhabitants, has Skaraborg Hospital (SkaS) as its main healthcare provider.

The facilities of Skaraborg Hospital are distributed over the cities of Skövde, Lidköping, Falköping, and Mariestad, which represent the major cities of this region. The hospital has approximately 800 beds and 4,700 employees in total (see Table 3 and Table 4 for more details), with the largest facility, Kärnsjukhuset (KSS) in the city of Skövde, having 450 beds. These four hospital sites work closely together, coordinating their resources to provide effective and high quality treatment to the healthcare consumers of the region.

<b>Distribution of personnel among a few professional categories:</b>	
Register Nurses	1870
Auxiliary Nurses	1239
Physicians	473
Administrative personnel	200

**Table 3 Distribution of personnel of some professional categories at SkaS (Skaraborgs Sjukhus 2008).**

Orthopaedics is one of the prioritised areas of the Swedish healthcare system and SkaS. However, the orthopaedic unit at KSS, which is the largest within the distributed SkaS facilities, struggles with long waiting times for consultation and surgery for non-emergency patients. In some cases, waiting times for consultation can exceed twelve months, and those for surgery may exceed six months. The initial work of the thesis focused on obtaining an understanding of the difficulties that the orthopaedic department deals with, and identifying areas where DES could be used to improve its activities.

<b>Healthcare services provided in 2007:</b>	
41 000	Bed days
19 300	Surgical operations
2 300	Child deliveries
204 300	Medical examinations

**Table 4. Approximate figures of healthcare services provided in 2007 at SkaS (Skaraborgs Sjukhus 2008)**

### ***4.3 The Orthopaedic Unit – Issues to Consider***

The orthopaedic department at KSS consists of several units and personnel groups, distributed into the following units:

- Ward 83- comprises 24 beds for elective (planned) patients
- Ward 84- comprises 24 beds for trauma patients

- Consulting clinic- a unit that deals with both outpatient consultations and inpatient preoperational procedures.

The above named units, especially the wards, collaborate closely with other units within SkaS and outside the hospital, these include:

- REHAB, which is a department comprising several rehabilitation wards.
- KAVA, which is a surgical emergency ward.
- The emergency unit (ED) at KSS.
- Geriatric care wards, which are municipally driven.

All these units affect each other, which means that decisions or policy changes in one of the units will have consequences for the others (see Figure 9). In addition, there are several other factors that influence the total capacity of the orthopaedic department, such as the number of physicians and operation teams, as well as the number of operating theatres, and so on. Furthermore, bottlenecks in one unit or facility will have a propagating effect on the other units, thus creating a complex pattern of lost capacity in the orthopaedic department.

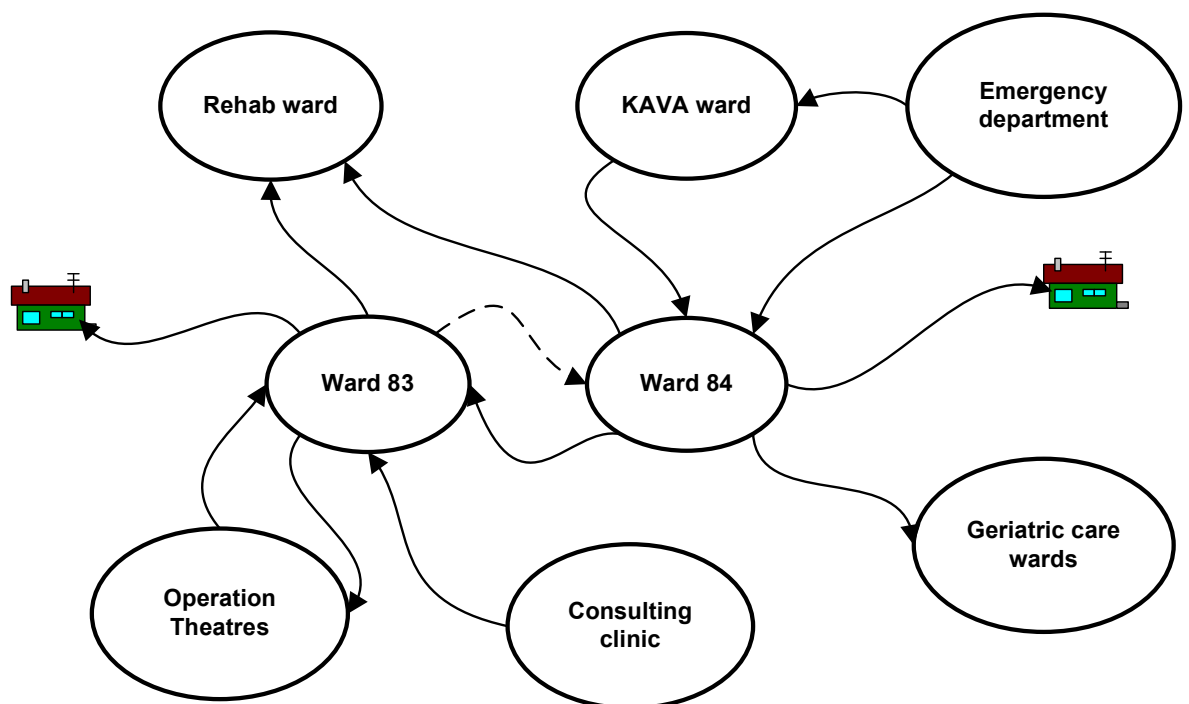


Figure 9. The general flow of patients between the different units

Wards 83 and 84 were facing policy and architectural changes that would have an impact on the capacity of the orthopaedic department. Furthermore, several complex relations to other units severely affect the wards' capacity, in terms of fewer beds available. There are several reasons for this situation. Firstly, an increasing number of trauma patients are passed on to ward 84. This is partly due to KAVA prioritising the surgery department, which leads to fewer bed places for the orthopaedic units, with the result that more patients from ED come directly to ward 84. This affects the bed capacity of ward 83 for elective patients. The reason is that ward 83 functions as a buffer for trauma patients when ward 84 is full. In general, trauma patients from ward 84 use approximately 20 per cent of ward 83's capacity.

Secondly, a significant number of trauma patients stay longer than necessary because it is difficult to find beds in the municipal nursing homes, which thus affects the total capacity of the wards. Thirdly, trauma patients arrive at random and their admittance must have priority over elective patients. When ward 84 does not have the capacity, ward 83 may need to refuse the admission of elective patients, leading to the loss of time booked for the operating theatre, the operation team, and the surgeon. All three are key elements in the orthopaedic department's process as well as critically scarce resources. During 2002, more than 50 elective patients could not be admitted because of a lack of ward places. In 2003 the situation improved, but this was mostly due to the geriatric care wards receiving patients more rapidly than due to the application of new strategies.

As previously mentioned, the wards also faced architectural changes which could worsen the situation. These changes were included in three redesign suggestions which would all considerably reduce the number of beds in the wards. From a total of 48 beds, 24 beds per ward, there would be 16 or 18 per ward, thus 32 or 36 beds altogether. This reduction was suggested in order to harmonise the size of wards 83 and 84 with the general size of the other wards in KSS. It is obvious that a reduction from 48 to 36 beds or even 32 would cause problems if no other measures were taken. In such a situation, simulation potentially offers a simple way of studying how reasonable the different suggestions are with regard to reducing the wards' capacity without increasing the waiting list of patients and without creating problems for another ward or unit.

## **4.4 Case study questions**

The overall objectives of this first study were to gain understanding of non-quantifiable issues such as difficulties related to, for example, communication, organisational barriers, data access and modelling challenges, and thereby refining more precise research objectives. This did not exclude that the individual project had its own more quantifiable objectives and questions to answer.

Several questions were raised regarding the possible improvement of the operations of the wards, which can be summarized as follows:

- How can the situation be improved so that the number of refused patients is minimized and the wards' occupancy level (i.e. utilisation, normally referred to as occupancy level when bed utilisation is considered) is maintained?
- What changes are necessary in order to manage an eventual down-sizing of the wards?

Consequently, in order to evaluate the different design suggestions, a DES model of wards 83 and 84 was built.

## **4.5 Model development**

Even for an experienced simulation analyst, a domain change from manufacturing to healthcare is not a trivial step. This circumstance and the fact that DES has not been previously used by SkaS, demanded a common foundation where both partners were able to understand the scope and needs of the project.

### **4.5.1 Software Package Used**

The Virtual System Research Centre at the University of Skövde has been using a number of software packages for the simulation of both 2D and 3D solutions for the last few decades. However, all of them primarily target the manufacturing domain. Nevertheless, the software chosen for this and the other case studies was Quest® from Dassault Systèmes DELMIA®. Quest is a 3D simulation software package which has the possibility of tailoring the logic and behaviour of the resources modelled. Unfortunately, the predefined objects and functionality have mainly been developed for industrial rather than healthcare systems. This resulted in considerable, tailor made adaptations, for example, new object functionality and complementing the relatively small number of pre-programmed

distributions. Quest® especially lacks several main discrete distributions, which was solved by writing new distribution expressions in Quest® programming language, SCL. In addition, an ExpertFit® software package was used for input analysis.

#### **4.5.2 Data Acquisition and Input Analysis**

Gathering information and data is always time consuming and complicated. The hospital uses both computerised and manual systems to store data. Some key patient data, which was kept in different computer systems, was not easily retrievable, and it demanded authorisation from senior management for this data to be accessible to the project.

The data was analysed and subsequently complemented with manually registered data which consisted of information from all the patients that had been admitted to wards 83 and 84 during 2002. The data used to model ward 83 was mainly collected during two 4-month periods, one in late winter and spring, and the other from late August to the middle of December. This is because ward 83 is closed during the summer period and only partly used during some other key weeks of the year. The following data was retrieved:

- The arrival pattern of elective patients to ward 83.
- The arrival pattern of trauma patients to ward 84.
- The time the different patient categories stayed in the wards.
- The percentage levels of the different diagnoses, gender, etc.

Additionally, some patient categories, belonging to specified diagnoses, were separated and individual distributions of these groups were identified. These distributions became useful when different policies were analysed. The available data, which describes the amount of time the patients stayed in the wards, was discrete, that is, only the day of arrival and discharge was gathered into the data systems. However, the pattern of arrival during the day is well known. Elective patients for hip or knee-joint plastic surgery arrive at 2 pm, the others at 7 pm, only between Sundays and Thursdays.

Trauma patients from KAVA are transferred to the orthopaedic department at 2 pm every day, while trauma patients from the emergency unit are transferred to the orthopaedic department between 7 pm and 8 am. The percentage of this last group of patients arriving



in the evening hours was estimated to be 40 per cent of the total in the group, while 60 per cent were estimated to arrive during the early morning hours. In addition to the discrete distribution, a continuous uniform distribution was used to give the exact hour when this category of patient would arrive. Whatever the circumstances in the orthopaedic wards, all trauma patients are admitted, even if they are allocated temporary bed places for a period of time. Patients from the two wards are generally ready to be discharged after lunch, which means between 1 pm and 4 pm. The simulation model assumes that all patients leave during this time period. The data was fed into statistical software, ExpertFit®, and discrete distributions were used. Table 5 shows the identified distributions.

DESCRIPTION	DISTRIBUTION
Daily arrival rate for elective patients, (Sunday – Thursday)	Binomial (15, 0.1908)
Daily arrival rate for trauma patients	Neg. Binomial (12, 0.7924)
LoS for elective patients	Neg. Binomial (7, 0.4867)
LoS for trauma patients	Neg. Binomial (3, 0.2906)

**Table 5. Identified distributions**

It should be remembered that the data for the inherent demand variation represented by the arrival rate of trauma patients was previously presented in Figure 5. The graph in Figure 5 shows the randomness of the day to day changes of the number of patients. Table 6 presents differences in average LoS among trauma patient for the seasons of 2007.

Average length of stay for trauma patients	Days
Spring	8,6
Autumn	7,3
Whole year	7,6

**Table 6. Average length of stay for trauma patients**

### 4.5.3 Modelling Logic and Using Graphics

Complex graphics is seldom of much use in simple models. The model built for this case study could have been made with a higher abstraction level than the actual wards. However, this was the hospital's first experience of DES and, therefore, it was necessary to build up

credibility for the simulation technology. Thus, a model with comprehensive logic and 3D graphics was developed. A print screen of the two wards is presented in Figure 10.



**Figure 10. Wards 83 and 84, modelled in Quest®, patient colour visualises gender**

Among the logic used was the way in which the rooms were allocated. For example, the rooms are three different sizes, single, double, or six-bed rooms. In addition, they can only be occupied by one gender at a time. This means that a male patient can only be placed in a room with other male patients, or in an empty room. Furthermore, patients having knee or hip joint plastic surgery have to be scheduled to a single or double room for 3-4 days after surgery to minimise the risk of infection. This leads to a complex pattern of moving patients from one room to another in order to make optimal use of the bed places.

Crucial to the model's logic was the movement of patients from ward 84 to ward 83 when there were no longer any beds available in ward 84. There are also situations when both wards are full and new trauma patients still arrive. In real life, this problem is solved by using temporary rooms. As soon as a regular bed place is empty in either ward 83 or 84, the patient in the temporary room is moved. The same activity was implemented in the model.

## **4.6 Validation of the model**

The validation procedure is crucial, not only for the future experiments or the use of the model, but also for credibility purposes. The project co-ordinator and the users need to be convinced that the model represents their activities correctly. There are several ways to validate a simulation model, for example, face validity, input-output transformations, historical data, and sensitivity analysis among others (Banks et al. 2001). The most suitable way in this particular case was to use the input-output transformation. A mapping was done between the model average and the two four-month periods in 2002. Two outputs from the runs were analysed: the occupancy level of the two different wards and the number of patients not admitted.

Initially, the results showed a clear mismatch between the model and the 2002 values. The main reason was that not all the data had been registered into the computer system. After correction and new identified distributions, the simulations gave the expected result. The occupancy levels increased and the number of refused patients was significantly raised in the model.

A significant variation was discovered in the average time that trauma patients stay in the ward during the spring and autumn season respectively (see Table 6). This led to further validation and calibration of the model. When the model was run with a distribution that reflected the spring values, it gave a similar result to the history spring data. The number of refused patients increased considerably, compared to using the whole year's data, and it reflected the wards' situation during the spring of 2002 very closely. This leads to the question of which type of distribution represents the future demands on the ward most. A comparison was made to the spring 2003 figures (see Table 7). This data reflects only three months of activity, since one of the units was closed for almost a month due to an outbreak of stomach flu. However, it provides some insight into a trend that was already being seen, of passing elderly patients onto geriatric care more quickly. The project group agreed to use the distribution based on the autumn data of 2002 as the one that best represented the foreseeable future.

<b>Runs based on:</b>	<b>Whole year data 2002, LoS 7.6 days</b>		<b>Spring data, LoS 8.6 days</b>	
	Nr. rescheduled	Occupancy level %	Nr. rescheduled	Occupancy level %
<b>Average</b>	10.3	79.2	26.6	85.7
<b>Stdev</b>	8.2	3.1	13.3	3.0
	<b>Autumn data, LoS 7.3 days</b>			
<b>Average</b>	4.7	77.8		
<b>Stdev</b>	4.23	2.8		
<b>Historical data</b>	<b>2002</b>		<b>2003</b>	
<b>Spring</b>	39	87.2	0	76.8
<b>Autumn</b>	7	81.0	-	-

Table 7. Validation runs results.

## 4.7 Experimentation and Analysis

Several sets of experiments were carried out. The objectives were to target the questions in section 4.4 by doing a set of “what if” analyses and then proposing system improvements. Comments on two sets of experiments follow.

### 4.7.1 First set of Experiments

The first set focused on passing contusion and fracture trauma patients from the orthopaedic wards to Rehab on the 4<sup>th</sup> day (see Figure 9), which would ease the stress on ward 84 and consequently on ward 83. Four different setups were conducted with the only difference being the number of active bed places in the wards. The number of beds for each ward and setup was 24, 20, 18 and 16. It must be remembered that autumn 2002 LoS data for elective and trauma patients was used, but divided into the different diagnoses, for example, contusion, fracture. In the results, presented in Figure 11, the number at the readings represents the number of bed places.

### 4.7.2 Second set of Experiments

The second set was based on completely discharging contusion and fracture trauma patients to a new, theoretical ward and reducing the number of beds in both wards 83 and 84 to a total of 16 beds by 2 equally distributed. There were plans of opening a new unit for this category of patients and there was consequently interest in planning its dimension. This study only provided a preliminary result on the dimension of a possible new ward. The new ward’s result, analysis and design suggestions will be presented in chapter 7 while this chapter will present the result and analysis of the scenarios affecting ward 83 and 84.

Four different scenarios were created with some significant differences:

1. The first setup used the whole year's average per cent of diagnosed fractures and contusions, which represented 22 per cent of the men and 38 per cent of the women.
2. The second one used the same per cent figures, but a new scheduling rule for the elective patients. Instead of rescheduling the patient when no place was available in ward 83, the system tried to find a place for the elective patient in ward 84.
3. The third scenario used scenario two, but in order to reduce the preventable variations, the scheduling of elective patients was evened out. This means that the simulation model no longer used the Binomial distribution presented in Table 5, instead a constant number of patients were enrolled at the ward from Sunday to Thursday, which represents a slight increase from 2.89 to 3 patients per day. Nevertheless, this increase means that 15 more patients receive a plastic knee or hip replacement within a 4 month period, which is a necessary increase considering the long waiting time.
4. The fourth scenario had all the settings found in the previous, but added a LoS reduction among elective patients equivalent to 1 day. The following section discusses why this reduction can be considered reasonable.

The results from these scenarios are presented in Figure 12. Every data point has a label identifying the total number of beds in both wards for that reading. The data points, which are averages, are based on 10 simulation runs. Each run represents 4 months or 120 days of operation and required slightly less than three hours of computing time. The long execution time is partly due to the complex logic used for room allocation. Correlation between the different scenarios was achieved by using the same line up of random seeds. Therefore, despite high variation in output measures within a scenario, the averages provide a good estimation of the quantitative differences of the outputs.

### 4.7.3 Analysis

An analysis of the results from these two sets of experiments shows that when the size of the wards are reduced, there is a clear tendency for them to become more vulnerable to variations in the pattern of incoming patients. This intuitive and yet uncommon knowledge, is well in line with the results presented in Green and Nguyen (2001). They show how systems which are affected by high rates of variability have difficulties in maintaining low waiting times and yet high utilization. This relation becomes more accentuated when systems are smaller. The equivalent to time delay, in the systems described by Green and Nguyen (2001), is the number of refused patients in the current scenarios. For instance, Figure 11 presents the occupancy level to the number of refused patients, for the first set of experiments. Additionally, the graph presents the same relation for the historical data that corresponds to the spring (8.6), autumn (7.3) and whole year (7.6) average of 2002, where the LoS values are shown next to the readings, see Table 6. In wards 83 and 84 during 2002, the occupancy level was higher in spring because the trauma patients stayed, on average, longer in the ward. However, the number of bed places remained static, that is, 48 places altogether.

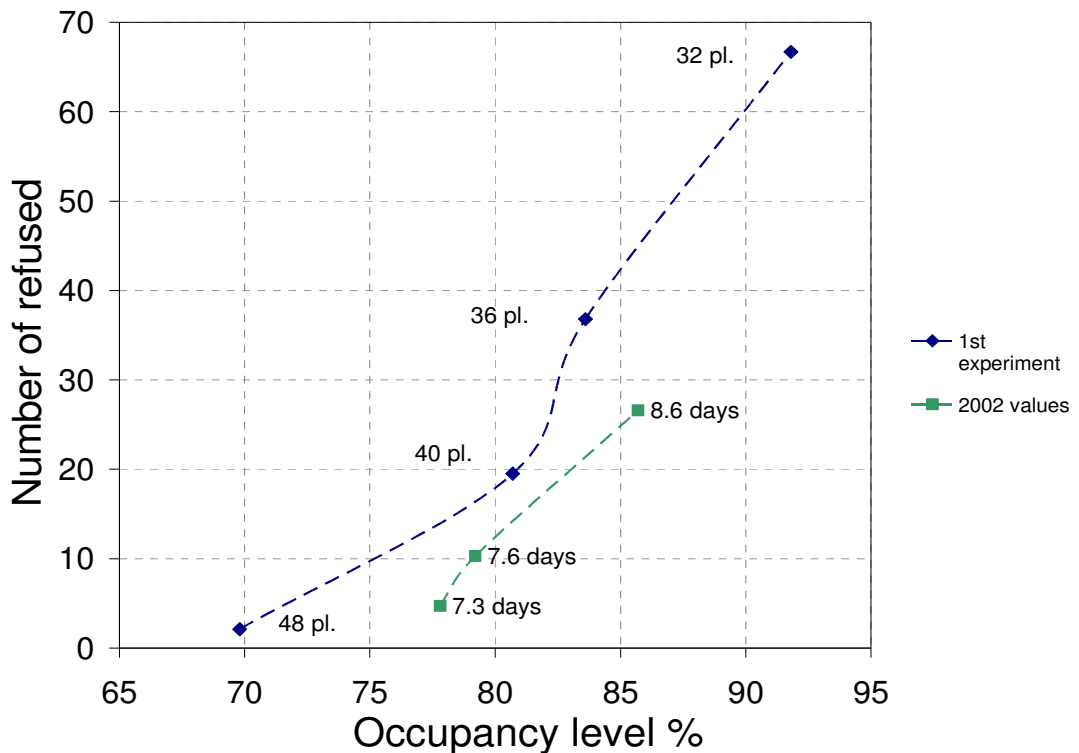


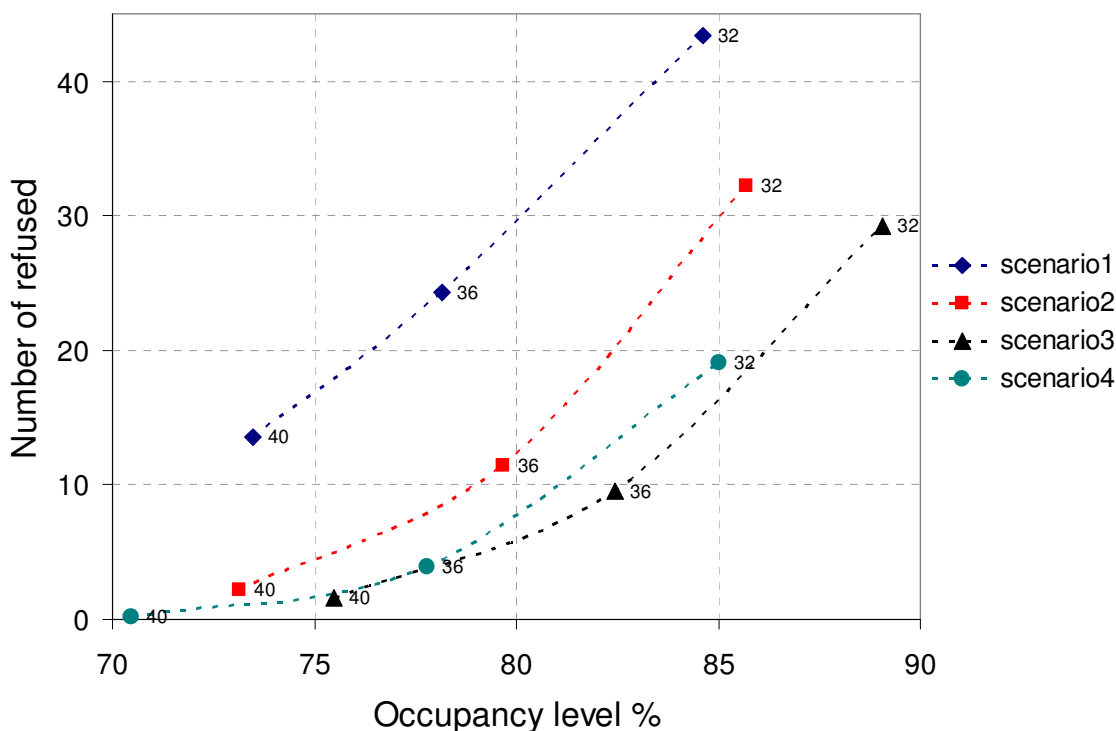
Figure 11. Results from the first set of experiments where contusion and fracture patients are moved on the 4<sup>th</sup> day, contrasted with 2002 values.

The second curve represents how the occupancy level ratio increased because the number of available bed places decreased, which consequently led to a higher demand per bed unit. The curve shows that the system becomes more sensitive to variations, resulting in a higher number of refused patients compared to the same level of occupancy in the original system. The relation is not linear and the curve represents the increasing rate of refusals when the number of bed places falls below 40. What can be seen is that despite the effort of moving contusion and fracture patients (which represented 22 per cent of the men and 38 per cent of the women) from the orthopaedic ward to the Rehab ward on the fourth day, it does not improve the occupancy level:number of refused ratio. This supports our statement that smaller systems are more sensitive to demand variations.

This tendency could be significantly improved if elective patients were allowed to be scheduled to ward 84 when there is a need, which the second study shows. The results of the second study are presented in Figure 12. The first two scenarios represent how different numbers of bed places affect the occupancy level and the number of refused patients. The lower red curve represents the system that allows patients from ward 83 to be scheduled to ward 84. This change makes a significant difference in terms of both better occupancy level and fewer refused patients. Moreover, it illustrates the importance of having, when possible, an open patient flow between wards, which is especially important when the wards are small. This new “virtual unit” can be used to level out a particular patient group whose numbers peak for a period of time, considering it would be unlikely that the numbers of all patient groups peaked at the same time. This approach of opening the barriers of systems is one way of making systems “bigger”. Instead of having two small systems with 20 bed places each, the linked wards function as one system with 40 bed places, and can therefore better handle demand variation.

In scenario three, the preventable variation has been reduced by having a constant arrival rate of elective patients scheduled from Sunday to Thursday (they arrive on these days and have surgery the day after). This constant rate is set at 3 patients, instead of the old average of 2.89, which is the main reason for the higher occupancy level of the wards. However, it is of value to note that the rate of refused patients does not increase and, instead, a slight decrease can be observed. Analysing the setup with 36 bed places, which is one of the propositions, it can be seen that even though the average of refused patients is slightly

under 10 in a 4 month period, there has been an increase of 15 operations in that time. This amounts to a total surplus of 5 operations above the current average. However, although this would compensate for the cancelled operations, it is not an acceptable situation considering that booking surgeons, operating teams and operation theatres for procedures that subsequently have to be rescheduled is a waste of resources anyway. Nevertheless, scenario three verifies the importance of reducing the preventable variations and identifies a second step in making a system more robust to demand variations.



**Figure 12. Results from second set of experiments – relation between number of refused, number of beds and occupancy level of wards 83 and 84 under the different scenarios**

Scenario four continues from the end of the previous setup. What happens if the elective LoS is reduced with one bed-day per patient? Well, the graph in Figure 12 shows that a significant improvement in reducing the number of refused patients is what happens. LoS stochastic distribution figures are very important for system dimension (Green and Nguyen 2001; Marshall et al. 2005). The question is how can this be achieved in real systems. With regard to the trauma patients in ward 84, it can be seen that LoS was reduced simply through better discharge methods and communication with municipal nursing homes (SoS 2005). However, this is not entirely useful for elective patients in ward 83, because they



are normally not discharged to nursing homes. A comparison between the groups' LoS figures shows that their average is almost identical (8.2 days for elective and 8.11 for trauma patients), but the variance is very different (15 for elective patients and 37 for trauma patients), suggesting that patients with extremely long LoS belong to the trauma patient category.

A literature review points to some guidelines that can be useful with regard to LoS reduction among surgery patients. Nilsson et al. (2000) describe the LoS outcome for the patients of two surgeons. They revealed that operation procedures had an important correlation to average LoS for the two groups of patients. One of the surgeons had an average operation time of only 30 minutes, but his patients had an average LoS of 5.1 days. The other surgeon had a longer average operation time, which was equivalent to 62 minutes, but his patients had a LoS of only 4.6 days. Consequently, operation techniques and procedures have an effect on the patients' LoS. Several authors in the field of Blood Management present another finding that supports this notion (Stulberg and Zadzilka 2007; Spahn et al. 2008). Blood Management techniques aim to reduce or eliminate the use of blood transfusions in surgery, because they represent a considerable cost for healthcare providers and an unnecessary risk for patients. In their book, Seeber and Shander (2007) present therapies, methods, tools and evidence based procedures, from over four decades of work, that eliminate, in almost all cases, the need for blood transfusions in elective surgical procedures. Besides the avoidance of blood transmitted diseases and the reduced cost from not having to store and handle blood units (Amin et al. 2004; Shander et al. 2007), the results are a faster patient recovery and therefore a lower LoS (Innehofer et al. 1999; Blumberg et al. 1996). The figures show a LoS reduction of between 0.8-5.1 days (Glenngård et al. 2005). Several hospitals in the US and clinics in Europe already have fully operational programs for blood management procedures, with very good results.

Several other minor experiments were also completed. The focus of these was that the main result variables, which are the number of patients in need of rescheduling versus the occupancy ratio of beds, were contradicting each other, despite various efforts to find a solution to this problem. Furthermore, it is increasingly difficult to find a high utilisation ratio when the system's inherent variation is high.

## **4.8 Discussions and contribution to framework**

It was clear that by having fixed discharging/rescheduling rules for different patient categories, the system could not obtain both low rescheduling figures and a high utilisation ratio at the same time. Additionally, it was apparent that by reducing the size of the units, it would become more difficult to tackle inherent system variations.

The first set of experiments highlighted how difficult and unrealistic the task of downsizing the two wards was, while maintaining similar conditions to the current operational ones. The second set of experiments took a more realistic approach in which several trauma patient groups were rescheduled to a new unit. Despite this, the new system design was unable to cope with the high variations. Three modifications were tested: 1) opening system boundaries, 2) lowering preventable variations, and 3) lowering LoS for elective patients.

The first two suggestions are more easily implemented, if a common consensus among the partners is reached. Unfortunately, this is not an easy task in reality. Opening the boundaries and accepting elective patients in the trauma ward is, intuitively, difficult for the staff. They are aware of the fluctuating arrival rate of trauma patients, and are therefore naturally cautious and do not want to take any chances. Uncertainty has that outcome. It creates the notion that we need to have extra margins. The second suggestion also causes problems due to personal and organisational working patterns and structures. Firstly, different specialists share the operating theatres, and they need to collaborate when new scheduling practices are introduced (Litvak et al. 2005; Persson 2007). Secondly, some of the specialist orthopaedic surgeons also work at other county hospitals and when they find an opening in their current work schedule, they inform the hospital which days they can operate in Skövde and procedures are rescheduled for them. Thirdly, the formal and informal power of surgeons makes it politically difficult to change routines without a very good foundation (CHSA 2002). For many, a simulation model result is not a sufficient foundation.

The third suggestion is much more difficult. Surgeons are reluctant to change their working routines, especially not their way of performing surgery. In Sweden, blood management approaches have been traditionally used by highly skilled and open minded surgeons. It is,

however, not the “new” techniques or methods applied in blood management that cause problems. It is instead the lack of awareness and the tradition and culture among physicians that do (Seeber and Shander 2007; Spahn et al. 2008). However, whether these changes would result in the estimated lowering of LoS is too early to say. The point is however that a reduction in LoS would certainly make a substantial difference.

The project group reached consensus that a more flexible approach was needed if the orthopaedic wards were to be downsized. One suggestion was to let several wards with patients in need of related care share a unit, which would level out the different wards’ variation. The planned new unit could possibly have that function. An important request from the managers of KSS was the need to add personnel into the models. Maintaining the right amount of personnel in the healthcare system is often discussed, both internally and in the media. In order to do this, patients need to be categorised according to their need of care. Furthermore, the many different processes carried out by the nurses and assistant nurses need to be correctly mapped and quantified.

In summary, two important conclusions were drawn from this first case study, and serve as guidance for the development of the framework of this thesis:

1. The need for the robust design of wards. Reducing the preventable variations. Compensating for small ward size by eliminating unnecessary boundaries and using more flexible structures.
2. The importance of being able to quantify the required staff level to the wards’ fluctuating conditions. In essence, it highlights the need to add patient dependency variations into the sources of systems’ variation that need to be addressed in order to efficiently deal with the design, planing and evaluation of inpatient wards.

Finally, one of the most valuable lessons of this first study was the insight it gave to both the researcher and the managers at KSS. Suddenly, KSS managers had the possibility of studying the future effects of planned changes and gain system understanding, without the risk of making expensive mistakes, which such changes could lead to. For example, the vulnerability of small systems to high system variations was something that was not fully understood by management. There was also a discussion about having a more flexible approach to staff, such as the idea of sharing staff among the wards, depending on the

needs of the different wards (staff pools). At the time, this was a politically incorrect idea, but economic changes were planned that would certainly change the current view. Discrete event simulation did function as a catalyst for constructive discussions and the submission of creative solutions.

A reflection, somewhat in retrospect, is that the healthcare management team, which was divided into upper management and ward management, was quite troubled and unfocused during the project. Although they were concerned, they were concerned with the wrong aspects of the project. A lot of energy was focused on whether the employees in their wards were getting a good evaluation and less energy was put into the actual aims of the project. Therefore, although they were collaborative, they did not ask for or follow up results and conclusions in the same energetic way that managers in industry would do. When they finally saw the conclusions they were astonished, partly because the results confirmed their beliefs and partly because they learned new things, but the sense was that they still did not know how to proceed. Whether this was the result of a waterfall project development (Banks et al. 2002) in contrast to the MAIPU approach described by (Eldabi 2000), or simply the inexperience of working together, is hard to say. This was not totally clarified until the contrary was experienced a few years later in the chronological, third case study, but presented in the following chapter of the thesis.

## **4.9 Summary**

This chapter presents the first of three case studies carried out at the Kärnsjukhus (KSS) hospital in the city of Skövde, which is the largest of the four hospital sites that belong to Skaraborg Hospital (SkaS). The overall objectives of this first study were to gain understanding of non-quantifiable issues such as difficulties related to, for example, communication, organisational barriers, data access and modelling challenges, and thereby refine more precise research objectives.

The case study took place in orthopaedic ward 83 for elective patients and ward 84 for trauma patients. These wards were facing a restructuring which would reduce the number of bed-places from the current 2 by 24 to the suggested 2 by 16 or 2 by 18 bed-places. The primary aim was to harmonise the size of the wards to the general ward size at KSS. Unfortunately, the situation in the wards demonstrated an existing deficit of bed-places that

had led to the bouncing (rescheduling of operation appointment at the last minute) of 47 elective patients during 2002. Every refused operation leads to a considerable waste of valuable resources and losses of reimbursement.

Several questions regarding the possible improvement of operations in the wards were raised. These questions can be summarized as:

- How can the situation be improved, so that the number of refused patients is minimized and the wards' occupancy level is maintained?
- What changes need to be done in order to manage an eventual downsize of the wards?

A DES model of wards 83 and 84 was built in order to evaluate different design suggestions. The results show that inherent variations, arising from the arrival rate of trauma patients, cause difficulties in maintaining both high levels of utilisation and low numbers of refused patients. The lessons from the project were the need to lower the preventable variations, the use of more flexible scheduling rules between wards in order to balance inherent variations and, when possible, reduce the LoS of patients. The quantified results of these actions are presented.

Finally, two conclusions are drawn from this first case study, which serve as guidance for the future development of the framework of this thesis:

- The need for the robust design of wards. Reducing the preventable variations. Compensating a small ward size by opening up unnecessary boundaries and using a more flexible structure.
- The importance of being able to quantify the required staff level to the wards' changing conditions. In essence, it highlights the need to add patient dependency variations into the sources of systems' variation that need to be addressed in order to efficiently deal with the design, planing and evaluation of inpatient wards.

## **5 Chapter five: Robust design of a maternity ward**

### ***5.1 Introduction***

In March 2006, three years after the initial case study, the Head of the Department of Obstetrics and Gynaecology at SkaS proposed to use DES during the design phase of the new maternity ward at KSS. What was striking after the initial meeting was the resolution and commitment that the project group showed. The project leaders were MDs and it did not take long before they understood the potential and benefits of simulation. Moreover, in contrast to earlier experiences, they felt that the simulation project, including how the model was built was their concern above all. This was a major change in attitude, awareness and knowledge compared to earlier years. The hospital directors had worked with their managers and head physicians in order to make them aware of well documented engineering tools. New, quality groups were formed, and Lean and  $6\sigma$  thinking was carefully implemented throughout the organisation. Several members of the quality groups took part in  $6\sigma$  and DES courses, as well as symposiums and lectures at the University of Skövde in collaboration with SkaS. This pleasant development was the result of the hard work and dedication of several individuals and the benefits are shared among many.

### ***5.2 Contribution to framework***

Nevertheless, the project at hand was well in line with the needs earlier identified in the orthopaedic wards, the need of designing robust systems, more specifically, a robust inpatient system that faces high levels of inherent demand variation. The quest was therefore to find and/or confirm design principles that are crucial in order to determine a suitable robust system design. Another significant contribution to the research framework of the thesis is the empirical confirmation of design principles it provides and the importance of considering the different sources of variation in the design of inpatient wards.

The chapter thus provides the background to the project, including system details, as well as the methodology used to evaluate the different design suggestions. It also provides references to other work that addresses similar units and compares the different approaches, advocating for a more enhanced, although simpler approach. The chapter finally presents

how the project's "what if" analysis can support the healthcare managers' decision making. The project is being extended (finalised August 2009), which is not part of the thesis. This work was initially presented by Urenda Moris et al. (2007) and is here described in more detail.

### **5.3 Background**

The main hospital, KSS, within SkaS, is the only one in the network with a maternity ward. This gives rise to particular needs, for example, some patients must travel quite long distances in order to get to the hospital, which is even more troubling during winter time. Another aspect is that couples appreciate the possibility of staying together at the hospital during the ante partum and post partum phases of the delivery process.

The maternity ward at SkaS is located in facilities which are 33 years old and it has a layout that supports a system process just as old. It is part of the Department of Obstetrics and Gynaecology and, together with the delivery ward and neonatal intensive care unit, forms the Perinatal centre. The maternity ward mainly consists of the ante partum and post partum units. The Perinatal centre has the following physical facilities:

- Antenatal reception – this is a pre-partum outpatient clinic.
- Triage, labour and delivery unit comprising 7 delivery rooms, 1 operating room and 4 so-called auxiliary rooms, each containing 2 beds, for triage and time before and after delivery.
- Ante partum unit of 4 beds – a patient is scheduled to this unit if she has contractions but the labour work has not started. If the patient does not go into labour, she may be sent home until labour begins.
- Post partum unit of 21 beds, divided in 3 rooms, each with four beds, and the remaining comprising 2 or 1 bed rooms. After delivery, the patient and the newborn(s) stay in this unit until they are fully recovered. This unit is physically located together with the ante partum unit.
- Neonatal Intensive Care Unit (NICU). Newborns requiring specialised care are admitted to this unit together with their parents. It has 7 family rooms, as well as the special care rooms for the newborns.

The patients are sorted into their respective Diagnosis Related Groups (DRG), see Table 8. Each DRG group contains patients with similar diagnoses and/or operation codes. The DRG system is not only used for aggregation purposes, but also has an important role in determining the reimbursement a hospital obtains for the patients, as the DRG group classification of a patient is correlated to the type of treatment and its cost. Sweden uses a classification system called NordDRG-se<sup>3</sup> (SoS 2008b).

<b>DRG 370-375 = Patient care that led to child delivery</b>	<b>DRG 376-377= Patient care where the patient comes during childbed/ breast feedings period</b>
<b>370 Complicated caesarean section operation</b>	376 Illness during puerperal period without operations
<b>371 Uncomplicated caesarean section operation</b>	377 Illness during puerperal period with operations
<b>372 Vaginal delivery with complicity</b>	<b>DRG 382-384 = Patient care where the patient went home without childbirth, the patient is still pregnant</b>
<b>373 Vaginal delivery without complicity</b>	382 Cease labour pains
<b>374 Vaginal delivery with sterilisation</b>	383 Other ante-partum diagnosis with medical complications
<b>375 Vaginal delivery with other operations than sterilisation</b>	

Table 8 DRG groups and their description

The relation between the DRG groups and their patient flow is described in Figure 13.

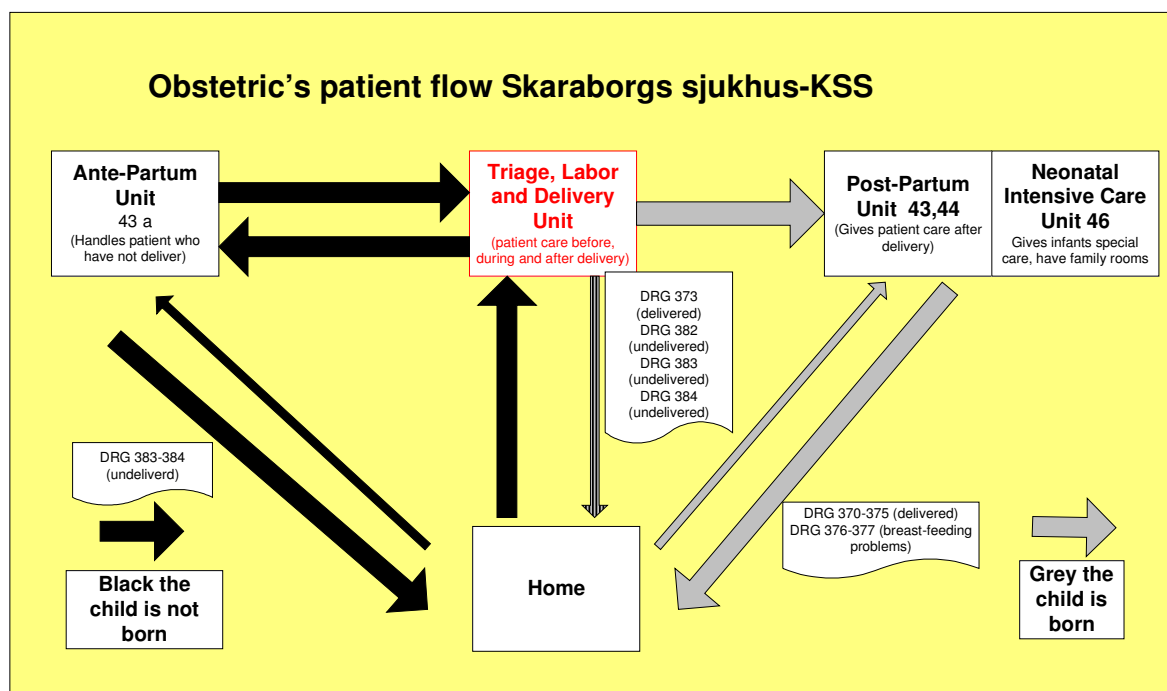


Figure 13. Description of the perinatal centre and the patient flow between the different units. Patient categories are identified by their DRG groups.

<sup>3</sup> NordDRG is a Nordic collaboration including, besides Sweden, also Finland, Norway, Denmark (until 2002), Iceland and Estonia. The suffix –se stands for Sweden.



The processes in the Perinatal centre are not supported efficiently enough by the maternity ward. There are several opportunities for improvement and problems have been identified.

1. The triage, labour and delivery unit is located on a different floor than the ante partum and post partum units; this results in a lot of patient transfers between the units and a poorer utilisation of personnel resources.
2. In a patient survey of 1999, a very low rating was given to the facilities. It revealed that the 4-bed rooms were the most unsatisfactory from a high service level perspective. The need for family rooms has increased since then.
3. The total number of childbirths varies considerably over 20-30 years, both nationally and regionally. There is also a high day to day variation, see Figure 14. The national variation since 1973 and the regional variation since 1990 are shown in Figure 15. (Seasonal variations over the year are not conclusive).

Considering these three issues, the process does not maintain an efficient use of the patients' rooms nor does it support the patients' demand of service.

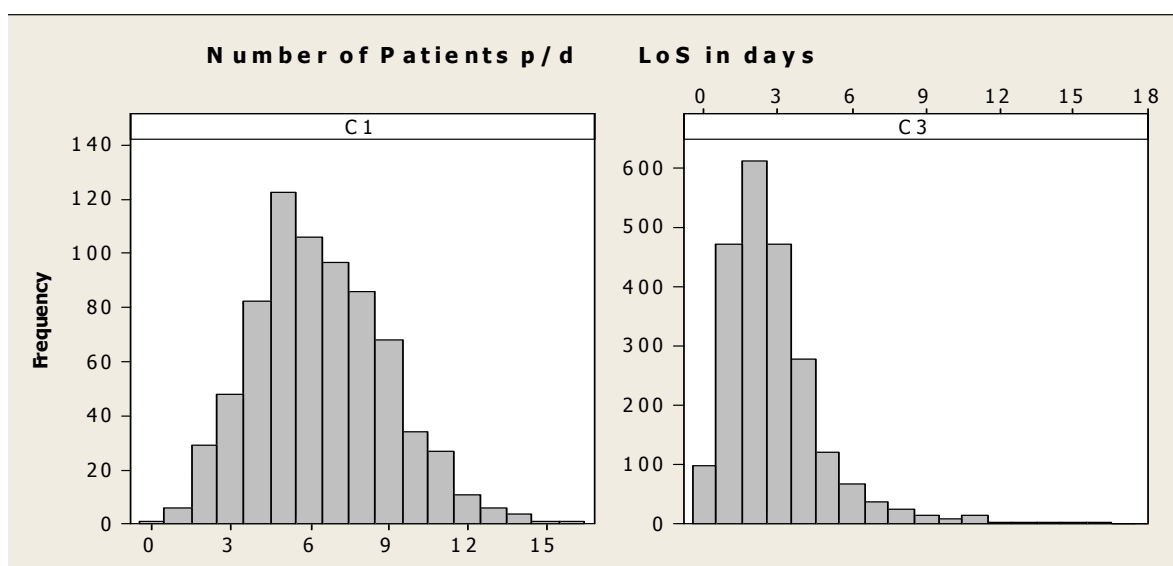


Figure 14. Arrival rate figures of patient's p/d for 2004-2005 and LoS figures for the same period

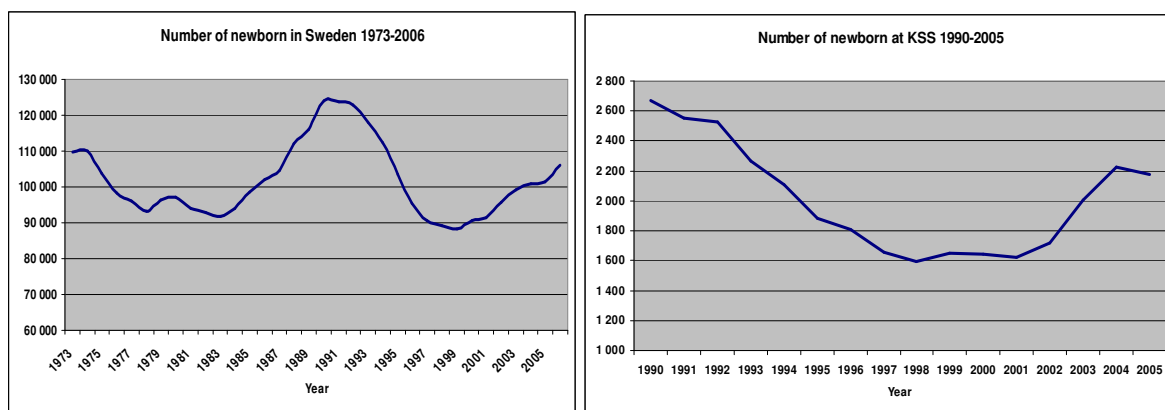
## 5.4 Project aims

The process of designing a new maternity ward started in 1998 and has been postponed on several occasions. One of the main reasons for these postponements has been the difficulty of presenting a final solution for the ward and how it would affect the rest of the unit. The unit management group had problems determining the number of rooms needed if a different process approach was implemented, difficulties estimating how a future increase

of birth rate figures would affect the ward, and how new policies would affect the need for family rooms. The main reason for these difficulties is the high level of variation the system encounters.

The project has several objectives that somewhat oppose each other:

- Higher service levels, letting more patients among DRG 370-375 have their own family room.
- A system that could withstand 20-30 years of birth-rate variations.
- A system that better supported the Perinatal unit's logistics, from both patient transport and personnel displacement.
- A system which has high, overall utilisation.



**Figure 15. Total number of births: in Sweden 1973 to 2006 and at SkaS 1990 to 2005**

The project group incorporated the head of the department, the head physicians of the clinics, an architect and the author. The project set-up was ideal, the group had access to all the resources they needed and the hospital's managers had prepared an information and data folder. Their interest, participation and collaboration were the prime conditions for a successful case study.

There are several published papers that deal with the configuration and process plans of maternity wards. However, their focus and scope is somewhat different. For example, (William 1998) presents a study conducted on the maternity process at Miami Valley Hospital. The women's hospital experienced capacity problems due to both new legislation that led to increasing LoS for the patients, and the steady increase in births. The solution

was found in better balancing the resources of the different areas, for example, expanding the triage area and reconfiguring other areas of the process. The same kind of solution is presented in Cochran and Bharti (2006). In their case, the scope of the problem was larger and included the entire obstetrics hospital. However, their solution addressed the same problem, to improve the overall utilisation of the system by improving the balance of the units. Their analysis and solution was found using a two step approach. Firstly, they made a queuing model to obtain a first approximation of the solution and, secondly, they completed the study by using DES. Both these projects had as their objective the higher utilisation of the overall system in order to achieve higher capacity. These aims were not the same as the objectives of the KSS project and their solution approach was also very different. The aim of William (1998) and Cochran and Bharti (2006) was to balance the system by changing labels on care rooms, which means that they added some resources from one part of the system and took it away from another part. However, this is not a robust solution as the system will maintain balance only if its conditions are stable, which is not the case, since crude birth rate variations tend to be high.

Variations in a healthcare system are difficult to deal with. One could try to address variations by using scheduling and different care approaches or policies, but this only partially applies in a maternity ward. It is impossible, in normal circumstances, to schedule when labour starts and, therefore, when patients arrive at the delivery unit. There are exceptions, for instance, it is possible to lower the mean and variation in the LoS distribution by implementing home visits for patients that leave the maternity ward earlier. This procedure was presented in (William 1998) and is being successfully used at KSS, reducing the total LoS with up to 36 per cent. Another successful approach, which highlights that improved medical care and system efficiency are not necessarily contradictory, originated at the maternity ward of the National Maternity Hospital in Dublin, Ireland (Thornton and Lilford 1994; O'Driscoll et al. 2003). This approach reflects the efforts achieved in reducing the length of labour experienced by women in the labour ward. The program has reduced the length of labour work for all categories of patients, to an exceptional limit of no more than 12 hours for first-time mothers. Other positive consequences of this program include healthier infants, mothers with higher self esteem, fewer caesarean deliveries and a lower rate of complications. These beneficial results have a positive effect on reducing LoS and variability.

However, these approaches do not solve all the variations in a maternity ward and they do not address the ones most relevant from a capacity point of view. The lesson is to minimise variation when possible and, when this is not possible, to build a system that is robust and able to deal with the variation (Benjamin et al. 1995; Ranjit 2001; Allen 2006). The objective is to determine the design most suitable for handling the system's variation over time, and while it might not be the optimal design at a particular stage it should be the best overall solution. Several papers address the use of the Taguchi approach in robust system design (Wild and Pignatiello 1991; Benjamin et al. 1995) and its combination with simulation based optimisation (Al-Aomar 2002; Kleijnen 2005). Most of the methodologies combine simple Taguchi techniques with more sophisticated Artificial Intelligence (AI) approaches. The combination of AI and DES for obtaining robust solutions is very relevant when we are dealing with a complex system with many design parameters and a large search space. Moreover, this is one of the research tasks that the Centre for Virtual Systems will undertake in the research project, OPTIMisation using Intelligent Simulation Tools (OPTIMIST) - Robust and Real-Time extensions, which starts in 2009. Nevertheless, this approach is not pertinent to our case. The relevant design parameters of our case are few and the presented approach of finding a robust solution therefore focuses on the combination of DES and Taguchi DoE.

The question is: How is a robust system for a maternity ward designed? The first case study, presented in chapter five, indicates the need to make more flexible use of available beds in other related units, so that the units can help each other when the need arises. The suggestion pointed out that the different units have patients with different variation patterns. The new "virtual unit" can be used to level out a particular patient group that peaks for a period of time, considering that it would be unlikely for all patient groups to peak at the same time. The main benefit would be that instead of over-dimensioning every unit, so that each one can cope with its own peak days, the "virtual unit", which is larger and more robust, can handle the total variation better than each unit for itself (Green and Nguyen 2001). A similar inherent demand variation, as the one observed in ward 84, is seen in the maternity ward at KSS, where the data range of admitted patients during a two year period was between zero to sixteen patients per day, see Figure 14. Three design suggestions were formed and a Taguchi analysis was conducted in order to measure the system's robustness.

## **5.5 Project development**

A typical DES project can incorporate a number of phases (Banks et al. 2001), for example, definition of problem and goal, model building, data acquisition, verification, validation, and so on. Naturally, these phases must not be treated as rigid sequential steps but should be seen as activities that are carried out in a concurrent way, and/or that demand more iteration before all the phases are properly defined and accomplished (Sadowski and Grabau 2000, Eldabi 2000). The availability of data, for instance, influences the selection of methods as well as the definition of a realistic goal. Furthermore, it is not unusual that an enhanced understanding of the system's behaviour changes the objectives of the project, giving birth to additional modelling and new experiments (Eldabi 2000). The problem definition and goal of the project have already been stated in previous sections, but it is important to realise that the different objectives will require a multi-criteria solution and evaluation in order to find the most robust system design.

This project was part of a larger context in as much as its results would affect other units. It therefore represented a first step, while the final solution for the entire obstetrics unit would be dealt with later. This meant that not all the future stakeholders were part of this first attempt, which implied that the solution, irrespective of its brilliance or not, was to be deferred until a full system approach could be determined.

### **5.5.1 Data acquisition and input analysis**

Most of the data required for this project was stored in the hospital's databases. The data contained information collected in 2004 and 2005, and was simply sorted according to different DRG groups, days, units, and so on. The only data collected manually during the project was the time that patients spent in the delivery rooms and the number of patients with illnesses or complications who were treated in ward 43 instead of the Neonatal unit. There is a plan that in the future this last patient category will be rescheduled to a new Neonatal unit with a broader patient mix (a future project, end of 2009).

The main identified distributions that drive the DES model are presented in Table 9. There are two distributions related to the triage and delivery unit. One is called Length of Labour (LoL) and refers to the time the patient is in one of the delivery/labour rooms where birth is given. The other distribution is the total time in the triage and delivery unit. This

distribution not only covers the time in the delivery rooms, but also the time before and after delivery, and whether the patient may use one of the auxiliary rooms in that unit. Another interesting aspect of the data is that most of the distributions fitted well and are easy ones to use. The Poisson distribution that models the behaviour of the arrival pattern is very suitable for our future analysis because it uses only the mean as the description parameter.

Days		External arrival rate per day (p/d),	
Saturday and Sunday		Poisson (5.531)	
Monday		Poisson (7.096)	
Remaining days		Poisson (6.729)	

DRG group	% of total	Total Length of Stay	Length of Labour (LoL) <i>(time in labour room and total time in triage-labour unit)</i>	
370	3.8	Lognormal (7.310, 5.180)	LoL –	
371	9.7	Poisson (4.069)	First time mothers: Weibull (8.796, 1.666)	
372	9.1	Poisson(3.939)	Others: Lognormal (5.778,2.784)	
373	64.6	Poisson(2.318)	Total time in Triage and Labour unit: Lognormal (14.049, 13.996)	
374, 375	1.1	Poisson(3.574)		
376, 377	1.2	Triangular(0,3,11)		
			Departure time	
382	2.9	37.5% Negative binomial (2, 034034)	DRG 370-377	DRG 382-384
383	5.6		Erlang_k(0.599,28)	Erlang_k(0.538,19)
384	2	62.5%Negative binomial (1, 0.39293)		

**Table 9 Identified distributions – input data analysis**

## 5.5.2 Model building, Verification and Validation

The model building phase proceeded in a relatively straight forward manner, mainly due to a patient flow that was easy to map and a software package which provided the possibility of building a model that is easy to change and reuse. The technical difficulties of the modelling were in 1) making the scheduling of patients, based on the dependency needs of their conditions, as pragmatic as in reality, and in 2) ensuring the model is easy to modify and reuse.

Animation and model output were used for verification and validation purposes. No major difficulties were encountered in the verification and validation process, mainly because of access to accurate data, the extended participation of system matter experts and no logical difficulties in the process model. The difficulties, in this case, concern whether one can

trust in the forecasting value of the model. The answer is somewhat ambiguous. One can say yes if one trusts the statistical birth rate prognoses (Statistics Sweden 2005, 2006a). However, it is difficult to know whether one can trust these prognoses, considering the constant yearly changes in the number of births, see Figure 15. During 2006 the number of births increased with 5 per cent compared to the 2004-2005 period. What is the purpose of the model if you cannot trust the future birth rates' prognoses? Well, the model was built mainly to check "what if" scenarios and, for that purpose, it is a suitable, verified, validated and accurate model.

Three model designs were built:

- The first model represents the obstetric unit of today. It is called the original model, see Figure 16.
- The second, represents a model of future concept one. The delivery unit is maintained, but the ante partum and post partum units are combined into one single ward. Patients with complications that were previously treated in Ward 43 are moved to the future Neonatal ward.
- The third model represents future concept two. The delivery unit is changed, the delivery and operating rooms are maintained, but the four auxiliary rooms (4 by 2 beds) are integrated into the maternity ward.

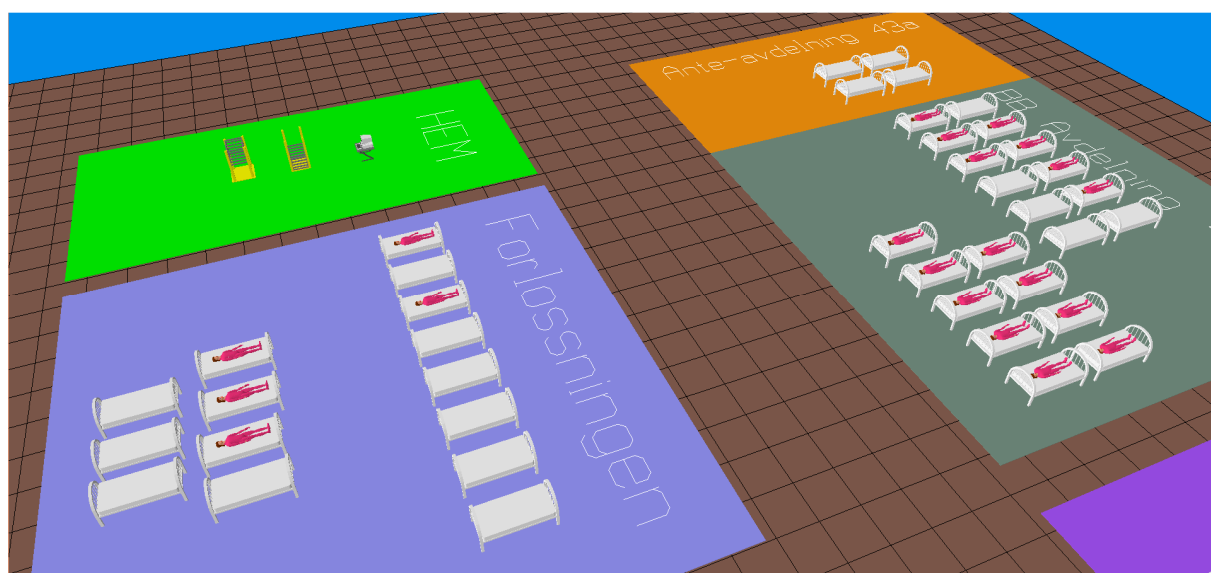


Figure 16. Original ward configuration

## 5.6 Experimentation and Analysis

Experiments and analyses are the truly valuable steps of a DES project and during this case study several experiments and analyses were conducted. Some of the experiments follow the common way of changing one design parameter at a time, while the main experiment follows the Taguchi DoE methodology (Ranjit 2001; Allen 2006).

The analysis and experimental phase is divided into two steps.

1. Identify best choice of system design (model 2 or 3) regarding both utilisation and service level, (used DoE).
2. Identify best compromise of number of rooms with regard to utilisation and service level, (used DoE results and additional analysis).

### 5.6.1 First set of experiments - DoE

Table 10 presents the identified system design and noise factors for the DoE study for which each parameter has three experiment levels. The experimental array combines both noise and design factors. This combined array, instead of the more traditional division of an inner array for design factors and an outer one for noise factors, has the advantage of reducing the total number of experimental runs (Sanchez et al. 1996; Roy 2001). In this case it provided the opportunity of using the L-9 and limiting to nine the number of experiment combinations. The factors and their corresponding levels were chosen in consensus with the project members.

Factors	Factor nature	Level 1	Level 2	Level 3
1 Birth rate	Noise	105%	115%	125%
2 LoS	Noise	90%	100%	110%
3 Nr. 2-beds rooms	Design	18	22	26
4 Auxiliary rooms, 2-bed places in each	Design	0 (model 3)	4 (model 2, config. 1)	8 (model 2, config. 2)

**Table 10 DoE factors and their levels**

In Table 10, factor 1 is birth rate, representing a noise factor with three levels indicating an increase of 5, 15 and 25 per cent over the 2004-2005 figures. The first level represents the level of the number of births for 2006. Level 2 indicates a 15% increase compared to the 2004-2005 figures, which is merely (on average) one extra patient per day. This equals the



variation levels one might find between two different months. Level 3 shows more of an increase but still not an unrealistic one; the same levels were experienced in the early nineteen nineties. The expected trend is that the number of births will increase in the region due to a higher population and its age structure (Statistics Sweden 2006b). This increase is considered to be correlated to an equivalent increase in the number of patients/visits from DRG 376-7 and 382-4. These DRG patients represent fewer than 12 per cent of the total number of patients arriving at the unit.

Factor 2 was selected to analyse how sensitive the system is to an increased average of the inherent variations in LoS. The factor's level was chosen with regard to the purpose of reducing the LoS, but also taking into account that a slight increase might result, for example, from more caesarean patients, something that would affect the overall LoS.

Factors 3 and 4 are related, and both are design factors. Factor 3 defines the total number of 2-bed rooms in the system (excluding delivery beds and operating rooms). Factor 4 defines how many of these rooms are used as auxiliary rooms (see second model configuration), which resembles the current situation. The auxiliary rooms are used when the patient arrives and for the period after the delivery until the patient is admitted into the maternity ward.

The model's performance responses of the system are the following:

- Occupancy level of the maternity ward's bed places (i.e. utilisation, normally referred to as occupancy level when bed utilisation is considered). When a room is used as a family room, (both parents stay) only one bed is considered to be occupied, lowering the occupancy level of the ward. The delivery beds are not included, but the model monitors whether all patients have access to a delivery room when the time comes to give birth.
- Service level is calculated as 100 per cent minus the percentage of patients among DRG 370-375 that share rooms with another patient during their stay in the maternity ward. This definition of service level was chosen by the management.

It should be remembered that a fraction of the patients are moved to a hypothetical future ward, and that the model uses full flexibility with regard to routing patients. This means

that when all the auxiliary rooms (in those cases the evaluated scenario/trial is based on a model with auxiliary rooms) are booked and the patient's labour has not begun, she would be scheduled to an available bed-place in the maternity ward.

The DoE experimental matrix chosen was the L-9 matrix which can accommodate four, three level factors and demands nine different trial conditions. As previously mentioned, DES models are stochastic, which means that the results from two simulations using the same trial conditions will differ. In order to obtain a correct measurement and picture of the future behaviour of the system, several runs on each trial condition are necessary. During these experiments, 10 samples on each trial condition were run. In addition, each sample was run for 135 days, with a warm-up period of 15 days.

Since the two result parameters are each other's opposite, for example, when utilisation increases the service level decreases, Overall Evaluation Criteria (OEC) were defined. The OEC contained both factors and was normalised so that the scale and Quality Characteristics (QC) of both factors had equal weight in the formula's results. By doing so it is possible to visualise the importance of the model configuration that gives the best performance disregarding the total number of rooms or the birth rate level. It must be remembered that the aim of this first step of experimentation and analysis is to find the best robust system approach among the proposed ones and not the optimum final configuration of resources. This is discussed during the second step of experimentation and analysis.

The DoE analysis uses a Signal to Noise (S/N) ratio to evaluate performance and improvement. A higher S/N ratio is always superior, irrespective of whether it is a minimisation or maximisation problem. Figure 17 clearly shows that the model without auxiliary rooms gives the best performance and is more robust in terms of providing a better service rate despite high utilisation levels. What it does not show is the number of rooms that is the best trade off between service and utilisation. If a 50-50 importance is given to utilisation versus service level, the optimum configuration is, according to the DoE levels, 22 rooms excluding auxiliary rooms. However, a 50-50 importance may not be realistic. The final evaluation is always a tradeoff where cost is a major decision parameter and managers at different levels of the organisation may have different opinions on the

necessity of a service level and the expected, future number of childbirths. This is why the analysis and experimentation phase proceeds to step 2.

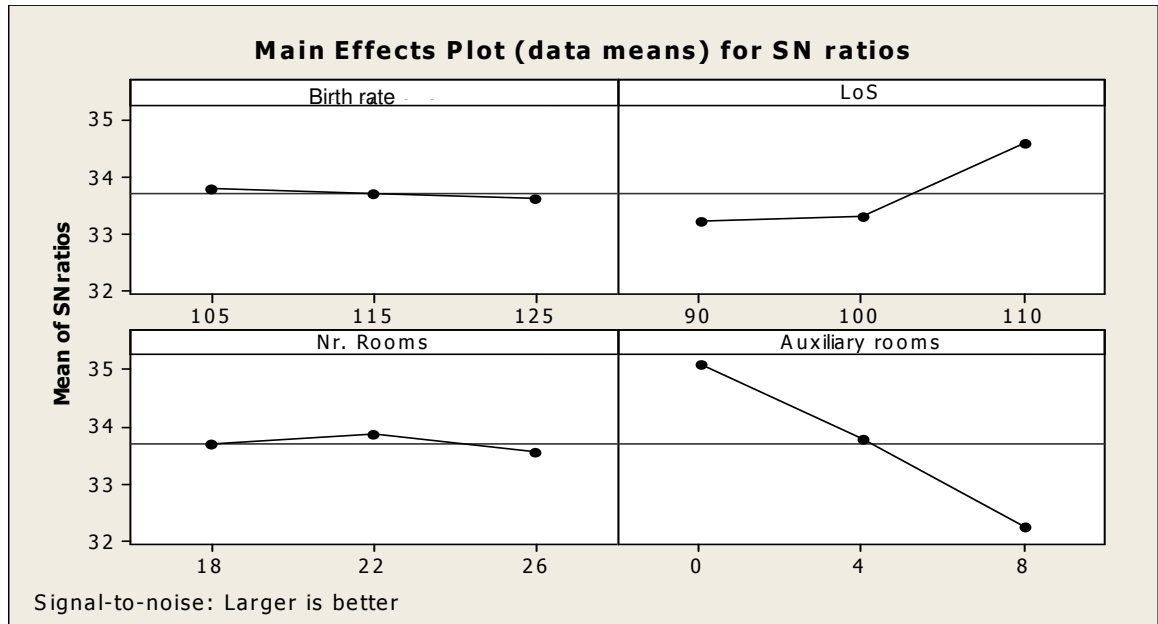
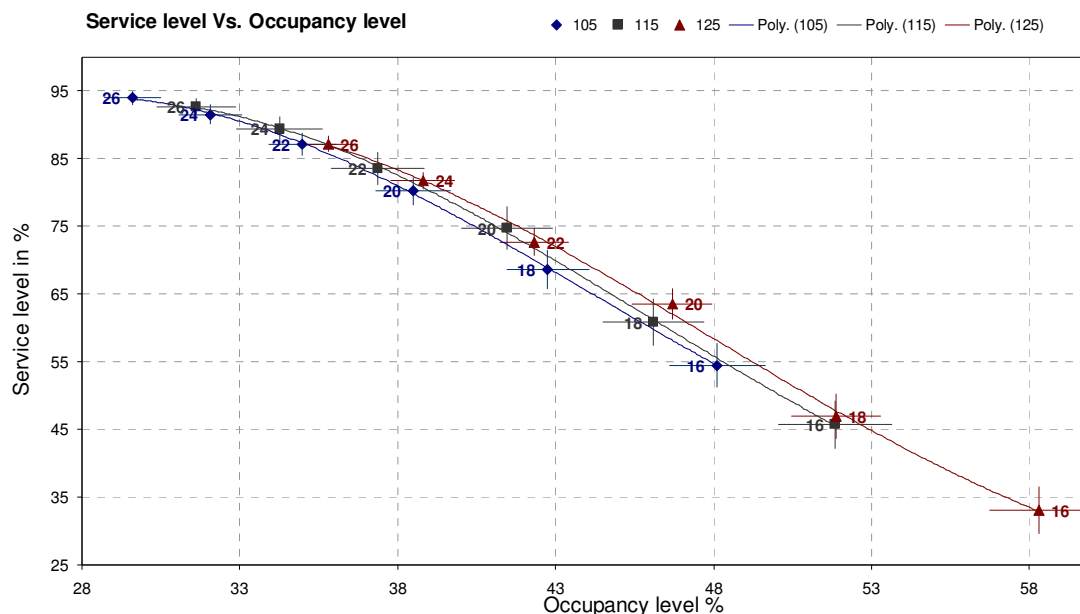


Figure 17. Main effects of factors and their SN ratios

### 5.6.2 Second set of experiments

The second step in the experimentation analysis phase focused on a more detailed study of model 3. The intention was to provide the hospital managers with a detailed graph that maps both service level and utilisation of different numbers of rooms and birth rate levels. This means that two of the earlier parameters were held constant, LoS was set at 100 per cent and the number of auxiliary rooms was set at zero. Each scenario was run for 135 days, with a warm-up period of 15 days and 10 replications. Six levels of rooms were analysed and three levels of birth rate.

The graph in Figure 18 shows the average results from the service level and occupancy level outputs for the different combinations of parameters. Each average result is represented by an output dot. The shape and colour of the dot shows whether the scenario represents 105-125 per cent of the birth rate, see the headings of Figure 18. The 95 per cent confidence interval (CI) of either the service level or occupancy level for every scenario can be seen by the length of the lines from the dot. The label on every dot shows the number of rooms used during the simulation scenario. Finally, three polynomial regression lines calculated from the outputs are shown.



**Figure 18. Results from model 3 analysis presenting both occupancy and service level**

The graph provides an overview of the service and occupancy levels of a future new maternity ward. It was of special interest to see whether it could stand a 25 per cent birth rate increase, both at the maternity ward and the delivery unit, which all configurations did. Moreover, the results illustrate that as a consequence of high inherent demand variation it is very difficult to attain both a high occupancy level and high service level. Compared to the utilisation figures for orthopaedic wards 83 and 84, the Maternity ward has much lower utilisation levels. At first glance it might seem that the occupancy level is far too low to be a responsible trade-off. The reason lies in the aim of attaining a service level far beyond that being offered in wards 83 and 84. It must be remembered that a service level of 50 per cent means that half of the patients in DRG 370-5 (88.3 per cent) can have access to a family room (2 beds) during their stay.

Beyond measuring whether a patient obtains access to a family room or not, a more important figure is whether she gets access to a bed without delay. Of interest in this matter is Green's discussion in (Green 2006). She presents interesting results about the relationship between the probability of delay by occupancy level and ward size, comparing her conclusions with empirical data from 148 obstetric units in New York State. Her findings indicate that the recommended maximum occupancy level, which is 75 per cent, is far too high for smaller maternity wards if they want to adequately guarantee a bed place without delay. For a ward of 35 to 40 bed places, a more adequate level is between 65-67

per cent occupancy. The delay time in obtaining a room was monitored during the simulation runs, but no patient had to wait more than two hours for a room in the maternity ward and all the patients had access to a delivery room when it was time to give birth. These results are supported by the work of Green in (2006).

Moreover, the results in Figure 18 show that the birth rate has a significant effect on the service level. This is especially clear when the number of rooms is reduced. In the eighteen room configuration, the service level decreases from 68.6 per cent to 46.9 per cent when the birth rate increases. Another observation is that the CI of the service level output increases considerably with a decrease in the number of rooms, while the CI for occupancy level is lower and has a lower rate of increase under the same conditions.

## ***5.7 Discussions and contributions to framework***

There are significant benefits in using DES for the purpose of designing robust solutions for healthcare systems. They provide insight into “what if” scenarios and a platform for discussion and decision making. The normal way of planning a new unit was based on heuristics and static values, but DES provides the opportunity of studying the variability in the system and therefore the possibility of dealing with that variability. Moreover, the combination of DoE analysis and DES was confirmed to be ideal for the evaluation of the system design without adding too much complexity to the analysis. This is especially true when the number of factors is reduced and Taguchi’s methodology is used. Tradeoffs are always made when a system is designed, but when the consequences of the decisions are known, they are easier to make.

The design suggestion of the maternity ward has several advantages. Firstly, it avoids small subsystems by not labelling rooms for different patient groups, thus creating a larger system instead of several smaller ones. In addition, this larger system is more robust to the inherent demand variations. Secondly, the two-bed room setup enables families to have their own room when the current demand permits it, without jeopardising the total ward capacity when the inherent demand increases. This simple configuration achieves both cost and space efficiency and is able to maintain a relatively high service level. The final size of the maternity ward has not been decided, but a first suggestion for the future obstetric unit with 18 two-bed rooms in the maternity ward has been presented by the architect (see

Figure 19). This suggestion is appealing because it would position both the delivery unit and the maternity ward on a common plane, resulting in an optimal logistical layout. It is clear however that a more permanent increase of childbirth numbers in the region would considerably affect the service level of the new obstetric unit, lowering it far under the aims of this project. Statistical prognoses (Statistics Sweden 2005, 2006a) indicate an increase of 11 per cent until 2024, followed by just a slight decrease in the years thereafter, but still showing higher childbirth numbers than the current figures.

On the other hand, considering that the occupancy levels of an 18 room maternity ward fluctuate between  $51.9 \pm 1.4$  with a 95 per cent CI and a birth rate of 125 per cent over the 2004-05 figures, it could be suggested that a larger ward, from a bed access point of view, is not necessary. Would single rooms be a solution? This was discussed. However, single rooms mean twice as many lavatories, and because they are considered family rooms, they would need to be somewhat larger than normal one-bed rooms to allow for a small extra bed. This option failed due to the lack of space.

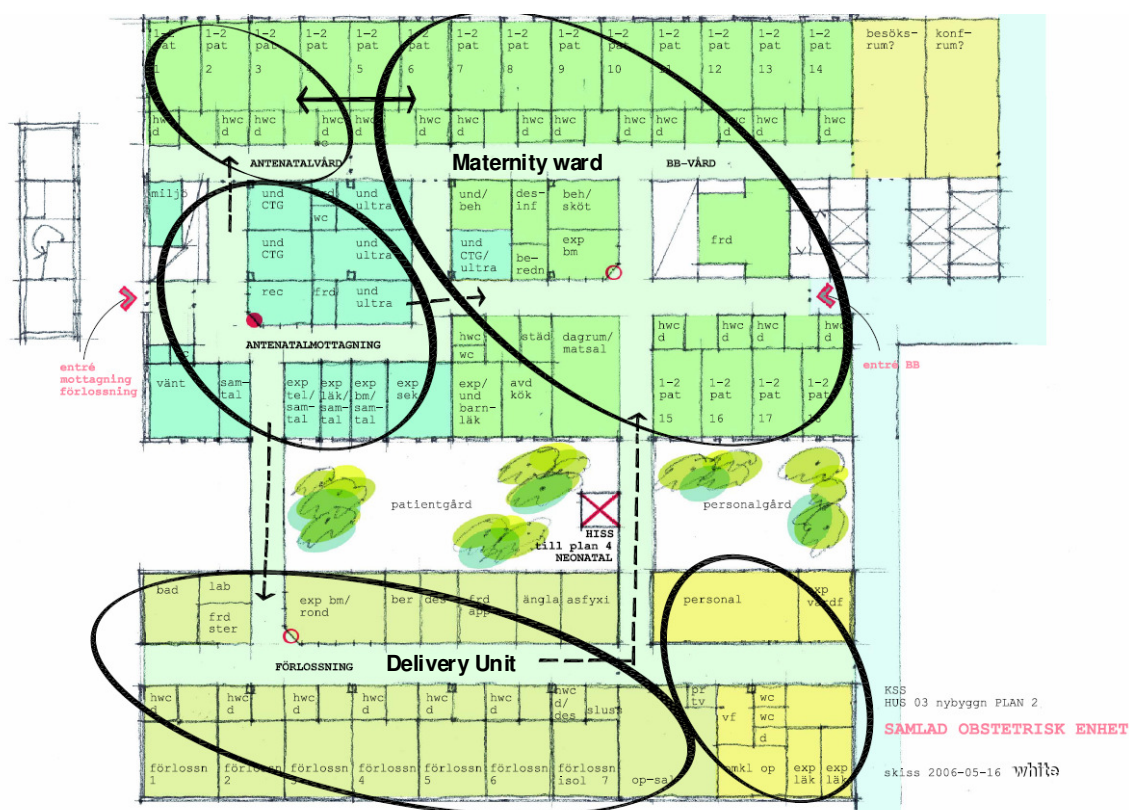


Figure 19 Suggestion for future Perinatal centre with an 18 room maternity ward.

The results give the head of the department the possibility to evaluate the different options and consequences of their decisions. The work with the obstetric unit now focuses on a fully integrated Perinatal centre, which means that the Neonatal intensive care unit and its family rooms will be integrated with the maternity ward to form a single unit. The objective is to gain the advantages proposed in this work, namely, to create a larger unit by integration and making use of family rooms as a general room type. A service level target of 80 per cent was finally agreed on. The construction of the new, intended Perinatal centre will start in January 2010 and be fully operational the following year. These facilities will represent a robust system with the ability to better handle inherent variations and the future service expectations of the population.

Nevertheless, this project confirmed earlier findings regarding robust system design and, at the same time, it corroborated a mindset change among the healthcare professionals. This new attitude provided a place for better understanding and collaboration, as well as an insight into how close and iterative conducting the work should be in order to gain acceptance and be successful. Moreover, the results gave the head of the Department of Obstetrics and Gynaecology the data background that supported submitting a legitimate request, based on facts, to the hospital's management body.

Its main contribution to the framework of this thesis, besides the system design lessons we learnt, lies in the nature of the operational conditions of the system, where a number of sources of variations explicitly set the design of the ward as a delicate task. The two objectives, obtaining both a satisfactory service level and occupancy level, turned out to be extremely difficult to achieve unless a robust system design was adopted. Empirical knowledge regarding the significance of taking into consideration system variations serves as a valuable lesson when we proceed to consider the importance of another source of variation in inpatient wards, namely patient dependency variation. Until now, with regard to system design, this source of variation has been neglected. This is unfortunate, since it is very important in order to develop an inpatient ward system or structure that supports both a satisfactory staff related service level for the patients and a satisfactory utilisation level of the staff.

## **5.8 Summary**

This chapter presents the second of three case studies. The project was carried out in a maternity ward at KSS. There were several study objectives. Firstly, to solve the problem at hand, which involved designing a new maternity ward that could both have a high service level and, at the same time, an occupancy rate equivalent to other comparable systems. Secondly, to confirm findings made in previous studies where robust system design was closely linked to more flexible approaches, including ways of making the ward “bigger”. Finally, to gain understanding of non-quantifiable issues related, for example, to communication, organisational barriers, data access and modelling challenges. There was still much to learn from the healthcare domain and much to improve in terms of better communication and collaboration.

The case study took place at a maternity ward in need of restructuring. The ward is part of the obstetric unit and, together with the delivery ward and the Neonatal intensive care unit, forms the Perinatal centre. Its facilities are based on a 33 year old layout that has been criticised for many years for its large four-patient rooms and poor logistical solutions. Despite the discontent, the development of the new maternity ward was postponed on several occasions due to a lack of groundwork on which to base a new design decision. DES was expected to remedy that situation.

Taguchi’s robust design methodology with a reduced DoE L-9 matrix was used together with DES to evaluate three design suggestions. The most robust design concept was then identified, and a second set of experiments evaluated, in more detail, different levels of both design and noise factors. The aim was to find the optimal trade-off between service and occupancy level. The final decision on the number of two-bed rooms is still pending. Occupancy level figures indicate an 18 room solution, but the aim of providing a high service level by offering family rooms may incline the decision toward an additional number of rooms. The final decision will be made after the future neonatal ward has been designed. The number of rooms in this unit will affect whether there is enough space for more rooms than the 18 initially suggested.

One of the important conclusions of the project was the maturity of the healthcare professionals and their organisation with regard to expectations and collaboration in a DES



project. The work done by the healthcare provider and the University in the time between the first project and this one has led to gratifying outcomes.

Finally, two design suggestions were made:

- Avoid small subsystems by not labelling rooms for different patient groups, leads to the creation of a larger system instead of several smaller ones. This larger system is more robust to the inherent demand variations.
- Using a two-bed room setup enables families to have their own room when the current demand permits it, without jeopardising the total ward capacity when the inherent demand increases. This simple configuration achieves both cost and space efficiency and is able to maintain a relatively high level of service.

This project confirmed earlier findings regarding robust system design and, at the same time, corroborated a mindset change among the healthcare professionals. This new attitude led to better understanding and collaboration, as well as an insight into how close and iterative conducting the work should be in order for it to gain acceptance and be successful. In addition, the results gave the heads of the Department of Obstetrics and Gynaecology the possibility to evaluate the different options and consequences of their decisions. Moreover, it gave them the data background that supported submitting a legitimate request, based on facts, to the Hospital's Management body.

Furthermore, it corroborated the importance of taking into consideration system variations and gave empirical evidence when we proceed to consider the significance of another source of variation in inpatient wards, namely patient dependency variation. Its importance, in order to develop an inpatient ward system or structure that supports both a satisfactory staff related service level for the patients and a satisfactory utilisation level of the staff, cannot be neglected.

## **6 Chapter six: Rehab case study – definition of modelling methodology**

### ***6.1 Introduction***

Solutions are the offspring of problems and needs, and they require reflection, time and effort to be elaborated. This is particularly true of the methodology presented in chapters six and seven. Although the main points were intuitive and not transcendental in any aspect, it took time and further studies before the whole methodology was formulated in print and documented. When issues are scrutinized there is an imperative need to look at them from different viewpoints, especially when the matter requires an interdisciplinary approach.

This chapter informs the reader of a case study conducted at a rehabilitation ward that used Beakta® for the measurement of the workload. The findings from this case study form the basis of a discussion about how to model the fluctuating dependency development of inpatients. It presents a modelling methodology for stochastic patient dependency variations, which has been applied in a DES model of a rehabilitation ward. The step by step presentation of the modelling methodology includes best practices and a detailed analysis description. This includes a discussion centred on the pros and cons of the PCS's activity study as data input for a DES model and relates it to the adequate level of the detail of the model. Moreover, it describes some of the organisational difficulties encountered.

### ***6.2 Development of framework***

The need and the foundations of the modelling methodology have been identified through the research steps previously described in the thesis. An understanding of how the different sources of variation affect inpatient systems has emerged through the ongoing literature review and the empirical experience gained from the two earlier case studies.

The literature review presented the need for the robust design of systems in order to cope with variation while maintaining an acceptable level of system efficiency. It highlights that a two step approach should be used, 1) reduce the preventable variations in the system and 2) design systems that are inherently robust. Moreover, it identifies that the design of efficient healthcare systems needs to address the different sources of variation affecting the

system, of which patient dependency variations and their effect on staffing requirements have previously been neglected. Additionally, the literature review gives the background of different tools currently used to define staffing levels at healthcare units, where PCSs, and Beakta® in particular, are identified as a methodology and tool with the relevant structure and data to complement DES in the objective of finding a *modelling approach that is able to help in the design, planing and evaluation of nurse staffing requirements in inpatient wards*.

Furthermore, the empirical evidence for the need of addressing the effects of variability as a major system design factor was presented in chapters four and five. Chapter four identified how an efficient running of wards 83 and 84 depended on several design and management rules:

- Firstly, the need to lower the preventable system variation through a more constant scheduling of inpatients' surgery.
- Secondly, the importance of having a more flexible routing between adjacent wards.
- Thirdly, that the size of the wards matters with regards to the efficient use of resources in systems affected by high levels of variation.

Moreover, two direct contributions from this exploratory case study are, firstly, the exposure of one of the prime concerns of healthcare organisations, namely, the need to determine the right staffing levels. This identified need made an important contribution in establishing the direction of the research work. Secondly, it provided a realistic scenario or benchmark which illustrated the values of the future modelling methodology including the lessons learnt from the earlier case studies.

Chapter five confirms how different sources of variability affect the efficiency of an inpatient system. It emphasises the need of robust system design in order to improve system efficiency. The findings in the chapter therefore corroborate the need to be able to model the different sources of variability, which includes the variability arising from patients' dependency fluctuation, in order to efficiently determine the adequate staffing level. Moreover, it confirmed that system flexibility is necessary and suggests system robustness through simplified routing and a new room configuration. The lessons learnt by

the first two case studies have contributed significantly to identifying the need and the context in which the proposed modelling methodology will play an important role in facilitating *the design, planning and evaluation of nurse staffing requirements in inpatient wards*.

The contribution of the subsequent chapters is highly relevant for the healthcare simulation community and for system design in particular. It presents a modelling approach for patients' dependency variation that is able to quantify the workload fluctuation of inpatient systems. For the first time, users of this methodology will see how the effects of patients' dependency fluctuation affect the systems' total workload and thus the staffing requirements. Through being able to model this source of variability, the system designer is able to view the outcome of how different system designs or management strategies affect the staffing needs. Moreover, the methodology gives a clear view of a future system's (or how a current system will change with a change in patient mix) workload variation and consequently its staffing needs. The modelling methodology also supports the possibility of making short term schedules and future forecasts, although this will, however, require additional work in techniques for patient grouping.

### **6.3 Modelling Approach**

How is a patient's dependency variation captured, quantified and modelled for use in a DES model? The aim in answering this question is to be able to evaluate and plan the need of personnel requirements in inpatient wards, as well as achieve the overall goal of designing robust inpatient systems which can handle inherent variability and provide optimal resource utilisation without jeopardising quality of care or personnel work satisfaction.

This section presents a modelling methodology for patients' dependency variations, which has been applied in a DES model of a rehabilitation ward. The step by step presentation of the modelling methodology includes a theoretical background, best practices and a detailed data analysis. The following steps and activities have been identified and are presented (illustrated by Figure 20) for orientation purposes:

1. **Acquire PCS data:** This first step requires an understanding of the different kinds of data, and their use, stored in the PCS's database. Most of the information is described in chapter four. However, throughout the chapter, this topic is put in a modelling context and exemplified, which should facilitate the reader's understanding.
2. **Aggregation of patients into patient modelling groups:** The second step is discussed in both this chapter and the following one. The aim is to make a relevant grouping of patients with similar characteristics. The task is not as straight forward as could be initially expected. The aggregations chosen are based on the information available and on established healthcare practice.
3. **Calculate dependency transition matrices for each group:** This is the main step of the proposed solution and the most extensive in terms of analysis and effort. The modelling methodology uses a Markov Chain Monte Carlo structure in order to simulate a patient's dependency variation. The inhomogeneous transition matrices used are based on PCS historical data and model the patient dependency transitions as state transitions. This step includes the adaptation and analysis of transition data before it is stored in the transition matrices.
4. **Modelling indicators and the determinant probabilities:** Before a patient's dependency level is translated into a workload figure, the combination of determinants for each indicator need to be fixed. This step uses indicator and determinant data from the PCS database and the already defined dependency level (see the former step) to establish the determinant values through a stochastic selection methodology.
5. **Use the activity study data:** After setting a patient's current dependency and determinant values, they are translated to workload equivalents. The activity study defines the workload equivalents for each determinant. The same study contains information about the percentage and distribution of the workload equivalents among the different staff categories over the 24 hours of the day.
6. **Implement the dependency level matrices and workload features into the DES model:** This step not only involves implementing the above mentioned data structures and stochastic procedures, it also deals with finalising the simulation model by adding the internal, arrival time distribution, the patient mix for the ward,

and additional logistics, such as arrival and discharge logistics. A novel discharge adaptation procedure is also described.

7. **Collect patient dependency and indicator variation data for validation purposes:** This last step identifies outputs from the simulation model, generated by the new modelling methodology, which need to be validated. The step describes how the validation can be carried out, and what measures can be taken if the result of the validation process is negative. This information is presented in chapter 7.

These are the main steps of the methodology. It should be remembered that using this proposed modelling methodology does not mean that the normal steps of a DES project methodology are excluded (Banks et al. 2001) or those of a healthcare adapted methodology, such as MAIPU presented by Eldabi (2000). The modelling and analysis activities of this methodology fit well within the previously suggested steps of a DES project. The intention is to help the simulation analyst solve modelling and data acquisition issues within each step, so that the patients' dependency variation can be successfully modelled. Before describing them in detail, some background information about the system's behaviour and the relation of the proposed method to other existing solutions could be beneficial.

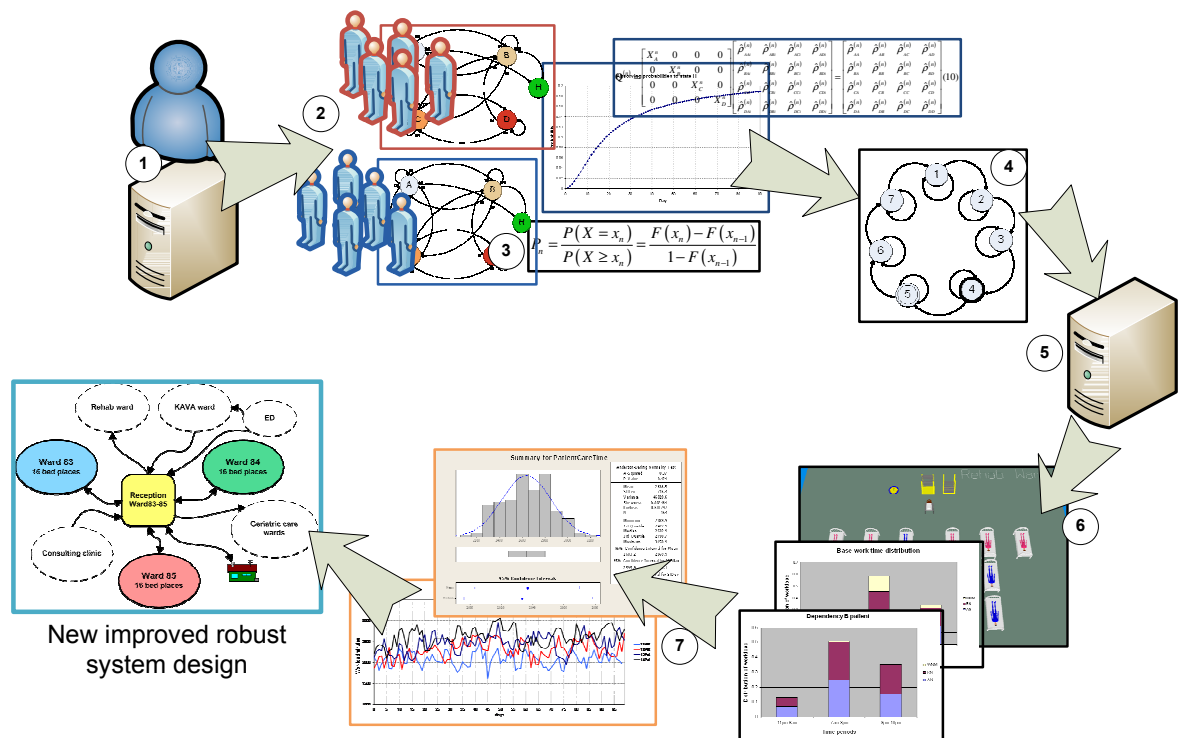


Figure 20. Modelling steps of suggested methodology leading to the analysis and design of a new improved system.

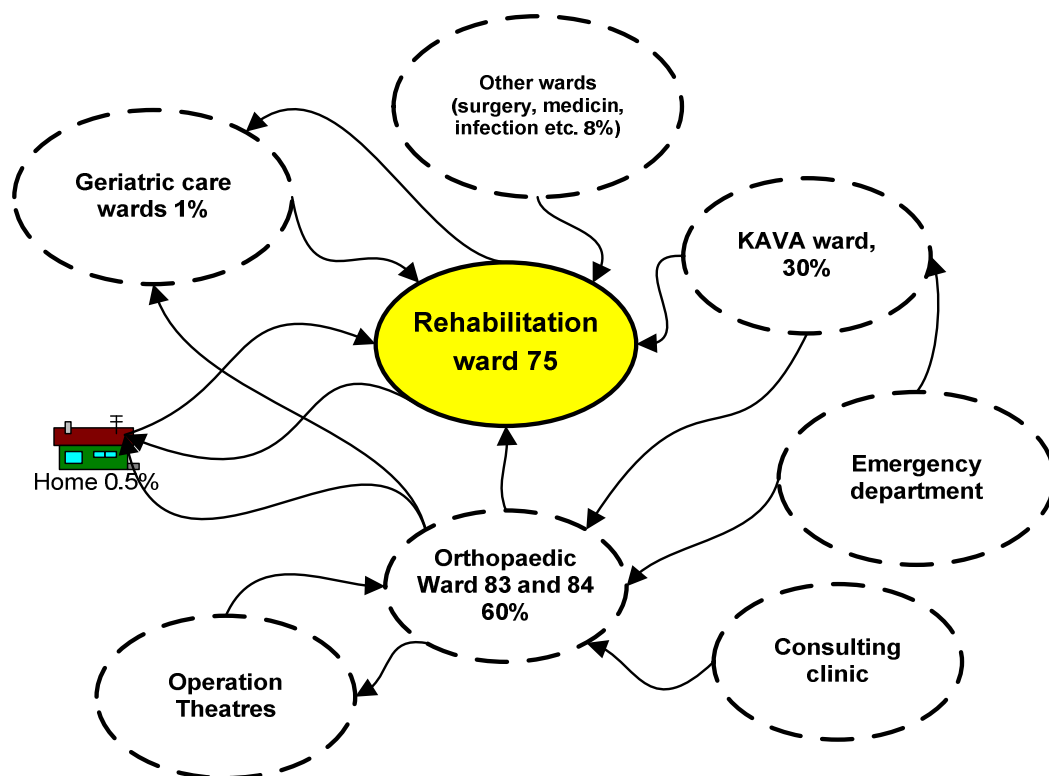
## **6.4 The rehabilitation ward at SkaS**

At the beginning of 2004, after successfully completing their first simulation study in two orthopaedic wards (see chapter four), the hospital management at KSS wanted to use simulation to address the question of adequate staffing levels. That aim had not materialised in their first simulation project, but was still a desirable goal.

There are four units at KSS that had implemented Beakta®. One of these units includes rehabilitation wards 75 and 76. They were considered the preferred choice because of their experience of PCSs, the proximity of the process to the orthopaedic wards already studied and the awareness of suitable personnel resources. Ward 75 was chosen of the two rehabilitation wards. The main reasons included the patient group size of the ward, which would provide better statistical estimations, that it only had elective patients, which was expected to give a more predictable patient workload, and it had a less stochastic environment, therefore ensuring, with greater certainty, that it would be a more representative activity study. The term “representative” refers to an activity study conducted over a representative period of time which therefore reflects the normal workload of the ward and accurate time measurements of the staff’s activities. Figure 21 shows the patient flows among units related to ward 75 and their contribution in percentage to the patient groups that obtain rehabilitation in ward 75.

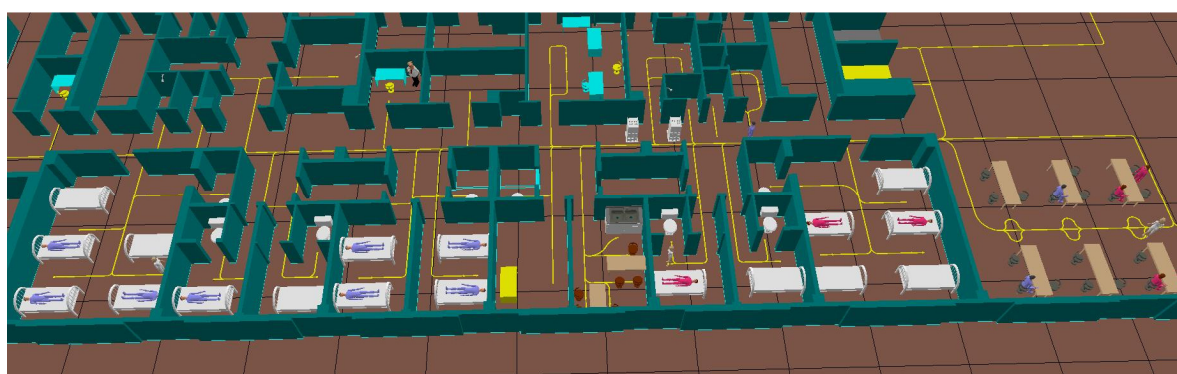
Several papers address the use and development of PCSs for rehabilitation units (Gender 1989; Dunbar and Diehl 1995; Sarnecki et al. 1998). The aim of the works presented in these articles is twofold. Firstly, the use of PCSs to calculate the right staff configuration and, secondly, the use of PCSs to monitor nursing care interventions in order to improve and facilitate the rehabilitation of patients through an adequate mix of staff and activity planning. It is argued that PCSs are inadequate with regard to achieving this last aim (Nelson et al. 2007) and there seems to be a consensus about using two types of systems, one for staff planning and another for rehabilitation planning. What can be stated is that there is not a total correspondence between a patient’s dependency level and his/her recovery progress. A patient might have recovered from the medical intervention or illness, but still requires a lot of support and help for other reasons than the ones related to the current hospital stay. The use of PCS in our case only focuses on staff planning and not on

determining the rehabilitation procedures. This is made clear by the fact that the physiotherapists working in ward 75 were not included in the activity study.



**Figure 21. The flows among units related to ward 75 and their patient contribution in percentage**

Ward 75 consists of 14 bed places distributed into two 3-bed (2 extra beds can be fitted in), two 2-bed, and four 1-bed rooms. A screenshot of a detailed simulation model of the ward is presented in Figure 22. The rooms are not shared among patients of opposite gender, but with flexible planning the different needs of patients can be met.



**Figure 22. Ward 75's room configuration and layout.**

The project was led by a member of the hospital management staff who worked directly under the hospital's executive director. Besides the management staff member, the project



group included the person responsible for the rehabilitation unit and the author. From a research point of view it seemed a promising start. However, healthcare organisations have a complex and difficult management and administration structure, as previously mentioned (CHSA 2002, pp. 50-54), which led to an intricate and frustrating problem concerning data acquisition.

#### **6.4.1 Organisational and juridical obstacles**

In Sweden, patient data is confidential and protected by a series of laws (changes in 2008 have alleviated the original problem). In order to obtain access to patient related data, personnel and researchers must sign a juridical agreement of confidentiality and only use the data for the specified purpose. Despite a valid agreement, the IT department's staff, who are in charge of the SQL database used by Beakta®, had very strong opinions about letting a person not employed by the hospital have access to the data. After six months and three meetings, which included senior management, they finally agreed to make a de-identified copy of the database available to the project. It seemed that the staff did not fully understand the laws and therefore did not know whether the juridical agreement was valid or not. However, it was interesting that despite the involvement of top management in the project access to the data was refused by the IT department's staff. This problem illustrated some of the difficulties in having several levels of command in an organisation and unanswered questions of responsibility.

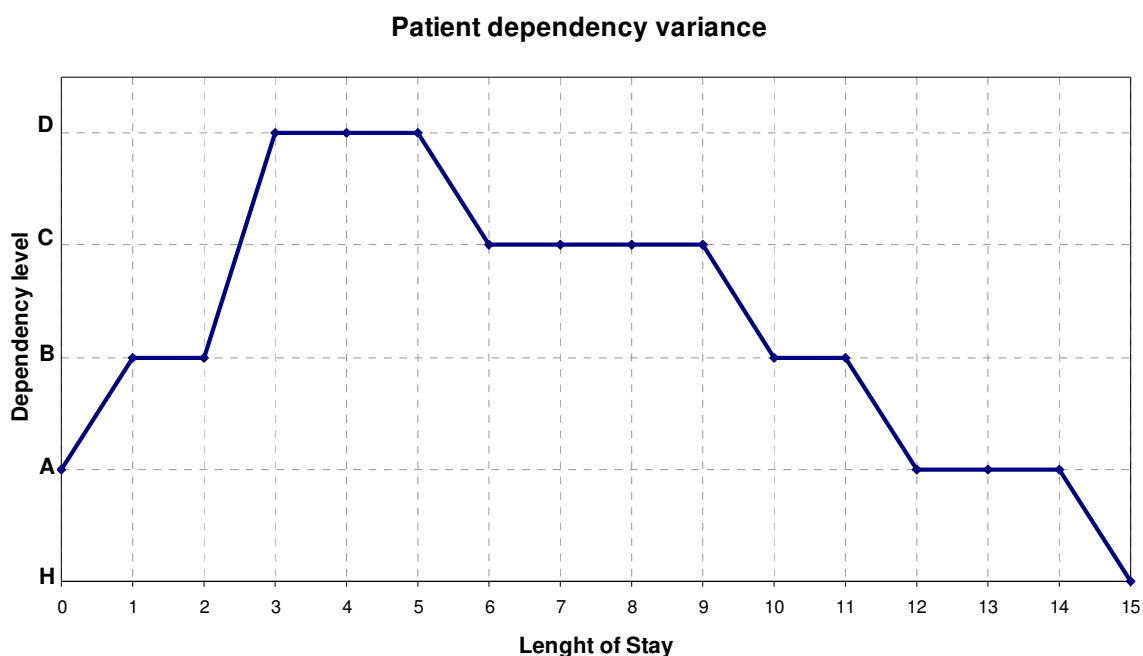
#### **6.4.2 Mapping and modelling difficulties**

While the problem of access to the patient data was being solved, the different tasks and working procedures of the staff, monitored by Beakta®, was mapped. The staff comprises three groups: the Ward's Nurse Manager (WNM), Registered Nurses (RNs) and Auxiliary Nurses (ANs). The Beakta® activity study monitors these three groups and quantifies their work according to *direct patient time, indirect patient time, personal time and administrative time*. The activity study also documents the proportion of time they use in these different areas and when (during the day) they perform the different activities. Although many activities were reasonably easy to map and schedule during a normal working day, the major difficulty emerged from the human ability to perform several tasks simultaneously and from the fact that many activities are shared among RNs and ANs.

During this early phase, and because the data was not available, an initial, detailed model was built in order to identify modelling challenges and difficulties. The development of this model required a high level of programming skills and time. One of the reasons for the additional modelling difficulties was that the software used, Quest®, was mainly developed for the modelling and simulation of industrial systems and therefore not adapted to the modelling requirements of the healthcare sector. At that time the model was not verified and validated but served as a benchmark for the future modelling approach.

### 6.4.3 Dependency variance modelling today

The patients' dependency variance, or "care pattern of need", over time refers to how the patients' need of nursing care develops during their stay in the ward. This dependency variance generates a nursing workload profile that is taken into account when planning the right amount of personnel, see Figure 23.



**Figure 23. Dependency profile of a specific patient, dependency H represents leaving the system**

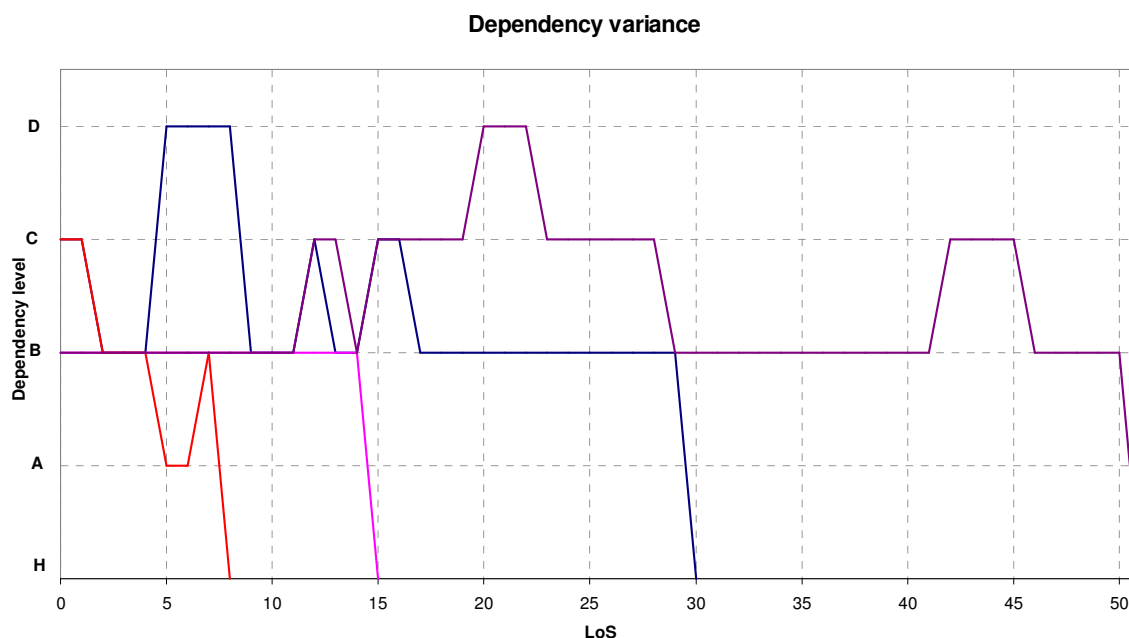
Dependency variance modelling is seldom (or ever) done with stochastic behaviour. Adan and Vissers (2002), for example, describe an integer programming, simulation model for patient mix optimisation that takes into consideration a predefined profile for specific patient groups. Warner (2006) employs a similar approach when he predetermines demand based care patterns using a forecast of demand application. In order to improve the

accuracy, Warner does suggest that several types of predetermined, demand based care patterns should be established for different diagnoses, ages, and so on. The aim is to aggregate the dependency profiles of patients into a manageable number of predefined profiles that are used for planning purposes. Each profile would, in advance, include the predetermined LoS of the patient and the amount of work the patient will generate during his/her stay. This procedure might be considered a necessary trade-off in order to define an optimal universal solution (Adan and Visser 2002) or to make a forecast in an uncertain domain (Warner 2006), but it does not correctly represent the system's fluctuating demand of resources and patient requirements.

A major deterioration of the DES results would occur if an average, standardised LoS for the patients was predetermined. This parameter is one of the most important stochastic parameters in defining the behaviour of an inpatient ward and its occupancy level (Green and Nguyen 2001; Marshall et al. 2005). LoS is highly stochastic in its nature for numerous reasons. Chapter 4 and 5 present, among others, the negative consequences of increased LoS in an orthopaedic trauma ward due to discharge difficulties, and the positive effects of decreasing LoS in a maternity ward due to implementing a home visit by a midwife. The reasons for a particular patient's LoS are more complex than a simple correlation to age and/or diagnosis. It includes multiple factors which are difficult and sometimes impossible to perceive. It would be wrong to expect that a static predetermined LoS profile would correctly forecast a patient's dependency behaviour. For example, Figure 24 illustrates the dependency variance and LoS of rehabilitation patients, in ward 75, with the same main diagnosis and DRG (they are categorised according to the NordDRG-se of 2003). The differences in LoS are striking! This distribution is not necessarily representative of other healthcare sector areas, but it illustrates how vulnerable and misleading a static LoS or predetermined pattern of dependency variance could be.

Related to this issue is the matter of combining patients into patient groups. PCSs normally group patients according to their respective DRG or their main diagnosis. Which of these is preferable? There is no categorical answer to this question. In Sweden, the healthcare sector uses a classification system called NordDRG-se (SoS 2008b) which groups patients who are clinically similar (diagnosis and treatment). The NordDRG system takes the patients' gender, age, main diagnosis and bi-diagnoses into consideration. This

classification system is in constant development and updated every year. Among the most important changes, considering the current discussion, are the new rehabilitation codes introduced in 2008, the so-called Nordic ASSEsment Score codes (NASS) (SoS 2008b). These codes were truly needed, considering the wide distribution among patients with the same NordDRG observed in this case study, which used the NordDRG system of 2003. This dispersion, observed in 2004, made using a patient's main diagnosis a better choice.



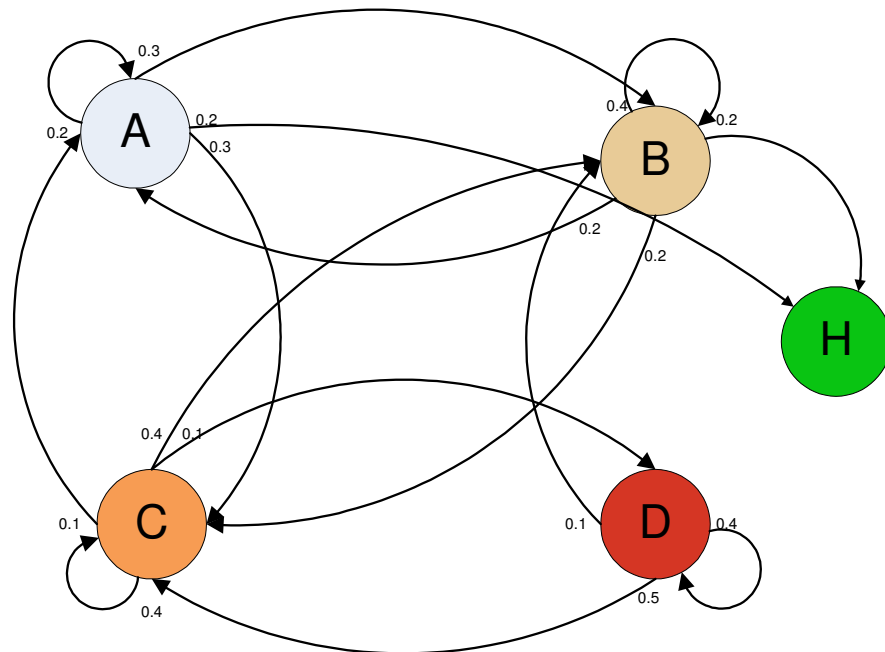
**Figure 24. Four patients' dependency variance and LoS**

Despite the choice of smaller groupings (DRG codes normally group numerous diagnoses), there was still a significant difference among patients. According to Per Sjöli, who works within the KKP project at KSS hospital, a better estimation of the patients' true dependency is made not only by considering the main diagnosis but also the number of different bi-diagnoses (co-morbidity) the patient might have (this is included in the changes introduced by the new NordDRG codes). The problem with such a decomposition to smaller patient groups is that it demands a larger set of data. Therefore, even if the optimum is to breakdown patient groups from DRG to diagnosis to combined main and bi-diagnosis, and take into consideration age, gender, and so on, it is not always feasible because the statistic sample this would result in could be too small. There are, however, techniques to minimise some of the negative effects of insufficient data points, which are presented in section 6.3.4 of this chapter.

#### 6.4.4 Stochastic modelling of dependency and LoS

An intuitive approach, when modelling dependency variations and LoS, is regarding patients' transitions between the dependency levels A to D as transitions between different states. The likelihood of changing from one state to another, and by doing so generating a different workload requirement, is governed by a probability. These transition probabilities, in their turn, are calculated from the observed dependency data of the patient's group.

Modelling the patients' dependency behaviour would consist of five states,  $S = \{A, B, C, D, H\}$ , where A to D are the dependency states used by the PCS, while state H (Home) represents the patient leaving the system. Figure 25 presents the possible movements between states of a predefined patient group on a given day. Every arrow represents a possible transition for a patient either staying in his/her current state or moving to another. The transition probabilities are shown by the figures connected to the arrow. This modelling approach models the patients' dependency transitions, not those of the system. Several patients, each following his/her own stochastic dependency transition, would represent the whole system's workload variation.



**Figure 25: Illustrates the states a patient can be in and the probability of transition between these states**

The patients' transition model can be described by a Markov process with one absorbing state. The process is absorbing because the patients entering state H do not leave that state

(from a modelling perspective). Irrespective of their starting state, they will all eventually reach state H and stay there. The modelling approach runs under the assumption that it obeys the Markov property. The Markov property states that the probability distribution of the state at time  $t + 1$  depends only upon the present state and not on any past states, leading to the following definition that can be found in (Winston 1994) and expressed below:

A discrete-time stochastic process is a Markov chain if, for  $t=0,1,2,\dots$  and all states,

$$P(\mathbf{X}_{t+1} = i_{t+1} | \mathbf{X} = i_t, \mathbf{X}_{t-1} = i_{t-1}, \dots, \mathbf{X}_1 = i_1, \mathbf{X}_0 = i_0) = P(\mathbf{X}_{t+1} = i_{t+1} | \mathbf{X}_t = i_t)$$

A patient starts in one of these states {A, B, C, D} and moves successively from one state to another until the patient reaches H where he/she remains. The starting state is decided by an initial probability distribution which defines the state the patient is in by time 0 and during his/her first 24 hours in the ward. The length of this time period depends on how often an evaluation of the patient's dependency is done; in our case, once per day counting from midday to midday. For each day, a random number generator is used together with transition probabilities to determine whether the patient stays in his/her current state or moves to a new state. This event is scheduled for every patient in the system on a daily basis. The stochastic dependency simulation modelling at patient level is built as a Markov Chain Monte Carlo (MCMC) simulation. The main difference is that instead of simulating one patient at a time, as traditionally in MCMC, the DES model simulates the dependency behaviour of all patients in the ward simultaneously. In addition, every patient follows the MCMC model that is adequate for his/her diagnosis. As will be shown, the time dependent Markov transitions can be viewed as an in-data structure for the DES model.

The initial distribution is expressed as follow:  $P(\mathbf{X}_0 = i) = q_i$ . This vector with the initial probabilities is represented according to  $\mathbf{q} = [q_A \quad q_B \quad q_C \quad q_D]$ . The sum of the vector's probabilities is 1.

A new day at the ward would represent the possibility of a new transition. These transition probabilities are modelled using an  $S \times S$  (States) matrix which contains the transition

probabilities between the system's states. The **transition probability matrix**  $\mathbf{P}^{(n)}$  for the

$$\text{proposed system would be as follow } \mathbf{P}^{(n)} = \begin{bmatrix} \rho_{AA}^{(n)} & \rho_{AB}^{(n)} & \rho_{AC}^{(n)} & \rho_{AD}^{(n)} & \rho_{AH}^{(n)} \\ \rho_{BA}^{(n)} & \rho_{BB}^{(n)} & \rho_{BC}^{(n)} & \rho_{BD}^{(n)} & \rho_{BH}^{(n)} \\ \rho_{CA}^{(n)} & \rho_{CB}^{(n)} & \rho_{CC}^{(n)} & \rho_{CD}^{(n)} & \rho_{CH}^{(n)} \\ \rho_{DA}^{(n)} & \rho_{DB}^{(n)} & \rho_{DC}^{(n)} & \rho_{DD}^{(n)} & \rho_{DH}^{(n)} \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Each transition probability is represented by  $p_{ij}^{(n)}$  where  $i$  is the current state and  $j$  is the future state and the suffix  $n$  represent the number of complete days the patient has been in the ward. The sum of the probabilities at each row must be 1 according to  $\sum_{j=A}^H p_{ij}^{(n)} = 1$ . The

last row  $[0 \ 0 \ 0 \ 0 \ 1]$  describes the transition probabilities from state H. That state is absorbing which means that there is a 100 per cent probability that the patient stays in H. In other words,  $p_{HH}^{(n)} = 1$  and all other probabilities are 0.

The transition probability matrix  $\mathbf{P}^{(n)}$  can also be defined in the following canonical form, in which there are  $R$  absorbing states and  $T$  transient states:

$$\mathbf{P}^{(n)} = \begin{bmatrix} \mathbf{Q}^{(n)} & \mathbf{R}^{(n)} \\ 0 & \mathbf{I} \end{bmatrix}, \mathbf{Q}^{(n)} \text{ represents the } T\text{-by-}T \text{ matrix of transient states. } \mathbf{R}^{(n)} \text{ is the}$$

nonzero  $T$ -by- $R$  matrix or absorbing probabilities.  $\mathbf{I}$  in our case is just 1 because  $R$  is 1.

If the initial vector  $\mathbf{q}$  is multiplied with the transition probability matrix  $\mathbf{Q}^{(1)}$ , which represent the transition probabilities between the initial day (day 0) to day one among transient states, the resulting vector  $\mathbf{q}^1$  would give the probabilities for the distribution of patients in the different states {A, B, C, D} during day one. If the system loses patients to the absorbing state {H}, between the days, the sum of the vector's probabilities will be less than one. Because the solution is concerned with a day to day calculation, our interest is to normalise the vector  $\mathbf{q}^{(n)}$ . The normalised vector represented by  $\mathbf{n}^{(n)}$  is calculated using the following equations:

$$\mathbf{q}^{(n)} = \mathbf{q}^{(n-1)} \times \mathbf{Q}^{(n)} \quad (1)$$

$\mathbf{q}^{(n)} = \begin{bmatrix} q_A^{(n)} & q_B^{(n)} & q_C^{(n)} & q_D^{(n)} \end{bmatrix}$  where the index  $n$  corresponds to the day of stay

$$\mathbf{n}^{(n)} = \frac{1}{\sum_{x=A}^D q_x^{(n)}} \mathbf{q}^{(n)} \quad (2)$$

$$\mathbf{n}^{(n)} = \begin{bmatrix} n_A^{(n)} & n_B^{(n)} & n_C^{(n)} & n_D^{(n)} \end{bmatrix} \quad (3)$$

The vector  $\mathbf{n}^{(n)}$  is used in equation (6) to adapt the transition probabilities to the absorbing state {H} (see section 6.4.7)

The proposed modelling is time-nonhomogenous, in contrast to a time-homogenous Markov chain where the transition probability matrix does not change over time. This means that  $\mathbf{P}^{(n)}$  is different to  $\mathbf{P}^{(n+1)}$ . The solution uses different transitions matrices for each day of the patient's LoS. Each particular patient diagnose group has its own matrices  $\mathbf{P}_{grp}^{(n)}$ . In other words,  $\mathbf{P}_{grp}^{(n)}$  represent the transition matrix for a particular patient group on day  $n$ , then  $\mathbf{p}_{grp}^{(n)} = \begin{bmatrix} \mathbf{P}^{(1)} & \mathbf{P}^{(2)} & \dots & \mathbf{P}^{(m)} \end{bmatrix}$  represents the vector of transition matrices, where  $m$  represents the maximum LoS.

Markov processes have been used to model patient health states according to several authors (Kapadia et al. 2000; Albornoz et al. 2006; Perez et al. 2006). There are two main differences between their approaches and the one presented here. Firstly, their dependency levels are not quantified in patient care time or personnel requirements. Although Kapadian et al. (2000) advocate that the results can be used to schedule nursing, their work includes no suggestion about how the model's results can be used to calculate personnel requirements. The second difference is that the transition matrix used by these authors is stationary (homogenous), in contrast to the nonhomogenous one used in this work. The advantage with a stationary matrix is that the solution gains analytical simplicity, but often at a substantial sacrifice in realism and flexibility. Many transition processes are not static. For example, the probability of leaving the system normally increases as a function of the number of days a patient has been registered in a ward (see the graph in Figure 30). Perez et al. (2006) point out that the inadequacy of fit of the Markov models presented in their article (only a 55 per cent fit for one of the patient groups) could be the result of the stationary assumption they made. There are more complex approaches in stationary



Markov models that are used to alleviate this problem, but they do not completely solve it and also add an additional modelling complexity (Sin and Kim 1994; Marshall and McClean 2003; Albornoz et al. 2006). Sonneberg and Beck (1993) and Plevritis (2005) claim that there are transition probabilities which are naturally time dependent and not easily described by a simple function, for example, the actual mortality rate over a life time. If the transition probabilities are time dependent, they suggest that instead of an analytical solution using one fundamental matrix the best choice is to use a MCMC which measures variability and embraces time dependent transitions. If time dependent transition probabilities are used, a simple way of implementing them in the model is to store the transition probabilities in an indexed table and retrieve them as the Markov model is evaluated.

At first glance it might seem as if the simulation modelling proposal is limited to a MCMC model. However, this modelling approach is just part of the total solution and because of its discrete and stochastic nature easily integrated in a DES model.

#### **6.4.5 Generation of transition probability matrixes**

The transition probability matrices are generated from the historical data of the ward's PCS. It is obviously important that the data is digitally stored in a database otherwise this task would require too much effort and be vulnerable to human error. The generation of matrices consists of taking the following steps:

##### **Create patient groups:**

As previously mentioned, the natural choices are using DRG classification or the main diagnosis as a grouping trigger. If more detailed information is available, which is not always the case, sub-grouping should be considered. These sub-groups would also take into consideration parameters such as age, gender or bi-diagnosis in order to generate data groups that better represent the patients' dependency behaviour. The limit for the number of groups that can be used is the amount of data available. A problem observed at rehabilitation ward 75 is the large number of different, main diagnoses. Despite more than two years of historical data, it was impossible to give each main diagnosis its own group. This led to the choice of making categories of related groups found under the same DRG and a category for mixed groups representing the less common diagnoses (see section 7.2.1).

**Arrange every patient’s dependency development:**

Table 11 presents part of a patient group’s dependency development viewed chronologically from the patient’s initial day until the patient leaves the ward. The initial day is the first day for all patients, and each patient has a separate row. The patient with the longest LoS will have the longest row of dependency classifications.

PatientID /Day	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
5859	C	C	C	B	B	C	B	B	B	B	B	B	C	B	B	C	H					
5867	C	B	B	B	B	B	B	A	A	H												
5879	B	A	A	B	B	A	B	H														
5896	B	B	A	A	B	B	B	A	A	A	A	A	A	A	H							
5960	C	B	B	B	B	B	B	B	B	B	A	B	B	B	H							
6827	D	C	C	C	C	C	C	C	C	B	B	C	B	B	B	B	B	B	B	B	B	H
6857	C	B	B	B	B	B	B	B	B	C	B	B	B	H								
6943	C	B	C	C	B	B	B	B	B	B	B	B	B	B	B	H						
6963	C	C	C	C	C	C	C	C	C	C	C	C	C	C	H							
6988	B	B	B	B	B	B	B	B	B	A	H											

**Table 11. Example of patient dependency development**

**Calculate the transition probabilities:**

The transition probabilities between states are based on the day to day variations observed in the historical data, presented in Table 11. An example of the point probability estimation can be seen in Table 12. The table shows the probabilities of state B for the corresponding days. TM 1 means the transition matrix between the initial day and day1. The data shows small variations between days, but also the increasing percentage of patients in state B that moves to state H. The lowest row shows the number of patients that were in state B before the transition.

TM	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
BA	0.18	0.10	0.04	0.08	0.14	0.03	0.11	0.09	0.06	0.12	0.10	0.09	0.11	0.06	0.07	0.03	0.06
BB	0.76	0.82	0.82	0.78	0.75	0.85	0.72	0.80	0.80	0.73	0.74	0.74	0.67	0.73	0.60	0.73	0.63
BC	0.07	0.08	0.14	0.13	0.10	0.05	0.12	0.08	0.06	0.09	0.13	0.12	0.06	0.10	0.11	0.14	0.16
BD	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
BH	0.00	0.00	0.00	0.01	0.00	0.06	0.05	0.03	0.08	0.06	0.03	0.05	0.17	0.10	0.22	0.11	0.16
No.	45	61	72	72	71	62	75	65	65	67	61	57	54	49	45	37	32

**Table 12. Transition probabilities for state B day 1 to 17.**

Besides the transition probabilities, the initial probabilities required are  $\mathbf{q} = [q_A \ q_B \ q_C \ q_D]$ . The initial probabilities show in which state the patients enter the system. Gender (1989) declares that all newly admitted patients to the rehabilitation

ward were given the highest dependency level (level 4) during their first day because of the expected, higher patient workload required. This was a result of the additional information, orientation and introduction procedures the patient and his/her family would need on the first day. A higher dependency level on the first day is confirmed in the data from ward 75. It must be remembered that estimating the patients' dependency level was done in retrospect in our case and the additional workload the new patient generated was easily observed and documented. Bear in mind that the dependency level is not totally correlated to the improvement progress of the patient's health (Nelson et al. 2007). The transition probabilities are based on the observable historical data, which is not necessarily equivalent to the historical observation. Each historically based transition probability is just a point estimation of the real value. The point estimation will come closer and closer to the real value when the number of observations increases. As a result it will always contain a certain level of uncertainty.

#### **Adaptation and storage of the transition probabilities:**

The final step concerns the adaptation and storage of the probabilities in their respective transition matrices. All the combinations in the  $S \times S$  matrices need to be stored in the corresponding data arrays used by the simulation model. After the creation of the patient groups, all the steps of this procedure can be, more or less, automated (this depends on which technique is used for calculating the point estimation, see section 6.4.6). Unfortunately, this procedure encounters a logical problem that demands slightly more analysis and adaptation of the data. The final row in Table 11 shows that by day 17 the number of patients in state B diminishes for natural reasons. The average LoS for this category is 17.5 days. If the total population in the group is 149 patients (which corresponds to more than two years of historical data), it is easy to understand that by the patient's 30<sup>th</sup> day only a few of the initial number of patients are still in the ward (the historical data presents 18 patients). This natural process means that we get fewer and fewer data points on which to base our transition probabilities. The most crucial probability is the rate by which the patients leave the system. Normally in DES models the patient's LoS is defined when the patient enters the system by a probability distribution. This is not the case in the proposed methodology, but it is still crucial to be able to model this parameter accurately and base it on the historical data. Several techniques can be used to make the point estimations of the transition probabilities

more accurate. These techniques with their pros and cons are more closely presented in the next section.

### 6.4.6 Point estimate data analysis

Before we discuss the problems around the point estimation of the transition probabilities, let us have a brief analysis of the historical data of a patient group in order to better understand the future discussion.

Data from one of the patient groups is presented in Figure 26 which reveals how the patients from this group are distributed in the different states and how these percentages evolve. A brief glance shows that the states with the most patients are B and C. The figure also reveals a quick transition of state percentage from state C to state B between the initial day and day 1. The source of this rapid change is the higher patient dependency level on the first day. Another clear trend is the increase of group H, which represents the percentage of patients leaving the system each particular day. The data also reveals that group D is very small.

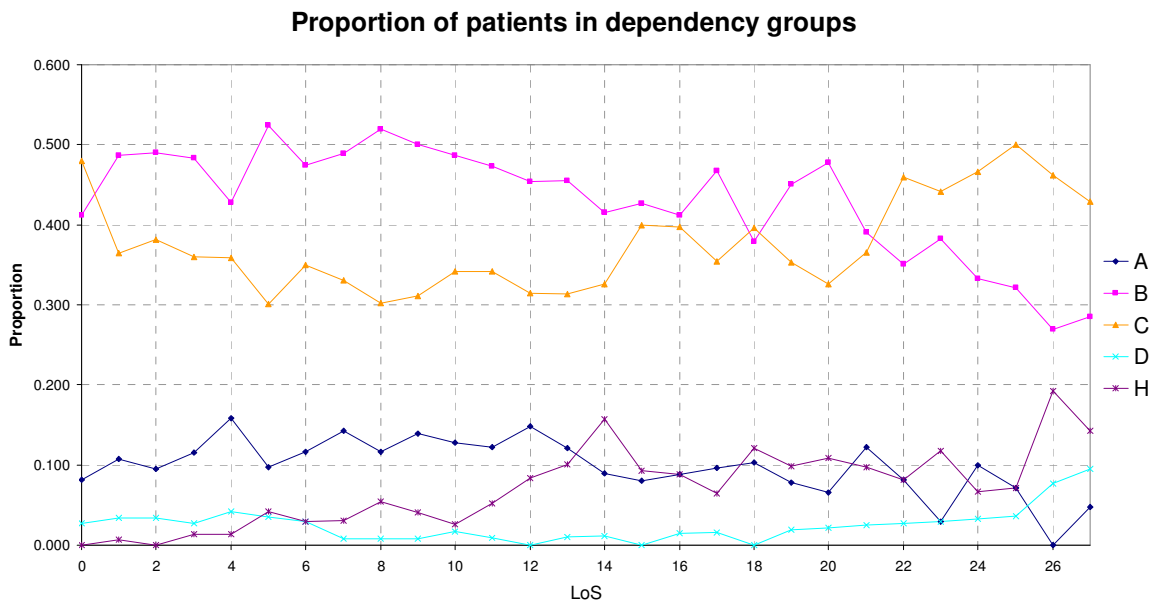


Figure 26. Proportion of patients, in the different dependency groups from arrival up to LoS of 27 days

What cannot be observed in Figure 26 are the transition probabilities between the groups. For example, this means that although the proportion of patients leaving the system is observed in group H, it does not reveal which state(s) these patients come from. Have they been in state B before leaving or do they come from several states and, in that case, what is

the percentage contribution from these respective states? A second point that is not observable in Figure 26 is how precise and valid these point estimations are. We simply do not see the total number of observations they are based on, and we have an intuitive understanding that point estimations are more precise or correct if they are based on a greater number of observations. The issue is whether this can be quantified and how can we use it to improve our estimations. In response, we are able to quantify the “quality” of the point estimations by using the confidence interval (CI) of the point estimates and, by doing so, obtain an assessment of the accuracy of the observation.

The unbiased point estimator of a transition proportion is calculated in a very straight forward manner using the following equation,  $\hat{\rho} = \frac{X}{n}$ , where  $X$  is the number of times the event occurs and  $n$  is the total number of binomial trials. For instance, during day seven the total number of patients  $n$  in state {B} is 75 and the number of patients  $X$  that stayed in state B is 54, giving  $\hat{\rho}_{BB}^7 = 0.72$  for that seventh day. The point estimator  $\hat{\rho}_{XX}^{(n)}$  is expected to approach the true value  $\hat{\rho}_{XX}^{(n)}$  as *sample size*  $\rightarrow \infty$ .

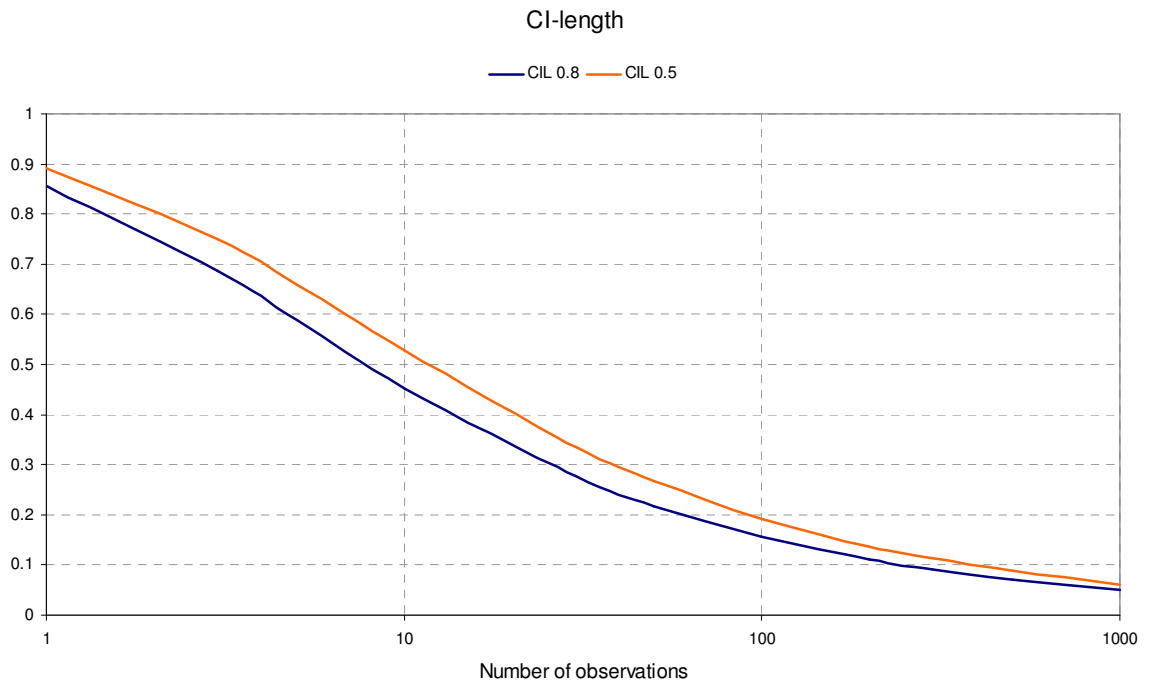
The interval estimation of a binomial proportion is, on the other hand, a highly discussed topic among statisticians because of the general misconceptions in the area. A common suggestion for the interval estimation of proportions (Montgomery and Rungar 1994) is the following equation:

$$\hat{\rho} - z_{\alpha/2} \sqrt{\frac{\hat{\rho}(1-\hat{\rho})}{n}} \leq \rho \leq \hat{\rho} + z_{\alpha/2} \sqrt{\frac{\hat{\rho}(1-\hat{\rho})}{n}}, \text{ where } \hat{\rho} \text{ is the unbiased point estimator.}$$

This equation is only suggested if the sample size  $n$  is not too small. However, there are not only different opinions about what is too small, Brown et al. (2001) also present figures concerning a high degree of uncertainty in the above equation’s adequate coverage of the CI. They recommend three alternative interval equations. The one used in this work is the so-called *Wilson interval* (Wilson 1927) as cited by Brown et al. (2001). The Wilson interval was selected for two reasons; it has a closed-form formula and a high accuracy level for small  $n$  (40 or less) (Brown et al. 2001). The formula is as follows:

$$\frac{\hat{\rho} + \frac{z_{\alpha/2}^2}{2n} - z_{\alpha/2} \sqrt{\frac{\hat{\rho}(1-\hat{\rho})}{n} + \frac{z_{\alpha/2}^2}{4n^2}}}{1 + \frac{z_{\alpha/2}^2}{n}} \leq \rho \leq \frac{\hat{\rho} + \frac{z_{\alpha/2}^2}{2n} + z_{\alpha/2} \sqrt{\frac{\hat{\rho}(1-\hat{\rho})}{n} + \frac{z_{\alpha/2}^2}{4n^2}}}{1 + \frac{z_{\alpha/2}^2}{n}} \quad (4).$$

The lower and upper limits of the CI are not necessarily symmetrically distributed. The CI length depends on both the number of observations  $n$  and on the  $\hat{\rho}$ -value. The chart in Figure 27 presents how the CI length of  $\rho = 0.5$  and  $\rho = 0.8$  depends on the number of observations they are based on. The scale is logarithmic and shows that beyond 300 observations the CI length is reduced to under 0.1, while approximately 50 observations increase the length to 0.2 - 0.25.

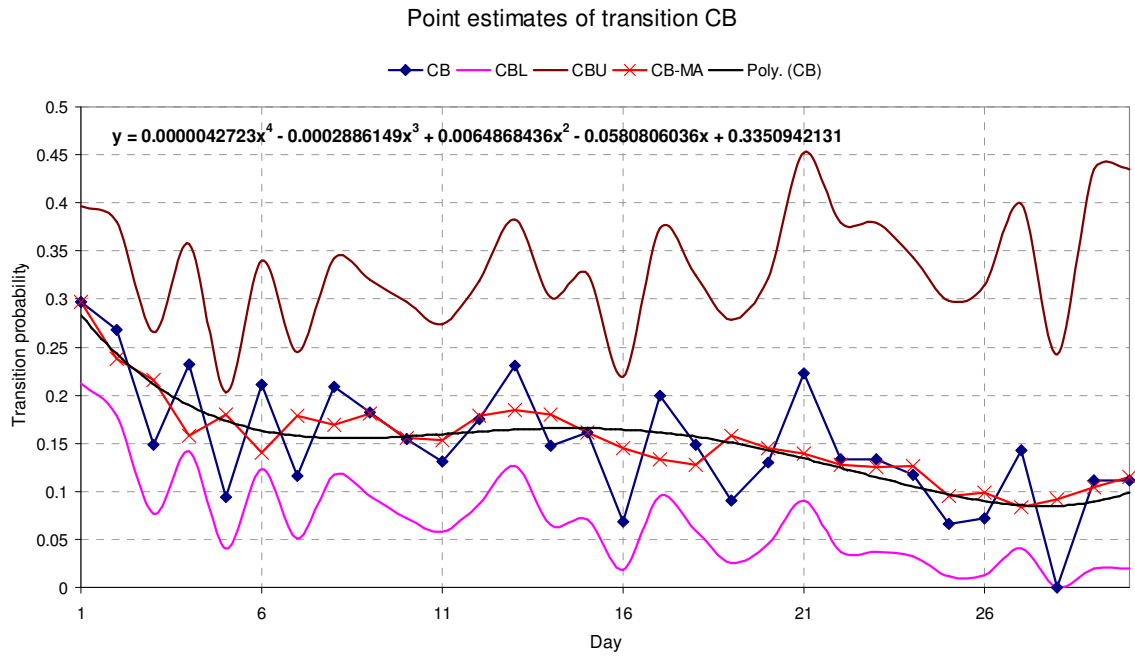


**Figure 27. CI length vs. number of observations for two point estimations**

Unless there are a considerable number of observations, it is obvious that the analysis of the point estimation data will need to deal with some level of uncertainty. This level of uncertainty will decrease with time because the ward's PCSs continuously obtain more data from new patients. Nevertheless, it is also important to remember that the reason for calculating the point estimations on a daily basis is to identify trend changes in the transition probabilities. These changes might be sudden, such as the ones that can be

observed for states B and C between the initial day and day 1, but they will be smoother or not present at all most of the time. This means that fluctuations arising from point estimations based on insufficient observations (and therefore affected by the inherent random variation of the system) can be identified and counteracted using smoothing techniques. Two such techniques, used for time series analysis, are centred moving average and polynomial regression. However, it should be remembered that the data analysis presented in this section does not aim to be a setup for all the steps and extended analyses the user would feel are necessary, but rather a pragmatic way of approaching the estimation of transition probabilities. It is also important to remember that the data analysis is based on some level of uncertainty, both in terms of the analysis of the data and the accuracy of the individual patient dependency evaluations conducted by the ward's personnel. Having excessive demands on the quality of the data estimations would result in a perpetual analysis phase and distrust of the results. It is therefore essential to bear in mind that the analysis aims to find a better way of representing the ward's behaviour than the one most currently being used.

Figure 28 shows the point estimations of transition probability  $\hat{\rho}_{CB}^{(n)}$  (CB in the chart) and the 95 per cent CI with a lower and upper border (CBL resp. CBU) for the first 30 days. The transition probabilities fluctuate from one day to another in a declining trend. What can also be observed is the increasing CI length (difference between CBU and CBL), which reveals that  $n$  is decreasing. Besides the above mentioned series, the chart presents two techniques of time series analysis. One is the centred moving average (CB-MA) and the other is a polynomial regression line (Poly. (CB)). Both these series represent a considerable smoothing of the initial point estimates, but without missing the observable trend and maintaining the results within the limits of the CI. Caution needs to be taken so that the time series analysis does not fall outside the CI of the point estimator and that it does not over-adapt to the random variations in the data. The higher the polynomial function or the lower the moving average window, the higher the risk of over-adapting to the sample data.



**Figure 28** Points estimates of transition probability  $\hat{\rho}_{CB}^{(n)}$ , CI Low (CBL) and Upper limit (CBU), Moving average smoothing (CB-MA) and Polynomial regression (Poly. CB)

Even though a centred moving average is easy to calculate (see examples of centred moving average techniques in Law and Kelton (2000) and NIST/SEMATECH (2007)), it requires more work in collecting and implementing the data into the transition matrices. The polynomial regression analysis, on the other hand, provides an expression that only depends on a single variable  $n$  which represents the patient's day of stay (see Figure 28 top left corner). It should be remembered that the regression analysis is only valid within the limits of the data taken into consideration. The strategy is to settle the estimation of probability on a final value when that limit is reached. This final value will represent the patient's behaviour from that day and forward. In other words, taking  $\hat{\rho}_{CB}^{(n)}$  as an example group:

$$\hat{\rho}_{CB}^{(n)} = \begin{cases} 4 \times 10^{-6} n^4 - 0.0003 n^3 + 0.0065 n^2 - 0.0581 n + 0.3351 & \text{if } n \leq 29 \\ 0.099 & \text{otherwise} \end{cases}$$

The determination of this final value requires heuristics and a graphical judgement of the point estimations. It is important to take into consideration that a point estimated on only a few data observations is of very little value. Therefore, one should not wait until the



number of observations is too low to make the final estimation, especially when no apparent future trend is observed.

Many of the transition probabilities in our Markov process contain only random variations but no trend variations over time, which enable a much simpler calculation of transition probabilities. Still, where do the trends come from? Well, one observed source is the initial transitions of states B and C. Another is discussed in the following section.

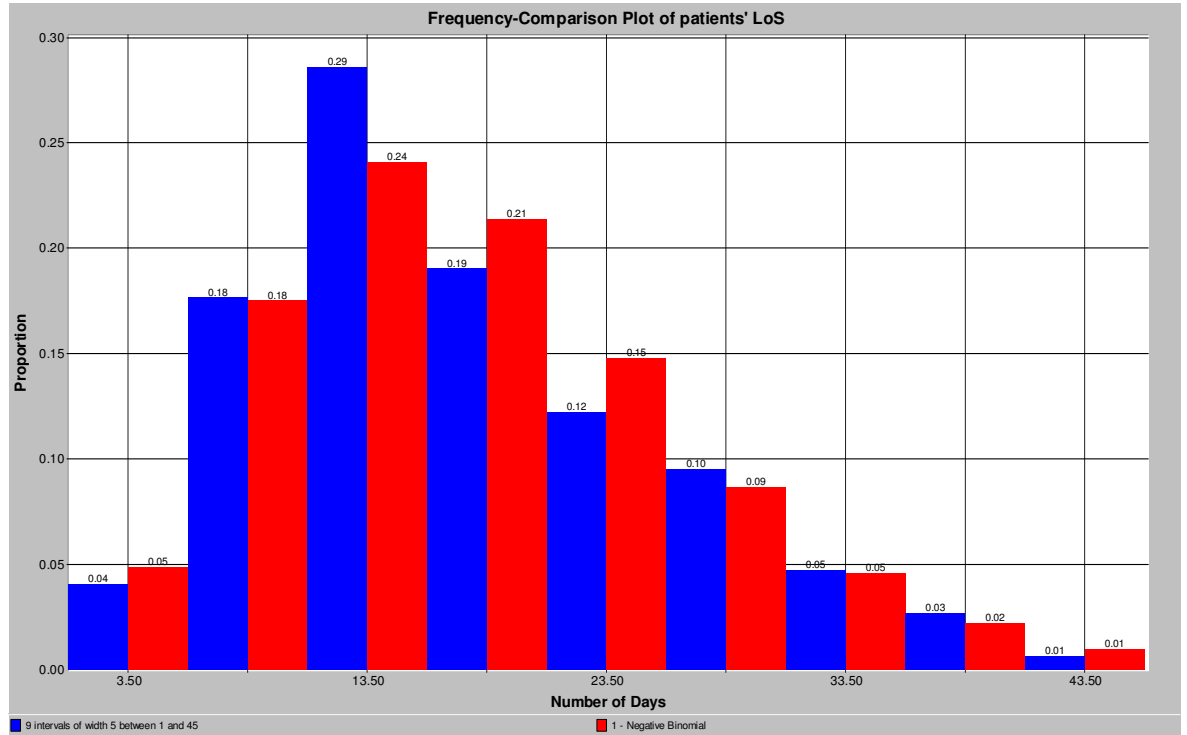
#### 6.4.7 Adaptation of transitions to fit LoS distribution

One way of improving the accuracy of the data analysis and contra resting the effects of few data points is by adapting the transition probabilities of  $\left\{ \hat{\rho}_{AH}^{(n)} \quad \hat{\rho}_{BH}^{(n)} \quad \hat{\rho}_{CH}^{(n)} \quad \hat{\rho}_{DH}^{(n)} \right\}$  to a fitted theoretical maximum likelihood estimated distribution of LoS. This distribution would be based on the total number of patients within the group and statistically verified through Goodness-of-Fit tests, consequently increasing the degree of certainty. This theoretically estimated distribution represents the usual way of defining the LoS variable in a simulation model.

The background of this approach is that the estimated distribution is a better, long term estimation of the system's true behaviour compared to the historical data. The assumption is that if we have a large number of observations the historical data would approach the estimated distribution. Figure 29 presents the sample group's data compared to the estimated distribution in a frequency diagram that shows how proportions of the patient sample with different LoS times are distributed. The LoS is normally chosen randomly from the estimated distribution when the patient enters the system.

This procedure is different to the one chosen in our approach. Using a Markov process to monitor the patients' dependency development and LoS means there is a stochastic probability which is randomly evaluated every day as to whether the patient will stay or leave. Our aim is to use the LoS fitted distribution to define the proportion of patients that leaves the system on day  $n$ , and not make that decision in advance. A second task to solve is how this proportion of day  $n$ , which is for the whole system, is individually adapted to the absorbing probabilities of leaving the system for each state  $\left\{ \hat{\rho}_{AHi}^{(n)} \quad \hat{\rho}_{BHi}^{(n)} \quad \hat{\rho}_{CHi}^{(n)} \quad \hat{\rho}_{DHi}^{(n)} \right\}$ ,

as previously discussed in section 6.4.6 (the term  $i$  introduced in the index of  $\hat{\rho}_{XHi}^{(n)}$  refers to *initial*. The final  $\hat{\rho}_{XH}^{(n)}$  is the product of  $\hat{\rho}_{XHi}^{(n)}$  and a scaling vector, see the following description).



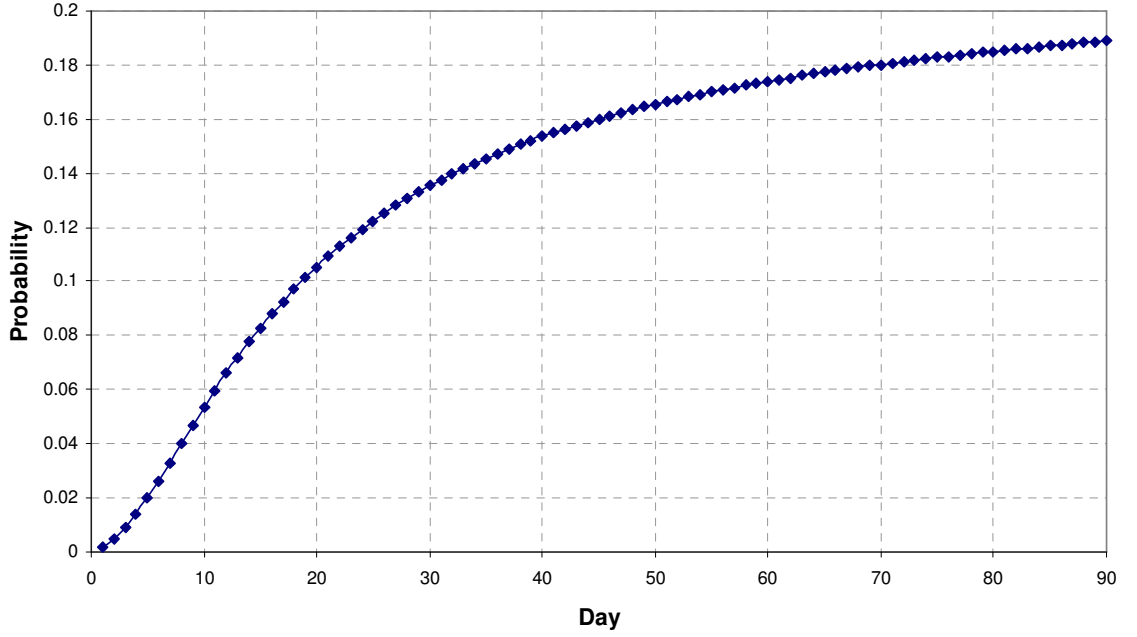
**Figure 29. Frequency Comparison plot between the historical data and the estimated distribution.**

The first aim, estimating the proportion per day that leaves the system according to the fitted LoS distribution, is represented by the following equation, where  $n$  represents the day of interest and  $F$  is the cumulative function of the fitted LoS distribution:

$$P_n = \frac{P(X = x_n)}{P(X \geq x_n)} = \frac{F(x_n) - F(x_{n-1})}{1 - F(x_{n-1})} \quad (5)$$

The increase of the absorbing probability to state H (leaving the system), observed in Figure 30, presents one reason for the time-nonhomogenous structure of the Markov process.

**Absorbing probabilities to state H**



**Figure 30. Absorbing probabilities to state H for patients in the sample group**

The second difficulty is how the absorbing probabilities are adapted to the value of equation (5). For the sake of simplicity, the relationship between the day to day probabilities to state H  $\left\{ \hat{\rho}_{AHi}^{(n)} \quad \hat{\rho}_{BHi}^{(n)} \quad \hat{\rho}_{CHi}^{(n)} \quad \hat{\rho}_{DHi}^{(n)} \right\}$  is maintained constant, but the magnitude of the probability is scaled so that the total sum is equivalent to:

$$P_n = \frac{F(x_n) - F(x_{n-1})}{1 - F(x_{n-1})} = X_H^{(n)} \left( n_A^{(n-1)} \hat{\rho}_{AHi}^{(n)} + n_B^{(n-1)} \hat{\rho}_{BHi}^{(n)} + n_C^{(n-1)} \hat{\rho}_{CHi}^{(n)} + n_D^{(n-1)} \hat{\rho}_{DHi}^{(n)} \right) \quad (6)$$

$\mathbf{n}^{(n)} = \begin{bmatrix} n_A^n & n_B^n & n_C^n & n_D^n \end{bmatrix}$  are the probabilities of patients in state  $\{A, B, C, D\}$  in day  $n$ .

$X_H^n$  : is the scale factor for day  $n$ , using equation (6) leads  $X_H^n$  to be:

$$X_H^n = \frac{P_n}{\left( n_A^{(n-1)} \hat{\rho}_{AHi}^{(n)} + n_B^{(n-1)} \hat{\rho}_{BHi}^{(n)} + n_C^{(n-1)} \hat{\rho}_{CHi}^{(n)} + n_D^{(n-1)} \hat{\rho}_{DHi}^{(n)} \right)} \quad (7)$$

This gives the scale vector  $\mathbf{X}_H^n = \begin{bmatrix} X_H^1 & X_H^2 & X_H^3 & \dots & X_H^m \end{bmatrix}$  where  $m$  is the maximum LoS.

The scale vector  $\mathbf{X}_H^n$  is in turn used to define the new absorbing probabilities, stored in the

T- by- R matrix  $\mathbf{R}^{(n)} = \mathbf{X}_H^n \times \begin{bmatrix} \hat{\rho}_{AHi}^{(n)} \\ \hat{\rho}_{BHi}^{(n)} \\ \hat{\rho}_{CHi}^{(n)} \\ \hat{\rho}_{DHi}^{(n)} \end{bmatrix} = \begin{bmatrix} \hat{\rho}_{AH}^{(n)} \\ \hat{\rho}_{BH}^{(n)} \\ \hat{\rho}_{CH}^{(n)} \\ \hat{\rho}_{DH}^{(n)} \end{bmatrix}$  (8). These new scaled probabilities in  $\mathbf{R}^{(n)}$  will

generate the need to rescale the transition probabilities in  $\mathbf{Q}^{(n)}$ . This means that states {A, B, C, D} have their own scale vectors  $\{X_A^n, X_B^n, X_C^n, X_D^n\}$ , but this time the scaling only affects the transition probabilities and not the absorbing probabilities in  $\mathbf{R}^{(n)}$ . The procedure follows a similar approach as the one in equation (7). It must be remembered

that the total sum of the probabilities for each day is always  $\sum_{j=A}^H p_{ij}^{(n)} = 1$ , leading to:

$$\mathbf{X}_{A-D}^{(n)} = \begin{bmatrix} 1 - \hat{\rho}_{AH}^{(n)} & 0 & 0 & 0 \\ 0 & 1 - \hat{\rho}_{BH}^{(n)} & 0 & 0 \\ 0 & 0 & 1 - \hat{\rho}_{CH}^{(n)} & 0 \\ 0 & 0 & 0 & 1 - \hat{\rho}_{DH}^{(n)} \end{bmatrix} \begin{bmatrix} \frac{1}{\sum_{x=A}^D \hat{\rho}_{Ax}^{(n)}} & 0 & 0 & 0 \\ 0 & \frac{1}{\sum_{x=A}^D \hat{\rho}_{Bx}^{(n)}} & 0 & 0 \\ 0 & 0 & \frac{1}{\sum_{x=A}^D \hat{\rho}_{Cx}^{(n)}} & 0 \\ 0 & 0 & 0 & \frac{1}{\sum_{x=A}^D \hat{\rho}_{Dx}^{(n)}} \end{bmatrix} \quad (9)$$

The resulting vector  $\mathbf{X}_{A-D}^{(n)}$  contains the scale factors for the  $\mathbf{Q}^{(n)}$  matrix. It follows that:

$$\mathbf{Q}^{(n)} = \begin{bmatrix} X_A^n & 0 & 0 & 0 \\ 0 & X_B^n & 0 & 0 \\ 0 & 0 & X_C^n & 0 \\ 0 & 0 & 0 & X_D^n \end{bmatrix} \begin{bmatrix} \hat{\rho}_{AAi}^{(n)} & \hat{\rho}_{ABi}^{(n)} & \hat{\rho}_{ACi}^{(n)} & \hat{\rho}_{ADi}^{(n)} \\ \hat{\rho}_{BAi}^{(n)} & \hat{\rho}_{BBi}^{(n)} & \hat{\rho}_{BCi}^{(n)} & \hat{\rho}_{BDi}^{(n)} \\ \hat{\rho}_{CAi}^{(n)} & \hat{\rho}_{CBi}^{(n)} & \hat{\rho}_{CCi}^{(n)} & \hat{\rho}_{CDi}^{(n)} \\ \hat{\rho}_{DAi}^{(n)} & \hat{\rho}_{DBi}^{(n)} & \hat{\rho}_{DCi}^{(n)} & \hat{\rho}_{DDi}^{(n)} \end{bmatrix} = \begin{bmatrix} \hat{\rho}_{AA}^{(n)} & \hat{\rho}_{AB}^{(n)} & \hat{\rho}_{AC}^{(n)} & \hat{\rho}_{AD}^{(n)} \\ \hat{\rho}_{BA}^{(n)} & \hat{\rho}_{BB}^{(n)} & \hat{\rho}_{BC}^{(n)} & \hat{\rho}_{BD}^{(n)} \\ \hat{\rho}_{CA}^{(n)} & \hat{\rho}_{CB}^{(n)} & \hat{\rho}_{CC}^{(n)} & \hat{\rho}_{CD}^{(n)} \\ \hat{\rho}_{DA}^{(n)} & \hat{\rho}_{DB}^{(n)} & \hat{\rho}_{DC}^{(n)} & \hat{\rho}_{DD}^{(n)} \end{bmatrix} \quad (10)$$

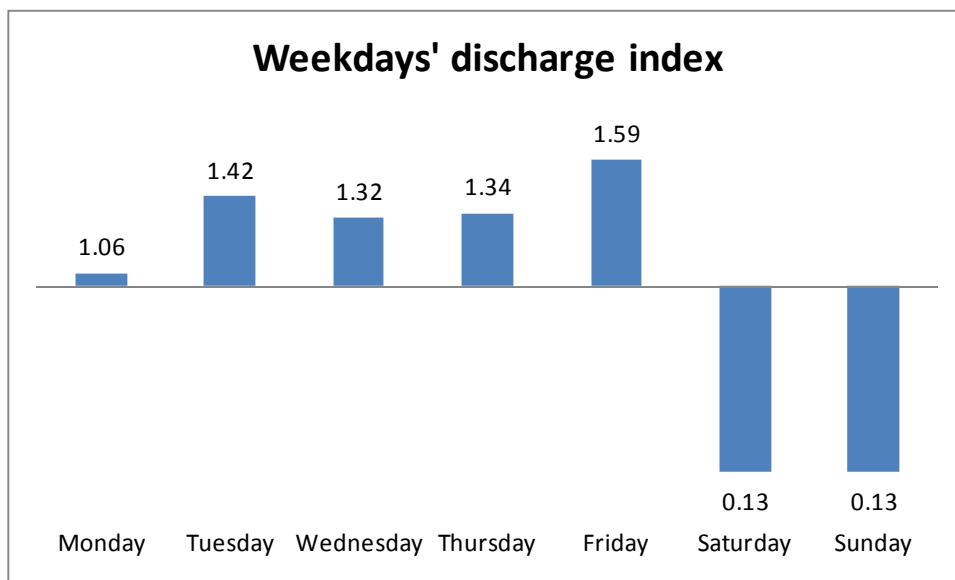
Or more simply stated  $\mathbf{Q}^{(n)} = \mathbf{X}_{A-D}^{(n)} \mathbf{Q}_i^{(n)}$  where  $\mathbf{Q}_i^{(n)}$  is the initial matrix for day  $n$  before scaled. Having calculated and scaled both  $\mathbf{R}^{(n)}$  and  $\mathbf{Q}^{(n)}$  the final values for

$\mathbf{P}^{(n)} = \begin{bmatrix} \mathbf{Q}^{(n)} & \mathbf{R}^{(n)} \\ 0 & \mathbf{I} \end{bmatrix}$  are settled. The scaling procedure is iterative, which means that in order to calculate  $\mathbf{P}^{(2)}$ ,  $\mathbf{P}^{(1)}$  needs to be defined first and so forth.

#### 6.4.8 Discharge modelling adaptation

One of the problems with the traditional way of modelling the LoS (using a fitted distribution) of a patient group is the difference in discharge percentages over the week. In many wards the patient must wait for the physician's evaluation before being discharged. This evaluation is normally done on the physician's scheduled round on weekdays and more seldom on weekends. The traditional way of modelling would not take the discharge routines into consideration, and the patient would leave the ward whenever the LoS ends. In the best of cases a patient's discharge could be postponed or brought back to the nearest weekday. This would obviously lead to an unnatural increase of the discharge percentage on Mondays and Fridays, something that would not accurately represent the system's behaviour. However, this problem can be avoided by using a non-stationary Markov process for modelling the LoS. The time dependent transition probabilities to state H  $\{\hat{\rho}_{AH}^{(n)} \quad \hat{\rho}_{BH}^{(n)} \quad \hat{\rho}_{CH}^{(n)} \quad \hat{\rho}_{DH}^{(n)}\}$  have been adapted to the Maximum Likelihood Estimation (MLE) of the LoS for the patient group based on equation (8). This adaptation does not take the discharge variations in the system into consideration. The current system has, for instance, only a 2 per cent per day discharge probability during weekends, while Fridays have the highest discharge probability (22.7 per cent) and Mondays the lowest (15.2 per cent) of the weekdays. If the discharge percentage is uncorrelated with the days of the week, every day of the week would have 14.3 per cent of the total number of the patients' discharges. In order to address this, a discharge factor is calculated and correlated to each day, see Figure 31. If the day has an equivalent of 14.3 per cent of the discharges, the factor would be equivalent to one and the transition probability for that particular patient and day  $\hat{\rho}_{XH}^{(n)}$  will not be altered. However, if there is a difference, the factor would be scaled to represent that difference. For example, the discharge factor for Fridays is 1.59, which represents there being a 59 per cent higher probability of discharge compared to the value of  $\hat{\rho}_{XH}^{(n)}$ . For a weekend, the factor would be 0.13, which represents an 87 per cent lower probability than the value of  $\hat{\rho}_{XH}^{(n)}$ . A change in  $\hat{\rho}_{XH}^{(n)}$  would obviously affect the other

transition probabilities of that day. Instead of recalculating the stored matrix, DES simulation gives the user the flexibility of adapting the probabilities locally using arithmetical calculations during the model execution, thus avoiding innumerable additional transition matrices. This modelling flexibility is what users appreciate and what enables the construction of more accurate models compared to the Markov ones (Simpson et al. 2009).



**Figure 31. Weekday's variation of discharge index**

The modelling procedure that is discharge adapted can also be used in models that do not take dependency levels into consideration. The implementation is much simpler and the Markov process contains only two states {stay in the system, leave the system} and one transition between these states. This transition is the one given in equation (5). The discharge adaptation can easily then be applied by simply multiplying the discharge day factor with the transition probability. The simplicity and increased modelling accuracy of this approach should be an incentive to change the traditional way of modelling the LoS of patients, especially if the discharge figures are highly correlated to the days of the week.

It may seem that calculating scaling vectors, polynomial and matrices demands a great deal of work. All analysis demands work, but most of it can be simplified if statistical and/or spreadsheet software with statistical packages are used. For example, with an extensive amount of historical data, the work of finding accurate estimators for the transition probabilities would become much easier and the results from Table 12 would be the final estimators. When that is not the case, a more elaborated procedure is needed. The aim of this section is to provide guidelines about how to estimate transition probabilities and,

irrespective of the amount of data available, achieve more accurate estimations. It presents two important contributions to knowledge:

*Firstly, the stochastic modelling of dependency level variation including how to adapt the fitted LoS distribution to the day to day probability of patient transitions to state H.*

*Secondly, the discharge modelling adaptation that can be used for models that do not take dependency modelling into consideration and is therefore a more generic modelling approach.*

#### **6.4.9 Indicator and determinant estimation**

A patient's dependency level does not entirely define the amount of nursing care per day he/she will require. There is a stochastic variation in every dependency level. That variation originates from the determinant value of each indicator. The patient is evaluated according to the seven indicators presented in section 3.5.1. There are three determinant values that a patient can obtain for every indicator. The sum of these determinants will establish the dependency level of the patient. The approach presented thus far is contrary to approach the healthcare personnel would have taken. We establish the dependency level first and then the corresponding determinant values. This approach becomes obvious and reasonable when we look at the number of possible combinations of indicators, which is  $3^7 = 2187$ . Thus, instead of having to use 2187 possible states, the present approach is limited to just {A, B, C, D, H}.

How is the dependency state then turned into an indicator combination? There are several procedures that facilitate the task and restrain the complexity. Firstly, each dependency state is defined in terms of the limits (min and max) of the sum of points from the seven indicators (see section 3.5.1). Secondly, there is abundant data that gives the probabilities for the determinant values 1-3 of the 7 indicators of each dependency level (see Table 13) and patient group. These two information sources provide viable restrictions on the number of possible outcomes and an easy way of determining the set of indicator values.

	1	2	3	4	5	6	7	1	2	3	4	5	6	7
	A							B						
1	0.964	0.996	1.000	0.071	0.973	0.969	0.879	0.241	0.955	0.884	0.006	0.270	0.872	0.732
2	0.996	1.000	1.000	0.384	1.000	1.000	0.991	0.575	1.000	0.998	0.108	0.536	0.996	0.991
3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	C							D						
1	0.005	0.452	0.693	0.000	0.004	0.472	0.602	0.000	0.059	0.103	0.000	0.000	0.015	0.353
2	0.075	0.963	0.977	0.041	0.060	0.978	0.986	0.029	0.750	0.735	0.000	0.044	0.721	0.853
3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

**Table 13. The table presents the cumulative probabilities for the determinant values (1-3) of the indicators (1-7) for the four dependency levels of our example group.**

The process of determining the seven indicator values starts after defining the dependency level of the patient. The next step is to randomly select a starting indicator for the determinant definition. In Figure 32 the random choice selected indicator 5 as the starting indicator and indicator 4 as the last one. A random value between 0-1 is generated for each of the indicators. The programming procedure tests if the random value is bigger than the accumulated probability value of determinant 1, which is illustrated in Table 13. If this is not the case, the procedure continues testing against the value of determinant 2 and so forth. As soon as the statement is fulfilled, the value from 1-3 is stored and the loop continues with the next indicator, where the process starts again and a new value is stored. When all the seven indicator values have been selected and stored the process is finalised.

There are two circumstances that demand a certain deviation from the above mentioned procedure and these are when the total sum of the indicators is below or above the values for the determined dependency group (see Figure 8). If the sum of the indicators is below the corresponding dependency level, the procedure continues after the last indicator value has been defined. For instance, in Figure 32 the last value was defined in indicator 4. If we assume the patient is at dependency level B, it means the minimum sum for the seven indicators is 10 and the maximum is 13. In this example the sum is 9, in other words below the minimum 10. The programming procedure continues in the following manner. The first value stored, corresponding to indicator 5, is recalculated using a new random seed. If the sum after the random selection of a new value for indicator 5 is within the limits of group B the procedure ends, otherwise the procedure continues with a recalculation of indicator 6 and so forth until the sum is within the specified limits. On the other hand, if the total sum exceeds the maximum limit, the programming procedure identifies this and adapts the remaining values of the indicators so that the sum is kept within those limits. This is why a



random selection of the starting indicator is used, that is, to avoid bias on the indicator values and contra rest the possible effects of the referred to adaptations.

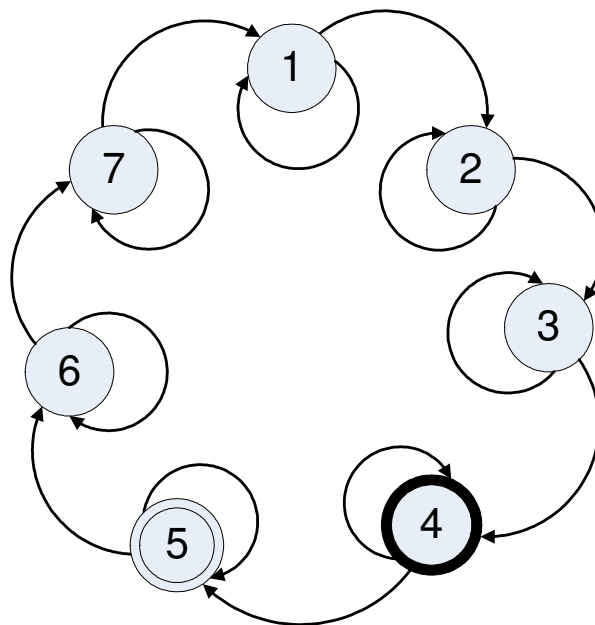


Figure 32. Determinant definition flow for the 7 indicators

#### 6.4.10 Calculation of workload requirements

Thus far the data used has been based on the day to day registrations stored in the PCS, but the most crucial data in the PCS is that stored during the activity study. This study provides information about the time equivalent of each indicator/determinant (see Table 1 ) as well as the proportion of work (per activity) performed by RNs, ANs, and the Ward Nurse Manager (WNM) respectively, and when that work is performed during the day (see Table 14).

The basics for workload calculations are simple. The central formula is:

$$\text{Total workload} = \sum_{i=1}^{\text{all patients}} \text{Direct patient care time}_i + \text{Remaining patient care time}_i + \text{Base work time} \quad (11)$$

As previously mentioned, direct patient care time and remaining patient care time are timed activities that can be addressed to a particular patient. However, the base work time represents activities that cannot be directly associated with a particular patient as well as activities related to ward administration. In order to clarify the relationship between the four grouped activities presented in Table 14 and the above central formula, we can state that *indirect care timed activities* can be addressed to either *remaining patient care time* or

*base work time* according to the above mentioned criterion. The sum of the *base work time* is equally shared among the patients irrespective of their dependency level.

As an example, let us assume a patient is assessed as having an indicator combination of {2 2 1 3 2 1 1}. The sum total of 12 means that he/she is considered to be at dependency level B. The indicators' value corresponds to the sum of  $46+7+4+19+12+8+3=99$  minutes of *direct patient care time* according to Table 1 in section 3.5.2. This sum, together with the relationship between the *direct care time* and the *remaining patient care time*, means (see Table 2) that the *remaining patient care time* is  $99 \times \frac{42}{58} = 71,7$  minutes. The *base work time*, according to the activity study, is 283.2 minutes per day and patient. Consequently, this example patient's contribution according to equation (11) is  $99 + 71.7 + 283.2 = 453.9$  minutes. The activity study reveals that the per cent of total workload time devoted by ANs is 52.89 per cent, for RNs it is 40.88 per cent while the WNM devotes 6.23 per cent, see Table 14.

How the different personnel categories make use of their time during the day is monitored by the PCSs. Table 14 presents the proportion of the various activities' workload each staff group covers and the percentage of the total workload the activity area represents. The results are based on a 24 hour workload measurement. A more interesting analysis might be to focus on the workload requirements of the day shift or between 7 a.m. and 7 p.m.. Whatever time frame of interest can easily be studied and accessed from the PCSs for the purpose of improving daily routines, identifying proper scheduling for the different shifts, and so forth. However, these issues are not in the scope of the current study, at least not at this level of abstraction. The simulation study examines the issues of planning a proper patient mix, as well as volume versus personnel mix and number in a stochastic system, but it does not address medical treatment issues or how the work routines should be set up.

	Activity area	AN	RN	WNM	%
<b>Direct care activities</b>	1.1 Hygiene/elimination	0.72	0.28	0.00	11.5
	1.2 Nutrition	0.61	0.39	0.00	1.2
	1.3 Observation/ examination	0.54	0.46	0.00	1.1
	1.4 Treatment	0.24	0.72	0.05	4.0
	1.5 Ambulation/ training	0.76	0.24	0.00	5.8
	1.6 Psychological and social support	0.63	0.37	0.01	2.3
	1.7 Communication and education	0.23	0.77	0.00	0.9
<b>Indirect care activities</b>	2.1 Own movement	0.60	0.39	0.02	8.4
	2.2 Transport outside ward	0.76	0.17	0.07	0.9
	2.3 Cleaning	0.91	0.09	0.00	1.6
	2.4 Food handling	0.92	0.08	0.00	7.5
	2.5 Rinse room work	0.83	0.17	0.00	1.2
	2.6 Store handling	0.97	0.03	0.00	1.0
	2.7 Drug dispense	0.01	0.99	0.01	6.2
	2.8 Pre and post medical work	0.06	0.94	0.00	1.1
	2.9 Report/briefing	0.37	0.60	0.02	7.9
	2.10. Nursing care planning	0.18	0.82	0.00	1.0
	2.11 Round	0.27	0.69	0.04	0.6
	2.12 Nursing care documentation	0.15	0.82	0.03	2.4
	2.13 Paper work	0.28	0.70	0.02	2.2
	2.14 Telephone	0.40	0.52	0.07	1.4
	2.15 Night duty/readiness	0.70	0.30	0.00	9.5
	2.16 PCS work	0.65	0.32	0.03	0.9
<b>Unit related work</b>	3.1 Paper work	0.23	0.17	0.60	1.9
	3.2 Telephone	0.09	0.30	0.62	0.6
	3.3 Personnel planning	0.35	0.31	0.34	3.3
	3.4 Co-worker dialogue	0.08	0.31	0.61	0.8
	3.5 Teaching and instruction	0.29	0.69	0.02	0.5
	3.6 Quality development	0.05	0.18	0.77	1.0
	3.7 Conference and education	0.34	0.17	0.49	1.9
	3.8 Computer problems	0.44	0.42	0.14	0.7
4.1 Personal time	0.61	0.35	0.04	8.9	
<b>Proportion of total workload</b>		<b>0.53</b>	<b>0.41</b>	<b>0.06</b>	

Table 14. Activity study results in terms of percentage of time taken for each activity and the proportion each staff group contributes with to cover the workload of the different activities.

The following sections deal with two different levels of abstraction used as modelling approaches and how the information from the activity study can be used. These sections present both the problems and possibilities of the different approaches and the importance of balancing the amount of time and effort put into the simulation models with the resulting benefits.

### ***6.5 Proper modelling abstraction level for implementation of modelling procedure***

The modelling aim of keeping the simulation model simple but not too simple has already been addressed in the thesis. The main objective is to select the minimum level of detail or maximum level of abstraction and still obtain accurate results that support the project's aims. Frantz (1995) describes model abstraction as "...a method for reducing the complexity of a simulation model while maintaining the validity of the simulation results..." and continues by defining a taxonomy to differentiate model abstraction techniques. He distinguishes three main groups within this taxonomy: model boundary, model behaviour, and model form modifications. Thus far, the current approach can be said to have already included several abstraction techniques, at least with regard to data representation. For example, in Table 14, which contains the activity study results, every activity area is an aggregation of several activities. Another example of abstraction is how the transition probabilities are modelled by using polynomial functions.

There are two main concepts of special interest regarding the adequate modelling approach; the scope and the level of detail of the model. The scope refers to the breadth of the model or, simply put, what should be included to meet the aims of the study. The level of detail refers to the depth of detail required for each element in the model (Robinsson and Bhatia 1995). It means that it is possible (recommended) to have different levels of detail in different aspects of the model. For example, a low level of graphical detail will most certainly not affect the model's accuracy or scope, although it might be damaging in terms of the model's credibility. On the other hand, if the aims of the DES study do not include patients' dependency and personnel workload variability, the model would be simpler but the scope would be much narrower. The work presented in this section deals with two models at two different abstraction levels or levels of detail, but with the same scope. The

simulated system of interest is the same and the aims are, if not identical, very similar. The preferred and fully implemented solution is the more abstract model.

Some of the questions that this section addresses are: How can the stochastic modelling approach be used and implemented in a simulation model? What is the adequate abstraction level for personnel workload modelling, considering the data available? What are its drawbacks and benefits? How can we maintain the scope and still have a minimum level of detail when addressing personnel requirements?

The background work of the two modelling approaches which use different levels of abstraction of the rehabilitation ward serves for the discussion and draws conclusions on the adequate level of complexity for the proposed method.

### **6.5.1 Detailed model**

Simulation models are always an abstraction of the real system of interest, but they differ in terms of the level of detail they include to mimic that system. The detailed model, even if it is an abstraction of the real system, is intended to include much of the intricate interaction between patients and staff. The simulation model was built to gain a better understanding of the dynamics of the system, including how under-staffing could affect a patient's waiting time to obtain help and how the ward's layout affected the percentage of staff time used for transport and logistical movements.

In order to be able to study these issues, a 3D model in the DES software Quest® was initially built. The model graphics, in terms of room size, room type and location were identical to the observed system, as were the staff shifts and their configuration (see Figure 22). The patient placement strategies, based mainly on gender and the consequence of moving patients between rooms, were also identical. The question was whether the data from the PCS would be sufficient to generate and schedule the great variety of patient and staff activities

#### **6.5.1.1 Data limitations**

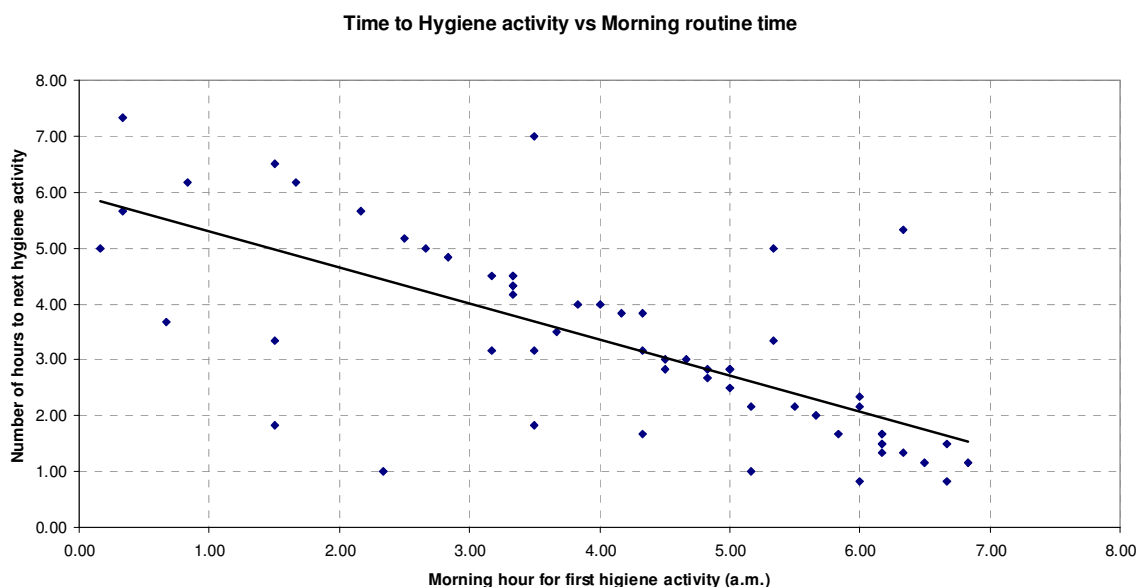
Data analysis is a demanding task even with structured data. The data analysis presented thus far has been limited to variations in dependency level and relations between the

indicator sums and their time representation. The time requirement for every indicator/determinator is based on the activity study, which also applies to the time requirements of the remainder of the activity areas presented in Table 14. The activity study, in its turn, is based on 7,700 timed observations, which represent 14 days of study by the different staff members. Every 10 minutes the personnel documented exactly which of the activities in Table 14 was being carried out. If a patient was part of that activity or if the activity was for the sake of a patient, the patient's identity was also registered. This data provides information about the distribution of a certain activity during the day and the aggregated time this activity might take. The reason for aggregated time is because the activity study monitored the activities at discrete times, every 10 minutes, which means, for example, one period represented 10 minutes whether that specific monitored activity took 1 or 19 minutes.

There are some drawbacks and limitations associated with the aggregation of time. It is not only obscures the multi-tasking facility of the personnel, but also the frequency of some activities. For example, a typical rehabilitation patient would receive medication 2-3 times every day. This is not evident in the data because delivering the drug to the patient takes only a fraction of the timed 10 minutes. Besides the time aggregation, the activity areas are a combination of several activities. For example, the activity area "Hygiene/elimination" includes helping the patient at his/her bed, as well as to, at and from the toilet or the shower. These aggregations are obviously necessary for practical reasons, but because the activity study does not capture every activity and its frequency it is not sufficient information for studying logistical issues. In the real system one would expect that the RNs and ANs have a greater number of activities but of shorter duration and, therefore, more movements between patients and/or in the ward's facilities. In addition, the only patient related activities recorded in the activity study were those where patients needed help from a member of the staff. This means that if a patient is, more or less, self-contained, his/her activities will not be monitored. While this does not affect the potential to use the model/data to evaluate the ward's personnel requirement, it impedes using the simulation model to identify the need of common resources, for example, determining the number of bathrooms and showers available and the possible waiting time before a patient can get access to them.

Part of this data limitation problem could be alleviated if the discrete interval time of the activity registration could be reduced or if the start and end times of the activity are recorded instead of having a discrete time registration (both Martorella (1996) and Gender (1989) describe PCS with activity study time registrations that register start and end times of activities). However, any of these changes would increase the administration workload of the activity study.

The data analysis of the detailed model adds a considerable workload compared to the aggregated model and contains numerous pitfalls. There are several events/activities that might at first seem random, but which, in reality, are correlated to other activities. For example, help with hygiene or elimination is random during part of the day and night, but definitively not during the morning activities. If the simulation operator chooses to model the events according to the Mean Time To Activity (MTTA) concept, he/she needs to limit that approach to the parts of the day when the activities are valid. Figure 33 presents the TT Hygiene for patients that received help with hygiene during the night shift, and it is clear that the next visit is correlated to the time remaining for the morning routines.



**Figure 33. Correlation between Time for Hygiene activity and the morning routines**

These activity patterns and their shifting nature during the day's 24 hours increase the data analysis workload enormously and create pitfalls which can cause mistakes in the modelling work.

### 6.5.1.2 Programming complications

Simulation software packages are usually user friendly up to a certain level of model complexity. In many simulation software packages the inexperienced user is able to build simple models with relatively little effort. However, as soon as the user needs to specify more complex behaviour or relations he or she is forced to leave the predefined package of functions and delve into the software's programming interface and language. Even at this stage the user will feel that he/she has some support and is not being left to drift aimlessly. Most software would have predefined pointers and functions that would make the simulation analyst's task simpler. With this assistance, as well as a general powerful and open programming environment, it is possible to build sophisticated models which represent real life systems extremely accurately. However, these possibilities come at a price and are not always entirely accessible to the end user

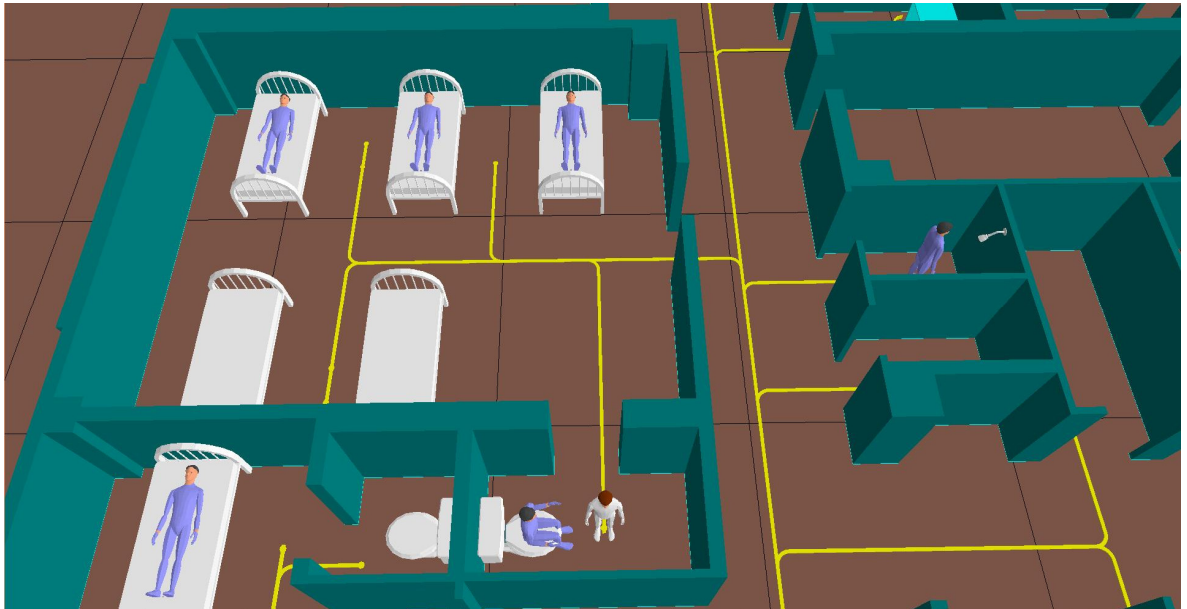


Figure 34. Snapshot of male patients' morning activities

The price is “blood, sweat and tears.” For example, the term debugging becomes a daily word and not an hour passes without more than a gentle reminder from your mailbox and now and then the sense of relief and fulfilment. The gratification is an artist's feeling of accomplishment, because even though the detailed model is of value, most of the resulting complexity is of no proper use. Chwif et al. (2000) identify three factors which cause many simulationists to increase complexity to an unnecessary level: the “show off” factor, the “include all” syndrome, and the “possibility” factor. The reasons for them seem to be human vanity, inexperience and technical advance.



The work approach of the detailed simulation model in this project was not an exception to the above mentioned behaviour and problems. The modelling of realistic human behaviour with its endless activities, interactions, selections and corresponding events, is very complex and time consuming. It demands tailor made procedures and functions for almost every element and scheduled activity. For each new implemented activity and its corresponding stochastic event generator, the complexity and data analyses increase and the software and program stability are put to the test. Debugging takes increasingly more of the modelling time, while monitoring programme modifications and their consequences demands a great deal of documentation and structure. The fact that most simulation software tools are not designed for healthcare systems, but for production systems, does not make the programming work easier. Instead, it forces the user to define and create his/her own functionality, graphics and distribution representation (the setup of predefined statistical distributions is, depending on the software, adapted to production systems and, therefore, the user is forced to implement, by programming, distributions better suited for healthcare systems). These statements, based on years of simulation modelling experience, reveal the extent of the difficulty of building models with a high level of detail.

### **6.5.1.3 Abortion of the modelling task and discussion**

During this loop of modelling, data analysis and verification of the simulation model, the task was aborted. The reasons for this decision became clearer during the progress of the project and were based on the previous discussion. Insight and experience prevailed over the artistic fulfilment of the work. Data limitations and programming complications led to assumptions and simplifications that questioned the validity of the model. Chwif et al. (2000) discuss that there is no clear relation between level of detail and model validity, on the contrary the article argues that a too high level of detail deteriorates the confidence in the simulation model.

This decision does not imply that PCS data cannot be used for detailed modelling, nor that a detailed simulation model of inpatient clinics is an impossible task. However, it does suggest that the activity study, with its discrete time interval measurement, needed to be adapted for the sake of the data requirement of the detailed model. In addition, the decision does indicate that the programming task is very great and that it requires a software

simulation package well adapted to healthcare systems and with an open architecture that enables the simulation analyst to model complex interactions, behaviour and events.

Of greater importance is to question the simulation analyst's motives and goals regarding the detail level of the model. Does it really meet the project's requirements in terms of time, cost and validity? Does the model's complexity add substantial value to the project's aim(s)? In this particular case it did not. There are examples of healthcare simulation models with a very high level of detail and 3D graphics, but these are used for the analysis of the architectural design of spaces in healthcare systems (Alvarado et al. 2003). Other detailed models are used to focus more on intangible parameters such as patient well-being, professional behaviour, organisational outcomes and patient centred design (Elf 2003). However, the reason for this work is different and not concerned with intangible aspects. It focuses on a methodology that helps the user model a healthcare system with an adequate staffing level, using existing data. It turns to the design of healthcare systems that are robust and can handle the stochastic variations they encounter. For these aims, the detailed model approach was inappropriate.

### 6.5.2 Aggregated model

There are two main aims for the simplification of simulation models. One of them is to decrease the work effort, time and cost of the project (Chwif et al. 2000; Madam et al. 2005). The second is to decrease the execution time of the simulation run (Johnson et al. 2005). Simulation runs based on models of the same scope but with a different detail level could differ as much as 10 times in execution time. If the analyst aims to use simulation based optimisation (SBO) or the output of the model to verify scheduling or planning alternatives in real time, a fast execution time is extremely important. For example, SBO may use several hundred simulation runs before an optimal solution is found, which means that every second of reduced time is therefore crucial. The relationship between abstraction and other activities of a simulation project are illustrated in Figure 35.

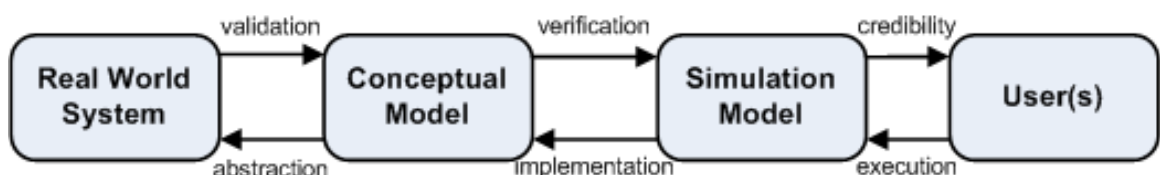


Figure 35: Model abstraction in the simulation process (Frantz 1995)

It is therefore reasonable that a simulation analyst does not include a needless detail level of features and complex logic in a model. On the other hand, abstraction of simulation models does contain several pitfalls, of which problems of validity and scope reduction are the most critical ones.

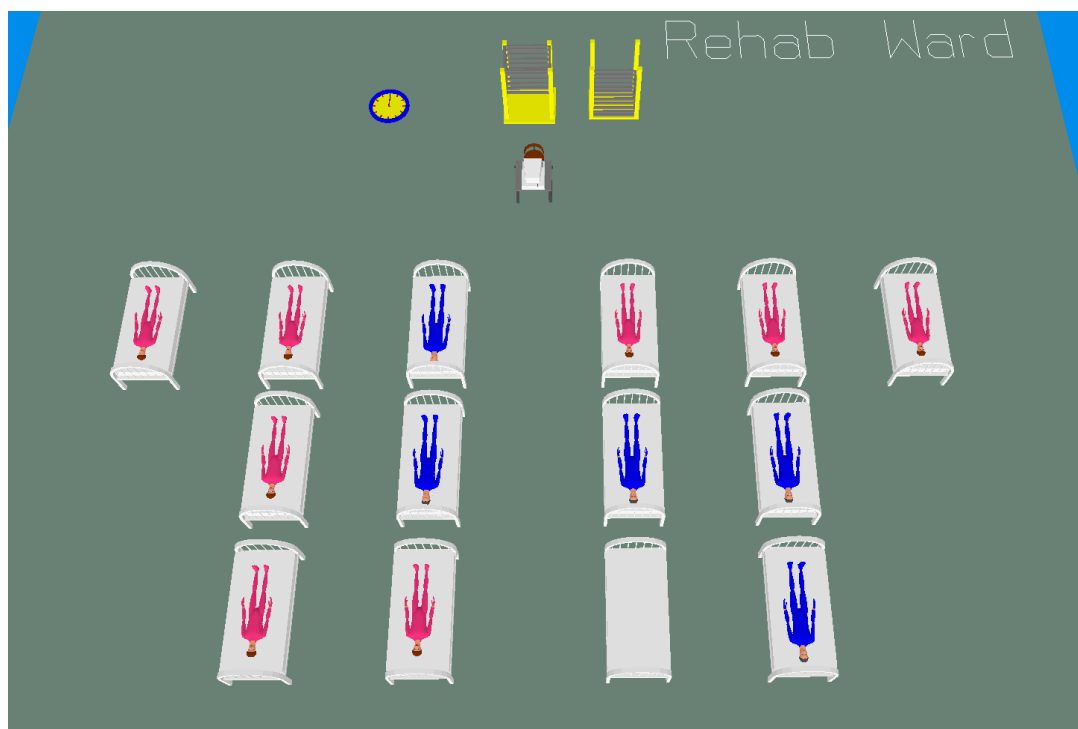
### **6.5.2.1 Model description and simplifications**

The aggregated model approach represents an abstraction of the detailed model, with a much lower level of detail, including simpler logic and graphical representation. It is called aggregated because the patient's behaviour and workload requirement is aggregated, compared to the previous approach. The aim and scope of the model is somewhat identical to the detailed model. It aims to be a tool that properly simulates the variation in workload for use in robust planning and designing of healthcare systems. The scope is the entire ward, including all patients, bed places and the patients' generation of nursing workload. It does not include numerous activities, graphics, behaviour and logic. If the aim of the model would have demanded such details, this would obviously have narrowed the scope, but that is not the case here, as will be shown.

The model contains only 19 elements of which 14 are bed places, see Figure 36. The connections are simplified to class connections, which facilitates patient routing and reduces program execution time. No considerations are made to patients' room placement and there is no personnel to help patients with any activities. The only scheduled activity is dependency level calculations, which are done once a day at noon. Immediately afterwards, the patients who received a dependency state H (home) leave the ward. At 3 p.m. new patients arrive in the ward.

The rehabilitation ward is never overcrowded and the patients that arrive are all scheduled. Consequently, there is a relatively loose way of scheduling new patients to the ward, which means that the ward nurse manager can adapt the number of patients they accept to the available workforce. This is not the case for many other units, including the orthopaedic trauma unit in ward 84 (see Figure 9) or the maternity unit, which means these units have a very severe planning problem. With regard to the rehabilitation ward, despite needing new patients who require rehabilitation, it is common with a postponement of at least one day before a new patient utilises an available place.

The type of patients, in terms of diagnosis, is randomly selected based on the percentage of different patient groups calculated from the historical data (see Table 15 in section 7.2.1). From the time the patient enters the ward, the simulation model will randomly select the coming day's dependency level of the patient, on the basis of the appropriate diagnosis related matrix, the patient's current dependency level and the time the patient has spent in the ward.



**Figure 36. Aggregated model view of the 14-bedplaces rehabilitation ward**

Compared to the detailed model, it is many times faster to develop and execute the aggregated model. The model calculates the total workload by aggregating each patient's workload contribution to a total sum. In addition, it distributes the workload time, in per cent, to different hour slots and allocates the workload time to the different staff categories according to the distribution found in the activity study. Other outputs are the number of patients treated and the ward's utilisation level. If the scope of the simulation model would have embraced other units, such as wards 83 and 84 (see chapter 4), it would have been possible to see waiting time for patients before they enter the wards as a viable output.

### **6.5.2.2 Scope and aggregation options**

There are several options regarding scope and information representation for the aggregated model which are easy to implement. For instance, in terms of workload management, it is much more interesting to study the morning and day shifts instead of the

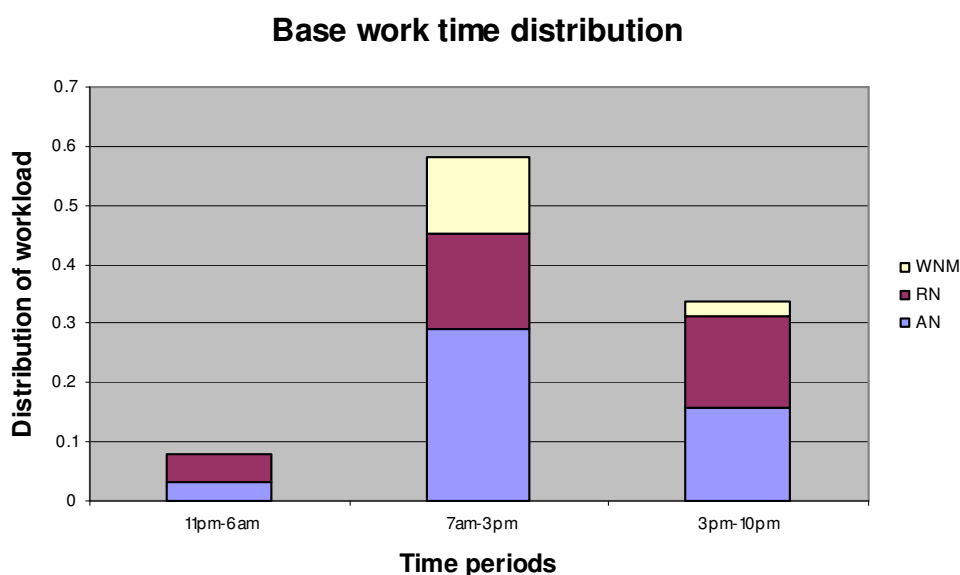
night shift. In inpatient clinics, the night shifts generally do not have a high workload. The staffing requirements during the night are based more on having enough personnel to guarantee patient well-being in case of a medical emergency. Table 14 shows that activity 2.15 “Night duty/readiness” represents 9.5 per cent of the staff’s total time. This figure corresponds to the average of the 24 hour activity study. If the time interval between 11:00 p.m. to 7:00 a.m. is studied, the figure for the same activity is 38.8 per cent, while for the busiest period of the day, the morning to lunch shift, it represents only 2.5 per cent of staff time. This is very important from a staffing and planning point of view.

There are two ways of addressing these differences in the periods of the day. One is limiting the scope of the simulation model by just studying the periods of the day that are of special interest. The second option is by dividing the day into periods of interest and studying how the patients’ requirements are met by the different staff categories in that particular period. For example, it was previously calculated that a specific dependency level B patient generated 453.9 minutes (7:33.54 hours) of total workload during a day, of which 283 minutes corresponds to the general *Base work time*. This *Base work time* represents the activities associated with the management and operation of a ward that are not patient specific. The total sum of these activities is distributed as a total figure of aggregated time among the different patients and represents an average of both weekdays and weekends.

The *Base work time* contains activities such as “personal time” and “night duty/readiness” which are not related to the actual workload. Personal time, for instance, includes the time the personnel use for all legitimate breaks (excluding lunch), for example, smoking and coffee breaks. The activity study reveals that 8.9 per cent represents personal time. This figure is in accordance with the legal regulation that stipulates personal time to be 8.3 per cent. Normally, this time allowance is aggregated and scheduled for longer breaks every three hours. The problem with taking these breaks into account is that they influence the pattern of the workload. A second problem related to the “personal time” is that it is based on the personnel’s scheduled time and not on their effective time. For example, a significant percentage of the night shift’s time relates to readiness. The 8.3 per cent allowed for personal time, according to regulation, is also calculated on to this readiness time, creating a significant personnel surplus despite the actual need being much lower.

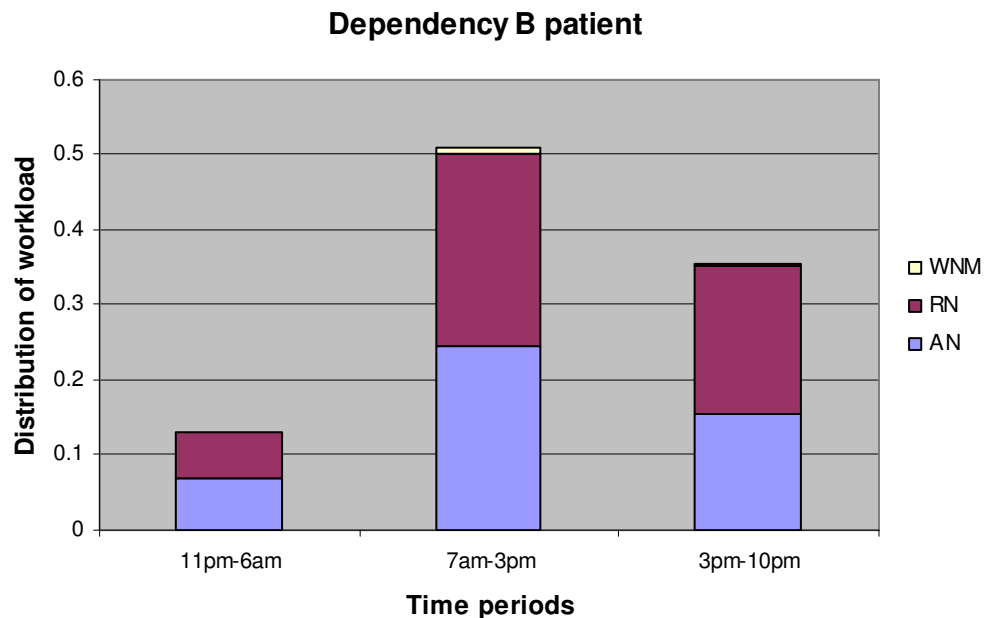
From an analysis point of view, it might be of more interest to study the real patient requirements and thereafter subtract both personal time and readiness from the activity study. The personal time can be added retrospectively, when the real workload has been determined and the different personnel schedules are established.

If the *Base work time* is recalculated without the mentioned entries, the aggregated time is reduced from 283.2 minutes to 196.3 minutes per patient and day. This new figure is still an average of both weekdays and weekends, for which the corresponding figures are 221 and 117 minutes respectively. The distribution of this time among the staff categories and periods of the weekday is presented in Figure 37 (during weekends a different distribution is used). It can be seen, for instance, that 13.2 per cent of the 221 minutes is part of the WNM's workload between 7 a.m. and 3 p.m.



**Figure 37. Weekdays' *Base work time* distribution**

The remaining time for our dependency B, example patient is the difference between 453.9 and 283.2, which consists of *Direct patient care time* and *Remaining patient care time*. The distribution of this aggregated time among staff categories and periods of the day differs between the different dependency groups (A-D) as well as between weekdays and weekends. Different percentage tables are therefore used with regard to calculating how the generated, patient workload requirement is distributed among staff categories and periods of the day. A graph of the dependency B group distribution for weekdays is presented in Figure 38.



**Figure 38. Distribution of patient's total workload requirement over periods of the day and staff categories for weekdays.**

The activity study's data registration gives the analysts the possibility to change the scope or detail level of the way in which the simulation model transforms the patients' dependency levels to workload for the different staff categories and periods of the day. The model's output provides the summarised workload per category for the periods of the day. The selection of the length and number of periods per day is freely determined by the project's members. Used in an intelligent way, it gives a more profound understanding of the variability of the system's workload.

### **6.5.2.3 Model limitations and discussions**

Changing from the detailed to the aggregated approach leads to some limitations, principally in the stochastic behaviour of the model. The stochastic limitation in the aggregated model is related to the internal arrival of patients with a corresponding diagnosis, to the dependency level development and the workload generation of indicator/determinant. This generated workload is then divided into periods and aggregated sums of several activities or blocks (*e.g. remaining patient care time, base work time*), without applying randomness. The detailed model, on the other hand, generates stochastic events based on the workload per direct care activity and other key activities per time period. The events would occur for the same time percentage as in the activity study, but

they would be randomly spread according to the MTTA concept, and the events activity time would also be a stochastic variable.

And that is the trade off, the price to pay for aggregating data, behaviour and elements. The question is whether it is worth it. On the one hand, the aggregated model will provide information on the workload generated from the patients' dependency variance and it will give this information relatively easily, structured, in time periods and per activity blocks. The model's results will provide healthcare managers with information about how a different patient mix, ward structures and scheduling options, not only affect the bed place requirements of the system but also the personnel requirements. It makes possible the building of systems that are better able to adapt to or tolerate variance, without having to accept either high costs or inappropriate service levels. The detailed model, on the other hand, would undoubtedly give a better and more realistic view of the ward's activities. However, this view would nevertheless be an abstraction of reality and the staff could still argue that it does not capture the true versatile nature of humans and, as a result of the data limitations, it would be based on too many assumptions. The answer to the question is therefore, it is worth it!

## ***6.6 Modelling Methodology Summary***

Thus far the focus has been on presenting a modelling approach, step by step. Practical suggestions and best practices have been argued and discussed. The aim has been to give the reader a logical approach while questions and uncertainties have been accumulated in the process. This section focuses on clarifying the methodology as a résumé of the earlier sections. The following steps and activities have been identified:

### **1. Acquire PCS data**

- Activity study data
- Patient dependency level evolution data
- Indicator/determinant distribution data

### **2. Aggregation of patients into patient modelling groups**

- Based on diagnosis, DRG, gender, age or other correlated factors
- Make a trade-off between significantly large groups for modelling feasibility and workload, and smaller groups for more distinct group behaviour



### **3. Calculate dependency transition matrices for each group.**

- Make use of a central moving average or other smoothing techniques to calculate point estimations.
- Adapt transitions to the absorbing state (leaving the system) of the LoS distribution probabilities calculated by equation (5 - 7)
- Balance the remaining transition probabilities following the procedure, as equations 6 and 7 present
- (The discharge modelling adaptation is not part of the transition matrices; it is done at vector level in the executing model.)

### **4. Modelling indicators and the determinant probabilities**

- Calculate determinant probabilities per indicator and dependency level (see Table 13)
- Follow the modelling suggestion illustrated in Figure 32 in order to avoid bias among the indicator combinations

### **5. Use the activity study data to:**

- Obtain the workload equivalents for each indicator/determinant value
- Identify how the personnel contribute, in proportion, to each activity (see Table 14)
- Identify how the patient requirements and personnel's workload are distributed during the 24 hours of the day, weekdays and weekends. Define a work profile, based on a percentage table, for the different personnel categories.

### **6. Implement the dependency level matrices and workload features into the DES model, complementing with:**

- The internal arrival distribution, which identifies the number of patients admitted daily (with its variation).
- The different patient groups' proportions (or mix) of the total amount.
- Arrival and discharge logic and, if necessary, complement with the discharge adaptation of the transition vector, see section 6.4.8.

### **7. Collect patient dependency and indicator variation data for validation purposes of:**

- LoS of transition matrices

- Proportions of dependency groups
- Workload level statistics

These are the main steps. Some of them need not be done in the presented order. However, once again it can be stated that it is not the intention of this modelling methodology to exclude the normal steps in a DES project methodology. The work represents a modelling methodology that aims to help the simulation analyst solve modelling and data acquisition difficulties in each step and thereby be able to successfully model patient dependency variation. In the proposed approach, the final step, which is step 7, concerns the validation of the matrices' outputs, focusing on three key measurement parameters: LoS, dependency level proportions and workload level, and variation. The results from the rehabilitation ward model and ways of verifying these measures are presented in the next chapter.

## **7 Chapter seven: Verification, validation and future system modelling**

### ***7.1 Introduction***

Does the proposed dependency modelling methodology adequately represent the patient's and, in extension, the system's behaviour? How is it used and what can be achieved by its use? The answers to these questions are crucial. The proposed modelling methodology needs to be verified and validated, as well as exemplified and evaluated. This chapter addresses these issues, describing the validation work and identifying the key parameters with which to validate from the methodology's perspective and suitable tests for these parameters.

The chapter aims to concretise the methodology through the use of a realistic scenario, an experimental case study. The study represents a conglomeration of system design best practices, as well as data and scenarios from the previous three case studies. The outcomes of the study exemplify why inpatient system variability is more than patient arrival patterns, and that efficient system design needs to deal with more than ward size, patient routing and patient mix levels. It needs to address staffing levels for both the efficient use of resources as well as patient well-being.

### ***7.2 Validation procedure***

Verification and Validation (V&V) are crucial steps in all simulation projects. Verification is concerned with determining whether the conceptual model has been correctly translated into a computer programme (see Figure 35). It involves, among other activities, debugging, programme trace and, if possible, control of functionality through the animated model. The verification of the model is partly discussed in section 6.5.2 and there is no further reference to it in this report. It only involves traditional activities and is not related to the modelling approach. Validation, on the other hand, is the process of determining whether a simulation model is an accurate representation of the real system or system of interest (Law and Kelton 2000; Banks et al. 2001). Validation is, in this case, of special interest because it is used to verify if the patient dependency modelling is adequate. Validation involves a verification of input modelling and its transformation to outputs or responses. It

might require a certain calibration of the model and it is not always totally conclusive (Kleijnen 1999), but it always aims to verify whether the simulation model is a credible representation of the system of interest.

The most important parameter to validate is the output(s) or response(s) of interest. A model which has been validated for a certain response is not necessarily valid if the aim or focus of the model changes. In our case, the most important response is the model's ability to represent the patient dependency variation, which includes both an average and variation pattern.

A widely used validation approach, presented in Banks et al. (2001), includes the following steps:

1. Build a model that has high face validity.
2. Validate model assumptions.
3. Compare the model's input-output transformation to the corresponding input-output transformations of the real system.

The first step, *face validity*, involves presenting the model, its results and animation, for users and others who are knowledgeable about the real system which is being simulated. Their opinion is invaluable, even though they might lack formal knowledge of operation techniques. Their comments and inputs might help to calibrate the model and/or improve its representation of the real system. It is important to get an acceptable "rating" from these system users in order to gain credibility for the simulation model and its ability to represent the real or future system. Here is a small, but still significant, difference between a more traditional validation procedure and the one presented in the thesis. While in a traditional project the aim is normally to improve a system's output (e.g. higher throughput, lower waiting times) or a new system design, the main aim of the rehabilitation project is to find a suitable modelling methodology. This means that the system's outputs are used to validate the modelling approach. This aim makes the validation procedure more concerned with the modelling assumptions and just one input-output transformation, namely, whether the patient mix generates the same patient dependency variation as in the real system. Therefore, prior to presenting the results from steps 2 and 3 of the validation procedure

described above, the following section provides a closer explanation of the system's patient mix.

### 7.2.1 Patient groups

Between February 2002 and September 2004, 676 patients were rehabilitated in ward 75. Most of the patients, over 50 per cent, suffered some sort of fracture. The identified groups are presented in Table 15 together with their main statistics. These statistics describe differences in LoS and in dependency. All the groups represent an aggregated number of diagnoses, but two of them, *remaining fractures* and *remaining diagnoses*, account for a larger aggregation representing more than 10 diagnoses. The *remaining diagnoses* group also contains patients with unidentified diagnoses. One of the groups is notable with regard to its LoS figures; *fitting/adaptation of arm/leg prosthesis*. This group has a very distinct ward scheduling. The patients arrive on Mondays and leave on Fridays, which constitutes four whole days. Their arrival pattern has therefore been modelled differently compared to the other groups. There is one exception in the data, a patient that stayed in the system for more than 150 days. It is not entirely unusual for some patients to stay for more than 80 days, but for modelling purposes the LoS is limited to 100 days, which covers 99.8 per cent of all patients.

Patient group name	Nr. of patients	%	LoS			Patient days (%) per dependency groups			
			Average	Median	Variance	A	B	C	D
Contusions	59	9	13.7	12	84.7	9.0	40.8	43.2	7.0
Hip and pelvis fractures	59	9	15.1	14.5	63.5	7.6	43.0	45.5	3.9
Femur fractures	149	22	17.6	15.5	78.5	10.8	47.6	39.2	2.5
Remaining fractures	142	21	17	13.5	185.6	10.3	48.9	35.0	5.8
Osteoarthritis	50	7	12.7	13	31.4	26.8	57.6	15.1	0.5
Fitting/adaptation of arm/leg prosthesis	39	6	4.2	4	1.3	15.8	43.0	34.2	7.0
Remaining diagnoses	178	26	15.1	11.5	170	9.2	48.7	33.4	8.7

Table 15. Identified groups and their LoS and dependency statistics

Every group is distinctively different in terms of LoS and dependency variation, but there are similarities. Fracture patient groups have a similar dependency development and several groups have very high variance figures. This means that changes in the patient mix do not necessarily affect the total system behaviour, but at the same time the high variance

stresses the need of an accurate, patient dependency modelling. The *femur fractures* patient group is used to present the different validation procedures and steps in input analysis. However, in the input-output transformation analysis, all the groups, in their corresponding mix, are illustrated in the validation procedure.

## **7.2.2 Key parameter measurement**

The following two steps in the validation procedure involve validating model assumptions and input-output transformation. These two steps embrace many different techniques and aspects of a simulation model. The focus here is on determining some of the procedures that are of special importance in the present case.

The task of input data analysis falls within the validation of model assumptions. The Markov process modelling approach used for patient dependency development is, from the DES model's perspective, considered to be part of its input data. It should be remembered that from a system design perspective our interest is to measure the system's workload, including its variance, and not how patients change the dependency level. It should also be considered that the dependency level provides the basis for the generation of the indicator/determinant before a workload figure is calculated.

Three parameters in need of validation have therefore been identified. Two of them are related to the Markov process modelling and can be considered to be input data for the model. They consist of the LoS matrix modelling including the newly proposed discharge modelling and the dependency level proportions for each patient group (see Table 15). The third parameter is the input-output transformation where the real system's workload and workload variation are compared to the model's equivalent.

### **7.2.2.1 Length of Stay and discharge modelling**

In sections 6.4.7 and 6.4.8, the modelling approach for LoS modelling and discharge adaptation is fully explained. This section limits its contents to presenting the outcomes of that modelling suggestion. The mathematical expressions in equations 6 to 10 (see section 6.4.7) present how the probability calculations of the matrices are adapted to the LoS distribution, visualized in Figure 29. This distribution has been fitted to the observed data and validated through the traditional input data analysis, which includes both heuristics and

analytical tests. The validation procedure described here is therefore more concerned with verifying whether the mathematical conclusions are correct and implemented correctly. Consequently, despite that there should not be any doubt about the equivalence in LoS generated by the LoS fitted distribution and the matrices, from a mathematical point of view, it is still of interest to confirm this modelling step in two different analyses. The first compares statistical indicators between the observed data and data from 20 simulation runs. The second approach performs the Kruskal-Wallis test of homogeneity between the observed data and the results from the different simulation runs.

The validation procedure used 20 simulation runs. Each one collected the LoS from the first 149 patients that entered the simulation model. The sample size was equivalent to the size in the historical data (see *Femur fractures* in Table 15). For each one of these runs, the average, the variation and the range (min and max LoS) were specified. The results varied between the runs due to the stochastic nature of the DES. The average of these 20 runs and a hypothesis test based on 19 degrees of freedom ( $n$  being 20) with a significance level of  $\alpha=0.05$  was calculated on each one of these parameters and compared to the historical data. The results, presented in Table 16 show that all of the statistical indicators fall well within the null hypothesis  $H_0 : |T| \leq t_{\alpha/2, n-1}$ .

	Observed data	Average from simulation runs	Standard deviation	$ T $	$t_{\alpha/2, n-1}$
<b>Average</b>	17.56	17.60	0.68	0.24	2.09
<b>Min</b>	2.00	2.05	1.00	0.22	2.09
<b>Max</b>	51.00	48.50	6.25	1.79	2.09
<b>Variance</b>	78.44	77.93	12.39	0.19	2.09

**Table 16. Summary of main statistics and hypothesis-testing**

Although this data provides an important verification of the resemblance between the historical data and the simulation results, it does not indicate whether the samples are distributed in a similar way. One way of further examining if the simulation runs' LoS outputs and the observed historical data are from the same distribution is by performing the Kruskal-Wallis test of homogeneity (Law and Kelton 2000). The Kruskal-Wallis hypothesis test is a nonparametric test that does not require normal distributed data. It is used to measure the median in  $k$  independent samples. The test ranks the different

observations and assigns a total rank value to each of the different samples. If the values between the samples are “close enough”, the test statistic will have a lower value than the critical levels for  $\alpha$ . We reject the  $H_0$  at level  $\alpha$  if  $T > \chi_{k-1,1-\alpha}^2$  where  $\chi_{k-1,1-\alpha}^2$  is the upper  $1 - \alpha$  critical value for the chi-square distribution with  $k-1$  degrees of freedom. As Table 17 presents, all the samples of the 20 simulation runs were considered to be from the same distribution as the observed data, irrespective of the alpha value.

Homogeneity test Kruskal-Wallis		Critical Values for the Level of Test (alpha)				
Test statistic	Degrees of freedom	0.25	0.15	0.1	0.05	0.01
15.42	20	23.83	26.50	28.41	31.41	37.57

Table 17. Homogeneity test between observed data and samples of 20 simulation runs.

The result presented here is from the discharge adapted logic and it confirms that the adaptation, on discharge per cent per day, does not affect the LoS distribution. The equivalent test was performed for a different set of simulation outputs, which did not use discharge adaptation. The results were as accurate, not more or less, but obviously missing the discharge correlation found in the real system. The discharge modelling approach can with advantage be used for simulation models that do not take patient dependency levels into consideration. It then only requires minor modifications to the traditional way of modelling (see section 6.4.8). These results verify the advantage that the transition matrices’ LoS modelling with discharge adaptation has over the traditional LoS modelling and at the same time serve as a V&V of the method.

### 7.2.2.2 Dependency level proportions

During their stay, the patients will contribute with a number of days in one or several dependency levels {A, B, C, D}. The total number of days that they contribute with depends on the LoS modelling, but the proportion of days in the different dependency levels depends on the transition probabilities between these levels. Since these probabilities are modelled through the use of central moving averages or polynomial regression (see section 6.4.6), there is a risk that the assumptions made in that modelling step are not accurate. The dependency level proportions will affect the system workload and is therefore crucial for the model’s objective. If a mismatch between the observable data and the model’s output is found, the simulation modeller has the opportunity to redo **step 3** in the modelling methodology (see section 6.4.5).



The validation results for the *femur fractions* group are presented in Table 18. They are based on the same 20 simulation runs as the data presented in Table 16. The result shows that dependency group A falls outside the  $H_0$  level for acceptance. The difference, even though it is measured as inconsistent with the null hypothesis, was considered as minor and ignored.

Dependency groups	Observed data	Average from simulation runs	Standard deviation	$ T $	$t_{\alpha/2, n-1}$
A	10.77	11.39	0.99	2.81	2.09
B	47.56	47.29	1.86	0.65	2.09
C	39.21	38.68	2.22	1.06	2.09
D	2.46	2.64	0.55	1.43	2.09

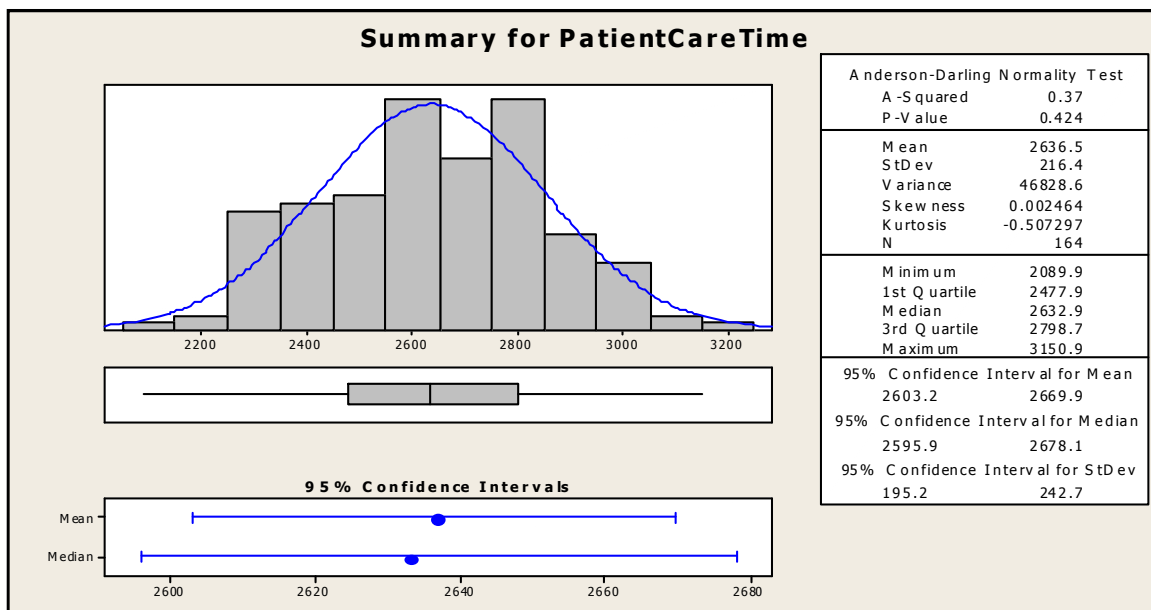
**Table 18. Calculations of dependency level proportions calculations and hypothesis-testing based on 20 simulation runs.**

### 7.2.2.3 Workload level and variation

The workload level and system variation constitute the most important and conclusive step in the validation procedure. Its results are obviously based on the V&V of the earlier steps and of the simulation logic. This step validates the complete simulation model. It contains all the patient groups and their stochastic arrival in per cent respectively and includes the groups' specific transition matrix. Moreover, it validates whether the input-output transformation of the simulation model is an accurate representation of the real system. The inputs are the patients entering the system according to the diagnosis mix presented in Table 15, and the output is the workload that these patients generate.

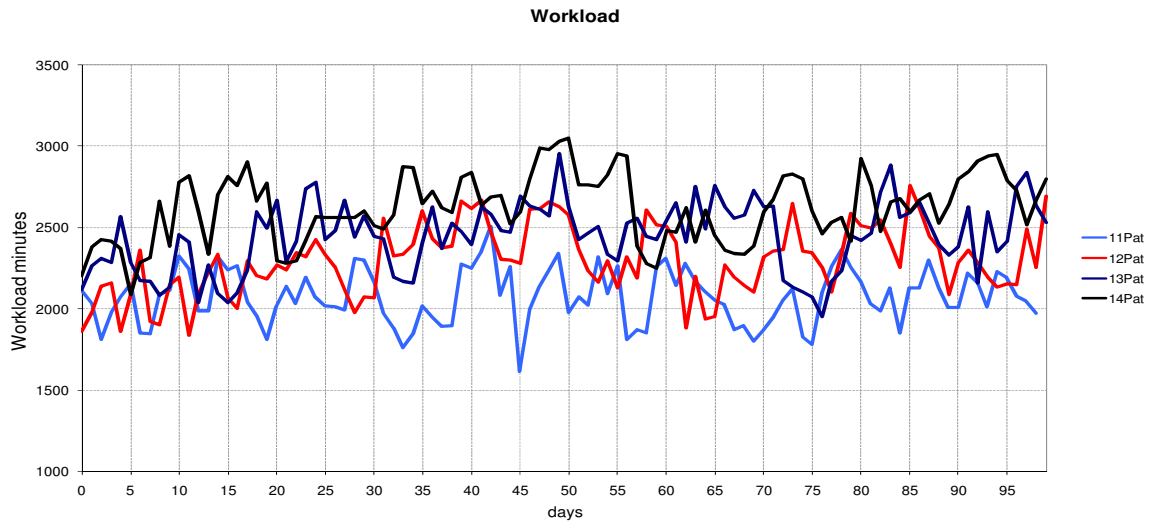
The observed data used for validation was obtained during 164 days between February 2002 and September 2004 when the ward had 14 patients registered, which corresponds to the maximum number of beds (other days when the number of patients was less than 14 were omitted). The data used was the ward's total workload of *Direct patient care time* and *Remaining patient care time*. *Base work time* was omitted because it is not affected by the patient's dependency level and is therefore a poor estimate for the variation of workload considered here. The workload measurements were taken from 3 p.m. to 3 p.m. (when new patients have arrived and old patients have left) and, consequently, 24 hours of workload

for each patient was always considered. The data analysis of this historical data is presented Figure 39.



**Figure 39. Summary of workload variations in minutes from 164 days of ward 75 where the number of patients is equivalent to 14, historical data.**

The time scale in Figure 39 represents minutes of workload, which have been generated by the historical dependency level of each patient. The total sum is the one visualized by the graph in Figure 39. What the graph emphasises is the importance of taking the dependency level and not only the number of patients into consideration. For instance, the fluctuation in workload for the same number of patients goes from a minimum of 2089.9 minutes to a maximum of 3150.9 minutes. The difference of over 1000 minutes of workload is not the theoretical maximum, which for 14 patients is almost 4000 minutes (4941 minus 1075 minutes). In Figure 40, approximately 400 days (about 100 days for each group) has been isolated from the historical ward data and divided into four groups representing the workload variation for days when the number of patients has varied from 11 to 14. The figure shows that the workload variations were considerable irrespective of the number of patients in the ward. A correlation analysis between the workload figure and the number of patients shows a sample correlation figure of  $r = 0.78$ , which suggests, however, that the number of patients is the highest correlating factor. Independent of these expected findings, there is still a considerable factor of workload randomness generated from the individual patient's dependency development.



**Figure 40. Workload variations in ward 75 for different numbers of patients and days**

The validation approach uses a set of five equally big samples (N=164) as the observed data visualized in Figure 39 and checks whether the main statistics falls within the null hypothesis. However, the work is limited to analysing data from days containing 14 patients for comparisons and practical reasons. The simulation runs from which the data was collected, presented the same patient number variation (a result of internal arrival and discharge figures) as the real system. From this data, only the days containing 14 patients were used. As Table 19 highlights, the results are well in line with the real system and all key parameters falls under the null hypothesis for  $\alpha = 0.05$ .

	Observed data	Average from simulation runs	Standard deviation	$ T $	$t_{\alpha/2, n-1}$
<b>Average</b>	2636.54	2643.10	40.60	0.36	2.78
<b>Median</b>	2632.92	2657.11	39.25	1.38	2.78
<b>Min</b>	2089.92	1914.01	365.26	1.08	2.78
<b>Max</b>	3150.92	3211.03	107.90	1.25	2.78
<b>Variance</b>	46828.57	47390.81	6785.00	0.19	2.78

**Table 19. Summary of main statistics and hypothesis testing for the workload measurement**

### 7.2.3 Validation conclusions

The aim of the validation work has been twofold. Firstly, to validate the adequacy of the modelling approach. Secondly, the work has focused on identifying the key parameters with which to validate from the methodology's perspective and to find suitable tests for these parameters. For future users, the focus should be put on the dependency level's

proportions and the input-output workload evaluation. The LoS' V&V procedure concludes that the modelling approach is valid if the correct theoretical distribution is identified (see Figure 29). On the other hand, the validation approach of the dependency level proportions is not as straight forward and might require some fine tuning. An important factor for accurate approximations of dependency level proportions is to have a sufficient amount of data. The sufficient amount is related to the quantity of historical data and to the grouping procedure. The data collected from ward 75 proved to be difficult to aggregate into homogenous groups because of the high variance in dependency and LoS among the same diagnosis and/or DRG groups. This is, for a layman, very surprising, but it was not for the healthcare professionals. It would be interesting to analyse whether the new rehabilitation codes introduced in 2008, the so-called NASS (SoS 2008b), are more accurate. Of even greater interest, due to its more generic solution, is the approach suggested by Marshall and McClean (2003), who make use of a more diversified number of factors for grouping which they term casual components (e.g. gender, age, admission method (reasons), destination, marital status). This relation between casual components and group behaviour can be analysed using data mining techniques. It is not until the variance is reduced and a more homogenous behaviour can be identified within the groups that this modelling approach can be used to more accurately predict the ward's short term workload development.

### ***7.3 Experimental case study***

Thus far, most of the discussion in chapters 6 and 7 has been centred on the theoretical background of the patient dependency modelling methodology. This section, on the other hand, exemplifies both the benefits and the limitations of this methodology in the design and analysis of future healthcare systems. It does this with the use of an experimental case study, which represents a conglomeration of system design best practices, data and scenarios from the previous three case studies. For example, some of the best practices used in this experimental case study are based on the maternity ward case study which clarified the need for robust system design in order to deal with high levels of inherent variations. The need to create a "bigger" system by a more flexible use of rooms and the correlation between occupancy level and service level are some of the issues addressed. Other best practices are highlighted in the orthopaedic study, especially in sections 4.7.2 and 4.7.3, and several suggestions for system improvement are simulated and analysed.

Another crucial piece of information was added by the rehabilitation case study. This study gave the basis for the dependency modelling methodology. Moreover, the data from the rehabilitation ward was modified and used in this experimental study. Wards 83 and 84 did not use Beakta® to trace the dependency level of their patients, but approximately 60 per cent of the patients enrolled in ward 75 came from wards 83 and 84. This means that they share the diagnosis of many patients. Therefore, although the dependency development and workload are different between the rehabilitation unit and the orthopaedic units, for illustration purposes the transition matrices from ward 75 were used with some important modifications, which are explained.

Finally, the scenario was based on the orthopaedic ward's case study. The discussion in chapter four states that the general size of wards at KSS was 16 or 18 beds, which meant that both wards 83 and 84 needed to be reduced. In section 4.7.2, a series of experimental settings are presented, and one of the changes, compared to the regular operation of the wards, was the rescheduling of contusion and trauma fracture patients to a hypothetical new ward. The aim was to be able to downsize the two wards. This section only reminded the reader of a proposed new ward but did not inform of its size or operational conditions. This information is provided in the current section.

The aim of the experimental case study was to visualise how the dependency modelling methodology implemented in a DES model can be used to address system parameters beyond the ward size, routing and scheduling options. It shows how it can be used to evaluate adequate staffing levels.

### **7.3.1 Experimental case system setup**

One of the enquiries made by the management of wards 83 and 84 concerned the number of bed places a new ward that nursed contusion and fracture trauma patients would need. In addition to the contusion and fracture patients from ward 84, the new ward was intended to care for fracture patients that were previously directly admitted from KAVA (surgical emergency care unit) to rehabilitation ward 75. The answer was therefore not as straight forward as it may have first seemed. Firstly, the arrival of patients is highly stochastic. They would be admitted directly from the ED to the new ward and could arrive at any hour of the day. Secondly, the LoS of fracture patients is highly variable, making it difficult to

obtain an even discharge pattern. These two sources of inherent variation made it impossible to design a single unit with enough bed places and that also maintained a high occupancy level. Even more troubling was determining adequate staffing levels for the unit.

The solution was found in addressing the issue with a bigger system approach, see Figure 41. The system contains three inpatient wards, each consisting of 16 bed places, which share a common reception. Ward 85 represents the projected new ward and is, together with ward 84, an inpatient trauma ward. The main concept is based on scenario three in section 4.7.2. Firstly, it consists of lowering the preventable variations by scheduling a constant number of three elective patients to ward 83 from Sunday to Thursday.

Secondly, it involves making it possible to schedule patients to one of the other two wards if, for example, ward 83 is full. This means that if necessary an elective patient might be temporarily placed in ward 84 or 85 and vice versa. The difference is that if the three wards are full, the elective patient's surgery will be rescheduled and the patient will be "bounced" from the system. A trauma patient arriving when the wards are full would be found a temporary bed place and stay in the system.

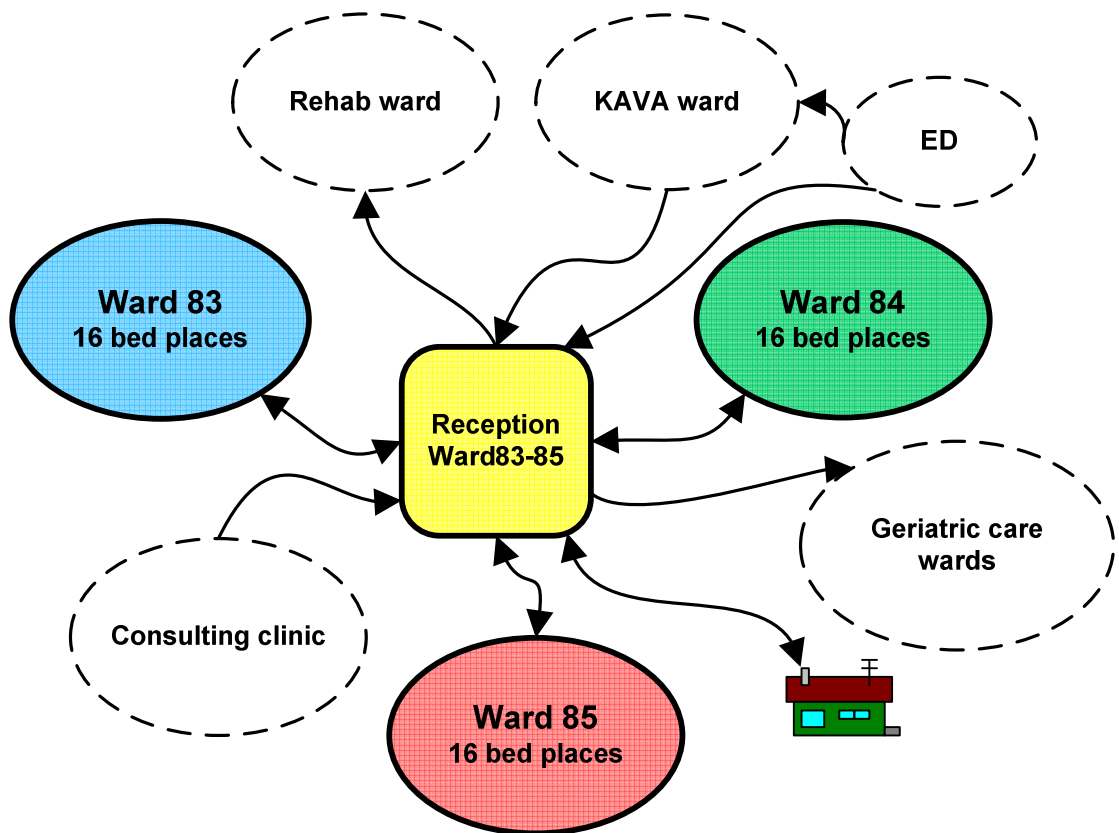


Figure 41 System configuration and flows, where ward 85 represents the new ward

Thirdly, the units would not only share bed places among each other, they are also intended to share personnel. This is not a polemic question because the personnel already practice work rotation between elective units and trauma units. In order to measure the amount and variability of the workload generated by the different patient groups and aggregated at unit level, the dependency matrices from rehabilitation ward 75 were used. These dependency matrices do not accurately represent the real workload generated by these patients, but are used here only for illustration purposes. If all inpatient units would implement a PCS, it would facilitate the simulation analysis of healthcare systems.

The dependency matrices used comprised the following. The elective patients in ward 83 made use of their equivalents in ward 75, which were the *osteoarthritis* dependency matrices. Two different matrices of patient groups were used for trauma patients in the new ward 85. For all trauma fracture patients, the matrices for the *femur fracture* patients were used, while for contusion patients the equivalent *contusion* matrices from the rehabilitation ward were used. The matrices for the *remaining diagnoses* were used for the remaining trauma patients in ward 83.

All the dependency matrices used from ward 75 were LoS modified. The LoS modifications were made by using the LoS distributions identified in chapter four for the different groups and the approach suggested in section 6.4.7. This means that all the patient groups received the same LoS figures as in the orthopaedic case study and they were given a workload estimation in line with the observations made from the rehabilitation case study. Finally, the total number of patients, dealt with in the three units, did increase because of the rescheduling of some patients from ward 75.

### **7.3.2 Experiments and analysis**

The objectives of the experiments were to illustrate the earlier findings in system design for variability and, more importantly, to show how the dependency modelling methodology can add a new dimension to system design, planning and evaluation. The new system design could now be evaluated from a workload balancing perspective.

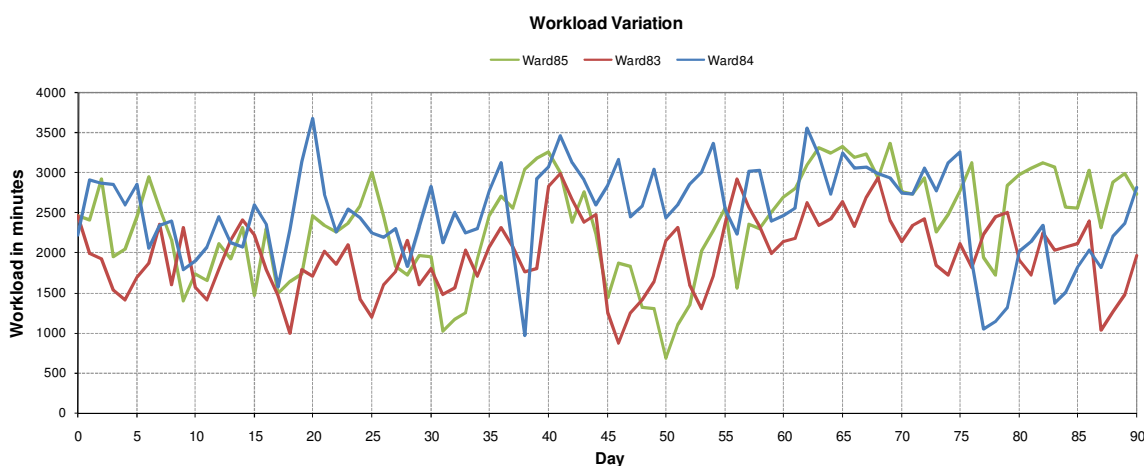
Two main scenarios were evaluated. The first had free scheduling of patients between the wards. This means that if a ward could not accommodate a newly admitted patient, the

model checked the other two wards for a bed place and, if one was found, the patient was moved. In the second scenario, only trauma patients had free scheduling, while elective patients could only be scheduled to ward 83. Occupancy figures and the number of rescheduled patients were the two guiding parameters. In Table 20, the average from ten simulation runs, each run equivalent to 120 days, during the two different setups is presented.

Autumn data, LoS 7.3 days:	Free scheduling of elective patients if no place is found in ward 83		Elective patients are only scheduled to ward 83	
	Nr. rescheduled	Occupancy level %	Nr. rescheduled	Occupancy level %
<b>Average</b>	2.1	79.5	17.5	79.2
<b>Stdev</b>	2.96	2.0	6.2	2.0

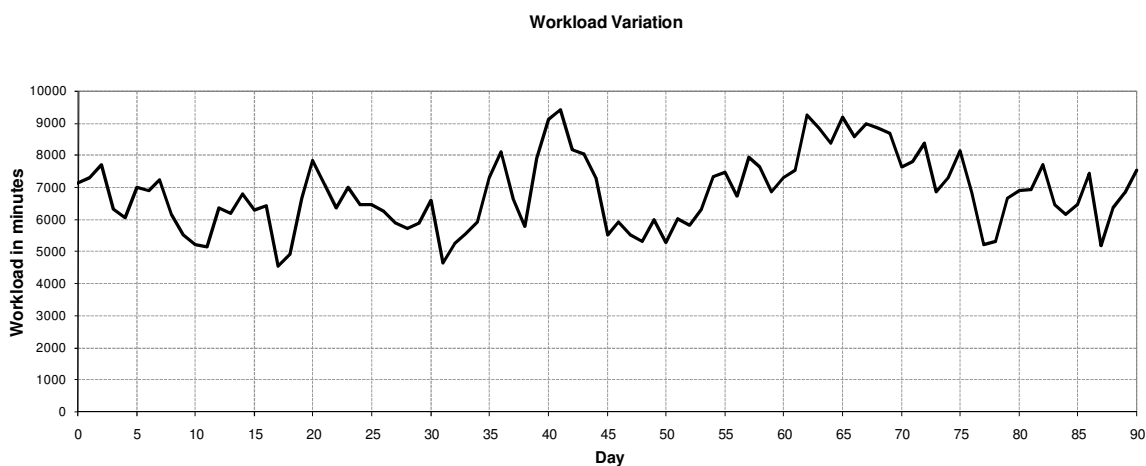
**Table 20 Results from the two scenarios**

The results were expected and well in line with our earlier analysis. Small health care systems are extremely vulnerable to inherent variations. It is therefore necessary to have a flexible and open structure in order to achieve efficiency and good service levels. Even more interesting are the results presented in Figure 42 and Figure 43. Figure 42, the day-to-day workload variance of the three wards is presented. As indicated, the workload variation is very high for each one of the wards, see Table 21. If these wards were operating individually, staffing would be extremely difficult. In addition, if the wards were staffed according to the average workload of the unit, the personnel would have either too little or too much to do, which would create unnecessary stress for the staff and a health risk for the patients.



**Figure 42. The day to day record of the three wards' workload variation, presented in minutes.**





**Figure 43. Aggregated workload variation for the three wards, presented in minutes**

Figure 43, on the other hand, shows the day by day workload for all wards aggregated under a single simulation run. Even though there still is a considerable day-to-day variance, the coefficient of variation is lower for the aggregated system than the one by one systems, see Table 21 (Table 21 presents the statistical results from the ten simulation runs of scenario one). Staffing for a larger system becomes easier than for each individual system on its own. The methodology visualises the results of a proposed system design and not only quantifies the workload average for the system's units, but also the variance in the workload. This information can be used to balance units against each other. The workload data can also be broken down in order to visualise the workload during particular hours of the day and can thus be used to design personnel shifts.

The workload information presented in these figures has not previously been considered in DES projects. One of the changes, in comparison to the old ward 84, consisted in taking some of the rehabilitation ward's fracture patients together with all the fracture patients from ward 84 and moving them to the proposed new ward. How does this move change the workload and workload variance of the unit? What is its dynamic and stochastic behaviour and how does it affect other units? These were questions that could not previously be addressed. Today, as a result of the proposed methodology and the combination of DES and PCS data, these questions can be answered.

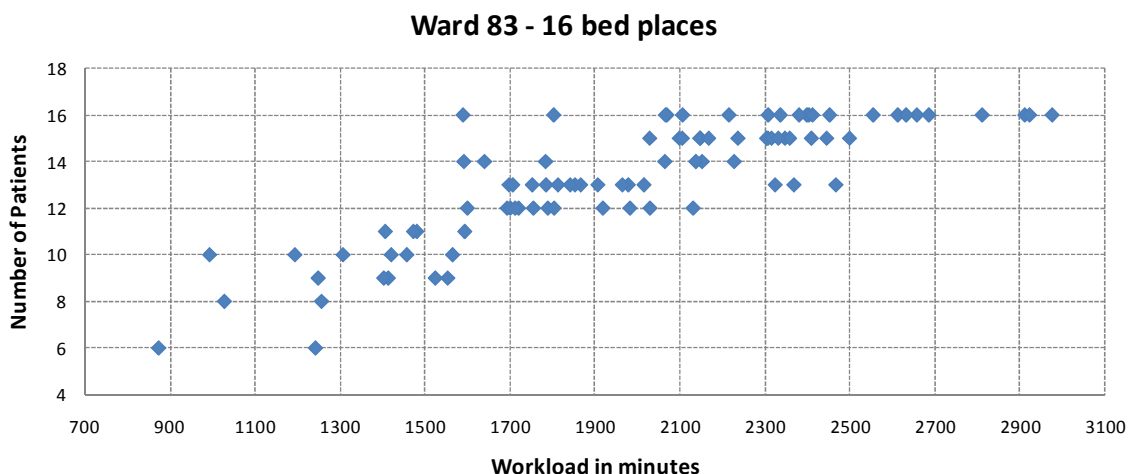
Table 21 shows that the coefficient of variation is slightly higher in wards 84 and 85 compared to ward 83. This is mainly due to the inherent variation in the arrival of patients.

The workload variation, observed in the system, is mainly correlated to the number of patients in the system, however, as previously mentioned, not exclusively.

Workload Variation				
	Average	Stdev	Coefficient of variation	Stdev
<b>Ward 83</b>	1870	442.5	0.219	0.019
<b>Ward 84</b>	2433	518.5	0.230	0.037
<b>Ward 85</b>	2217	486.7	0.247	0.024
<b>Whole system</b>	6520	846.9	0.153	0.014

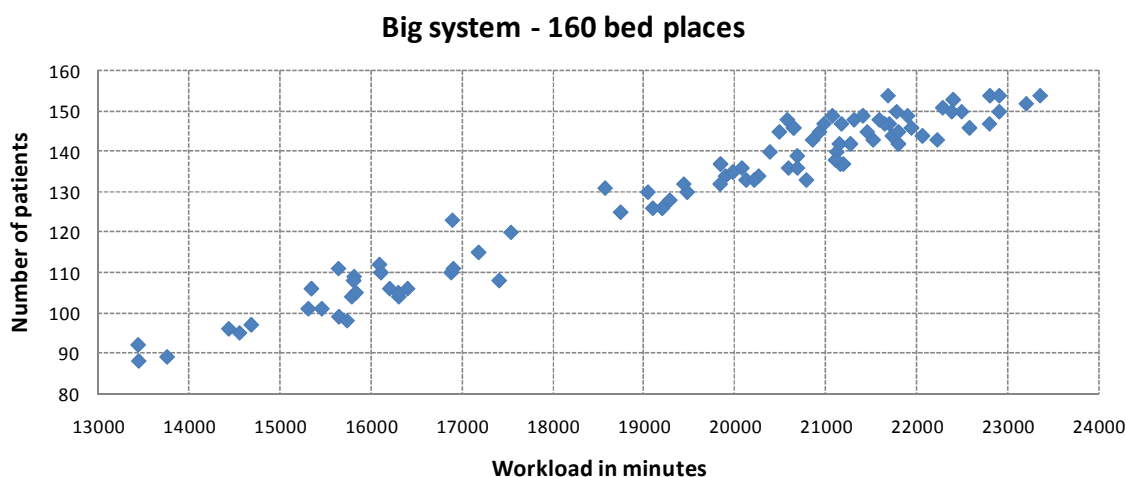
**Table 21. Workload variation for the wards individually and aggregated for the whole system, figures in workload minutes**

In Figure 44, a sample of 90 days' statistics from ward 83 is presented. Each value represents one day's relation between the number of patients and the unit's workload. This selection shows a sample correlation value equivalent to  $r = 0.84$ , which is a relatively high figure, higher than previously observed in the rehabilitation unit. One reason for this difference might be the disparity in homogeneity between the patient groups in the system. The rehabilitation ward consisted of several different patient categories with different dependency patterns, while the patient categories in ward 83 were limited to two large groups with more similar patterns. Another reason can be related to the shorter LoS, which could affect certain dependency trends. However, the results in Figure 44 particularly indicate that besides the number of patients in the system the unit's workload is dependent on the inherent dependency variations.



**Figure 44. The relation between the number of patients and the system's workload in ward 83.**

A different conclusion can be drawn when we study a larger system. For example, the data presented in Figure 45 represents data from a system equivalent to ten times the size of ward 83 with the comparable workload and number of patients, but scaled to a ward of 160 bed places. A correlation analysis between the number of patients and the workload of the system results in a sample value of  $r = 0.977$ , which is far higher than the one for the individual ward 83.



**Figure 45. The relation between the number of patients and the system's workload.  
System size 160 beds**

In other words, the correlation between the number of patients and the system's workload is stronger when the system becomes bigger. This conclusion suggests that the dependency modelling methodology is more suitable for analysis and evaluation when the systems studied are smaller or a conglomeration of interdependent systems. Can linear regression be the best choice when bigger systems are studied? It would depend on whether the workload data for the system is available and on the purpose of the analysis. The output data for this proposed system is based on historical data, but the new system's workload is a result of simulating the dependency modelling methodology. This modelling methodology has the advantage of facilitating the design and evaluation of systems, making it easy to test and try different mixes, patient volumes and system setups. The dependencies' matrices are a key factor in visualising the outcomes of those changes without any difficulty. Moreover, the methodology enables the design of a more balanced system and determines the adequate size and amount of resources irrespective of the systems' size or detail level of the study. This flexibility and versatility is not easily found in regression analysis.

### **7.3.3 Experiments conclusions**

The aim of the experimental case study was to clarify how the dependency modelling methodology complements a regular DES study by adding workload levels as a system design and evaluation parameter. Inpatient system variability concerns more than patient arrival patterns. Additionally, robust and efficient system design needs to deal with more than ward size, patient routing and patient mix levels. Staffing levels are crucial for both the efficient use of resources and patient well-being.

The results and analysis of the study show what has been stated earlier; size matters when it comes to dealing with system variations. More importantly however are the findings that a robust system design can help to alleviate the effects of high inherent variation. The first scenario with flexible routing shows higher occupancy figures and a lower number of rescheduled patients. Additionally, the workload coefficient of variation is lower when the three wards were considered as a single work pool and not as three separate units. This emphasises that with regard to addressing workload variation and its corresponding staff levels, a holistic system approach needs to be used.

The sense of suddenly being able to “see” the system’s workload variation and being able to take it into consideration in the design of healthcare systems is truly rewarding. The application of the methodology extends even further however. It gives the user the hour by hour workload of the units and, therefore, the possibility of shift planning. The future aim is to use it at an operational level, for day-to-day staffing. It would resemble the solution proposed by Warner (2006) but with the difference of using stochastic dependency matrices instead of pre-defined dependency patterns. Some basic requirements would include using a PCS system on a daily basis at the ward. This would enable the grouping of patients in order to reduce the inherent dependency variations and LoS variations (the data collected from ward 75 proved difficult to aggregate into homogenous groups). Additionally, the inherent variation of patients’ arrival would need to be reduced in comparison to the highly stochastic arrival pattern shown in the trauma wards.

## **7.4 Conclusions**

The results from chapters six and seven mainly concern a modelling methodology for inpatients’ dependency evolution and how that affects the workload of an inpatient ward.

The presented modelling approach is unique and constitutes a feasible and practical approach for simulation analysts who aim to address the question of adequate staffing levels for inpatient wards. One of the presented solutions, discharge modelling, is more generic and can be used without the need of dependency data or a staffing aim. The methodology opens opportunities for system improvement and robust system design that not only focuses on bed capacity, but also on the most important question, namely, staff requirements and adaptation. If the hospital uses PCSs in order to measure patients' dependency levels, the needed data is accessible and can be analysed and structured to answer the questions at hand. However, besides these two novel contributions, the chapters have addressed relevant topics, both in terms of system or domain understanding and with regard to modelling considerations and sound analysis practices.

Moreover, the experimental case study has visualised some of the potentials of the modelling methodology and summarised some of the findings from previous chapters. It has foremost stressed the need to take a system approach, avoiding preventable variations and designing robust systems that can handle the inherent variations. The combination of DES and PCS data has opened the possibility to “see” workload fluctuations in inpatient wards and understand how system design can alleviate its effects. Finally, it has emphasised how the combination of these tools can be used to deal with variability in the design, planning and evaluation of staffing requirements of inpatient units.

## **7.5 Summary**

Chapter seven focuses on the verification and validation of the proposed dependency modelling methodology. It attempts to answer questions concerning whether the proposed dependency modelling methodology adequately represents the patient's and thereby the system's behaviour, as well as how the methodology is used and what can be achieved by its use. The chapter is therefore divided in two main parts. The first part concerns the verification and validation procedures and results. It describes the validation work and identifies the key parameters needed to validate from the methodology's perspective and suitable tests for these parameters.

The second part focuses on answering questions related to how it is used and what the benefits and limitations of its use are. It addresses these issues through the use of an

experimental case study. The study represents a conglomeration of system design best practices, data and scenarios from the previous three case studies. The outputs from the study help to concretise the ideas presented in chapter 6 and exemplify them through a set of experiments and analysis. Moreover, the chapter reveals that the combination of DES and PCS data opens the possibility to “see” workload fluctuations in inpatient wards, as well as understand how system design can alleviate its effects. The study also serves as the source of a discussion on the methodology’s limitations, as well as the possibilities it brings to DES studies and future work.

## **8 Chapter Eight: Conclusions**

### ***8.1 Introduction***

This final chapter concludes with a presentation of the main results and reflections on the major findings of the dissertation. Moreover, it outlines the main contributions as well as how the aims and objectives of the research were met. The chapter finally presents a discussion of possible further work and research goals.

### ***8.2 Dissertation results in summary***

Healthcare providers in Sweden and elsewhere are facing increasing costs and reduced funding. Healthcare managers are thus forced to take into consideration the fluctuating demands and high variability facing their systems whilst, at the same time, making an efficient use of resources. This requires a novel approach to healthcare system design.

The dissertation presents a top-down approach, in which a system view is highlighted as fundamental for understanding and improving healthcare system design. It deals with the vulnerability of small systems to variations which need to be reduced, as well as designing systems that are inherently insensitive to variations. Hospital systems consist of complex interrelations between relatively small units, each of which is vulnerable to stochastic demand variations. This is not unique to Swedish healthcare. The findings of the case studies highlight the importance of designing flexible systems, and thereby creating bigger virtual units. The studies also show how variations can be reduced and how this would result in considerable improvements.

Nevertheless, the most important scalable resource within the healthcare organisation is its human resources. The right level of nurse staffing is of major importance for the patients' wellbeing and survival, which puts the planning and scheduling of human resources as one of the foremost aims of the organisation. Many tools for staffing and personnel planning in healthcare systems have been developed over the years, but none of them take into consideration the variable nature of the development of patient dependency levels and the resulting workload variations. Although DES techniques, in principle, have all the features

for modelling the variation and stochastic nature of healthcare systems, DES has not been previously used for workload studies of inpatient wards.

The main contribution of this work is therefore how a combination of DES and PCSs data can be used to model workload variations and consequently be applied to plan nurse staffing requirements in systems with high variability. One of the attractive details of this modelling methodology is that it makes use of data that is already available at many wards. The presented work provides step by step guidance in how the analysis and modelling task should be carried out. In addition, it embraces a thorough discussion on the adequate modelling level and advocates a more aggregated modelling approach which gives the simulation model an exquisite simplicity while still maintaining the necessary scope. The work also defines a novel modelling approach for dependency level and LoS modelling of a patient's dependency evolution, including an adaptation to the ward's discharge figures.

The validation of the modelling methodology confirms the accuracy of the LoS and discharge adapted matrices of the Markov process and the determinant (workload levels for different activities) transformation definition into workload time for the ward. Even more importantly than visualising the ward's workload variation is the possibility of taking this variation of nurse staffing needs into consideration when a robust staffing solution is evaluated. The modelling approach opens the way for a set of analyses and system design evaluations. For example, it enables an evaluation of how a different patient mix and/or scheduling of patients affects the need for nursing staff. Other analysis opportunities include the possibility of determining the size of a work pool with which to meet the need of additional staff in a number of adjacent wards with shifting workloads. If the patient group variance is moderate, it can be used to forecast how the workload will evolve in the following days.

### **8.3 Meeting the research aim and objectives**

The research aim of this work has been *to identify how and why DES can be effectively utilised to design, plan and evaluate inpatient healthcare systems and their nurse staffing requirements*. This aim has involved a more holistic view of inpatient healthcare system design. The project was motivated by a need for better system design and better staffing methods in inpatient care. These reasons are highlighted by the literature review and



grounded in empirical experience. This research involves presenting *why* we need better staffing methods in inpatient care (*e.g. sections 1.2.4.4, 2.4, 4.8 and chapters 6 and 7*), *why* it is important to address inpatient healthcare system design in general (*e.g. 1.2, 2.3, chapters 4, 5, 6, 7*), and *why* DES is the tool of preference (*section 2.5*). However, although the *why* of this research is easily appreciated by the reader, the *how* of things requires a more thorough reflection. This work has a predominant *how* approach because it deals with contemporary challenges and proposes feasible solutions to these challenges. The healthcare challenges are identified through a literature review and empirical studies. They involve both domestic (Sweden) and international issues and maintain a system view in which cultural and management involvement are not ignored (*see e.g. sections 1.2.3, 2.2.1, 2.6, 4.8 and 5.7*). Moreover, the work presents a number of design suggestions for inpatient wards, stressing *design, planing and evaluation* principles that have been corroborated through several case studies, thus answering the *how* approach several times.

There is nevertheless one *how* that is more prominent due to its unique and single appearance in this research area and that is *how* patient dependency fluctuation is modelled and linked to workload and staff calculations in a DES model. The answer is presented thoroughly in *chapter 6*, as well as verified and validated in *chapter 7*. It is presented in a step by step modelling methodology that covers every modelling step systematically and provides practical suggestions to problems that might be encountered. Without an answer to *how* to model patient dependency fluctuation, there is no answer to *how* to design inpatient systems that properly consider the fluctuating staffing requirements. Without an adequate modelling methodology there is no point in addressing the challenge of adequate staffing levels through the use of simulation. Such a staffing level avoids unnecessary mortality or poor quality treatment for patients while at the same time it is cost effective and does not lead to work overload and absenteeism among nurses.

Does this contribution lose credence because it is based on existing PCSs data? None at all, on the contrary, it gains importance because it is more feasible to attain. It is simple and easy to apply and truthfully answers the interrogative *how*. All simulation models, whether they are DES or of another kind are dependent on data. This data might be estimated or accurate, but the access to the data, in itself, does not diminish its value. Moreover, the

modelling methodology gives the modeller a step by step procedure on data modification, adaptation, regression suggestions, verification, validation, aswell as an accurate stochastic modelling of it, including the identification of a suitable modelling abstraction level in a thorough discussion on its pros and cons.

The work objectives include a number of sub-objectives presented in *section 1.3*. These sub-objectives support the main aim of the thesis and they have been carefully covered by this work. A brief cross reference between the thesis report and the sub-objectives follows:

- To carry out an exploratory pilot study and a complementary literature survey (*see chapters 2, 3, and 4*).
- Identify the principles and suggestions of the best robust system design practices based on the results of the experimentation and analysis of the case studies (*see chapters 4, 5 and 7*).
- Develop an appropriate modelling methodology for inpatient dependency variation, specifying the data requirements and data sources (*see chapters 6 and 7, as well as sections 4.8 and 5.7*).
- Use a system approach to highlight the benefits of an improved modelling methodology (*see chapters 4 and 5 to appreciate the importance of a solution that takes a system view and chapter 7 to picture how the modelling methodology is used to address variability in patient dependency as a modelling parameter for the design of the new system*).

Consequently the results confirm that the research aim of this work - *to identify how and why DES can be effectively utilised to design, plan and evaluate inpatient healthcare systems and their nurse staffing requirements* - has been met

## **8.4 Contributions of this research**

The main contributions and scientific novelty of this research are:

- **The development of a modelling methodology for inpatients' dependency variation and the study of how this variability gives insight into the effects of the workload variation of an inpatient ward.**

The modelling approach presented is unique and constitutes a feasible and practical method for simulation analysts who aim to address the question of adequate nurse staffing levels for inpatient wards. The modelling methodology makes it possible to model the stochastic and dynamic behaviour of patients' dependency evolution. It uses PCS data that is already available in many wards. Moreover, it contains analytical suggestions for increasing its accuracy, despite an eventual insufficient amount of data. Discharge modelling, one of the modelling components of the presented solution, is more generic and can be used without the need of dependency data or a staffing aim. The methodology opens opportunities for system improvement and robust system design that not only focuses on bed capacity, but also on the most important question, namely, staff requirements and their adaptation to the workload variation of the system.

- **The identification of principles and practical suggestions for the robust system design of inpatient wards using DES for the first time in the Swedish healthcare system.**

Although the system design findings during this work were not unique in themselves, the design's introduction and how it has been applied in the Swedish healthcare system are clearly distinctive. This work, which began in 2002, represents the first DES research work carried out in Sweden in this domain. As such, it has strongly contributed to helping healthcare professionals understand how demand variations affect healthcare systems and put stress on their resources, making it difficult to run them efficiently. The work has not only provided practical understanding of system behaviour, it has also made DES acceptable as a suitable tool for healthcare system design. Furthermore, it has highlighted the importance of designing flexible systems, and thereby creating bigger virtual units. Finally, the work has stressed the need to reduce preventable variations and exemplified how these changes, if implemented, would result in substantial improvements in the efficient use of healthcare resources.

## **8.5 Further work**

Three main topics for further work are presented below in order of priority. Since the topics are interrelated, the best results/improvements will be achieved if all three are addressed simultaneously. The ultimate goal is to be able to use simulation and the dependency matrices to operationally plan for both short and long term staffing. The vision

is to, in a not too distant future, have dynamic patient dependency data which monitors the different DRG patient groups' entire stay. This data would include not only LoS information, but also a dynamic and variable dependency evolution. In addition, it could be grouped so that patients with similar patterns would be identified. A holistic system approach is possible with this information. Applying this approach, the resources for the complete process flow of the patient can be planned, optimised and adapted to the variable needs of the system.

### **8.5.1 Improvement in patient grouping**

One of the problems this work faced was the difficulty in aggregating rehabilitation patients into homogenous groups. The task of defining homogenous groups is important for the more accurate prediction of the development of the ward's short term workload and subsequently the use of this information for short term staffing. A future analysis focusing on how well the new rehabilitation codes, introduced in 2008, the so-called NASS, classify more homogenous groups would be of great interest. Even of greater importance, due to its more generic solution, is making use of a more diversified number of factors for grouping, such as gender, age, admission method (reasons), destination, marital status, and so on. The relation between these so called "casual" components and group behaviour can be analysed using Bayesian belief networks and data mining techniques.

### **8.5.2 National dependency matrix libraries**

The transition probabilities of the dependency matrices are dependent on historical data. This data is expected to be both valid and considerable in order to accurately represent the true transition probability. Moreover the data analysis leading to the calculation of the transition probabilities can be both difficult and time consuming. Considering then that each inpatient group will need its own group of dependency matrices, it represents a significant work effort. It is for all these reasons that a national approach should be undertaken to define dependency matrices for inpatient care. These matrices should use data from certified wards, which follows the evidence based procedures that the National Board of Health and Welfare have stated.

### **8.5.3 IT solutions and operational scheduling**

Thanks to the combination of DES and PCS data and the time dependency of the matrices of the transition probabilities, it is possible to carry out a real time update of the simulation model and predict the system's coming events. The purpose would be to use the prognosis

from the model to schedule patients, staff and/or other resources, aiming at an optimal real time operational solution. Similar solutions are currently being developed in projects with an industrial focus. A future aim is therefore to tailor make similar solutions for the healthcare sector. However, although technological changes and improvements are being constantly implemented, the main goals of the healthcare providers should never be forgotten: saving lives, alleviating pain, giving emotional support, and curing people. These tasks are dependent on qualified staff and on the right staffing levels for a changing workload.

## 9 References:

- ADAMS-WENDLING, L. (2003) Clocking care hours with workload measurement tool. *Nursing Management*, 34(8), pp. 34-39.
- ADAN, I.J.B.F and VISSERS, J.M.H. (2002) Patient mix optimisation in hospital admission planning: a case study. *International Journal of Operations & Production Management*, 22(4), pp. 445-461.
- AL-AOMAR, R. (2002) A Robust Simulation-based multicriteria optimization methodology. In: *Proceedings of the 2002 Winter Simulation Conference, December 2002*. San Diego, California, pp. 1931-1939.
- ALBORNOZ, V., HINRICHSEN, M., MIRANDA, P. and PEÑA, P. (2006) Use of Markov chains in prediction of the dynamics of patient behaviour in a cardilogic intensive care unit. *Ingeniare*, 14(2), pp. 153-158.
- ALLEN, T.T. (2006) *Introduction to Engineering Statistics and Six Sigma*, 1<sup>st</sup> ed. Springer-Verlag London Limited.
- ALVARADO, R.G., ABUFARDE, F.B., MORENO, P.R. and BRAVO, M.P. (2003) Modeling of activities: an approach to the virtual representation of human behaviors in architectural spaces tested in emergency units. *International Journal of Architectural Computing*, 1(3), pp.333-345.
- AMIN, M., FERGUSSON, D., WILSON, K., TINMOUTH, A., AZIZ, A., COYLE, D. and HÉBERT, P. (2004) The societal unit cost of allogenic red blood cells and red blood cell transfusion in Canada. *Transfusion*. 44(10), pp. 1479-1486.
- ANELL, A. (2005) Swedish healthcare under pressure. *Health Economics*. 14(S1). pp. S237-S254.

- ARTHUR, T. and JAMES, N. (1994) Determining nurse staffing levels: a critical review of the literature. *Journal of Advance Nursing*, 19(3), 558-565.
- BANKS, J., CARSON II, J.S., NELSON, B.L. and NICOL D.M. (2005) *Discrete-Event System Simulation*, 4<sup>th</sup> ed. New Jersey: Pearson Prentice Hall.
- BARNES, C.D., QUIASON, J.L., BENSON, C. and MCGUINNESS, D. (1997) Success stories in simulation in health care. In: *Proceedings of the 29<sup>th</sup> conference on Winter Simulation, December 1997*. Atlanta, Georgia, pp. 1280-1285
- BEAKTA (1995) *Beakta att sätta värde på omvårdnaden*, Beakta AB, Lund.
- BENJAMIN, P.C., ERRAGUNTLA, M. and MAYER, R.J. (1995) Using simulation for robust system design. *Simulation*, 65(2), pp.116-128.
- BERGMARK, Å. (2008) Market Reforms in Swedish Healthcare: Normative Reorientation and Welfare State Sustainability. *Journal of Medicine and Philosophy*, 33(3), pp. 241-261
- BERGQVIST, I. and EDBERG, A-K. (2005) Vårdtyngsmätning inom anesthesiologisk omvårdnand- testing av ett instrument. *Vård i Norden*, 25(2), pp.9-14.
- BLUMBERG, N., KIRKLEY, S.A. and HEAL, J.M. (1996) A cost analysis of autologous and allogenic transfusions in hip-replacement surgery. *American Journal of Surgery*, 171(3), pp.324-30.
- BOTTER, M. (2000) The use of information generated by a Patient Classification System. *Journal of Nursing Administration*. 30(11), pp. 544-551.
- BRAHA, D. and MAIMON, O. (1998) The Measurement of a Design Structural and Functional Complexity, *IEEE Transactions on Systems, Man, and Cybernetics*, 28(3), pp. 527-534.

- BROWN, L.D., CAI, T.T. and DASGUPTA, A. (2001) Interval Estimation for a Binomial Proportion. *Statistical Science*, 16(2), pp 101-133.
- BUERHAUS, P.I., STAIGER, D.O. and AUERBACH, D.I. (2000) Implications of an Aging Registered Nurse Workforce. *Journal of American Medical Association*, 283(22), pp. 2948-2954.
- CHSA, CENTRUM FÖR HÄLSO OCH SJUKVÅRDSANALYS (Centre for Healthcare Analysis) (2002) *Jakten på den goda styrningen – En kunskapsöversikt kring styrning och organisation inom hälso- och sjukvård* (eng. The hunt for the good management – A knowledge survey around management and organisation within the healthcare system), 2002, Gothenburg. Centrum För Hälso– Och Sjukvårdsanalys, report nr. 3, ISBN 91-974223-2-0.
- CHWIF, L., PEREIRA BARRETTO, M.R. and RAY, J.P. (2000) On simulation model complexity. In: *Proceedings of the 2002 Winter Simulation Conference, December 2000*. Orlando, Florida, pp. 449-455.
- COCHRAN, J.K. and BHARTI, A. (2006) Stochastic bed balancing of an obstetrics hospital. *Health Care Management Science*. 9(1), pp. 31-45.
- DE GROOT, H. (1989a) Patient Classification System Evaluation. Part1: Essential system elements. *Journal of Nursing Administration*, 19(6), pp. 30-35.
- DE GROOT, H. (1989b) Patient Classification System Evaluation. Part two, system selection and implementation. *Journal of Nursing Administration*, 19(7), pp. 24-30.
- DUNBAR, L.J. and DIEHL, B.C. (1995) Developing a patient classification system for the pediatric rehabilitation setting. *Rehabilitation Nursing*, 20(6), pp. 328-332.
- EDWARDSON, S.R. and GIOVANNETTI, P.B. (1994) Nursing workload measurement systems. In: FITZPATRICK, J.J. and STEVENSSON J.S. eds. *Annual review of nursing research, Volume 12, June 1994*, Springer, pp.95-123.



- ELDABI, T. (2000) Simulation modelling: problem understanding in healthcare management. Thesis (PhD), Brunel University.
- ELDABI, T., ZAHIR, I. and RAY, J.P. (2002) A proposed approach for modelling health-care systems for understanding. *Journal of management in Medicine*, 16(2/3), pp. 170-187.
- ELF, M. (2003) *Modelling and simulation as planning tools for improvement of the quality of healthcare environments-towards a conceptual model of a care process*, Thesis for degree of licentiate of architecture, Chalmers University of Technology.
- FRANTZ, F. (1995): A Taxonomy of Model Abstraction Techniques. In: *Proceedings of the 1995 Winter Simulation Conference, December 1995*. Arlington, Virginia, pp. 1413-1420.
- GALAN PERROCA, M. and EK, A-C. (2007). Utilization of Patient classification systems in Swedish hospitals and the degree of satisfaction among nursing staff. *Journal of Nursing Management*, 15(5), pp. 472-480.
- GENDER, A.R. (1989) Developing and computerizing a patient classification system in a rehabilitation nursing setting. *Rehabilitation Nursing*, 14(2), pp. 58-63, 72
- GIOVANNETTI, P. (1979) Understanding patient classification systems, *Journal of Nursing Administration*, 9(2), pp. 4-9.
- GLENNGÅRD, A.H., PERSSON, U. and SÖDERMAN, C. (2005) Costs associated with blood transfusions in Sweden – The societal cost of autologous, allogenic and perioperative RBC transfusion. *Transfusion Medicine*, 15(4), pp. 295-306.
- GREEN, L. (2006) Queueing analysis in healthcare. In: R. HALL, ed. *Patient Flow: Reducing Delay in Healthcare Delivery*. Los Angeles, California: Springer, pp. 281-307.
- GREEN, L.V. and NGUYEN, V. (2001) Strategies for Cutting Hospital Beds: The Impact on Patient Service. *Health Services Research*. 26(2), pp. 421–42.

GROOTHUIS, S., VAN MERODE, G., FLOS, B. and BROEREN, S. (2001) An analysis of nursing systems through simulation. In: *Proceedings of the International conference of health sciences simulation, January 2001*. Phoenix, Arizona, pp.12-27

GUARY, E.G.A. and KLEIJNEN, J.P.C. (1998) Risk analysis of robust system design. In: *Proceedings of the 1998 Winter Simulation Conference, December 1998*. Washington, DC, pp. 1533-1540.

HALLIN, B. and SIVERBO, S. (2003) *Styrning och organisation inom hälso- och sjukvård*, Sweden, Studentlitteratur AB.

HEURGREN, M. (2000) Case Costing and DRG Development--Labour Costs for Inpatient Nursing. In: *Proceedings of 16th PCSE Conference, September 2000*. Groningen, Netherlands.

HURST, K. (2002) *Selecting and Applying Methods for Estimating the Size and Mix of Nursing Teams*. Department of Health, London and Leeds University Health Sciences and Public Health Research Institute, Leeds.

INNERHOFER, P., WALLECZEK, C., LUZ, G., HOBISCH-HAGEN, P., BENZER, A., STÖCKL, B. HESSENBERGER, G., NUSSBAUMER, W. and SCHOBERSBERGER, W. (1999) Transfusion of buffy coat-depleted blood components and risk of postoperative infection in orthopedic patients. *Transfusion*, 39(6), pp. 625-632.

JACK, E.P. and POWERS, T.L. (2004) Volume Flexible Strategies in Health Services: A Research Framework, *Production and Operations Management*, 13(3), pp. 230-244.

JACOBSON, S., HALL, S. and SWISHER, J. (2006) Discrete-Event Simulation of health care systems. In: R. HALL, ed. *Patient Flow: Reducing Delay in Healthcare Delivery*. Los Angeles, California: Springer, pp. 211-252.

JOHNSON, R.T., FOWLER, J.W. and MACKULAK, G.T. (2005) A discrete event simulation model simplification technique, In: *Proceedings of the 2005 Winter Simulation Conference, December 2005*. Orlando, Florida, pp. 2172-2176.

JUN, J.B., JACOBSON, S. H. and SWISHER, J. R. (1999): Application of discrete-event simulation in healthcare clinics: A survey, *Journal of Operational Research Society*, 50(2), pp. 109-123.

JÄGSTAM, M. (2004) *On efficient use and integration of discrete event simulation in industry*, Ph.D. Thesis (PhD), Faculty of Computer Science and Engineering, De Montfort University.

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 Operation Research*, 124(2), pp. 353-359.

KARNON, J. (2003), Alternative decision modelling techniques for the evaluation of healthcare technologies: Markov processes versus discrete-event simulation. *Health Economics*, 12(10), pp.837-848.

KLEIJNEN, J.P.C. (1999) Validation of models: statistical techniques and data availability. In: *Proceedings of the 1999 Winter Simulation Conference, December 1999*. Phoenix, Arizona, pp. 1280-1285.

KLEIJNEN, J.P.C. and GAURY, E. (2003) Short term robustness of production management systems: A case study. *European Journal of Operational Research*, 148(2), pp. 452–465.

KLEIJNEN, J.P.C. (2005) Supply chain simulation tools and techniques: a survey. *International Journal of Simulation & Process Modelling*, 1(1/2), pp. 82-89.

KLEVMARKEN, A. and LINDGREN, B. (2008) *Simulating and Ageing Population - A microsimulation approach applied to Sweden*, 1<sup>st</sup> ed. Emerald Group Publishing Ltd. Bingley, UK.

KOPACH-KONRAD, R., LAWLEY, M., CRISWELL, M., HASAN, I., CHAKRABORTY, S., PEKNY, J. and DOEBBELING, B.N. (2007) Applying Systems Engineering Principles in Improving Healthcare Delivery. *Journal of General Internal Medicine*, 22(S3), pp. 431-437.

KUMAR, A.P. and KAPUR, R. (1989) Discrete Simulation Application – scheduling staff for emergency room. In: *Proceedings of the 1989 Winter Simulation Conference, December 1989*. Washington, DC, pp. 1112-1120.

LE LAY, A., DESPIEGEL, N., FRANÇOIS, C. and DURU, G. (2006) Can discrete event simulation be of use in modeling major depression? *Cost Effective and Resource Allocation*, 4(1), pp.1-19, Available from: <http://www.resource-allocation.com/content/4/1/19> [Accessed September 2008]

LEVENSTAM, A. K. and BERGBOOM ENGBERG, I. (1993) The Zebra system – a new patient classification system. *Journal of Nursing Management*, 1(5), pp. 229-237.

LEVENSTAM, A. K. and BERGBOOM ENGBERG, I., (1997) How to translate nursing care into cost and staffing requirements: part two in the Zebra system, *Journal of Nursing Management*, 5(2), pp. 105-114.

LEVENSTAM, A. K. and BERGBOOM ENGBERG, I. (2002) Changes in patients' need of nursing care reflected in the Zebra system. *Journal of Nursing Management*, 10(4), pp. 191-199.

LITVAK, E., BUERHAUS, P.I., DAVIDOFF, F., LONG, M.C., MCMANUS, M.L. and BERWICK, D.M. (2005). Managing unnecessary variability in patient demand to reduce nursing stress and improve patient safety. *Journal on Quality and Patient Safety*, 31(6), pp. 330-338.

- LOWERY, J.C. (1996) Introduction to simulation in health care. In: *Proceedings of the 1996 Winter Simulation Conference, December 1996*. Coronado, California, pp. 78-84.
- MADAN, M., SON, Y., CHO, H. and KULVATUNYOU, B. (2005): Determination of Efficient Simulation Model Fidelity for Flexible Manufacturing Systems, *International Journal of Computer Integrated Manufacturing*, 18(2-3), pp. 236-250.
- MALLOCH, K and CONOVALOFF, A. (1999) Patient classification systems, Part 1. The third generation. *Journal of Nursing Administration*, 29(7-8), pp. 49-56.
- MARSHALL, A.H. and MCCLEAN, S.I. (2003) Conditional Phase-type distributions for modelling patient length of stay in hospitals. *International Transactions in Operations Research*. 10(6), pp. 565-576.
- MARSHALL, A., VASILAKIS, C. and EL-DARZI, E. (2005) Length of Stay-Based Patient Flow Models: Recent Developments and Future Directions. *Health Care Management Science*, 8(3), 213-220.
- MASTERSON, B.J., MIHARA, T.G., MILLER, G., RANDOLPH, S.C., FOEKNER, M.E. and CROUTER, A.L. (2004) Using Models and Data to Support Optimization of the Military Health System: A case Study in an Intensive Care Unit. *Health Care Management Science*, 7(3), pp. 217-224.
- MARTORELLA, C. (1996), Implementing a patient classification system. *Nursing Management*, 27(12), pp 29-31.
- MCGUIRE, F. (1998) Simulation of Healthcare. In: J. BANKS, ed. *Handbook of Simulation - Principles, Methodology, Advances, Applications, and Practice*. New York, NY. John Wiley & Sons Inc. pp. 605-627.
- MCLAUGHLIN, C.P. (1996) Why variation reduction is not everything: a new paradigm for service operations. *International Journal of Service Industry Management*, 7(3), pp. 17-30.

- MONTGOMERY, D.C. and RUNGER, G.C. (1994) *Applied Statistics and Probability for Engineers*. New York: John Wiley & Sons. Inc.
- NELSON, A., POWELL-COPE, G., PALACIOS, P., LUTHER, S.L., BLACK, T., HILLMAN, T., CHRISTIANSEN, B., NATHENSON, P. and COLEMAN GROSS, J. (2007) Nurse staffing and patient outcomes in inpatient rehabilitation settings. *Rehabilitation Nursing*, 32(5), pp. 179-202.
- NG, A., URENDA MORIS, M., SVENSSON, J., SKOOGH A. and JOHANSSON, B. (2007), FACTS Analyser: an innovative tool for factory conceptual design using simulation. In: *Proceedings of the Swedish Production Symposium, August 2007*, Gothenburg, Sweden.
- NIEBEL, B.W. and FREIVALDS, A. (2008). *Methods, standards, and work design*, 12<sup>th</sup> ed. New York: Mc Graw-Hill.
- NILSSON, H., HEURGREN, M., WESTANDER, F. and EDSTRÖM, B. (2000) National project for implementing Case costing in Swedish Health and Medical Care. In: *Proceedings of 16th PCSE Conference, September 2000*. Groningen, Netherlands.
- NILSSON, H. (2002) Case-Costing as a tool for healthcare management- results and examples from an ongoing work in Sweden. In: *Proceedings of 18th PCSE Conference, October 2002*. Innsbruck, Austria.
- NIST/SEMATECH (2007) *e-Handbook of Statistical Methods*. Available from: <http://www.itl.nist.gov/div898/handbook/> [Accessed 15/01/2007].
- NOON, C.E., HANKINS, C.T. and MURRAY, J.C. (2003) Understanding the Impact of Variation in the Delivery of Healthcare Services. *Journal of healthcare management*, 48(2), pp. 82-97.
- O'DRISCOLL, K., MEAGHER, D. and ROBSON, M. (2003). *Active management of labour: the Dublin experience*. 4. ed. Edinburgh: Mosby.

PEREZ, A., CHAN, W. and DENNIS, R.J. (2006) Predicting the length of stay of patients admitted for intensive care using a first step analysis. *Health Services and Outcomes Research Methodology*, 6 (3-4), pp. 127-138.

PERROCA, M. G. and EK, A-C. (2007) Utilization of patient classification systems in Swedish hospitals and the degree of satisfaction among nursing staff. *Journal of Nursing Management*, 15(5), pp. 472-480.

PERSSON, M. (2007) *Modelling and analysing hospital surgery operations management*. Thesis for degree of licentiate of engineering, Blekinge Institute of Technology.

PLEVRITIS, S.K. (2005) Decision analysis and simulation modeling for evaluating diagnostic test on the basis of patient outcomes. *American Journal of Radiology*: 185(3), pp. 581-590.

PROCTOR, P., COMPTON, W.D., GROSSMAN, J. and FANJIANG, G. (2005). *Building a Better Delivery System: a New Engineering/Healthcare Partnership*, Committee on Engineering and the Healthcare System. Washington, DC: National Academy of Engineering and Institute of Medicine, National Academy Press.

RANJIT, R.K. (2001) *Design of experiments using the Taguchi approach*. New York, John Wiley & Sons, Inc.

RAUHALA, A., KIVIMAKI, M., FAGERSTROM, L., ELOVAINIO, M., VIRTANEN, M., VAHTERA, J., RAINO, A.K., OJANIEMI, K. and KINNUNEN, J. (2007) What degree of work overload is likely to cause increased sickness absenteeism among nurses? Evidence from the RAFAELA patient classification system. *Journal of Advanced Nursing*, 57(3), pp. 286–295.

ROBINSON, S. and BHATIA, V. (1995) Secrets of successful simulation projects. In: *Proceedings of the 1995 Winter Simulation Conference, December 1995*. Arlington, Virginia, pp. 61-67.

ROY, R.K. (2001) *Design of Experiments Using the Taguchi Approach: 16 Steps to Product and Process Improvement*, Wiley, New York, NY.

RHOTEN, D. (2004) Interdisciplinary research: trend or transition. *Items and Issues - Social Science Research Council*, 5(1-2), pp. 6-11.

SÁNCHEZ, P.J. (2006) As simple as possible, but no simpler: a gentle introduction to simulation modeling. In: *Proceedings of the 2006 Winter Simulation Conference, December 2006*. Monterey, California, pp. 2-10.

SADOWSKI, D.A. and GRABAU, M. R. (2000) Tips for successful practice of simulation. In: *Proceedings of the 2000 Winter Simulation Conference, December 2000*. Orlando, Florida, pp. 26-31.

SALAR, SWEDISH ASSOCIATION OF LOCAL AUTHORITIES AND REGIONS (2005a) *Kommer vi att ha råd med sjukvården?* 2005. Stockholm. Swedish Association Of Local Authorities And Regions, ISBN 91-7164-094-0.

SALAR, SWEDISH ASSOCIATION OF LOCAL AUTHORITIES AND REGIONS (2005b) *Swedish Healthcare in an International Context – comparison of care needs, cost, and outcomes*. 2005, Stockholm. Swedish Association of Local Authorities And Regions, Stockholm, ISBN 91-7164-049-5.

SALAR, SWEDISH ASSOCIATION OF LOCAL AUTHORITIES AND REGIONS (2008) *Quality and Efficiency in Swedish Health Care – Regional comparisons 2007, 2008*. Stockholm: Swedish Association of Local Authorities and Regions, and The National Board of Health and Welfare, ISBN 978-91-7164-320-9.

SANCHEZ, S.M., SANCHEZ, P.J., RAMBERG, J.S. and MOEENI, F. (1996) Effective engineering design through simulation. *International Transactions in Operational Research*, 3(2), pp. 169-185.



SANCHEZ, S.M., OGAZON, T., FERRIN, D.M., SEPÚLVEDA, J.A. and WRAD, T.J. (2000) Emerging issues in healthcare simulation. In: *Proceedings of the 2000 Winter Simulation Conference, December 2000*. Orlando, Florida, pp. 1999-2003.

SARNECKI, A.J., HAAS, S., STEVENS, K.A. and WILLEMSSEN, J.A. (1998) Design and implementation of a patient classification system for rehabilitation nursing. *Journal of Nursing Administration*, 28(3), pp. 35-43.

SEEBER, P. and SHANDER, A. (2007) *Basics of blood management*. 1<sup>st</sup> ed. Oxford, UK, Blackwell Publishing Ltd.

SHANDER, A., HOFMANN, A., GOMBOTZ, H., THEUSINGER, O.M. and SPAHN, D.R. (2007) Estimating the cost of Blood: past, present and future directions. *Best practice and Research Clinical Anaesthesiology*, 21(2), pp. 271-289.

SHULDHAM, C.M. (2004) Commentary. Nursing skill mix and staffing. *Journal of Nursing Management*, 12(6), pp. 385–387.

SIMPSON, K.N., STRASSBURGER III A., JONES, W. and DIETZ, B. and RAJAGOPALAN, R (2009) Comparison of Markov model and Discrete Event Simulation techniques for HIV-Disease. *Pharmacoeconomics*, 27(2), pp. 159-165.

SIN, B. and KIM, J.H. (1995) Nonstationary hidden Markov model. *Signal Processing*, 46(1), pp. 31-46.

SKARABORGS SJUKHUS (2008) *Skaraborgs Sjukhus, Årsredovisning 2007*. (Annual report). Available from:

<http://www.vgregion.se/upload/SkaS/om%20sjukhuset/%C3%85rsredovisningar/%C3%85rsredovisning%202007%20slutversion%20-%206%20feb%2008.pdf?epslanguage=sv>

[Accessed Mars 2009]

SOLIMAN, F. (1998) Automation of patient dependency systems. *Journal of Medical Systems*, 22(4), pp. 225-236.

SONNEBERG, F.A. and BECK, J.R. (1993) Markov Models in medical Decision Making: A Practical Guide. *Medical Decision Making*, 13(4), pp. 322-338.

SoS, SOCIALSTYRELSEN (THE NATIONAL BOARD OF HEALTH AND WELFARE) (2005) *Samverkan vid in- och utskrivning av patienter I sluten vård*, Stockholm: Socialstyrelsen författningssamling, SOSFS 2005:27, 2005.

SoS, SOCIALSTYRELSEN (THE NATIONAL BOARD OF HEALTH AND WELFARE) (2008a), *Lägesrapporter 2007 – Hälso- och sjukvård*, Stockholm, Socialstyrelsen, SBN: 978-91-85999-10-1

SoS, SOCIALSTYRELSEN (THE NATIONAL BOARD OF HEALTH AND WELFARE) (2008b) *Vägledning till NordDRG Svensk fullversion 2008*, Stockholm, Socialstyrelsen, 2008-106-2.

SPAHN, D. R., MOCH, H., HOFMANN, A. and ISBISTER, J.P. (2008) Patient blood management: the pragmatic solution for the problems with blood transfusion. *Anesthesiology*, 109(6), pp. 951-953.

STATISTICS SWEDEN (2005) *The fertility projection model – description of the calculation of number of births*. Stockholm: Statistic Sweden, ISSN 1104-4047

STATISTICS SWEDEN (2006a) *The future population of Sweden 2006–2050*. Stockholm: Statistic Sweden, Demographic Reports 2006:2.

STATISTICS SWEDEN (2006b) *Background material about demography, children and family 2006:1, Population growth in municipality groups and the influence on total fertility*. Stockholm: Statistic Sweden, ISSN 1104-4047.

STULBERG, B.N. and ZADZILKA, J.D. (2007) Blood Management Issues-Using Blood Management Strategies. *The Journal of Arthroplasty*, 22(4, 1), pp. 95-98.

SWEDEN. MINISTRY OF HEALTH AND SOCIAL AFFAIRS (2007) *Health and medical care in Sweden*. Stockholm: Fact Sheet no. 16. Available from: <http://www.ud.se/sb/d/574/action/browse/c/114/c/y2007> [Accessed 15/01/09].

SWEDISH INSTITUTE (2007) *Swedish healthcare*, Stockholm: The Swedish Institute, report FS 76.

TAGUCHI, G. (1986) *Introduction to Quality Engineering*, Asian Productivity Organisation, Tokyo.

THE FEDERATION OF THE SWEDISH COUNTY COUNCILS (Landstingsförbundet)(2000) *Patientklassificering inom omvårdnadsområdet – metodbeskrivning and användning för KPP och andra syften*. (In English: Patient Classification in the nursing Care Area Description of the Method and Use for Cost per Patient and Other Purposes). Stockholm: Landstingsförbundet, Report no. 2. KPP project.

THORNTON J.G. and LILFORD, R. J. (1994) Active management of labour: current knowledge and research issues. *British Medical Journal*, 309, pp.366-369.

URENDA MORIS, M., DE VIN, L.J. and ERIKSSON, P. (2004) Introducing discrete event simulation for decision support in the Swedish healthcare system. In: *Proceedings of the 2004 Western Multi Conference: Health Sciences Simulation, January 2004*. San Diego, California, pp. 48-53.

URENDA MORIS, M., LEZAMA, T. and DE VIN, L.J. (2007) Robust design of a Maternity ward supported by discrete event simulation - a case study. In: *Proceedings of FAIM 2007 Conference, June 2007*. Philadelphia, Pennsylvania, pp. 107-114.

VISSERS, J.M.H., VAN DER BIJ, J.D.H. and KUSTERS, R.J. (2001a) Towards decision support for waiting lists: An operations management view. *Healthcare Management Science*, 4(2), pp. 133-142.

VISSERS, J.M.H., BERTRAND, J.W.M. and DE VRIES, G. (2001b) A framework for production control in healthcare organizations. *Production Planning and Control*, 12(6), pp. 591-604.

WORLD HEALTH ORGANIZATION (2006) *World Health Report 2006 - Working together for health*. Geneva: The World Health Organization, 2006.

WALLEY P., SILVESTER, K. and STEYN, R. (2006) Managing Variation in Demand: Lessons from the UK National Health Service. *Journal of Healthcare Management*, 51(5), pp. 309-322.

WARNER, M. (2006) Personnel staffing and scheduling. In: R. HALL, ed. *Patient Flow: Reducing Delay in Healthcare Delivery*. Los Angeles, California: Springer, pp. 189-210.

WILD, R.H. and PIGNATIELLO, J.J. (1991) An experimental design strategy for designing robust systems using discreteevent simulation. *Simulation* 57(6), pp. 358-368.

WILLIAM, J.C. (1998) Birth of a new maternity process. In: *Proceedings of the 1998 Winter Simulation Conference, December 1998*. Washington, DC, pp. 1429-1432.

WINSTON, W.L. (1994) *Operation research: Application and algorithms, 3<sup>rd</sup> ed.* (Chapter 19) Duxbury Press, Belmont California.

YIN, R.K. (2007) *Fallstudier: design och genomförande* (Original title: Case study research). Malmö, Sweden, Liber AB.

## Earlier Publications

URENDA MORIS, M., DE VIN, L.J. and ERIKSSON, P. (2004) Introducing discrete event simulation for decision support in the Swedish healthcare system. In: *Proceedings of the 2004 Western Multi Conference: Health Sciences Simulation, January 2004*. San Diego, California, pp. 48-53.

URENDA MORIS, M., LEZAMA, T. and DE VIN, L.J. (2007) Robust design of a Maternity ward supported by discrete event simulation - a case study. In: *Proceedings of FAIM 2007 Conference, June 2007*. Philadelphia, Pennsylvania, pp. 107-114.