Modeling and Management of Variation in the Operating Theatre

Pieter S. Stepaniak

Modeling and Management of Variation in the Operating Theatre

Pieter Stepaniak

ERASMUS UNIVERSITEIT ROTTERDAM

Cover Design: Simone Vinke

Photography: Christiaan Heijster, Captain Boeing 737

Lay out and Print: Simone Vinke, Ridderprint BV, Ridderkerk

ISBN: 978-90-5335-351-8

Publication and distribution of this thesis was made possible by HighCare Interim & Advies, Rotterdam The Netherlands and Catharina Hospital Eindhoven, The Netherlands.

Modeling and Management of Variation in the Operating Theatre

Modeleren en managen van variatie in de operatiekamer

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de
rector magnificus

Prof.dr. H.G. Schmidt

en volgens besluit van het College voor Promoties.
De openbare verdediging zal plaatsvinden op

donderdag 16 december 2010 om 09.30 uur

Pieter Szymon Stepaniak
geboren te Breda

ERASMUS UNIVERSITEIT ROTTERDAM

Promotiecommissie

Promotoren:
Prof. dr. ir. G. de Vries
Prof. dr. J. van de Klundert

Overige leden: Prof. dr. A.P.M. Wagelmans
Prof. dr. G.M. Koole
Prof. dr. F. Dexter

Dla mój mama
W pamięć o moim ojcu

Dedicated to the brave men of the 1st Polish Armoured Division who, under command of general Stanisław Maczek, liberated parts of our country from German occupation and oppression during 1944 and 1945.

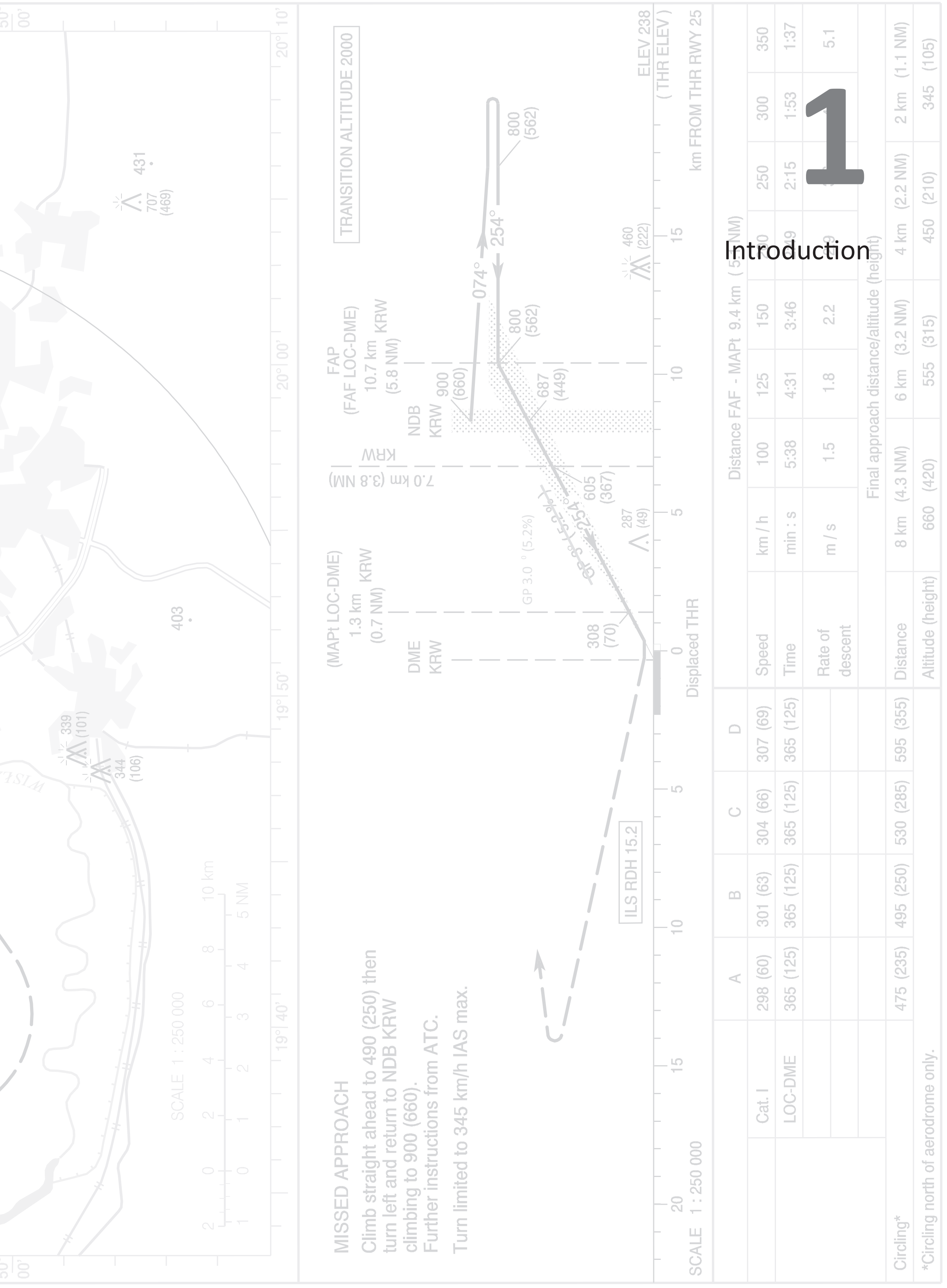
Contents

INSTRUMENT APPROACH CHART - ICAO

AERODROME ELEV 241m
HEIGHTS RELATED TO THR RWY 25 - ELEV 238m

APP	121.075
	134.675
TWR	123.250

KRAKÓW / Balice
ILS or LOC
RWY 25 (CAT A/B/C/D)



1 Introduction

Introductory remarks

After having worked in the profit industry, I continued my career in 2004 as a manager of operating rooms (ORs) in a large general teaching hospital in Rotterdam. My experiences in industry management taught me to work efficiently, effectively and to excel in service to every customer and prospect. With this experience in mind I started my new job on the first of January 2004. A job in an environment filled with costly equipment and a range of highly skilled professionals such as surgeons, residents, anesthesiologists and OR staff: a multi-million euro business within a hospital. Last but not least, a business with customers: patients who needed care. Prior to starting the job, I had assumed that processes were already efficient and effective, as a result of the relatively high labor and investment costs. Being a pilot, I fully realize what a valuable resource airspace is, particularly when subject to high traffic demand. Since airspace is a fixed volume, as is the case with OR capacity, managing it is a vital activity for satisfying the needs of the aircraft operators in the most efficient and equitable manner using a sophisticated decision support system. As none of this appeared to be the case in the OR environment, I conjectured that it must be possible to run the OR more efficiently, effectively and in a more patient-centered way.

The following examples serve to illustrate my impressions. Due to poor case scheduling, OR staff is forced to stand around idly, and expensive nursing, anesthesia and support staff are wasted on some of the days. On other days, the OR staff works beyond regular working hours to finish the workload on that day. Surgeons/anesthesiologists arrive too early or too late in the OR and teams are not always ready at the scheduled time. Capacity in the OR is sometimes insufficient for patients who arrive in the emergency department, which causes scheduled patients to be denied surgery that day, or for staff to work late. Such situations frequently result in nurses, doctors, management and patients becoming extremely frustrated. When looking at an OR both in an era in which both cost-containment and quality of health care are considered of prime importance, hospitals simply have to utilize ORs effectively and efficiently. In 2007, these experiences and impressions motivated me to start studying how to control the enormous variation in activities in the OR. I started by looking at the variations in case durations, surgical processes, and scheduling processes.

As this thesis will demonstrate, a fundamental understanding of the variation and proper control in the OR makes it possible to improve its efficiency and effectiveness, and therefore also improve the quality of care provided to the patients.

Introduction

Background

The Institute of Medicine (IoM) report [1] “Crossing the Quality Chasm: A New Health System for the 21st Century” described many problems in the quality of the United States health care delivery system. The report suggests that: “Health care should be:

- Safe: avoiding injuries to patients from the care that is intended to help them
- Effective: providing services based on scientific knowledge to all who could benefit, and refraining from providing services to those not likely to benefit (avoiding underuse and overuse, respectively)
- Patient-centered: providing care that is respectful of and responsive to individual patient preferences, needs, and values, and ensuring that patient values guide all clinical decisions
- Timely: reducing waits and sometimes harmful delays for both those who receive and those who give care
- Efficient: avoiding waste, including waste of equipment, supplies, ideas, and energy
- Equitable: providing care that does not vary in quality because of personal characteristics such as gender, ethnicity, geographic location, and socioeconomic status”.

The view the IoM has on quality is quite similar to the view expressed in Dutch Quality of Care Institutions Act [2]: The health care provider offers ‘responsible’ care. Responsible care implies care of a high standard, that is appropriate care provided in an effective, efficient and patient-centered way and that meets the patient’s actual needs (Article 2). To achieve appropriate care, the organizations must demonstrate that there is a planned effort to maintain and improve the quality of care in a systematic way. According to the act, a systematic way means that at least three steps are to be followed: “(a) the quality of care should be measured, for example by means of satisfaction surveys or quality indicators; (b) the results of such measurements are to be evaluated against explicit standards or goals; and (c) based on this evaluation, the organization is required to make the necessary changes in care processes or in their quality policy”. Such a quality management approach is intended to provide for a continuous process of quality assessment and improvement of care [3,4]. According to the Dutch Research for Man and Environment, coordination and cooperation in health care and patient safety score relatively low. The efficiency of health care in the Netherlands is not optimal and quality is not a driving force in the health care market [5].

The meaning of timelines, efficient and effective health care

According to the IoM, *timely access* to care is “reducing waits and sometimes harmful delays for both those who receive and those who give care”. One of the most serious problems has to do with timely access to hospital services. Problems involving access to care manifest themselves in a variety of forms, including rejection of patients seeking services. For instance, when a patient accesses the hospital, he or she is likely to encounter waits, delays, and cancellations. If the patient requires surgery, it is not uncommon to experience waits due to stacking of cases in the OR, or to be delayed by more than one day, even on the day of surgery itself. The start of the surgery schedule in the morning is often delayed, putting pressure on the timeliness of the surgeries of the scheduled patients, more so when non scheduled patients arrive from the Emergency Department. As a result, OR staff may need to work overtime.

Having to work frequently beyond regularly scheduled hours due to badly scheduled ORs can lead to both overtime costs and intangible costs, the latter resulting from dissatisfaction and reduced motivation on the part of the staff [6,7]. Another effect of delay in care delivery in the OR is the extension of a patient’s length of stay this entails. Prolongation of the stay of a patient implies that the occupied bed cannot be given to another patient, and hence fewer patients can be served. Not only do increases in the length of stay therefore result in extra cost and/or loss of revenue, they are also major sources of both patient and provider dissatisfaction with the present care delivery system. The variations caused by the various aforementioned problems has been reported to cause poor patient flow, emergency department overcrowding and hence limited access to care, nurse understaffing/overloading, diminished quality of care and high health care cost [8].

Efficient care “avoids waste, including waste of equipment, supplies, ideas, and energy” [1]. For the sake of argument, consider a situation where all patients have the same disease, the same degree of illness, and respond identically to therapy. Let us further assume that all patients are elective patients and all medical practitioners and health care systems are standardized. In this highly stylized situation, 100% efficiency in health care delivery might be attainable. Within the boundaries of knowledge and technology, there would be zero waste [9]. In reality, patients vary, have different diseases, and respond differently to therapy, etc. The natural variation influences the delivery process. Controlling the variation can, however, be useful in making the processes more predictable, and hence increasing efficiency.

Effective care “is based on providing services based on scientific knowledge to all who could benefit, and refraining from providing services to those not likely to benefit (avoiding underuse and overuse, respectively)"[1]. Instruments to achieve organizational and workforce excellence are, for example, lean thinking and Six Sigma. Lean is an important dimension of quality; all work that doesn't add value for the customer is defined as waste. Six Sigma is a methodology that uses data and statistical analysis to measure analysis and improve a company's operational performance by identifying and eliminating defects to enhance customer satisfaction [10]. Performance measurement is then needed in, for example, identifying and tracking progress against organizational goals, identifying opportunities for improvement and comparing performance in benchmarks.

The meaning of health care quality and how do achieve it

The Donabedian model [11] of structure-process-outcome is generally used as the basis for much of the work addressing quality and outcomes. Donabedian framed the concept of quality assurance in terms of three types of measures (Figure 1): structure (what do we need to have to be able to achieve quality), process (what do we need to do to achieve quality), and outcomes (what do we need to achieve).

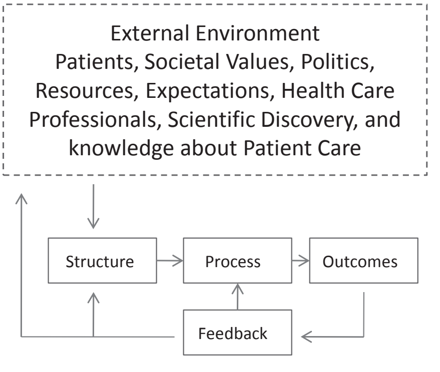


FIGURE 1. DONABEDIAN'S MODEL FOR QUALITY

Donabedian notes that any efforts to improve quality need to recognize that health care is embedded in, and greatly influenced by, the larger external environment. Structure relates to static characteristics such as facilities, equipment, and personnel [12]. Process consists of activities involved in the process of delivering health care services, including the technical and interpersonal actions of health providers and patients as well as organizational processes within the health care system [13]. In other words, process looks at what takes place during care, while outcomes assess the effect on care in regards to a patient's health. Donabedian suggests that each dimension can be judged independently or in conjunction. Furthermore, he says that if both structure and process are adequate, one can assume the outcome will be positive.

From surgical planning and process to outcome

Does increased timeliness, and efficient and effective scheduling of surgical cases lead to improved outcomes?

Surgical delay has been shown to be an important determinant of patient satisfaction across the continuum of preoperative-operative-postoperative care [14]. Delays in scheduled surgical cases affect patient satisfaction even more than the intraoperative anesthesia experience [15]. Delays in surgery resulting from cancellations, bumping of cases and poor scheduling can have a significant impact on quality of care for scheduled cases as well [16]. Delays only add to the patient's inherent anxiety associated with surgery and engenders anger and frustration. The operating room, by its very nature, is an extremely stressful, uncertain, dynamic, and demanding environment where staff members need to manage multiple highly technical tasks, often simultaneously [16,17]. Other factors also impact the system within the OR. Examples are individual, group and organizational performance issues such as team- and time management, interpersonal skills, leadership, workload distribution, dynamic decision making, human machine interface, problem detection, capture of errors (slips, mistakes, fixation bias), loss of situational awareness, high mental and physical workload, fatigue, environmental stress, production pressure and personal life stress [18]. Moreover, the dynamics of the OR are complex because they form a point of intersection among multiple groups with their own agendas and requirements. Since ORs are relatively scarce resources, poor scheduling and misuse of ORs can provide opportunities for conflict and competition.

OR staff carry out their sometimes long working days under time pressure. The Joint Commission on the Accreditation of Healthcare Organizations has identified time pressures to start or complete the procedure as one of four contributing factors to increased wrong site surgery [19]. Similar to other professions, the undue pressures of time that result from falling behind create stress that can lead to cutting corners or inadvertent error. Relative to other hospital settings, errors in the operating room can be catastrophic (i.e. wrong site surgery, retained foreign body, unchecked blood transfusions). In some cases these errors can result in high-profile consequences for the patient, surgeon or hospital [20]. In other words, poor scheduling and the subsequent induced variation in processes reduces outcome.

Because the approach in this thesis is both from an operational research and logistics point of view, it is necessary to position this thesis within a planning framework. This positioning will then be described and illustrated.

Planning framework

The flow of activities in the OR through surgical case planning, directing, and controlling, and then back to planning again can be formalized by a planning and control cycle. Because there are some differences between industry and service-oriented industries [21,22,23,24] a production control framework for hospitals has been developed [21]. Characteristic for this framework is that patients, processes and chains are the basis for organizing care and it deals with balancing effective, efficiency and timely care. The framework is based on an analysis of the design requirements for hospital production control systems [25,26] and builds on the production control design concepts developed [27]. It is then applied in the context of the OR. In this thesis the decisions made on the first four levels of the model are given. The focus of the thesis is on the fifth level of the production control framework as applied to the OR. This level concerns the actual scheduling of patients, given planning rules and service requirements for the coming days or weeks. It is concerned with the processes used in facilitating day-to-day activities that need to be performed to deliver timely, effective and efficient care for the patient.

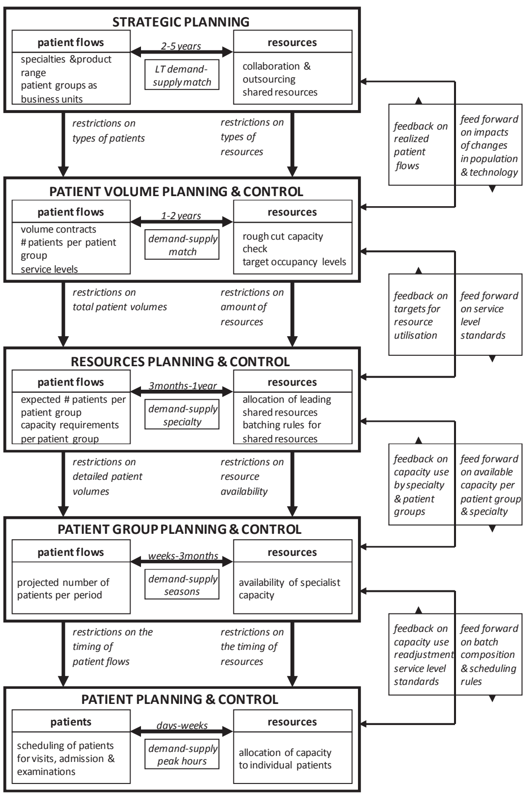


FIGURE 2. PRODUCTION CONTROL FRAMEWORK [21]

Aim of the study

Based on the time required to construct schedules as well as the quality of resulting schedules [28,29] evidence indicates that case scheduling in practice often is performed poorly [9,30]. Additionally, methods which improve the reliable estimate of surgical cases naturally lead to improved timeliness, efficiency, and effectiveness of OR processes [31,32,33,34,35]. As stated earlier, poor surgical case scheduling and the subsequent induced variation in processes reduces outcome. Reasoning along these lines, W. Edwards Deming concluded that the real enemy of quality is variation in processes. A main objective in operations management is therefore to identify sources of variation [10]. Though variation exists in every process and always will, controlling the identified variation helps managers and clinicians to improve efficiency by aligning the health service delivery processes towards the desired results [8]. Because the OR is a leading source [21] controlling variation of OR schedules and processes has the second order effect of reducing variation and improving quality in subsequent processes throughout in the hospital. Indeed, an OR scheduling process which reduces the census variability of the OR, can improve the flow of surgical patients to downstream inpatient units, resulting in a more even and predictable patient care burden [36]. Furthermore accurate preoperative scheduling of surgical episodes is critical to the effort to minimize variability in the length of the surgical day and maintain on time starts for cases to follow [36].

This thesis addresses the issue of improving the outcome of healthcare in the OR by modeling and managing variation in medical operations, thereby leading to an increase in the timeliness, efficiency and effectiveness of health care. This can be realized by using a planning and control-based activity that focuses continuously on controlling variation. The result is twofold: First through the feedback loop in the production control framework it results in a better control of hospital activities [21]. Second, it may help to achieve meaningful, sustainable improvement of quality of care in the OR and consequently in the subsequent delivery health care system. Or to put it in the perspective of the process part of Donabedian's model: this is how to improve the outcome of healthcare.

Thesis outline

As argued earlier, methods that improve the reliable estimate of surgical case durations lead to improved timeliness, efficiency, and effectiveness of OR processes. For this reason, three studies are performed, described in chapters two, three and four.

Chapter two describes how case scheduling can be improved. Gains in OR scheduling may be obtained by using accurate statistical models to predict surgical and procedure times. The three main contributions of this paper are the following: (i) the validation of Strum's results on the statistical distribution of case durations, including surgeon effects,

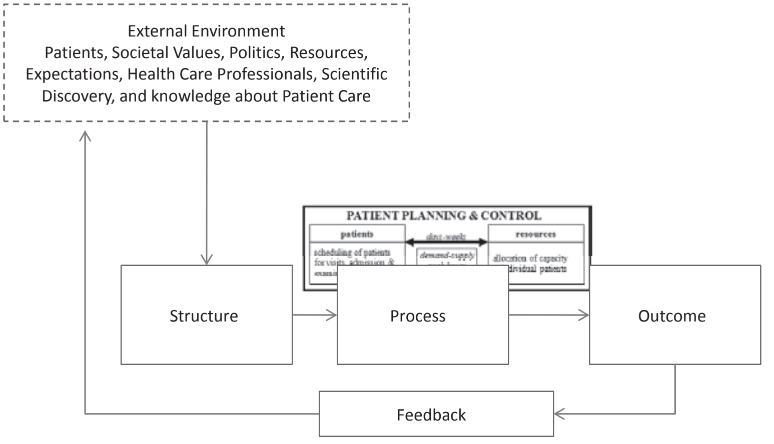


FIGURE 3.SYNTHESIS OF THE PRODUCTION CONTROL FRAMEWORK AND DONABEDIAN'S MODEL FOR QUALITY

using OR databases of two European hospitals, (ii) the use of expert prior expectations to predict durations of rarely observed cases and (iii) the application of the proposed methods to predict case durations, with an analysis of the resulting increase in OR efficiency. We retrospectively review all recorded surgical cases of two large European teaching hospitals in the period 2005-2008, involving 85,312 cases and 92,099 hours in total. Surgical times tend to be skewed and bounded by some minimally required time. We compare the fit of the normal distribution with that of two- and three-parameter lognormal distributions for case durations of a range of CPT-anesthesia combinations, including possible surgeon effects. For cases with very few observations, we investigate whether supplementing the data information with surgeon's prior guesses helps to obtain better duration estimates. Finally, we use best fitting duration distributions to simulate the potential efficiency gains in OR scheduling. This chapter appears in:

Pieter S. Stepaniak, Christiaan Heij, Guido H. H. Mannaerts, Marcel de Quelerij, and Guus de Vries. Modeling Procedure and Surgical Times for Current Procedural Terminology-Anesthesia-Surgeon Combinations and Evaluation in Terms of Case-Duration Prediction and Operating Room Efficiency: A Multi-center Study. Anesth Analg 2009;106:1232-45.

Chapter three analyzes the possible dependence of surgeon factors like age, experience, gender, as well as modeling team composition on procedure time. The effect of these factors is estimated for over 30 different types of medical operations in two hospitals, by means of Analysis Of Variance (ANOVA) models for logarithmic case durations. The practical significance of several factors on surgical procedure times is tested by comparing the quality of predicted case durations for 2009. This chapter appears in:

Pieter S. Stepaniak, Christiaan Heij and Guus de Vries. Modeling and prediction of surgical procedure times. Stat Neerl 2010;64:1-18.

Chapter four describes the effect of scheduling consecutive similar cases and the use of a fixed team on OR case duration and turnover time. This study on the impact of similar consecutive cases on the turnover-, surgical-, and procedure time tests the perception that repeating the same manual tasks reduces the duration of these tasks. We hypothesize that when a fixed team works on similar consecutive cases the result will be shorter turnover and procedure duration as well as less variation as compared to the situation without a fixed team. To test our hypothesis, two procedures were selected and divided across a control group and a study group. Patients were assigned at randomly to the study or control group. This chapter appears in:

Pieter S. Stepaniak, Christiaan Heij, Wietske Vrijland, Marcel de Quelerij and Guus de Vries. Working with a fixed OR team through the day on consecutive similar cases and the effect on OR case duration and turnover time: By random assigning patients to study and control days. Accepted for publication, Arch Surg (24/09/2009).

Chapter five analyzes the managerial part of the OR. Within the daily dynamics of performing different cases in different ORs the first signs of variation are visible when one or more scheduled cases are completed later or earlier than scheduled. This results in gaps within OR schedules. Actions taken to fill these gaps are, for example, rescheduling cases to different ORs, and scheduling emergency cases. These actions are performed by the Operating Room Coordinator (ORC). The ORC observes the daily variation and takes the above-mentioned actions such that scheduled and non-scheduled cases are performed without ending too late in too many ORs at the end of the day. There are observed differences among the personalities of the four ORCs with regard to their willingness to accept taking on more risk concerning their daily planning. The hypothesis is tested that the relationship between the personality of each of the four ORCs and the risk an ORC is willing to take of cases running late influences OR efficiency. Sometimes cases are canceled at the end of the day by OR management on the argument that it is not cost-effective to proceed with a surgery case after regular working hours. It will be shown whether it is more cost-effective to proceed with a new case which has a chance of finishing in overtime hours than to postpone the case. This chapter appears in:

Pieter S. Stepaniak, Guido H. H. Mannaerts, Marcel de Quelerij, and Guus de Vries. The Effect of the Operating Room Coordinators Risk Appreciation on Operating Room Efficiency. Anesth Analg 2009;108: 1249-56.

In its influential 'Crossing the quality chasm', the Institute of Medicine [1] identifies six quality dimensions of health care, among which are efficiency and timeliness. The six dimensions together, make quality improvement a complex matter, as interventions which yield improvement regarding one dimension may have a negative effect regarding another: the quality dimensions form conflicting objectives. In this research we simultaneously address efficiency and timeliness of care in the operating theatre.

We formally model the real time surgery scheduling to minimize a weighted sum of cancellation of scheduled cases, overtime cost, moving scheduled cases from the day to the service operating room and scheduling emergency/acute cases after an imposed time limit. Stepaniak et al. [2] show how risk attitudes of OR planners influence the quality of scheduling. We formally model heuristics which are based on different risk attitudes and analyze their mutual performance. More generally, we analyze Monte Carlo based optimization methods and use recent actual data from the St. Franciscus Gasthuis, Rotterdam.

Pieter S. Stepaniak, Ronald van der Velden, Albert Wagelmans and Joris van de Klundert. Quality improvement: balancing the risks of overtime and cancellation of scheduled cases. Submitted for publication.

The final concluding remarks, practical implications and reflection are described in chapter seven.

References

1. Institute of Medicine & Committee on Quality health care in America. *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington DC, *National Academic Press*, 2001.
2. Ministry of Health, Welfare and Sports in the Netherlands. Care Institutions Quality Act. http://www.minvws.nl/international/documentation/quality of care. Last assessed September 1st 2010.
3. Berwick DM. *Crossing the boundary: changing mental models in the service of improvement*. *Int J Quality in Health Care* 1998;10:435–41.
4. Sluijs EM, Wagner C. Progress in the implementation of quality management in Dutch health care: 1995–2000. *Int J Quality Health Care* 200315:223-34.
5. Westert GP, Berg MJ van den, Koolman X, Verkleij H. *Dutch Health Care Performance Report 2008*, RIVM.
6. Stachota P, Normandin P. Reasons registered nurses leave or change employment status. *J Nurs Adm* 2003;33:111-18.
7. Thomson TP, Brown H. Turnover of licensed nurses in skilled nursing facilities. *Nurs Econ* 2002;20:66-69.
8. McLaughlin CP, Kaluzny AD. Continuous Quality Improvement in Health Care, Third Edition Center for Health Services Research, *University of North Carolina at Chapel Hill*, Chapel Hill, North Carolina 2006.
9. Litvak E, Long MC. Cost and quality under managed care: Irreconcilable differences? *Am J Manag Care* 2000;6:305-12.
10. Tannat G. *Design for six sigma: launching new products and services without failure.* Gower Publishing Limited 2002.
11. Donabedian A. The quality of care: how it can it be assessed? *JAMA* 1988;260:1743-48.
12. Joshi MS, Nash DB, Ransom SB. *The Healthcare Quality Book*. Chicago, IL: Health Administration Press 2005.
13. Kovner A. Futures. In A. R. Kovner (Ed.), *Jonas's health care delivery in the United States* (5th ed., pp. 1995;532-53). New York: Springer Publishing.
14. Tarazi EM, Philip BK. Friendliness of or staff is top determinant of patient satisfaction with outpatient Surgery. *Am J Anesthesiology* 1998;25:154-7.
15. Brown DL, Warner ME, Schroeder DR and Offord KP. Effect of intraoperative anesthetic events on postoperative patient satisfaction. *Mayo Clinic Proc* 1997;72:20-5.
16. Reason J. "Safety in the operating theatre part 2: Human error and organizational failure." *Qual Saf Health Care* 2005;14:56-60.
17. Silen-Lipponen M, Tossavainen K, Turunen H and Smith A. Potential errors and their prevention in operating room team work as experienced by Finnish, British and American nurses. *Int J Nurs Prac* 2005;11:21-32.
18. Weinger M, Englund C. Ergonomic and human factors affecting anesthetic vigilance and monitoring performance in the operating room environment. *Anesthesiology* 1990;73:995-1021.
19. ACOG Committee Opinion: Patient safety in the surgical environment. *Obstet Gynecal* 2006; 107:429-33.
20. Makary MA, Sexton BJ, Freischlag JA, Millman EA et al. Patient safety in surgery. *Ann Surg* 2006;243:628 –35.
21. Vissers J, Beech R. Health operations management. Patient flow logistics in health care 2005.
22. Morton A. Analysis: What's the difference between a hospital and a bottling factory. *Br Med J* 2009;339:2727.

23. Royston G. Shifting the balance of care into the 21st century. *Eur J Oper Rer* 1998;105:267-276.
24. Delesie L. Bridging the gap between clinicians and health managers. *Eur J Oper Rer* 1998;8:28-35.
25. Vries G de, Bertrand JWM, Vissers JMH. Design requirements for healthcare production control systems. *Prod Plan Control* 1999;10:559-69.
26. Vissers JMH, Bertrand JWM, De Vries G. A framework for production control in health care organisations. *Prod Plan Control* 2001;12:591-604.
27. Bertrand JWM, Wortmann JC, Wijngaard J. *Production control: A structural and design oriented approach*. Elsevier, Amsterdam 1999.
28. Beaulieu H, Ferland JA, Gendron B, Michelon P. A mathematical programming approach for scheduling physicians in the emergency room. *Health Care Manage Sci* 2000;3:19-200.
29. Carter MW, Lapierre SD. Scheduling emergency room physicians, *Health Care Manage Sci* 2001;4:347-60.
30. McManus ML, Long MC, Cooper A, Mandell J, Berwick DM, Pagano M, Litvak E. Variability in surgical caseload and access to intensive care services. *Anesthesiology* 2003;98:1491-96.
31. Wickizer TM. Effect of hospital utilization review on medical expenditures in selected diagnostic areas: An exploratory study. *Am J Public Health* 1991;81:482-84.
32. Dexter F, Epstein RH, Marsh HM. A statistical analysis of weekday operating room anesthesia group staffing costs at nine independently managed surgical suites. *Anesth Analg* 2001;92:1493-98.
33. Dexter F, Traub RD. Sequencing cases in the operating room: Predicting whether one surgical case will last longer than another. *Anesth Analg* 2000;90:975-79.
34. Dexter F, Traub RD, Macario A. How to release allocated operating room time to increase efficiency: Predicting which surgical service will have the most underutilized operating room time. *Anesth Analg* 2003;96:507-12.
35. Lapierre SD, Batson C, McCaskey S. Improving on-time performance in health care organisations: A case study. *Health Care Manage Sci* 1999;2:27-34.
36. Litvak E, Buerhues PI, Davidoff F, Long M, et al. Managing unnecessary variability in patient demand to reduce nursing stress and improve patient safety. *J Qual Pat Safety* 2005;31:330-38.
37. Litvak E. Optimizing Patient Flow by Managing its Variability. In “Berman S. (ed.): Front Office to Front Line: *Essential Issues for Health Care Leaders*”. Oakbrook Terrace, IL: Joint Commission Resources, 2005, pp. 91-111.

TEP01
EHAM
TO T/D
0010Z
WIND
009° / 19K

2

Modeling procedure and surgical times for CPT-anesthesia-surgeon combinations and evaluation in terms of case duration prediction and OR efficiency

A multi-center study

Pieter S. Stepaniak[ł], Christiaan Heij[¤], Ph.D., Guido H.H. Mannaerts[±], M.D. Ph.D., Marcel de Quelerij[¥], M.D., Guus de Vries[ł], Ph.D.

Published Anesth Analg 2009;109:1232-45

[ł] Institute of Health Policy and Management, Erasmus University Rotterdam
[¤] Econometric Institute, Erasmus School of Economics, Erasmus University Rotterdam
[±] Department of General Surgery, St. Franciscus Hospital Rotterdam
[¥] Department of Anesthesiology, St. Franciscus Hospital Rotterdam

Abstract

Background
Gains in OR scheduling may be obtained by using accurate statistical models to predict surgical and procedure times. The three main contributions of this paper are the following: (i) the validation of Strum's results on the statistical distribution of case durations, including surgeon effects, using OR databases of two European hospitals, (ii) the use of expert prior expectations to predict durations of rarely observed cases and (iii) the application of the proposed methods to predict case durations, with an analysis of the resulting increase in OR efficiency.

Methods
We retrospectively review all recorded surgical cases of two large European teaching hospitals in the period 2005-2008, involving 85,312 cases and 92,099 hours in total. Surgical times tend to be skewed and bounded by some minimally required time. We compare the fit of the normal distribution with that of two- and three-parameter lognormal distributions for case durations of a range of CPT-anesthesia combinations, including possible surgeon effects. For cases with very few observations, we investigate whether supplementing the data information with surgeon's prior guesses helps to obtain better duration estimates. Finally, we use best fitting duration distributions to simulate the potential efficiency gains in OR scheduling.

Results
The three-parameter lognormal distribution provides the best results for the case durations of CPT-anesthesia (surgeon) combinations, with an acceptable fit for almost 90% of the CPTs when segmented by the factor surgeon. The fit is best for surgical times and somewhat less for total procedure times. Surgeons' prior guesses are helpful for OR management to improve duration estimates of CPTs with very few (less than ten) observations. Compared to the standard way of case scheduling using the mean of the three-parameter lognormal distribution for case scheduling reduces the mean over-reserved OR time per case up to 11.9 (11.8-12.0) minutes (55.6%) and the mean under-reserved OR time per case up to 16.7 (16.5-16.8) minutes (53.1%). When scheduling cases using the three-parameter lognormal model the mean over-utilized OR time is up to 20.0 (19.7-20.3) minutes per OR per day lower than for the standard method and 11.6 (11.3-12.0) minutes per OR per day lower as compared to the biased corrected mean.

Conclusion
OR case scheduling can be improved by employing the three-parameter lognormal model with surgeon effects and by using surgeon's prior guesses for rarely observed CPTs. Using the three-parameter lognormal model for case duration prediction and scheduling significantly reduces both the prediction error and the over-utilized OR time.

Introduction

The OR is a major production unit in each hospital. For hospitals, the main two operational risks of ORs consist of high idle times (i.e., under-utilized OR time) and work outside regular hours (i.e., over-utilized OR time). Frequent work beyond scheduled hours does not only lead to overtime costs, but also to intangible costs resulting from dissatisfaction and reduced motivation of staff. Overtime work is one of the primary reasons for nurses to terminate their employment [1], and scheduling conflicts are a major cause of nursing staff turnover [2]. Therefore, efficient OR management should aim for maximal use of available OR time while preventing frequent overtime work [3]. OR schedules depend crucially on estimated case durations, and statistical models may help to improve these estimates to support management in the cost-efficient use of expensive surgical resources.

In the following, we provide a brief review of some relevant results in the literature on case duration distributions and case scheduling. Early results found that OR waiting times follow a two-parameter lognormal distribution [4] and that OR operation times follow a distribution that is normal [5] or lognormal [6]. Knowledge of the probability distributions of case durations has advanced markedly in the past decade [7,8,9]. The single most important source of variability in surgical procedure times is surgeon effect. Type of anesthesia, age, gender, and American Society of Anesthesiologists risk class were additional sources of variability [7]. In another study Strum [8] tested surgeries with two component procedures. The conclusion is that dual CPT surgeries were better modeled by the lognormal distribution than by the normal distribution. Surgical procedure times are frequently distributed with non-zero start times that require a lognormal model with a shifted parameter for best model estimates [9]. Decision rules based on the skewness and coefficient of variation of the data can be used to identify the correct alternative 78% of the time, but do not do any better than a single rule based on the skewness [9]. The way in which the lognormal location parameter is estimated affects the ability of goodness-of-fits tests to correctly recognize the model and the accuracy of percentile point values derived from the estimated model [10].

In an empirical study [11] is shown that surgical time and total procedure time are lognormal distributed. Surgical procedure time fits the lognormal distribution for 93% of all CPT codes whereas surgical time fits normal distribution for about 80% of all CPT codes studied. For some of the scheduled cases there is few or no data available, making statistical modeling difficult. These cases can disproportionately affect decision making under uncertainty because not sufficient data-driven recommendations could be obtained. A number of studies have tried to solve this problem of few or no cases [12,13]. Dexter [12] validated a practical way to calculate prediction bounds and compared the OR times of all cases, even those with few or no historic data for surgeon and the scheduled procedure(s).

The conclusion of this study is that when historic data are available, they should be used in combination with the scheduled OR time. Historic data provide value in estimating the proportional variation in OR time. Finally, the scheduled OR time alone is nearly as good a predictor of the expected mean OR time of a new case as the Bayesian method.

In another study [14] elective case scheduling at hospitals and surgical centers at which surgeons and patients choose the day of surgery, cases are not turned away, and anesthesia and nursing staffing are adjusted to maximize the efficiency of use of OR time. In this study two patient-scheduling rules are investigated: Earliest Start Time or Latest Start Time. The achievable incremental reduction in overtime by having perfect information on case duration versus using historical case durations was in this study only a few minutes per OR. The differences between Earliest Start Time and Latest Start Time were also only a few minutes per OR. There are cases which have a high probability of taking longer than scheduled. Increasing the case's scheduled duration could than reduce over-utilized OR time [15]. Dexter [15] studied surgeons' and schedulers' case scheduling behavior to evaluate whether such a strategy would be useful. The impact of inaccurate, scheduled case duration on staffing costs and unpredictable work hours can be reduced by allocating appropriate total hours of OR time (i.e., staffing) for the cases that will get done, regardless of the inaccuracy of the scheduled durations of those cases.

There are many other studies related to optimally scheduling cases [16,17,18,19,20,21,22,23]. All these studies contribute to optimizing the use of scarce and costly operating rooms. Based on the above studies we can conclude that gains in OR scheduling efficiency may be obtained by using accurate statistical models to predict surgical and procedure times. Therefore the three main contributions of this paper are the following: (i) the validation of Strum's results on the statistical distribution of case durations, including surgeon effects, using OR databases of two European hospitals, (ii) the use of expert prior expectations to predict durations of rarely observed cases and (iii) the application of the proposed methods to predict case durations, with an analysis of the resulting OR efficiency.

Material and methods

In this section we first present our database. Then we describe our methods.

Data

We retrospectively reviewed all recorded surgical cases from two large teaching hospitals in the period 2005-2008 (total 85,312 cases).

Because there are differences in case duration based on type of anesthetic used, we classify the CPT codes by type of anesthesia: general, local, and regional [(8, 24)]. Monitored anesthesia care is not a type of anesthesia used in the hospitals under study. We use the following definitions. Surgical Time: the time from incision to closure of the wound. Procedure Time: time when patient enters the operating suite until the patient leaves the OR. To detect the influence of sample size on the Shapiro-Wilk[1] test, we divided the sample size into very small , small , medium ($30 \leq n < 200$), and large ($n \geq 200$). In Table 1 we present the dataset for hospital A. For every case frequency interval, the number of CPT codes, the number of cases, and the total hours spent for these cases in the period 2005-2008 is shown.

TABLE 1. OPERATING ROOM HOSPITAL A AND B

	Hospital A			Hospital B		
Case frequency	CPT-codes	Cases	Hours	CPT-code	Cases	Hours
n ≥ 200	53	22,512	20,248	35	20,417	18,324
30 ≤ n < 200	347	16,388	19,973	287	16,587	19,184
10 ≤ n < 30	454	4,012	6,266	419	3,271	4,641
n < 10	318	1,004	1,717	201	1,121	1,746
Total	1172	43,916	48,204	942	41,396	43,895
Case frequency (1 CPT code)						
n ≥ 200	53	22,512	20,248	35	20,417	18,324
30 ≤ n < 200	253	11,621	13,707	198	13,984	15,187
10 ≤ n < 30	431	3,174	4,518	337	2,912	3,681
n < 10	287	541	823	189	762	1,117
Total	1024	37,848	39,296	759	38,075	38,308

	Hospital A		Hospital B	
Number of CPT-Anesthesia combinations	CPT	Cases	CPT	Cases
Anesthesia	685	17,561	579	15,664
General	147	4,310	134	4,408
Local	340	22,045	229	21,324
Regional	1172	43,916	942	41,396
1 CPT-Anesthesia combinations (case frequency ≥ 10)				
General	433	15,007	294	14,783
Local	157	4,050	85	2,543
Regional	147	18,250	191	19,987
Total	737	37,307	570	37,313

[1] For further information concerning the Shapiro-Wilk test we refer to the statistics section

Total cases amount to 44,223, of which 289 (0.7%) cases were omitted due to incomplete data. In 15 cases (0.03%) the operation was canceled, although the patient received anesthesia, and in three cases a donor procedure was performed. In our analyses we have 43,916 cases (1,172 CPT-anesthesia combinations), with hours totaling 48,204.

There were 37,848 cases (39,296 hours) with one CPT-anesthesia combination, 5,177 cases (7,312 hours) with two CPT codes and 891 cases (1,596 hours) with more than two CPT codes. Average cases per year with 2-CPTs: 1,294 (median 1,305 min 1,165 max 1,401). For CPTs with more than two codes the average is 222 cases (median 221, min 201 max 247). To eliminate a potential confounding factor[2], in our study we considered only surgical procedures with a single CPT code. Therefore, we confined our analysis to 37,307 cases with a case frequency of ≥ 10 (737 CPT-anesthesia combinations).
We broke down the CPTs according to the various surgeons (Table 2).

TABLE 2. NUMBER OF SURGEONS/ANESTHETISTS

	Hospital A	Hospital B
Specialty		
Eye surgery	3	3
Orthopedic	4	4
Ear-nose-throat	2	2
Neurosurgery	2	-
Urology	3	3
General surgery	6	6
Obstetrics and genecology	6	6
Jaw surgery	2	-
Plastic surgery	2	-
Total	30	24
Number of anesthetists	9	10

There are 30 surgeons and 6,349 CPT-anesthesia-surgeon combinations (43,916 cases, 48,204 hours, Table 3).

If we differentiate to combinations with at least 10 cases per surgeon and 1 CPT-anesthesia code, 1,341 CPT-anesthesia-surgeon combinations remain (32,347 cases, 34,512 hours).

[2] In our analyses, we use only surgeries with one CPT code (as in Strum [12]) to avoid possible confounding factors. Procedures with for example two CPT codes (CPT1 and CPT2) can be performed in different ways. First CPT1 and then CPT2, or vice versa. The sequence can then be a confounding factor.

TABLE 3. NUMBER OF UNIQUE CPT CODE-ANESTEHSIA -SURGEON COMBINATIONS

Specialty	Hospital A	Hospital B
Eye surgery	61	48
Orthopedic	970	1141
Ear-nose-throat	71	59
Neurosurgery	41	0
Urology	174	137
General surgery	2108	1244
Obstetrics and genecology	2420	1844
Jaw surgery	47	0
Plastic surgery	457	0
Total	6349	4473
Cases	43,916	41,396
Hours	48,204	43,895

Irrespective of the number of CPT codes: of the 1,172 CPT codes there are 318 CPT-anesthesia combinations (1,004 cases, 1,717 hours) which were performed less than ten times in a period of four years. Of the 43,916 cases scheduled, for 46 cases (0.1%) the actual procedure code was different than the scheduled code. In 132 cases (0.3%) the actual surgeon was different than the scheduled surgeon.

In Table 1 the data set for hospital B is presented. Total cases amount to 41,916, of which 520 (1.2%) cases were omitted due to incomplete data. The analysis is limited to 41,396 cases (942 CPT-anesthesia combinations, 43,895 hours). There were 38,075 cases (38,308 hours) with only one CPT code, 2,707 cases (4,531 hours) with two CPT codes and 614 cases (1,056 hours) with more than two CPT codes. Average cases per year with 2-CPTs: 676 (median 687 min 634 max 699). For CPTs with more than two codes the average is 153 cases (median 151, min 143 max 166). As in hospital A, we considered only cases with one CPT code and each CPT-anesthesia combination with a case frequency of 10 or more. We confined our analysis to 37,313 cases (570 CPT-anesthesia codes). There are 24 surgeons (Table 2) and 4,473 CPT-anesthesia-surgeon combinations (41,396 cases, 43,895 hours Table 3).

If we differentiate to combinations with at least 10 cases per surgeon, 1,147 CPT-anesthesia-combinations remain (30,274 cases, 32,927 hours). Of the 41,396 cases scheduled, in 28 cases (0.07%) actual procedure code was different than the scheduled code. In 89 cases (0.2%) the actual surgeon was different than the scheduled surgeon. Next we describe in detail what we have studied and how the study is performed.

Fitting the normal, two- and three-parameter lognormal models for 1-CPT-anesthesia-surgeon combinations with case frequency ≥ 10

We repeat Strums' work [9,10,11] for the normal, two- and three-parameter log-normal modeling of surgical procedure times. Repeating Strums' work is important scientifically because replication of research is a way to refine our understanding of modeling surgical cases. The three-parameter lognormal model is of interest because surgical procedure times are frequently distributed with non-zero start times that require a lognormal model with a shift parameter for best model estimates [9,11]. A non-zero start time means that minimum surgical procedure times, even for the simplest procedures, are strictly positive. As is assumed [11], the percentage of cases that fit the lognormal model can be even higher when segmented by the factor surgeon. Therefore we validate whether performed procedure times and surgical times of CPT-anesthesia-surgeon combinations fit a normal, two-parameter or three-parameter lognormal distribution.

The general formula for the lognormal model can be described as follows:

$$f_X(x;\ \mu, \sigma, \theta) = \frac{1}{(x-\theta)\sigma\sqrt{2\pi}} e^{-\frac{(\ln(x-\theta)-\mu)^2}{2\sigma^2}}$$

for x > θ, where θ = shift parameter for duration data θ > 0. The case where θ equals zero is called the two-parameter lognormal model. For the three-parameter lognormal model, we estimate the shift parameter by using a modified version of the approach of Spangler [10]. The shift parameter describing the location or origin of the random variable is important for decision making because it provides a lower bound on values of the random variable [10]. First for every CPT-anesthesia (surgeon) combination we calculate the natural logarithm of surgical time and procedure time. We then use the bisection method to estimate the shift parameter. That way we estimate three parameters for each combination of surgeon(s) and procedure(s). The bisection method we used is as follows:

Set LOWER=0

Set UPPER = smallest observed value

Initial Guess= (LOWER+UPPER)/2

Subtract GUESS from all observed values, take the logarithm, and estimate the mean and standard deviations. Then recalculate the Shapiro Wilk *p*-value (= p_{new}). We repeat this iteratively using bisection to find the shift parameter that results in the largest value of the p-value. We choose to stop the iteration if $(p_{new} - p_{old}) / p_{new} \times 100\% < 1\%$ or if $p_{new} < p_{old}$. If the final *p*-value is larger than 0.05, we do not reject the hypothesis of the normal model.

Estimation with Specialist Prior Guess

If very few data are available ($n < 10$), it may help to use prior information to get more reliable estimates of the time distribution. Therefore we present a method to estimate the mean procedure time from prior and actual data for procedures with less than 10 cases. As is commonly known, Bayes's theorem provides a mechanism for combining a prior probability distribution for the states of nature with sample information to provide a revised (posterior) probability distribution about those states of nature. These posterior probabilities are then used to make better decisions. Our approach differs with that of Dexter [12,25] in the way that we use the surgeon's prior statement on the distribution in terms of quantiles of the operation time. To get the prior information required, we asked surgeons to make a prior statement on the distribution of the procedure time for cases with a frequency less than 10.

For a given procedure, we asked surgeons in the period October – December 2008 before they started the scheduled case to make an estimation in terms of quantiles (25%, 50%, 75% and 95%) of the time distribution of a procedure. With this information we were able to update our uncertainty in the light of new evidence. In the analyses we use the two-parameter lognormal model where the mean and variance is calculated from a weighted mean of the actual data and the prior data. Further we assume that the specialists do not remember the previous operation times, so that all realized times (past and current) can be treated as containing similar information. Next we explain our model for using prior information in mathematical terms.

Let T denote the procedure time and let ln(T) be its natural logarithm. Assuming a two-parameter lognormal model for the procedure time, it follows that ln(T) is normal with mean μ and variance σ^2. We now have to combine the prior and actual data information to estimate the mean μ. Let m be the prior mean and s the prior variance of ln(T). Let the 25%, 50%, 75%, and 95% quantiles of Ln(T) be denoted by T25, T50, T75, and T95, then the normal distribution implies that:

$$\begin{pmatrix} T25 \\ T50 \\ T75 \\ T95 \end{pmatrix} = m \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} + s \begin{pmatrix} -0.675 \\ 0 \\ 0.675 \\ 1.645 \end{pmatrix} \qquad (1)$$

For example, with two prior estimates made by specialist 1 and 2 we estimate the following model:

$$\begin{pmatrix} T25 \ (specialist\ 1) \\ T50 \ (specialist\ 1) \\ T75 \ (specialist\ 1) \\ T95 \ (specialist\ 1) \\ T25 \ (specialist\ 2) \\ T50 \ (specialist\ 2) \\ T75 \ (specialist\ 2) \\ T95 \ (specialist\ 2) \end{pmatrix} = m \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} + s \begin{pmatrix} -0.675 \\ 0 \\ 0.675 \\ 1.645 \\ -0.675 \\ 0 \\ 0.675 \\ 1.645 \end{pmatrix} \quad (2)$$

where [-0.675 0 0.675 1.645] is the vector with corresponding z-values. This vector provides the regressor needed to estimate location and scale of the lognormal prior distribution corresponding to the quantiles. The vector is repeated for each specialist.

When we have data for j specialists, hence 4j times, we get 4j equations with given values on the left-hand side and with unknown values of m and s. This can be seen as a regression model with two unknown parameters, m (the constant term) and s (the slope). By applying regression, it is well known that the constant term m will be the sample mean of the Ln(Quantiles) values and that the slope s can also be computed quite easily.

The prior mean of the operation time is $\exp(m + 0.5\, s^2)$ and the prior variance is $(\exp(s^2) - 1) \exp(2m + s^2)$. The prior standard deviation is $[(\exp(s^2) - 1) \exp(2m + s^2)]^{0.5}$.
We take the value of m as the prior mean xs^* and the value s^{-2} of as τ. The posterior mean is then given by the formula:

$$\mu^* = xs^* \frac{\tau}{\tau + n_k} + \overline{x_k} \frac{n_k}{\tau + n_k} \qquad (3)$$

The resulting weight is $w = \tau / (\tau + n)$and the posterior mean is equal to:

$$\mu^* = wxs^* + (1 - w)\overline{x} \qquad (4)$$

Note that this is the posterior mean of the log times. The mean of the actual times is given by $\exp(\mu^* + 0.5\sigma^{*2})$, where σ^{*2} is the posterior variance. The prior variance is s^2 and the data variance is σ^2. An intuitive method is to weight these two values in the same way as was done for the mean, so that

$$\sigma^{*2} = ws^2 + (1 - w)\sigma^2 \qquad (5)$$

Combining these results, we get: the posterior for the operation times is lognormal with mean μ^* and variance σ^{*2}. The mean of the operation times is then given by:

$$\exp(\mu^* + 0{,}5\sigma^{*2}) \qquad (6)$$

Improving coupling between estimates of scheduled time and the actual procedure time
When reserving OR time for a procedure, the OR management needs to balance the costs of reserving too much time against the costs of reserving too little [26]. If too much time is allocated to a case, expensive OR capacity is likely to be wasted, leading to a decrease in OR utilization [12,13,14,16,17,21,22,23,27]. With too little allocated capacity to a surgical case, the OR schedule must be modified, resulting in idle OR times in operation rooms and increased demand for anesthesiologists, nurses and support staff. Improving coupling between estimates of scheduled time and the actual time reduces the prediction error of a scheduled surgical case. By using a simulation we compare the effect on the prediction error of scheduling cases when applying three different case modeling methods. The first method of estimating scheduled case durations is based on taking the trimmed mean time of the last 10 case durations.

The second method uses the bias-corrected scheduled OR time. This method is based on the following linear regression based on data 2005-2007: *Actual OR time = intercept + slope * (scheduled OR time)*. This regression shows how much better it is for purposes of choosing how long to schedule a case (as compared to lower/upper prediction bounds or times remaining in cases) to use statistically based methods as compared to simple adjustment of the scheduled OR time. The last method uses the mean of the three-parameter lognormal model.

To make it possible to compare the outcome of the three methods, only procedures are used with a case frequency of 10 or more, with 1 CPT-anesthesia code and fitting the three-parameter lognormal model. We use the data available (from the sample). Historical data from 2005-2007 is used and then the window is expanded to include predictions made on each day in 2008 using data from 2005-2007 and from 2008 till the day prior to making the prediction. The originally scheduled sequence of cases was not changed. For instance, when scheduling an Inguinal hernia repair (Lichtenstein) on January 2nd 2008, only historical data up to and including January 1st 2008 is used. The actual time on January 2nd is used for scheduling this procedure in for January 4th. The difference between the actual OR time of a procedure is compared with the scheduled procedure time as calculated by each the of the three methods. If the actual procedure time is larger than the scheduled time, that procedure is under-reserved. Otherwise it is over-reserved. For each method, the number of under- or over-estimated procedures is counted as the mean under- and overestimated time per case. Differences in the mean under- and over-estimated time per case between the three methods were tested with an paired *T*-test.

OR Inefficiency

Operating room inefficiency is defined as the sum of under-utilized OR time and over-utilized OR time, multiplied by the relative costs of overtime [16,23,26]. Under-utilized time is hours of staffed operating time at straight time wages, but not used for surgery, set-up or clean-up of the OR. Over-utilized time is hours after operating room time, staffed at overtime. The relative cost of overtime in our study is 1.50. The cost per hour of over-utilized OR time includes: indirect costs, intangible costs, and retention and recruitment costs incurred on a long-term basis due to staff working late. Due to fixed OR capacity in our hospital (8AM – 4 PM), the short-term objective in maximizing OR efficiency is to reduce over-utilized OR time [15]. In hospital A for example the mean end time of all ORs running after 4 PM is 4.19 (±17) minutes [3].

We analyzed the effect of the different methods of case duration prediction on the OR efficiency. In the first method we use the trimmed mean of the last 10 case durations, in the second method we used the bias-corrected scheduled OR time and the mean of the three-lognormal model in the last method. Case scheduling with original cases in 2008 were used. For each method add-on elective cases with their concomitant turnover times were daily scheduled. *Best Fit Descending* was used, an off-line algorithm in which add-on elective cases are sorted based on longest to shortest with fuzzy constraints. Cases were considered in the order specified by the algorithm. If no OR had sufficient open time available for the case, but sufficient open time was available in the OR with the most remaining time provided the scheduled duration of the case was shortened by ≤ 15 min, then the case was assigned to the OR with the most remaining time [28].

For all cases (2 or more CPTs, and procedures with case frequency < 10) that are not meeting the criteria, we used the actual case duration as the scheduled duration (i.e., perfect retrospective knowledge). After scheduling the cases and knowing the actual OR times of these same cases the mean over-utilized OR time was calculated considering each OR-day to be independent of all others. Differences in the mean over-utilized OR time between the three methods were tested with an paired *T*-test.

Statistics and software

The null hypothesis of the Shapiro-Wilk test statistic (W) is that a sample is from a normally distributed population. Thus $p < 0.05$ for W rejects this supposition of normality. Most authors agree that this is the most reliable test for non-normality for small to medium-sized samples [29,30,31,32,33,34,35,36,37].

To perform the Shapiro-Wilk test we use StatsDirect statistical software. Further we used SPSS15, Excel 2007, and COBOL. Normal probability plots where examined visually for those CPT-anesthesia-(surgeon) combinations that were not well-fitted by either the normal or lognormal models. We analyzed QQ-PP and Box plots to confirm the results of the Shapiro-Wilk test. Examination of the calculated skewness and kurtosis, and of the histogram, boxplot, and normal probability plot for the data may provide clues as to why the data failed the Shapiro-Wilk. In our database, start and end of anesthesia time, surgical time and procedure time are recorded exactly (to the minute). D'Agostino [29] pointed out that the Shapiro-Wilk test can be affected by rounding.

Results

Fitting the normal, two- and three-parameter lognormal models

In some of the procedures we found outliers. In the database there is a so-called "remark field" in which unexpected events during an OR are filled in. The outliers we encountered were due to logistical problems (16 times) in the OR, surgeon arriving late (12 times) and OR team not ready (4 times). These outliers can be seen as incidental therefore we removed these data. Table 4 shows the results of fitting CPT-anesthesia groups to the normal and the two-, three-parameter lognormal models for both hospitals separately.

TABLE 4. TABULAR COMPARISON OF SAMPLE SIZE AND SHAPIRO-WILK GOODNESS-OF-FIT P-VALUES FOR THE NORMAL MODEL FOR PROCEDURE TIME AND SURGICAL TIME (1 CPT-ANESTHESIA COMBINATIONS)

Normal model Hospital A							
Procedure time		p < 0,05		p ≥ 0,05		CPT codes	
Small	10≤ n<30	271	36.8%	160	21.7%	431	58.5%
Medium	30≤ n<200	153	20.8%	100	13.6%	253	34.3%
Large	n ≥200	34	4.6%	19	2.6%	53	7.2%
Total		458	62.1%	279	37.9%	737	100.0%
Surgical time							
Small	10≤ n<30	204	27.7%	227	30.8%	431	58.5%
Medium	30≤ n<200	124	16.8%	129	17.5%	253	34.3%
Large	n ≥200	22	3.0%	31	4.2%	53	7.2%
Total		350	47.5%	387	52.5%	737	100.0%
Hospital B							
Procedure time							
Small	10≤ n<30	194	34.0%	143	25.1%	337	59.1%
Medium	30≤ n<200	138	24.2%	60	10.5%	198	34.7%
Large	n ≥200	34	6.0%	1	0.2%	35	6.1%
Total		366	64.2%	204	35.8%	570	100.0%
Surgical time							
Small	10≤ n<30	146	25.6%	191	33.5%	337	59.1%
Medium	30≤ n<200	100	17.5%	98	17.2%	198	34.7%
Large	n ≥200	15	2.6%	20	3.5%	35	6.1%
Total		261	45.8%	309	54.2%	570	100.0%
2-Parameter lognormal model Hospital A							
Procedure time							
Small	10≤ n<30	179	24.3%	252	34.2%	431	58.5%
Medium	30≤ n<200	112	15.2%	141	19.1%	253	34.3%
Large	n ≥200	21	2.8%	32	4.3%	53	7.2%
Total		312	42.3%	425	57.7%	737	100.0%
Surgical time							
Small	10≤ n<30	121	16.4%	310	42.1%	431	58.5%
Medium	30≤ n<200	89	12.1%	164	22.3%	253	34.3%
Large	n ≥200	14	1.9%	39	5.3%	53	7.2%
Total		224	30.4%	513	69.6%	737	100.0%

Hospital B							
Procedure time		p < 0,05		p ≥ 0,05		CPT codes	
Small	10≤ n<30	101	17.7%	236	41.4%	337	59.1%
Medium	30≤ n<200	86	15.1%	112	19.6%	198	34.7%
Large	n ≥200	29	5.1%	6	1.1%	35	6.1%
Total		216	37.9%	354	62.1%	570	100.0%
Surgical time							
Small	10≤ n<30	84	14.7%	253	44.4%	337	59.1%
Medium	30≤ n<200	64	11.2%	134	23.5%	198	34.7%
Large	n ≥200	9	1.6%	26	4.6%	35	6.1%
Total		157	27.5%	413	72.5%	570	100.0%
3-Parameter lognormal model							
Hospital A							
Procedure time							
Small	10≤ n<30	86	11.7%	345	46.8%	431	58.5%
Medium	30≤ n<200	41	5.6%	212	28.8%	253	34.3%
large	n ≥200	16	2.2%	37	5.0%	53	7.2%
Total		143	19.4%	594	80.6%	737	100.0%
Surgical time							
Small	10≤ n<30	67	9.1%	364	49.4%	431	58.5%
Medium	30≤ n<200	37	5.0%	216	29.3%	253	34.3%
Large	n ≥200	13	1.8%	40	5.4%	53	7.2%
Total		117	15.9%	620	84.1%	737	100.0%
Hospital B							
Procedure time							
Small	10≤ n<30	52	9.1%	285	50.0%	337	59.1%
Medium	30≤ n<200	31	5.4%	167	29.3%	198	34.7%
Large	n ≥200	19	3.3%	16	2.8%	35	6.1%
Total		102	17.9%	468	82.1%	570	100.0%
Surgical time							
Small	10≤ n<30	38	6.7%	299	52.5%	337	59.1%
Medium	30≤ n<200	35	6.1%	163	28.6%	198	34.7%
Large	n ≥200	8	1.4%	27	4.7%	35	6.1%
Total		81	14.2%	489	85.8%	570	100.0%

If we look at hospital A for the CPT-anesthesia combinations, then procedure times fit the normal model 37.9% and surgical time 52.5%. The fits for the two-parameter lognormal model ($p \geq 0.05$) are respectively 57.7% and 69.6%. For the three-parameter lognormal model, the fits for procedure time ($p \geq 0.05$) are 80.6% and 84.1% for surgical time. If we differentiate to CPT-anesthesia-surgeon combinations, then procedure times fits the two-parameter lognormal model ($p \geq 0.05$) in 70.4% of the combinations. The results for surgical times are 79.6% (Table 5. For the three-parameter lognormal model, the fits for the procedure time are 87.6% and 90.7% for surgical time. The results for hospital B are roughly in line with those for hospital A.

TABLE 5. TABULAR COMPARISON OF SAMPLE SIZE AND SHAPIRO-WILK GOODNESS-OF-FIT P-VALUES FOR THE NORMAL MODEL FOR PROCEDURE TIME AND SURGICAL TIME (1 CPT-ANESTHESIA-SURGEON COMBINATIONS)

Normal model							
Hospital A							
Procedure time		$p < 0.05$		$p \geq 0.05$		CPT codes	
Small	10≤ n<30	329	24.5%	322	24.0%	651	48.5%
Medium	30≤ n<200	371	27.7%	310	23.1%	681	50.8%
Large	n ≥200	8	0.6%	1	0.1%	9	0.7%
Total		708	52.8%	633	47.2%	1,341	100.0%
Surgical time							
Small	10≤ n<30	228	17.0%	423	31.5%	651	48.5%
Medium	30≤ n<200	329	24.5%	352	26.2%	681	50.8%
Large	n ≥200	7	0.5%	2	0.1%	9	0.7%
Total		564	42.1%	777	57.9%	1,341	100.0%
Hospital B							
Procedure time							
Small	10≤ n<30	318	27.7%	226	19.7%	544	47.4%
Medium	30≤ n<200	301	26.2%	295	25.7%	596	52.0%
Large	n ≥200	6	0.5%	1	0.1%	7	0.6%
Total		625	54.5%	522	45.5%	1,147	100.0%
Surgical time							
Small	10≤ n<30	219	19.1%	325	28.3%	544	47.4%
Medium	30≤ n<200	261	22.8%	335	29.2%	596	52.0%
Large	n ≥200	5	0.4%	2	0.2%	7	0.6%
Total		485	42.3%	662	57.7%	1,147	100.0%
2-Parameter lognormal model							
Hospital A							
Procedure time							
Small	10≤ n<30	201	15.0%	450	33.6%	651	48.5%
Medium	30≤ n<200	189	14.1%	492	36.7%	681	50.8%
Large	n ≥200	7	0.5%	2	0.1%	9	0.7%
Total		397	29.6%	944	70.4%	1,341	100.0%
Surgical time							
Small	10≤ n<30	101	7.5%	550	41.0%	651	48.5%
Medium	30≤ n<200	167	12.5%	514	38.3%	681	50.8%
Large	n ≥200	5	0.4%	4	0.3%	9	0.7%
Total		273	20.4%	1068	79.6%	1,341	100.0%

Hospital B							
Procedure time		$p < 0.05$		$p \geq 0.05$		CPT codes	
Small	10≤ n<30	221	19.3%	323	28.2%	544	47.4%
Medium	30≤ n<200	127	11.1%	469	40.9%	596	52.0%
Large	n ≥200	4	0.3%	3	0.3%	7	0.6%
Total		352	30.7%	795	69.3%	1,147	100.0%
Surgical time							
Small	10≤ n<30	124	10.8%	420	36.6%	544	47.4%
Medium	30≤ n<200	146	12.7%	450	39.2%	596	52.0%
Large	n ≥200	3	0.3%	4	0.3%	7	0.6%
Total		273	23.8%	874	76.2%	1,147	100.0%
3-Parameter lognormal model							
Hospital A							
Procedure time							
Small	10≤ n<30	77	5.7%	574	42.8%	651	48.5%
Medium	30≤ n<200	84	6.3%	597	44.5%	681	50.8%
Large	n ≥200	5	0.4%	4	0.3%	9	0.7%
Total		166	12.4%	1175	87.6%	1,341	100.0%
Surgical time							
Small	10≤ n<30	57	4.3%	594	44.3%	651	48.5%
Medium	30≤ n<200	64	4.8%	617	46.0%	681	50.8%
Large	n ≥200	4	0.3%	5	0.4%	9	0.7%
Total		125	9.3%	1216	90.7%	1,341	100.0%
Hospital B							
Procedure time							
Small	10≤ n<30	81	7.1%	463	40.4%	544	47.4%
medium	30≤ n<200	95	8.3%	501	43.7%	596	52.0%
Large	n>200	3	0.3%	4	0.3%	7	0.6%
Total		179	15.6%	968	84.4%	1,147	100.0%
Surgical time							
Small	10≤ n<30	57	5.0%	487	42.5%	544	47.4%
medium	30≤ n<200	89	7.8%	507	44.2%	596	52.0%
Large	n ≥200	2	0.2%	5	0.4%	7	0.6%
Total		148	12.9%	999	87.1%	1,147	100.0%

We tried to understand why surgical time fits the normal and two- and three-parameter lognormal models better than is the case for procedure time. Procedure time consists of three main activities: administering anesthesia, preparation the patient for surgery, and performing the actual surgery. For the cases under study, the proportion of surgery time is on average 75% of the total procedure time. For preparation time and anesthesia time, these percentages are respectively 18% and 7%. While preparing the patient for surgery, relatively more persons of the OR staff are involved in various activities and protocols as compared with administering anesthesia and surgery. To gain a better understanding of this, for every CPT-anesthesia code we tested both the anesthesia time and preparation time for the two-parameter lognormal model.

With $p \geq 0.05$, 92.5% of anesthesia time is lognormal distributed, whereas 17.6% of the preparation time shows a fit to the lognormal model. Hence preparation time is modeled badly as compared with anesthesia time. This could explain why procedure time is less modeled for the lognormal model than surgical time.

Table 6 is a paired comparison of the two- and three-parameter lognormal models and the normal model using the Friedman test. We compared the normal model with the two-parameter lognormal model and the three-parameter model. The two-parameter lognormal model was superior to the normal model for modeling procedure time and surgical time. The three-parameter lognormal is superior to the two-parameter lognormal model and normal model. Surgical time is estimated better than procedure time when modeling with both the two- and three-parameter lognormal models and the normal model.

TABLE 6. FRIEDMAN RESULTS FOR PAIRED COMPARISON OF THE SHAPIRO- WILK P-VALUE FOR THE NORMAL, LOGNORMAL AND THREE-PARAMETER LOGNORMAL FOR PROCEDURE AND SURGICAL TIMES

Hospital A				
Hypothesis	Ln2(pt) = Ln2(st)	N (pt) = N (st)	Ln2(st) = N(st)	Ln2(pt) = N(pt)
Friedman test statistic	41	335	247	298
Kendall's W	0.015	0.228	0.051	0.095
Rank sum	1315-1847	1121-1901	1598-1279	1761-1357
p value	0.0039	0.0011	0.0038	0.0117
Hospital B				
Hypothesis	Ln2(pt) = Ln2(st)	N (pt) = N (st)	Ln2(st) = N(st)	Ln2(pt) = N(pt)
Friedman test statistic	28	241	161	202
Kendall's W	0.151	0.625	0.094	0.597
Rank sum	895-981	1002-1354	1204-906	1364-994
p value	0.0084	0.0033	0.0128	0.0094
Hospital A				
Hypothesis	Ln3(pt) = Ln3(st)	Ln3(st) = Ln2(st)	Ln3(pt) = Ln2(pt)	
Friedman test statistic	121	501	498	
Kendall's W	0.011	0.064	0.054	
Rank sum	1528-1978	1705-1125	1814-1134	
p value	0.0021	0.0004	0.0003	
Hospital B				
Hypothesis	Ln3(pt) = Ln3(st)	Ln3(st) = Ln2(st)	Ln3(pt) = Ln2(pt)	
Friedman test statistic	87	435	477	
Kendall's W	0.131	0.101	0.314	
Rank sum	904-1041	1304-879	1415-987	
p value	< 0.000	< 0.000	< 0.000	
Hospital A				
Hypothesis	Ln3(st) = N(st)	Ln3(pt) = N(pt)		
Friedman test statistic	89	298		
Kendall's W	0.0047	0.0059		
Rank sum	1487-897	1948-1007		
p value	< 0.000	< 0.000		
Hospital B				
Hypothesis	Ln3(st) = N(st)	Ln3(pt) = N(pt)		
Friedman test statistic	147	198		
Kendall's W	0.0089	0.0021		
Rank sum	1487-879	1546-921		
p value	<0.000	<0.000		

Estimation with Specialist Prior Guess

In the results section we focus (arbitrarily) on the Total Thyreoidectomy procedure (Table 7). The results of other procedures are found in Table 8. The two procedure times (261 and 198 minutes) are realized after combining the prior statements of the specialists with the previously realized times. As we have data for two specialists, and therefore 8 times, we get 8 equations with given values on the left-hand side, the values in the column "Ln(Quantiles) (Table 7) ", and with unknown values of *m* and *s*. In Table 7 we show the output of SPSS. The *R*-squared of this regression is 0.85, indicating a good fit.

TABLE 7. ESTIMATION OF THE MEAN PROCEDURE TIME FROM PRIOR AND ACTUAL DATA (TOTAL THYREOIDETOMY)

Actual	Ln(Actual)	Specialist	Quantiles	Ln(Quantiles)
268	5.591	1	100	4.605
126	4.836	1	150	5.011
378	5.935	1	170	5.136
311	5.740	1	200	5.298
172	5.147	2	110	4.7
361	5.889	2	160	5.075
162	5.088	2	200	5.298
261	5.565	2	220	5.394

Dependent variable: Ln(Quantiles)

Descriptive statistics of actual data (SPSS)

	Coefficients		*t*	Sig.	
	B	Std. Error			
Constant	4.947	0.047	105.804	0	
Z vector	0.287	0.049	5.835	0.001	
	N	Minimum	Maximum	Mean	Std. Dev.
Time	9	126	378	248.56	89.972
Ln(Time)	9	4.836	5.935	5.453	0.382

The outcomes are m = 4.947 and s = 0.287. In other words, the prior statements of the specialists can be translated as a lognormal model with a mean of 4.947 and a standard deviation of 0.287. The prior mean of the operation time is 147 minutes, and the prior variance is 1.847. The prior standard deviation is 43 minutes. We take the value of m = 4.947 as the prior mean xs* and the value of $1/s^2 = 1/0.2872^2 = 12.14$ as τ. The resulting weight is w = 0.574, and the posterior mean (equation 4, Methods) is equal to 5.162. The prior variance s^2 is 0.0824, and the data variance is 0.1459. Weighing these two values as was done for the mean (equation 4, Methods) gives a value of 0.109 for σ^{*2}. Combining these results, we get the posterior for the operation times which is lognormal distributed with mean μ^*= 5.162, and variance σ^{*2}= 0.109. The mean of the operation times is then 184 minutes. Note that the prior mean was 147 minutes, and the data mean time was 249 minutes. The posterior mean of 184 lies closer to the prior mean than to the data mean. This is because the prior distribution has a relatively small standard deviation (43 minutes) as compared to that of the data (90 minutes) and because the number of data points (9) is small. If we wish to determine, for instance, a 95% upper bound for the operation time, then this is done by estimating the 95% bound for the log-times. In our example, the log-time has normal posterior with μ^* = 5.162 and variance σ^{*2} = 0.109, so that σ^* = 0.330. Then the 95% upper bound for the log-time is μ^* + 1.645 σ^* = 5.705. The bound for the time itself is then exp(5.705) = 300 minutes.

Table 8 present the results for the data mean (SD), prior [*mean time*, *SD*] and posterior [*mean time*, *SD*] for 30 procedures. From this table we see that the posterior mean is a weight of the data mean and the prior mean. The variance of the posterior mean always lies between the data variance and prior variance.

TABLE 7. RESULTS OF THE ESTIMATION OF THE MEAN PROCEDURE TIME FROM PRIOR AND ACTUAL DATA

Procedure	A	B	C	D	E	F	G	H
Osteochondritis dissecans knee	4	2	41	16	55	12	51	14
Corticotomy chin	4	2	80	30	74	22	77	25
Strumectomy	9	2	211	53	200	46	205	48
Nephrectomy (abd/via lumb, open)	9	2	214	43	205	49	210	48
Parathyreoidectomy	8	1	158	28	182	44	174	38
Nerve repair	5	3	177	61	141	39	158	47
Recto-Sigmoidresection Hartman	8	1	176	26	143	48	161	34
Extirpatie glandula submandibularis	9	2	93	22	72	22	83	22
Stomach resection/Cholecystectomy	5	1	198	23	186	39	187	36
Distal Tibia	4	2	128	48	81	24	86	30
PTFE loop	4	2	191	101	171	87	172	96
Liver surgery	4	2	154	61	128	39	134	44
Sigmoid resection (open)	3	2	232	62	225	46	226	51
End. Stageringslymfadenectomy	8	2	156	25	159	40	157	40

Latissimus dorsi lap upper body	4	1	275	45	279	40	277	41
Total Thyreoidectomy	7	2	249	90	147	43	184	62
Transversectomy	7	2	145	42	168	38	165	39
Lymph dissection	9	2	171	55	91	40	136	47
Ileastomy	3	2	149	45	91	40	111	42
Rec. AP colon intraperitoneaal	6	2	95	54	88	26	91	53
PKLND,Stag. lymfadenectomy	8	2	274	158	228	60	238	110
Arhrodese knee	3	1	239	45	224	34	229	39
Intracondylaire Humerus	4	1	107	30	95	44	93	37
Debulking	4	2	260	74	124	51	153	70
Ureterimplantation	4	1	203	90	205	57	194	71
Nissen fundoplication	4	1	234	36	239	41	235	40
Rectosigmoidresection	4	1	231	70	172	21	173	25
Enteroanastomosis	3	1	215	175	137	37	156	57
Reconstructive surgery by Roux Y	5	2	332	134	300	61	311	78
Endoscopic hemicolectomy	8	3	224	79	216	48	219	57

Explanation A	Historical number of cases
B	Number of prior estimates
C	Data mean (min)
D	Historic data mean (min)
E	Prior mean (min)
F	Prior SD
G	Posterior mean
H	Posterior SD

Improving coupling between estimates of scheduled time and the actual procedure time In hospital A (Table 9), under the standard method the mean over-reserving per case is 22.9 (22.5-23.3) minutes while the mean under-reserving is 21.6 (21.3-22.0) minutes. The result of the regression is: a*ctual OR time = 18.16 + 0.88 * (scheduled OR time) with s*tandard error of the *constant* 0.30 and *slope* 0.04 (p < 0.0001), R-squared 0.55. Applying the biased regression then the mean over-reserving per case is 16.3 (16.2-16.5) minutes while the mean under-reserving is 12.6 (12.4-12.7) minutes. For the three lognormal model the results are 12.9 (12.7-13.0) over-reserving and 9.6 (9.5-9.7) under-reserving. The mean over- and under-reserving between the three methods are significant (p < 0.001). The results for hospital B are in line with hospital A.

TABLE 8: IMPROVING COUPLING BETWEEN ESTIMATES OF SCHEDULED AND THE ACTUAL TIME

Hospital A					
History: 2005-2007, cases used for scheduling 2008 (n= 10,664)					
	Method				
	1	2	3	Effect	
Method	Standard	Regres.	3-par	(1)-(2)	(1)-(3)
Total hours of over-reserving	2,284	2,401	1,041	-5.1%	54.4%
# cases over-reserving	5,992	8,815	4,845	-47.1%	19.1%
Average over-reserving/case (min)	22.9	16.3	12.9	28.8%	43.7%
SD (min)	21.4	9.4	8.4		
Total hours of under-reserving	1,534	387	765	74.8%	50.1%
# cases under-reserving	4,255	1,849	4,780	56.5%	-12.3%
Average under-reserving/case (min)	21.6	12.6	9.6	41.7%	55.6%
SD (min)	18.8	7.6	5.4		
Hospital B					
History: 2005-2007, cases used for scheduling 2008 (n= 8,794)					
	Method				
	1	2	3	Effect	
Method	Standard	Regres.	3-par.	(1)-(2)	(1)-(3)
Total hours of over-reserving	1,867	1,314	747	29.6%	60.0%
# cases over-reserving	5,031	5,631	4,294	-11.9%	14.6%
Average over-reserving/case (min)	22.3	14.0	10.4	37.2%	53.4%
SD (min)	21.8	8.3	7.1		
Total hours of under-reserving	1,874	949	924		
# cases under-reserving	3,713	3,163	4,078	14.8%	-9.8%
Average under-reserving/case (min)	30.3	18.0	13.6	40.6%	55.1%
SD (min)	26.7	10.1	8.2	62.2%	69.3%
Over-reserving: Scheduled time - actual time > 0 (hrs)					
Under-reserving: Scheduled time - actual time < 0 (hrs)					

OR Inefficiency

In hospital A 12,138 cases were scheduled. The mean over-utilized OR time (minutes) per OR per day for the standard method is 23.4 (22.7 - 24.0), for the biased corrected mean time 16.6 (16.1 – 17.2) and the three-lognormal 6.6 (6.2 – 6.9). For hospital B 8,794 cases were scheduled. The mean over-utilized OR time per OR per day for the standard method is 30.6 (29.6 – 31.5), for the bias- corrected mean time 22.2 (21.4 – 22.9) and for the three-lognormal model 10.6 (10.1 – 11.2).

Discussion

Modeling the distribution of OR cases is one of the key steps in a planning process. In our study the focus is more on decision making before the day of surgery. In other studies [12,16,25]. the focus is toward decisions on the day of surgery. These do not involve average OR times, but rather lower prediction bounds, upper prediction bounds, and especially times remaining in cases. Both focuses are helpful in effectively schedule and efficiently use expensive surgical resources. We find that the percentage of cases fitting the normal, two- and three-parameter lognormal models is higher for surgical time than for total procedure time (the opposite is true for Strum [11]). The evidence supports the idea that type of surgery is the most important single source of variability amongst surgeries [7]. Using the bisection method and applying the three-parameter lognormal model fits procedure time and surgical time better than the two-parameter lognormal model without shift parameter. This can be explained by the fact that the two-parameter lognormal model is a limitation of the three-parameter lognormal model. When segmenting to the factor surgeon, the fits are even higher for the two- and three-parameter lognormal models.

One could ask why the fits are better with CPT-anesthesia-surgeon segmentations. As an a priori hypothesis, Strum et al. suggest that this may be due to surgeon work rates [7]. If Strum et al. are correct, then segmentation into surgeon-specific groups should result in more homogeneous work rates and thus a better fit to the lognormal. Another reason is that due to further segmentation, the number of available cases reduces and because of this reduction of cases the *p*-values will increase. This could also explain why the lognormal model fits for the CPT-anesthesia-surgeon combinations are higher. We confirm as in other studies that small groups have a better fit than the medium and large groups. This lack of discrimination relates to the design of the statistical tests. D'Agostino, Shapiro, and others [29,31,32,33,34,35,36,37] discuss the fact that goodness-of-fit tests become more discriminating as the sample sizes increase. Conversely, it may be obvious that samples with $n < 10$ for example may indiscriminately fit almost any model.

If few data are available, the use of prior information given by the surgeon may lead to a better estimation of the case duration. Because the posterior distribution contains all the information we need to make statistical decisions, we can use it for predicting case durations and case scheduling. The uncertainty of the posterior data is less than when using only the data without prior information. On the other hand, if the amount of historical data for a specific procedure increases, the usefulness of the prior information will decrease. This is because with an increasing number of observations, the sample mean will determine the outcome. Our approach differs in some respects from the classical one as discussed [12]. This is caused by the fact that we have prior data that are quite informative and that can be translated in terms of a log-normal prior distribution. In the classical approach, the prior on the two parameters μ and σ^2 consists of three parts:

- For given σ, the (conditional) prior for μ is normal. The prior for σ is inverted gamma.
- The (unconditional, marginal) prior for μ is a t-distribution
- The (unconditional, marginal) posterior for μ is (another) t-distribution

Our prior information is not directly related to mean and variance, but can be translated to mean and variance of the normal distribution (of the log-times). So, we combine a normal prior with a normal distribution of the observed data. However, in applying the calculation rules to get the posterior, we employ the classical framework. So, this is not fully consistent. However, the central formulas (nr 2, 4) have a direct intuitively appealing interpretation that applies also in our framework: we take a weighted average of the prior and data information, and the weights are inversely proportional to the uncertainty involved in both types of information: proportional to $t = 1/s^2 = 1/($*prior variance*$)$ and to $n = 1/(1/n) = 1/($*data variance*$)$. If we wish to keep closer to the classical set-up, we need to estimate the parameters and of the prior (*inverted Gamma*) distribution of the standard deviation. These two parameters can be estimated by considering all other types of operation and modeling the resulting set of (inverted) sample variances for all these types of operations as in Dexter [13]. The (marginal) posterior of the mean (of the log-times) then becomes *t* instead of normal.

Further our results for CPTs with few data may potentially be useful if the data from the two hospitals were compared to findings in another study [12]. The latter paper did not find the Bayes method to have important value for the mean. The overall effect for every case including those with multiple CPT would be needed. Finally we find that compared to the standard way of case scheduling using the mean of the three-parameter lognormal distribution for case scheduling reduces the mean over-reserving OR time per case up to 53.1% and the under-reserving OR time up to with 55.6%. Using the three parameter lognormal model for case scheduling causes a lower mean over-utilized OR time up to 20.0 (19.7-20.3) minutes per OR per day as compared to the standard method and 11.6 (11.3-12.0) minutes per OR per day as compared to the bias-corrected scheduled OR time.

Limitations

The prior information could be misleading when the prior variance is too small, since specialists may underestimate the variance. Surgeon case durations for specific procedures may change progressively, for example as a result of subtle changes in the demographics of a patient population [15]. We asked specialties if they were aware of these changes. No specialty recognized that these changes had occurred in the past four years. We assumed that the specialists do not remember the previous operation times so we treated all realized times (past and current) as containing similar information. In practice, surgeons may or may not actually remember historical case durations.

Although the studied procedures have a relatively low occurrence and are performed by different surgeons, we believe that the effect of the memory of an individual surgeon on the results may exist but will be very small. In the simulation for case duration prediction and efficiency gains we omitted procedures not fitting the three-parameter lognormal mode and procedures with a case frequency less than ten.

Because of this the real efficiency gains may be over-estimated. Although in the hospitals under study 86% of all cases consist of 1-CPT code, we cannot make general conclusions or statements regarding the impact of improving case duration prediction on the efficiency of use of OR time, but only as related to the cases under study.

Conclusion

OR case scheduling can be improved by employing the three-parameter lognormal model with surgeon effects and by using surgeon's prior guesses for rarely observed CPTs. As compared to standard case scheduling practices and the biased corrected method using the three-lognormal model for case scheduling both significantly reduce the mean under- and over-estimated OR time per case as well over-utilized OR time.

References

1. Stachota P, Normandin P. Reasons registered nurses leave or change employment status. *J Nurs Admin* 2003;33:111-18.
2. Thomson TP, Brown H. Turnover of licensed nurses in skilled nursing facilities. *Nurs Econ* 2002;20:66-69.
3. Stepaniak PS, Mannaerts GHH, Quelerij M de, Vries G de. The effect of the operating room coordinatORs risk appreciation on operating room efficiency. *Anesth Analg* 2009;108:1249-56.
4. Rossiter CE., Reynolds JA. Automatic monitoring of the time waited in out-patient departments. *Med Care* 1963;1:218-25.
5. Barnoon S. Wolfe H. Scheduling a multiple operating room system: A simulation approach. *Health Serv Res* 1968;3:272-85.
6. Hancock WM, Walter PR, More RA, Glick ND. Operating room scheduling data base analysis for scheduling. *J Med Syst* 1988;12:397-409.
7. Strum DP, Sampson AR, May JH, Vargas LG. Surgeon and type of anesthesia predict variability in surgical procedure times. *Anesthesiology* 2000;92:1454-66.
8. Strum DP, May JH, Sampson AR, Vargas LG, Spangler WE. Estimating times of surgeries with two component procedures: comparison of the lognormal and normal models. *Anesthesiology* 2003; 98:232-40.
9. May JH, Strum DP, Vargas LG. Fitting the lognormal distribution to surgical procedure times. *Decision Sci* 2000;31:129-48.
10. Spangler WE, Strum DP, Vargas LG, May JH. Estimating procedure times for surgeries by determining location parameters for the lognormal model. *Health Care Manage Sci* 2004;7:97-104.
11. Strum DP, May JH, Vargas LG. Modeling the uncertainty of surgical procedure times: comparison of log-normal and normal models. *Anesthesiology* 2000;92:1160-7.
12. Dexter F, Ledolter J. Bayesian prediction bounds and comparisons of operating room times even for procedures with few or no historical data. *Anesthesiology* 2005;103:1259-67.
13. Dexter F, Traub RD. Statistical method for predicting when patients should be ready on the day of surgery. *Anesthesiology* 2000;93:1107-14.
14. Dexter F, Traub R. How to schedule surgical cases into specific operating rooms to maximize the efficiency of use of operating room time. *Anesth Analg* 2002;94:933-42.
15. Dexter F, Macario A, Ledolter J. Identification of systematic under-estimation (bias) of case durations during case scheduling would not markedly reduce over-utilized operating room time. *J Clin Anesth* 2007;19:198-203.
16. Dexter F, Epstein R.H, Traub R, Xiao Yan. Making management decisions on the day of surgery based on operating room efficiency and patient waiting times *Anesthesiology* 2004;101:1444-53.
17. Macario A, Dexter F: Estimating the duration of a case when the surgeon has not recently performed the procedure at the surgical suite. *Anesth Analg* 1999;89:1241-5.
18. Zhou J, Dexter F. Method to assist in the scheduling of add-on surgical cases-upper prediction bounds for surgical case durations based on the log-normal distribution. *Anesthesiology* 1998; 89:1228-232.
19. Dexter F, Epstein RH, Marsh HM. A statistical analysis of weekday operating room anesthesia group staffing costs at nine independently managed surgical suites. *Anesth Analg* 2001;92:1493-8.
20. Dexter F, Macario A. Forecasting surgical groups total hours of elective cases for allocation of block time: Application of time series analysis to operating room management. *Anesthesiology* 1999; 91:1501-8.

21. Dexter F, Epstein RH. Operating room efficiency and scheduling. *Anesthesiology* 2005;18:195-8.
22. Friedman DM, Sokal SM. Increasing operating room efficiency through parallel processing. *Ann Surg* 2006;243:10.
23. McIntosh C, Dexter F, Epstein RH. Impact of service-specific staffing, case scheduling, turnovers, and first-case starts on anesthesia group and operating room productivity: tutorial using data from an Australian hospital. *Anesth Analg* 2006;103:1499-516.
24. Dexter F, Dexter EU, Masursky D, Nussmeier NA. Systematic review of general thoracic surgery articles to identify predictors of operating room case durations. *Anesth Analg* 2008;106:1232-41.
25. Dexter F, Epstein R, Lee R, Ledolter J. Automatic updating of times remaining in surgical cases using Bayesian analysis of historical case duration data and “Instant Messaging” updates from anesthesia providers. *Anesth Analg* 2009;108:929-40.
26. Strum DP, Vargas LG, May JH. Surgical subspecialty block utilization and capacity planning: a minimal cost analysis model. *Anesthesiology* 1999;90:1176-85.
27. Dexter F, Macario A. When to Allocated operating room time to increase operating room efficiency. *Anesth Analg* 2004;98:758-62.
28. Dexter F, Macario A, Traub RD. Which algorithm for scheduling add-on elective cases maximizes operating room utilization? Use of bin packing algorithms and fuzzy constraints in operating room management. Anesthesiology 1999;91:1491-500.
29. D’Agostino, RB. Tests for the normal distribution. In “*R. B. D’Agostino & M. A. Stephens (Eds.), Goodness-of fit techniques*” . New York: Marcel Dekker, Inc., 1986;367-419.
30. Conover WJ, Practical Nonparametric Statistics (3rd edition). Wiley 1999.
31. Shapiro SS, Wilk MB. An analysis of variance test for normality. *BIoMetrika* 1965;52:591-9.
32. Royston JP. The *W* Test for Normality. *Applied Statistics* 1982;31:176-180.
33. Royston JP, Shapiro-Wilk. Normality test and P-value. *Applied Statistics* 1995;4:44.
34. Madansky A. *Testing for normality, Prescriptions for working statisticians*. New York, Springer-Verlag, 1988, pp 15-31.
35. Pearson ES, D’Agostino RB, Bowman KO. Tests for departure from normality: comparison of powers. *BIoMetrika* 1977;64:231-46.
36. Shapiro SS, Wilk MB. An analysis of variance test for normality. *BIoMetrika* 1965;52:591-611.
37. Shapiro SS, Wilk MB, Chen HJ. A comparative study of various tests for normality. *J Am Stat Assoc* 1968;63:1343-72.

230K

3

Modeling and prediction of surgical procedure times

Pieter S. Stepaniak[ł], Christiaan Heij[¤] , Ph.D., Guus de Vries[ł] , Ph.D.

Stat Neerl 2010;64:1-18

[ł] Institute of Health Policy and Management, Erasmus University Rotterdam
[¤] Econometric Institute, Erasmus School of Economics, Erasmus University Rotterdam

Abstract
Accurate prediction of medical operation times is of crucial importance for cost efficient operation room planning in hospitals. This paper investigates the possible dependence of procedure times on surgeon factors like age, experience, gender, and team composition. The effect of these factors is estimated for over 30 different types of medical operations in two hospitals, by means of *ANOVA* models for logarithmic case durations. The estimation data set contains about 30,000 observations from 2005 till 2008. The relevance of surgeon factors depends on the type of operation. The factors found most often to be significant are team composition, experience, and daytime. Contrary to widespread opinions among surgeons, gender has nearly never a significant effect. By incorporating surgeon factors, the accuracy of out-of-sample prediction of case durations of about 1,250 surgical operations in 2009 is improved by up to more than 15 percent as compared to current planning procedures.

Introduction

Operating rooms (ORs) are among the most expensive surgical resources in hospitals [1]. In an era of cost-constrained health care, efficiency increases if a larger number of surgical operations can be performed within the available OR time [2]. The OR management of medical institutions needs to balance the costs of reserving too much time, with resulting idle time of the OR, against the costs of reserving too little time. In the last case, the OR schedule must be modified, resulting in an increased demand for anesthesiologists, nurses, and support staff. Therefore, accurate prediction of case durations helps in effective OR scheduling, it reduces waiting times for patients and idle times of medical and other staff, and thereby it improves the quality of health care delivered in other services throughout the hospital.

Surgical procedure times are inherently unpredictable, and the amount of uncertainty varies greatly among different types of operations. Hospitals employ standard classifications of operations, in terms of so-called current procedure terminologies (CPTs). The type of anesthesia also affects case durations, as shown in Strum et al. [3,4] and Dexter et al. [5].

The purpose of this paper is to identify factors affecting case durations and to exploit these factors to improve case duration predictions. The empirical analysis is based on extensive data bases of surgical operations in two teaching hospitals in The Netherlands. The OR management in these two hospitals often receives arguments brought forward by surgeons, anesthetists, and OR staff, as to why surgical cases should be planned shorter or longer than usual due to a range of factors. The factors mentioned most frequently to slow down procedure times are the following: composition of the surgical team (presence of residents, that is, physicians receiving specialized clinical training), lack of experience (low recent work rate for this CPT), gender (female surgeons would be more precise and more careful, and hence slower), age (younger surgeons are less experienced), and time of the day (fatigue in the afternoon). Some of these factors have been analyzed before for hospitals in the US, for instance, in Strum et al. [(3)]. As labor regulations and working habits are quite different in Europe, it is of interest to study the effect of these factors within an European setting.

The main results are the following. For several CPTs, some of the factors contribute significantly (at the 1% significance level) to operation times. This holds true most notably for relatively complex surgical operations, for instance, those involving endoscopic and laparoscopic procedures. Team composition, work rate, and daytime are the most commonly relevant factors. Age matters only for two CPTs, and gender for none of the CPTs (and at the 5% significance level only for a single CPT, cataract in hospital A, where female surgeons work faster than their male colleagues). The practical relevance of these factors is demonstrated by improved out-of-sample prediction of case durations for 2009. As compared to current

OR planning procedures, which are based on the last ten cases of each CPT, the accuracy is improved by 10-15%. Even if the more advanced three-parameter lognormal model for case durations is taken as benchmark, incorporation of significant surgeon factors leads to improvements of the same order of magnitude.

The paper has the following structure. First we present, next the statistical model for case durations is discussed. Finally the results in terms of relevant factors and the gains in predictive accuracy are described .

Data

Surgical procedure times

The data are obtained from surgical databases of two large teaching hospitals in The Netherlands, covering about 100,000 operations in the period from January 2005 till August 2009. The data from 2005 till 2008 are used in estimation, leaving out the data of 2009 for predictive evaluation purposes. The two hospitals, that will be labeled as A and B, differ in several aspects, such as covered specializations, organizational structure, OR protocols, OR logistics, and intensity of teaching. Therefore, the two hospitals will be analyzed separately, but with similar methods. For each operation, the database contains information on the type of operation (the CPT-anesthesia combination), on the procedure and surgical times, and on several surgeon factors (as will be discussed in the next subsection). Case durations depend on the type of anesthetic used [3,5] distinguishing three types of anesthesia, that is, general, local, and regional. The procedure time is defined as the time passing from entry into the operating suite until leaving the OR, This includes the surgical time, that is, the time passing from incision to closure of the wound. The attention will be focused on procedure times, as these are the relevant durations for OR planning. These times will also be denoted as surgical procedure times, indicating that these times include the surgical operation itself as well as the required OR procedures preceding and following the operation.

For the period 2005 till 2008, the database of hospital A contains over 44,000 cases for nearly 1,200 CPT-anesthesia combinations, with total OR time of about 50,000 hours. For various reasons, the actually employed dataset is much smaller and contains 17,516 cases for 29 CPT-anesthesia combinations and a total OR time of about 20,000 hours. The main reason for this data reduction is that CPTs are excluded if they occur relatively infrequently or if they are always performed under similar circumstances. More precisely, in order to be included in the analysis, a CPT-anesthesia combination should exhibit sufficient variation in surgeon factors to allow for an analysis of the effect of these factors. Therefore, for every CPT-anesthesia combination, the imposed minimal requirements are at least 150 cases in total and at least 25 cases for every surgeon involved. Further, about 15% of the cases consist of composite operations involving multiple CPTs. These operations are excluded to avoid possible confounding factors, following Strum et al. (5). Composite operations do not only occur rather infrequently in a fixed composition, but other factors such as the order of the operations may also affect the composite case durations. Minor other reasons for exclusion are operations with incomplete data (less than 1%), and special operations like donor procedures and operations not started or not completed (less than 0.1%).

A similar data selection strategy is followed for hospital B. This database contains about 42,000 cases for about 1,000 CPT-anesthesia combinations, with total OR time of about 45,000 hours. The actually employed dataset, after applying the selection strategy discussed before, contains 12,030 cases for 25 CPT-anesthesia combinations and a total OR time of about 16,000 hours. The total number of included CPT-anesthesia combinations in hospitals A and B is 32, with 22 common ones for hospitals A and B, 7 for hospital A alone, and 3 for hospital B alone. Table 1 shows the included CPTs and contains information on the procedure times. The last four columns show the total number of surgeons and residents involved in each CPT, as well as the number of cases performed in the morning and in the afternoon.

TABLE 1A. DATA OF HOSPITAL A (2005-2008)

	Cases	Mean	Median	SD	Min	Max	Sur-geons	Res.	Cases AM	Cases PM
Ablatio mamma	152	85.0	73.0	18,5	12	198	5	7	79	73
Acetabuloplastic	675	91.0	83.0	13.7	26	166	5	5	286	389
Appendectomy, open	462	73.4	63.0	23.5	32	240	7	15	201	261
Arcomion resection	774	69.0	62.0	12.8	19	199	5	0	388	386
Arthr knee surgery	722	42.6	40.0	15.1	18	163	5	0	293	429
Arthr. nettoyage knee	417	40.0	35.0	10.6	20	87	5	0	183	234
Arthr. tot/part. menisc	1,248	40.9	35.0	12.1	15	147	5	0	608	640
Bi/trimalleolar fracture	189	88.0	91.0	11.2	7	132	6	6	77	112
Cataract	3,219	27.9	35.0	8.2	12	86	3	0	1537	1,682
Diagn. D & C Hyster.	426	44.4	40.0	21.1	3	108	5	0	198	228
End. appendectomy	154	97.8	88.0	20.6	48	172	7	6	59	95
End. tot. prostatec.	294	236.7	189.0	39.0	55	383	3	3	150	144
Femur fracture	342	67.0	64.0	9.2	18	99	7	12	186	156
Genisis total knee	952	72.9	66.0	31.4	11	131	5	0	514	438
Hemicolectomy	152	182.0	188.0	17.3	83	426	5	5	67	85
Hernia inguinalis	764	70.6	62.0	19.8	31	155	7	13	340	424
HNP lumbale	613	73.5	64.0	20.4	40	219	3	0	251	362
Ileus surgery	167	99.0	94.0	14.7	43	177	4	3	109	58
Laminectomy lumbale	340	87.3	76.0	24	40	222	2	0	171	169
Laparoscopic chol.	800	123.2	103.0	34.6	53	340	4	14	443	357

Lap. sterilisation	182	61.0	48.0	15.9	5	94	5	5	71	111
Mammared. both	431	102.3	89.0	34.1	55	227	4	0	198	233
Man. placenta rem.	281	40.0	45.0	22.0	12	236	6	4	108	173
Scopic dec. shoulder	401	45.2	40.0	9.3	11	37	5	0	179	222
Sectio caesarea	961	60.2	54.0	13.7	26	171	6	7	393	568
Total hip arthroplasty	1,221	97.6	84.0	24.6	15	332	5	0	77	644
Trans. Res. prostate	533	69.4	64.0	23.1	5	121	4	0	278	255
Ureterorenoscopy	212	78.2	71.0	35.0	20	221	3	0	89	123
Uterus extirpation	432	98.0	91.0	19.3	12	154	5	2	191	241
Total	17,516						30	19	8,223	9,293

TABLE 1B. DATA OF HOSPITAL B (2005-2008)

CPT	Cases	Mean	Median	SD	Min	Max	Sur-geons	Res.	Cases AM	Cases PM
Ablatio mamma	687	82.0	78.0	21.0	13	201	8	6	358	329
Acetabluloplastic	804	97.0	89.0	16.9	38	169	5	5	194	610
Appendectomy, open	547	91.0	80.0	20.6	4	171	8	13	202	345
Arcomion resection	678	64.0	60.0	15.3	13	187	7	4	498	180
Arthr. knee surgery	200	39.5	35.0	17.9	18	4	5	0	103	97
Arthr. nettoyage knee	214	37.0	35.0	11.7	35	17	5	0	120	94
Arthr. tot/part. menis	300	46.8	43.0	15.1	23	103	3	0	161	139
Bi/trimall. fracture	156	98.0	88.0	13.2	6	210	7	6	90	66
Cataract	1,541	26.1	25.0	10.1	32	70	4	0	639	902
Cholestectomy open	1,110	87.0	81.0	15.3	6	198	7	6	698	412
Colon resection	430	169.0	150.0	14.3	10	201	4	2	199	231
Diagn. D & C Hystero.	688	47.4	44.0	18.4	5	101	4	0	310	378
End. appendectomy	269	88.7	70	17.2	15	163	6	5	127	142
End. total prostatec.	301	243.0	171.0	31.3	9	375	3	0	125	176
Femur fracture	298	108.2	95.0	31.7	8	222	5	0	129	169
Hernia inguinalis	268	75.4	71.0	21.7	4	124	7	16	151	117
Ileus surgery	151	108.0	100.0	17.1	11	191	4	8	67	84
Laminect. lumbale	294	85.5	80.0	18.8	20	125	2	7	139	155
Lap. cholestecomy	305	119.7	104.0	25.2	20	218	4	6	128	177
Mammareduc. Both	564	114.1	100.0	14.4	12	227	4	0	291	273
Man. placenta rem.	405	50.6	45.0	26.4	9	117	6	4	233	172
Scopic dec. shoulder	401	45.2	40.0	9.3	11	137	5	0	179	222
Small bowel resection	684	101.0	89.0	21.1	16	242	5	3	385	299
Trans. Resec. prostate	414	64.1	61.0	24.8	14	162	3	0	221	193
Uterus extirpation	321	101.8	96.0	22.6	19	172	5	0	140	181
Total	12,030						25	12	5,884	6,146

Surgical factors

The literature review of Dexter et al. [5] identifies 48 papers reporting significant factors affecting the perioperative time, that is, the total time required for a patient's surgical procedure, including ward admission, anesthesia, surgery, and recovery. There are multiple reports of the effects on OR times of operative procedures, perioperative team composition including primary surgeon, and type of anesthetic, in that sequence of importance. Strum et al. [3,4] mention the work rate of the surgeon as the most important source of variability in surgical procedure times, with type of anesthesia, age, gender, and American Society of Anesthesiologists risk class as secondary sources of variability. The age of the surgeon is mentioned in Van Houdenhoven [6]. As described in the Introduction, several of these surgeon factors were also brought forward by surgeons, anesthetists, and OR managers in hospitals A and B. In total, the following five factors will be taken into account.

Gender

A popular belief is that female surgeons are more precise and more careful in performing operations, resulting in longer case durations. The gender of the surgeon is indicated by the dummy variable 'Female' (with value 1 for females and 0 for males). For the CPTs of Table 1, the total numbers of female and male surgeons in hospital A are respectively 7 and 23, and in hospital B these numbers are respectively 7 and 18.

Age

In general, older surgeons are more experienced and they may therefore work more efficiently. This effect is mentioned, for instance, in Van Houdenhoven (6). It could also be that surgeons work fastest in the middle period of their career, as older surgeons may become tired more quickly. However, because of the limited number of surgeons, a distinction in two age categories is preferred. The age of surgeons who are active in hospitals A and B ranges between 30 and 60 years. The two age groups are indicated by the dummy variable 'Age', with value 1 if 45 or above and 0 if younger than 45. For the CPTs of Table 1, the total numbers of surgeons above and below 45 years of age are respectively 14 and 16 in hospital A, and in hospital B these numbers are respectively 13 and 12. For a team of surgeons performing an operation, the age is defined as the age of the oldest surgeon in the team.

Workrate

For a given CPT and surgeon, the work rate is related to the number of similar operations that this surgeon has performed in the recent past. A higher work rate means that the surgeon is more experienced in this kind of operation and that case durations may become shorter [5]. Again, because of the limited number of surgeons, a distinction in two classes of work rate is preferred. The work rate is defined to be high if the surgeon performed a similar

CPT at most three weeks ago, and it is defined to be low if this was more than three weeks ago. This rate is indicated by the dummy variable ‘Work rate’, with value 1 for a high rate and 0 for a low rate. For the CPTs of Table 1, the percentage of operations with a high work rate is 81 for hospital A and 84 for hospital B. For a team of surgeons performing an operation, the work rate is defined as work rate of the leading surgeon of the team.

Team

For all procedures of Table 1, the OR surgeon team always consists of a surgeon who is assisted by at least one other surgeon or a resident. Residents are surgeons who receive specialized clinical training in the hospital. It is common belief that the presence of a resident has an increasing effect on case durations, because the resident receives on the job training during the operation. The team composition is indicated by the dummy variable ‘Team’, with value 1 if the team consists of surgeons only and 0 if a resident is part of the team.

Daytime

Some people work better in the morning, others in the afternoon, in the evening, or at night. A recent study [7] shows differences in brain excitability, that is, people who say that they feel best during a certain part of the day tend to have a brain that is most easily excitable during that part of the day. As an operation is a team effort of the involved surgeons and assisting staff, it is not easy to combine the daytime effect for each individual in a joint team effect. Still, it is of interest to know whether the time of the day has an effect on case durations. The time of an operation is indicated by the dummy variable ‘Daytime”, with value 1 for the afternoon (operations starting at 12.00 PM or later) and 0 for the morning (operations starting before 12.00 PM). It might be that case durations are longer in the evening and at night, due to less availability of surgeons and staff. However, such operations are very rare in the two hospitals under consideration, and there is insufficient information to test for separate evening and night effects. Therefore, operations taking place during the evening or at night are excluded due to insufficient data.

Model for surgical procedure durations

Distribution of case durations

The literature on surgical procedure times deals nearly exclusively with the situation in the US. Early results report a lognormal distribution for OR waiting times [8] and a normal [9] or lognormal [10] distribution for OR case durations. Insight in the distribution of case durations has advanced markedly in the past decade [11,12,13,14]. The empirical study of Strum et al. [12] indicates a lognormal distribution of surgical procedure times. Strum et al. [11] consider composite operations consisting of two different surgical procedures and conclude that the lognormal distribution fits such case durations better than the normal distribution. As surgical procedures require a positive start-up time, the shifted lognormal distribution (also called the three-parameter lognormal, written as 3-logN) is explained in Strum et al. [13] and a modified approach is applied in Stepaniak et al. [14]. For the far majority of CPTs, this distribution provides a better fit than the normal and lognormal distributions. Let the procedure time (in minutes) of a given CPT be denoted by T, then the 3-logN distribution for can be written as

$$\log(T-\alpha) = \beta + \varepsilon, \varepsilon \sim N(0, \sigma^2)$$

Here $\alpha > 0$ is the shift parameter, and ε denotes an unobserved random error term causing unpredictable variation. Stated otherwise, after shifting by α, the logarithmic procedure times are normally distributed with mean β and standard deviation σ. The procedure time is always larger than α, and the median is equal to $\alpha + \exp(\beta)$ .The effect of surgeon factors on case durations is modeled by replacing β in the above model by parameters depending on the factors, similar to what is done in analysis of variance (ANOVA) models. If all five factors discussed are included, the model becomes:

$$\log(T-\alpha) = \beta_{PT} + \varepsilon, \varepsilon \sim N(0, \sigma^2)$$

$$\beta_{PT} = \beta_0 + \beta_1 Gender + \beta_2 Age + \beta_3 Workrate + \beta_4 Team + \beta_5 Daytime$$

We call this the ANOVA model. This model is estimated for each CPT and each hospital separately, allowing for different surgeon factor effects according to the hospital and the type of surgical procedure. Although it may be possible to cluster some of the CPTs in Table 1 in groups with identical parameters, this will not be pursued here, because the OR planning system is based on individual CPTs. For a given CPT and hospital, the error terms associated with all corresponding case durations in the database are assumed to be independent and identically distributed. The various hypotheses on surgeon factors discussed previously can be expressed in terms of the following hypotheses on the parameters of the above model:

$$\beta_1 > 0,\ \ \beta_2 < 0,\ \ \beta_3 < 0,\ \beta_4 < 0$$

Further, it is expected that surgeon factors become more important as the complexity of surgical procedures increases. A procedure is complex if it requires highly trained OR staff performing very specific operational procedures and if the risk of perioperative complications is larger than what is usual for routine procedures.

Estimation and prediction

For each CPT of Table 1, the ANOVA model for procedure times is estimated for both hospitals separately, using data from the period 2005-2008. Factors that do not vary are removed from the model. For instance, if all surgeons for a CPT are male, then the effect of gender cannot be estimated for this CPT. To start, all factors that do vary for the CPT are included in the model. Next, backward elimination is used for stepwise removal of insignificant factors. In the end, if all remaining factors are significant, each of the other factors is tested once more for significance when added to the other factors. In addition, the significance of interaction effects between the factors is tested (as none of these interactions is significant, these results will not be reported). All tests employ the same significance level, which is 10%, 5%, or 1%.

To evaluate the practical relevance of the identified significant surgeon factors, the models that are estimated with data for 2005-2008 are used to predict the case durations in the period from January till August 2009. The prediction model is kept fixed, even though the parameters could be re-estimated after each relevant CPT operation in 2009. This choice conforms to practical planning constraints, which demand that models are kept fixed, for instance, for periods of twelve months. The forecast study is restricted to the CPTs for which at least one factor is significant at the 1% significance level.

Three prediction methods are compared. The first is the method that is currently employed in the OR management of both hospitals. The predicted time is simply the average of the ten most recent durations of this CPT. The second method predicts the procedure time to be the median of the 3-logN distribution (without factors), that is α + exp(β). The third method predicts the case duration to be equal to the median of the ANOVA model, that is,

$$\alpha + \exp\,(\beta_0 + \textstyle\sum_j \beta_j F_j$$

including only those factors Fj for which the estimate of βj is significant (at the 1% level). Predicted case durations are compared with the actual procedure times, and the accuracy is evaluated in terms of absolute prediction errors (in minutes). The significance of the difference in mean absolute errors of two methods is tested by the paired t- test.

Results

Surgeon factors

For each hospital and CPT, the significant surgeon factors are obtained by the backward selection strategy described in Section 3.2. The results are summarized in Table 2, which shows how often each factor is found to be significant for significance levels of 10%, 5%, and

1%. For instance, in hospital A, the effect of the factor 'Gender' can be analyzed for 22 CPTs, as for the other 7 CPTs the gender does not vary among the surgeons. The gender effect is significant (and negative) for 3 CPTs at the 10% level (with a median effect of -1.8%), for 1 CPT at the 5% level (with a median effect of -8.2%), and never at the 1% level. In hospital B, gender is never found to be significant, not even at the 10% level. This means that there is no support whatsoever for the commonly expressed opinion that female surgeons would work slower. The gender effect is very weak, and at most it indicates faster work of female surgeons.

Age effects are found to be often significant at the 5% level, mostly with faster work of older surgeons, but the effect is significant at the 1% level only for two CPTs (with a time reduction of about 10% for older surgeons). Work rate effects are significant in several cases, with varying sign at levels of 10% and 5%, but with a consistent time saving effect at the 1% level (of about 5%) for high work rates. The team composition is significant in many cases, and in the far majority of cases the presence of a resident in the team causes longer procedure times (of about 15%, at the 1% level). Daytime effects are significant in many cases, mostly with slower work in the afternoon.

TABLE 2. SURGEON FACTOR EFFECTS (NUMBER OF CPTS WITH POSITIVE AND NEGATIVE EFFECT, AND MEDIAN PERCENTAGE EFFECT ON PROCEDURE TIME)

	Coding (1/0)	CPTs nr		$p < 0.1$			$p < 0.05$			$p < 0.01$	
Hospital A			+	-	Median	+	-	Median	+	-	Median
Gender	1= Female	22	0	3	-1.8	0	1	-8.2	0	0	-
Age	1 = Older	29	3	15	-3.9	2	11	-4.1	0	1	-8.7
Work rate	1 = High	23	5	4	3.5	3	4	-2.9	0	4	-5.3
Team	1=No res.	15	0	11	-10.6	0	9	-13.7	0	7	-15.3
Daytime	1 = PM	29	16	6	3.0	12	2	4.7	1	1	-0.5
Hospital B											
Gender	1= Female	18	0	0	-	0	0	-	0	0	-
Age	1 = Older	25	6	15	-4.3	3	14	-5.7	0	1	-9.9
Work rate	1 = High	17	1	4	-2.1	0	2	-7.3	0	2	-7.3
Team	1=No res.	14	2	6	-7.3	1	5	-13.2	0	4	-14.1
Daytime	1 = PM	25	17	3	3.9	14	2	5.5	3	0	7.5

Table 3 shows the estimated surgeon factor effects for each CPT separately, 29 for hospital A and 25 for hospital B. The effects are shown only if they are significant at the 10% level. The number of significant factors varies among CPTs. For each of the 22 CPTs that are performed

at both hospitals, the sign and size of the effects are often quite the same in both hospitals, even though the effects of some factors cannot be estimated at both hospitals, that is, if the factor does not vary for the CPT under consideration. For instance, for the CPT ablatio mamma, the age affect in hospitals A and B is respectively -1.9% and -3.5%, the team effect is -12.9% and -12.5%, the daytime effect is 8.6% and 7.4%, and the work rate effect is significant only for hospital A (at the 5% level) and not for hospital B (at the 10% level).

Age and Daytime are the factors found most often to be significant. Work rate and team composition are also significant in many cases, and the largest percentage effects are found for these two factors. Gender is nearly never of any importance. The only significant gender effect at the 5% level is for cataract in hospital A, where female surgeons work 8% faster than male surgeons. The CPTs that have at least two significant factors at the 1% level correspond to relatively complicated surgical procedures requiring special skills: ablatio mamma, open appendectomy, endoscopic appendectomy, endoscopic total prostatectomy, laparoscopic cholestectomy, and laparoscopic sterilization. For many of these complicated procedures, the work rate and team composition effects on procedure times are considerable, up to 20%. As compared to less demanding CPTs, complex procedures require more time both for on the job training of residents and for activating specialized skills if the surgeon did not practice these skills within the preceding three weeks.

TABLE 3A. PERCENTAGE EFFECT OF SIGNIFICANT SURGEON FACTORS ON PROCEDURE TIME HOSPITAL A

(shown only if significant at 10%; * and ** denote significance at 5% and at 1%)

CPT	Nr cases	Gender	Age	Work rate	Team	Daytime
		(1= Female)	(1 = Older)	(1 = Higher)	(1 = No res)	(1 = PM)
Ablatio mamma	152	-	-1.9	9.2*	-12.9**	8.6 **
Acetabuloplastic	675	-	-3.2*	-	-	-7.0*
Appendectomy, open	462	-0.7	-8.7 **	-	-10.6**	2.5*
Arcomion resection	774	-	-4.1*	-	-	5.0*
Arthr knee surgery	722	-	-2.9*	-	-	1.0
Arthr. nettoyage knee	417	-	-4.1*	-	-	-3.1
Arthr. tot/part. menisc	1,248	-	-	-	-	6.2*
Bi/trimalleolar fracture	189	-	-	-	-8.6	4.1*
Cataract	3,219	-8.2*	-	-	-	-
Diagn. D & C Hyster.	426	-	-2.5	-	-	-0.4
End. appendectomy	154	-	-3.9*	-7.3**	-13.7**	8.1*
End. tot. prostatec.	294	-	-	-8.9**	-20.3**	8.5*
Femur fracture	342	-	-	6.1*	-4.1	3.4
Genisis total knee	952	-	-	-	-	3.6
Hemicolectomy	152	-	3.1	-	-4.3*	2.5*
Hernia inguinalis	764	-	-6.2*	-	-3.8*	2.4
HNP lumbale	613	-	-2*	-	-	4.4*
Ileus surgery	167	-	-	6.3	-	1.7*
Laminectomy lumbale	340	-	-	3.5	-	6.8*
Laparoscopic chol.	800	-	-7.6*	-3.2**	-19.2**	-3.3
Lap. sterilisation	182	-1.8	-8.5*	-2.9**	-15.3**	-
Mammared. both	431	-	-3.8	-	-16.8**	-5.3
Man. placenta rem.	281	-	-	-	-	5.4*
Scopic dec. shoulder	401	-	5.4*	-	-	-
Sectio caesarea	961	-	4.3*	-	-	-
Total hip arthroplasty	1,221	-	-2.4*	-	-	-
Trans. Res. prostate	533	-	-	4.1*	-	-9.6**
Ureterorenoscopy	212	-	-	-	-	-
Uterus extirpation	432	-	-9.2	-	-	-

TABLE 3B. PERCENTAGE EFFECT OF SIGNIFICANT SURGEON FACTORS ON PROCEDURE TIME
HOSPITAL B

(shown only if significant at 10%; * and ** denote significance at 5% and at 1%)

CPT	Nr Cases	Gender	Age	Work rate	Team	Daytime
		(1= Female)	(1 = Older)	(1 = Higher)	(1 = No res)	(1 = PM)
Ablatio mamma	687	-	-3.5*	-	-12.5**	7.4*
Acetabluloplastic	804	-	-5.4*	-	-	4.0*
Appendectomy, open	547	-	0.6	-	-	2.6**
Arcomion resection	678	-	-8.6*	-	-	6.0*
Arthr. knee surgery	200	-	-6.5*	-	-	-
Arthr. nettoyage knee	214	-	-8.7*	-	-	7.1*
Arthr. tot/part. menis	300	-	-	-	-	1.3
Bi/trimall. fracture	156	-	-5.8*	-	1.2*	-
Cataract	1,541	-	6.5*	-	-	5.5*
Cholestectomy open	1,110	-	7.6*	-	-14.3**	-6.4*
Colon resection	430	-	4.8	-	-15.3**	3.8
Diagn. D & C Hystero.	688	-	-8.5*	-	-	-4.6
End. appendectomy	269	-	5.6*	-5.8**	2.2	8.2**
End. total prostatec.	301	-	-9.0*	-8.7**	-	8.3*
Femur fracture	298	-	-4.1*	-	-	-
Hernia inguinalis	268	-	-	-	-	2.2
Ileus surgery	151	-	2.1	-2.1	-	2.7
Laminect. lumbale	294	-	-4.3*	-1.9	-	-
Lap. cholestecomy	305	-	-9.9**	-	-13.8**	7.5**
Mammareduc. Both	564	-	-5.7*	-	-	2.7*
Man. placenta rem.	405	-	-	-	-2.1	3.5*
Scopic dec. shoulder	401	-	-	-	-	-4.2*
Small bowel resection	684	-	-7.8*	-	-1.9*	6.7*
Trans. Resec. prostate	414	-	-4.0*	2.7	-	4.8*
Uterus extirpation	321	-	-1.1	-	-	-

Summarizing, the largest effects are obtained for work rate and team composition for complicated CPTs. In most cases (and at the 1% level always), procedure times are relatively shorter for older surgeons, for a high work rate, and for teams without resident. Gender has hardly any effect. In most cases, procedure times are shorter in the morning than in the afternoon, but for some CPTs this effect is reversed.

The mixed daytime effect can be due to the fact that this effect is measured jointly for the full OR team involved in the operation and without information on the time preference of the members of the team. A small-scale study was performed to investigate this further. Ten surgeons of hospital A and also ten surgeons of hospital B were asked whether they have any preference for performing operations in the morning or in the afternoon. Of these 20 surgeons, 9 prefer the morning, 10 the afternoon, and one surgeon has no preference. In total, the 19 surgeons with a preference are active in 64 CPTs. For each surgeon and CPT, the average case duration in the morning is compared with that in the afternoon. Of the 64 surgeon-CPT combinations, the fastest work was delivered in 48 cases in the preferred daytime and in 16 cases in the non-preferred daytime. This effect of preferred daytime on case durations is significant (the p-value according to the binomial distribution with a success probability of 50% is smaller than 0.01%). For hospital A (B), the fastest work was delivered in 23 (25) cases in the preferred daytime and in 7 (9) cases in the non-preferred daytime, corresponding to a p-value for the absence of daytime effects of less than 1% in both cases. As daytime preferences are not known for many of the surgeons involved in the CPTs of Table 1, this factor could not be incorporated in the analysis of surgeon factor effects in Tables 2 and 3. However, the small-scale study indicates that it may help to incorporate surgeon preferences in OR planning.

Prediction

In order to evaluate the practical usefulness of surgeon factors in predicting case durations, the attention is restricted to CPTs for which at least one surgeon factor is significant at the 1% level. This holds true for eight CPTs in hospital A and seven CPTs in hospital B, five of which occur at both hospitals. The ANOVA models, estimated with the data of 2005-2008 and with the estimated factor effects of Table 3 that are significant at the 1% level, are used to predict the procedure times for the period from January till August 2009. The total number of predicted case durations is 683 for hospital A and 575 for hospital B.

Table 4 summarizes the results of three prediction methods, that is, the current method (average of last ten cases), the three-parameter lognormal model without factors (3-logN), and the ANOVA model. The table shows the mean and standard deviation of the absolute prediction errors, that is, the differences between the predicted time and the actual case duration. The differences in mean absolute prediction errors of the three methods are evaluated both in absolute terms (in minutes) and in relative terms (as percentage of the median procedure time for each CPT over the prediction period).

As an illustration, Figure 1 shows the absolute prediction errors and the differences of these errors of the three prediction methods for the 71 endoscopic appendectomy operations that took place in hospital A between January and August 2009. The current method predicts the procedure time as the average of the last ten case durations of this CPT, and this estimate is updated after each operation in 2009. The 3-logN predictions are obtained from the *ANOVA* model without factors, estimated with data from 2005 till 2008 and with fixed parameters for 2009. Finally, the *ANOVA* predictions are also obtained from a model estimated with data from 2005 till 2008 and with fixed parameters for 2009. This model includes factors only if they are significant at the 1% level. Table 3 shows that the included factors are work rate (with coefficient -0.073) and team composition (with coefficient -0.137). Figure 1 shows that the smallest prediction errors are obtained for *ANOVA*, and that 3-logN is second-best. The predictions of *ANOVA* are better than the current method in 67 out of 71 cases, and they are better than 3-logN in 53 out of 71 cases. The differences in absolute forecast errors of the three methods are all significant (at the 5% level) when tested by the paired *T*-test.

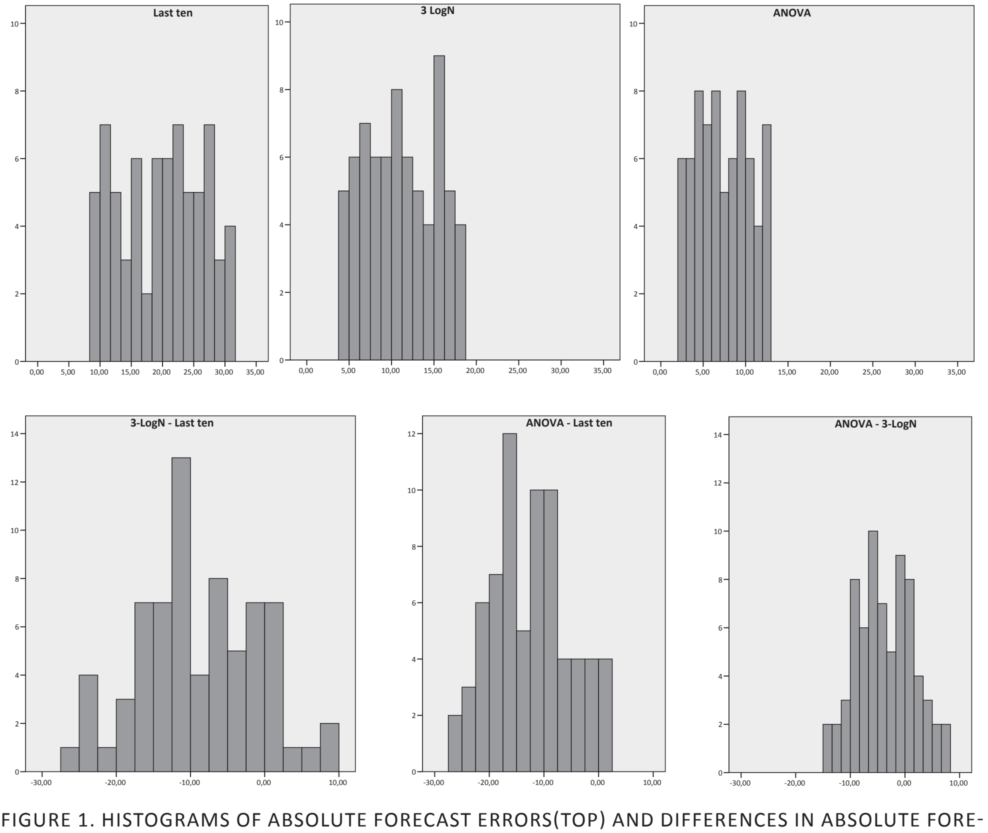


FIGURE 1. HISTOGRAMS OF ABSOLUTE FORECAST ERRORS(TOP) AND DIFFERENCES IN ABSOLUTE FORECAST ERRORS (BOTTOM) FOR 71 PREOCEDURE TIMES OF ENDOSCOPIC APPENDECTOMY

Table 4 shows that, in all of the considered 15 CPTs in hospitals A and B, the 3-logN predictions are more accurate than the currently employed method. The same holds true for the *ANOVA* predictions, except for transurethral resection of the prostate in hospital A. As compared to the current method, the forecast improvements of 3-logN are up to 10%, and those of *ANOVA* are up to 18%. The *ANOVA* predictions are better than the 3-logN predictions in the far majority of cases (11 out of 15), with gains of up to 15%. For three CPTs in hospital B, 3-logN is slightly better than *ANOVA* (up to 2%), and for one CPT in hospital A, 3-logN is 6% better than *ANOVA*. The paired t-test finds that, for hospital A, *ANOVA* improves significantly on 3-logN (at the 5% level) for 7 out of 8 CPTs, and the reverse holds true for the remaining CPT. For hospital B, *ANOVA* is significantly better than 3-logN for 4 out of 7 CPTs, and the difference is not significant for the other 3 CPTs.

When averaged over the eight considered CPTs in hospital A, the gain in prediction accuracy is 5 minutes (5%) for 3-logN as compared to the current method, 10 minutes (11%) for *ANOVA* as compared to the current method, and 5 minutes (7%) for *ANOVA* as compared to 3-logN. For hospital B, the prediction gains are 4 minutes (4%) for 3-logN as compared to the current method, 8 minutes (8%) for *ANOVA* as compared to the current method, and 4 minutes (4%) for *ANOVA* as compared to 3-logN. On average, the standard deviation of the prediction errors is smallest for *ANOVA* (3.7 minutes in hospital A and 4.1 minutes in hospital B), as compared to 3-logN (4.7 in A and 4.5 in B) and the current method (5.9 in A and 5.2 in B). Although these differences are not large, reduction of uncertainty is important in OR planning. It is a nice finding that the improved prediction accuracy of *ANOVA*, which is based on more elaborate models involving surgeon factors, is combined with reduced forecast uncertainty. Stated otherwise, the smaller prediction bias of *ANOVA* comes without any cost of increased variance.

TABLE 4. PREDICTION ERRORS OF PROCEDURE TIMES (ABSOLUTE ERRORS AND DIFFERENCES BETWEEN METHODS), JANUARY - AUGUST 2009

CPT		Absolute errors (minutes)						Diff (minutes)			Diff (% of median procedure time)		
Hospital A	Cases	Last 10		3-LogN		ANOVA		3LogN	ANOVA	ANOVA	3LogN	ANOVA	ANOVA
		mean	SD	mean	SD	Mean	SD	Last10	Last10	3LogN	Last10	Last10	3LogN
Ablatio mamma	51	19.3	6.3	15.7	6.4	6.5	3.3	-3.6	-12.8	-9.2	-4.6	-17.5	-12.6
Appendectomy, open	67	16.2	5.2	12.3	4.4	7.7	2.9	-3.9	-8.5	-4.6	-6.2	-13.5	-7.3
End. appendectomy	71	19.9	6.6	11.0	4.2	7.4	3.1	-8.9	-12.5	-3.6	-10.1	-14.2	-4.1
End.tot.prostatec.	71	24.6	8.2	14.2	6.4	6.6	4.5	-10.4	-18.0	-7.6	-5.5	-9.5	-4.0
Lap. Cholestect.	142	18.2	4.2	15.9	3.2	9.9	6.5	-2.3	-8.3	-6.0	-2.2	-8.1	-5.8
Lap. Sterilization	68	12.3	4.3	11.2	4.2	4.9	2.6	-1.1	-7.4	-6.3	-2.3	-15.4	-13.1
Mammared. both	132	21.8	9.5	17.3	5.8	9.8	2.9	-4.5	-12.0	-7.5	-5.1	-13.5	-15.6
Trans. res. Prostatec.	81	9.4	3.2	6.3	2.8	10.2	3.9	-3.1	0.8	3.9	-4.8	1.3	6.1
Total / Average	683	17.7	5.9	13.0	4.7	7.9	3.7	-4.7	-9.8	-5.1	-5.1	-11.3	-7.1
Hospital B													
Ablatio mamma	94	18.2	7.2	16.2	5.2	6.0	3.2	-2.0	-12.2	-10.2	-2.6	-15.6	-13.1
Appendectomy, open	73	14.2	3.8	10.5	3.2	11.2	6.1	-3.7	-3.0	0.7	-4.6	-3.8	0.9
Cholestecomy open	203	13.2	3.9	8.2	2.9	10.1	4.1	-5.0	-3.1	1.9	-6.2	-3.8	2.4
Colon resection	63	19.2	4.9	12.9	4.3	14.2	5.2	-6.3	-5.0	1.3	-4.2	-3.3	1.7
Endoscopic appendectomy	62	16.5	5.3	14.7	4.3	8.2	3.2	-1.8	-8.3	-6.5	-2.3	-10.6	-8.3
Endoscopic total prostatectomy	41	24.4	6.8	18.2	6.7	8.3	3.1	-6.2	-16.1	-9.9	-3.5	-9.2	-5.8
Laparoscopic cholestectomy	39	17.3	4.2	14.2	5.2	7.1	3.5	-3.1	-10.2	-7.1	-3.0	-9.8	-6.8
Total / Average	575	17.6	5.2	13.6	4.5	9.3	4.1	-4.0	-8.3	-4.3	-3.8	-8.0	-4.1

Conclusion

Depending on the type of operation (CPT) and on the hospital, procedure times may depend on several surgeon factors. In particular, for complex operations, factors like relevant work rate experience of the surgeon and composition of the surgical team may have large effects. The effect of team composition goes up to 20%, and when combined with work rate, the total effect goes up to 30%. Other relevant factors are age of the surgeon and time of the day. Gender has nearly never any effect, and the only effect that is significant (at the 5% level) is found for cataract, where female surgeons work 8% faster than male surgeons. A predictive out-of-sample analysis for case durations in 2009 shows that surgeon factors help in predicting case durations.

As compared to the methodology currently employed in both hospitals, mean absolute prediction errors are reduced by up to 18 minutes and up to 18% of the median procedure time. The most significant gains are obtained for relatively complex CPTs, especially those involving endoscopic and laparoscopic procedures. As the complexity of surgical procedures shows an ever increasing trend, surgeon factors may become even more important in the future.

The practical implementation of (*ANOVA* or other) prediction models is done best after consultation of surgeons, OR management, and other staff involved in the operation room activities. As hospitals differ widely in aspects like surgical experience with different specializations, organizational structure, OR protocols and OR logistics, the effect of surgeon factors will differ among hospitals. Therefore, it may be best to estimate separate models for each hospital. The results of this paper show several differences between the two considered hospitals, although the type of effect is quite the same in many cases, especially for complex procedures.

The achieved improved forecast accuracy can be of great help for operation room planning. Reduction of case duration uncertainty will have positive benefits in terms of patient health care and human resource planning in hospitals.

References

1. Vissers J, Beech R. *Health operations management: Patient flow logistics in health care*, Routledge Abingdon, Oxon 2005.
2. Strum DP, Sampson AR, May JH. Vargas LG. Surgeon and type of anesthesia predict variability in surgical procedure times. *Anesthesiology* 2000;92:1454-66.
3. Stepaniak PS. Mannaerts GHH, Quelerij M de, Vries G de. The effect of the operating room coordinators risk appreciation on operating room efficiency. Anesth Analg 2009;108:1249-56.
4. Dexter F, Dexter EU, Masursky D, Nussmeier NA. Systematic review of general thoracic surgery articles to identify predictors of operating room case durations. *Anesth Analg* 2008;106:1232-41.
5. Strum DP, May JH, Vargas LG. Modeling the uncertainty of surgical procedure times: Comparison of log-normal and normal models. *Anesthesiology* 2000;92,1160-67.
6. Van Houdenhoven M. *Healthcare logistics: The art of balance*. Ph.D. Thesis, Erasmus University Rotterdam, Scriptum Publishers. 2007, Schiedam.
7. Tamm AS, Lagerquist O, Ley AL, Collins DF. Chronotype influences diurnal variations in the excitability of the human motor cortex and the ability to generate torque during a maximum voluntary contraction. *J Biol Rhythm* 2009;24:211-24.
8. Rossiter CE, Reynolds JA. Automatic monitoring of the time waited in out-patient departments. *Med Care* 1963;1:218-25.
9. Barnoon, S, Wolfe H. Scheduling a multiple operating room system: A simulation approach. *Health Serv Res* 1968;3:272-85.
10. Hancock WM, Walter PF, More RA, Glick ND. Operating room scheduling data base analysis for scheduling. *J Med Syst* 1988;12:397-409.
11. Strum DP, May JH, Sampson AR, Vargas LG, Spangler WE. Estimating times of surgeries with two component procedures: Comparison of the lognormal and normal models. *Anesthesiology* 2003;98:232-40.
12. May JH, Strum DP, Vargas LG. Fitting the lognormal distribution to surgical procedure times. *Decision Sci* 2000;31:129-48.
13. Spangler WE, Strum DP, Vargas LG, May JH. Estimating procedure times for surgeries by determining location parameters for the lognormal model. *Health Care Manage Sci* 2004;7:97-104.
14. Stepaniak PS, Heij C, Mannaerts GHH, Quelerij M de, Vries G de. Modeling procedure and surgical times for CPT-anesthesia-surgeon combinations and evaluation in terms of case-duration prediction and operating room efficiency. *Anesth Analg* 2009;109:1232-45.

4

Working with a fixed OR team on consecutive similar cases and the effect on case duration and turnover time

Pieter S. Stepaniak[ł], Wietske W. Vrijland[±], M.D., Ph.D., Marcel de Quelerij[¥], M.D., Guus de Vries[ł], Ph.D., Christiaan Heij[¤], Ph.D.

Accepted for publication Arch Surg (September 24th 2009)
Scheduled for publication December 2010

[ł] Institute of Health Policy and Management, Erasmus University Rotterdam
[±] Department of General Surgery, St. Franciscus Hospital Rotterdam
[¥] Department of anesthesiology, St. Franciscus Hospital Rotterdam
[¤] Econometric Institute, Erasmus School of Economics, Erasmus University Rotterdam

Abstract
If variation in procedure times could be controlled or better predicted, the cost of surgeries could be reduced through improved scheduling of surgical resources. This study on the impact of similar consecutive cases on the turnover-, surgical-, and procedure time tests the perception that repeating the same manual tasks reduces the duration of these tasks. We hypothesize that when a fixed team works on similar consecutive cases the result will be shorter turnover and procedure duration as well as less variation as compared to the situation without a fixed team. To test our hypothesis, two procedures were selected and divided across a control group and a study group.

Patients were assigned at randomly to the study or control group. For the inguinal hernia repair we find a significant lower preparation time and 10 minutes less procedure time in the study group, as compared to the control group. Variation in the study group is lower as compared to the control group. For the laparoscopic cholecystectomy only preparation time is significantly lower in the study group as compared to the control group.

For both procedures there is a significant decrease in turnover time. Scheduling similar consecutive cases and performing with a fixed team results in lower turnover times and preparation times for the studied cases. The procedure time of the inguinal hernia repair decreases significantly and has practical scheduling implications. For more complex surgery like the laparoscopic cholecystectomy there is no effect on procedure time.

Introduction

When looking at an OR in an era of cost-constrained health care, it is of economic importance for medical institutions to effectively schedule expensive surgical resources and to use the efficiently. Variation and thus uncertainty in procedure times complicates surgical scheduling and reduces operational efficiency. If variation in procedure times could be controlled or better predicted, the cost of surgeries could be reduced through improved scheduling of surgical resources.

OR schedules depend crucially on estimated case durations, and statistical models may help to improve these estimates to support management in the cost-efficient use of expensive surgical resources [1,2]. Therefore we need to model and thus predict surgical procedure times more accurately. More accurate predictions of case durations ultimately helps to achieve meaningful, sustainable service quality improvement in the OR and, as a consequence in the delivery system through:

- decrease of unused costly OR capacity/equipment which can be used for new patients
- decrease of the number of cancellations of scheduled surgeries

The standard personnel for an OR includes a surgeon (with or without an assisting surgical resident) and a surgical nurse, an anesthesiologist, a nurse circulator, and a nurse providing anesthesia. In general teaching hospitals it is common that during the course of a day personnel is switched across various ORs. Because the personnel switches across ORs the switch may also be across type of procedure and OR team.

Although every staff member prepares daily for the procedures they are assigned for, some staff members noticed that they needed adaptation/ familiarization time when starting as a team at the beginning of the surgical procedure or during it. This formed an obstacle team members to attaining a smooth work-flow. Some studies [3,4,5,6,7,8,9,10,11] focus on working with teams or redesigning processes in the OR. For instance: Operating room turnover time and daily caseload can be improved by analyzing the routine tasks of the operating team and minimizing inefficiencies.

A coordinated multidisciplinary process redesign can significantly reduce operating room turnover time [3]. Results demonstrate that a coordinated multidisciplinary process redesign can significantly reduces room turnover time as well as anesthesia induction and emergence time [4]. Recent publications have focused on increased operating room throughput without increasing total OR time. Reorganizing the perioperative work process for total joint replacements sustainably increased OR throughput [6]. Studies in both operations management and health care have found that performance on a procedure improves with increased experience [12].

The implication of the so-called "learning curve" or experience curve is that "practice makes perfect" and organizations "learn by doing" [13]. Yet, some studies show homogeneous learning curves across sites [14] and others show heterogeneity across sites [15,16,17,18]. These varying results may be due to differences in the extent of social and organizational changes provoked by a new technology or practice, which give rise to differences in user acceptance and behavior [19,20]

Much of the world's OR capacity uses consistent teams throughout the day. We took advantage of a unique feature of our OR organization to estimate how much time could be saved in a simple and a complex procedure by establishing consistent teams where previously none had existed. We constructed a study of batch processing of similar procedures in the OR using a fixed OR team. A batch of procedures consist of the same procedures which are performed during the day by the same fixed OR team in the same operating room. We hypothesize that this concept reduces the adaptation / familiarization time for a specific procedure and hence the OR time.

Material and methods

Selection procedures

We defined the following inclusion criteria for the selected procedures: (1) the procedure is not yet part of an OR program with consecutive similar operating procedures, (2) the procedure is done by a surgical resident and an experienced surgeon, (3) one procedure must be relatively low in complexity of performance, the other relatively high in complexity (as defined by the surgeons/anesthesiologists)

Based on the inclusion criteria we limited the study to two procedures: (1) Inguinal hernia repair (according to the Lichtenstein technique), under spinal anesthesia, (2) laparoscopic cholecystectomy, under general anesthesia

Study design on study days

This design for the study group is based upon three central factors: (1) a similar type of procedure for all involved patients groups, (2) a fixed OR team on the day of surgery and (3) a well-defined routine protocol for all members of the operative team. This protocol is explicitly discussed within the OR team at the beginning of the day

To test our hypothesis we make a distinction between control days and study days. The control days are the data flow of patients during the study period without applying the consecutive concept. The study days are on Fridays in the odd weeks. Then two laparoscopic cholecystectomies are performed and four inguinal hernia repairs. The spinal anesthesia for the inguinal hernia repair is given in the holding. In our hospital, a standard type of mesh is used in inguinal hernia repair. This mesh is fetched by the operative nurse prior to the first operation that takes place in that particular OR

Before staring the study, within the team we considered to performing at least three laparoscopic cholecystectomies on every study day by the same fixed team. Based on the available data we knew that every laparoscopic cholecystectomy takes about two hours meaning that the same OR team has to be in the OR for at least 6 hours (with turnover time approximately 7 hours). Consequently team members noted that there were possible extra risks involved when performing this program of three surgeries for instance fatigue in team members. Because we want to perform a study with no compromise at all on the quality of care for our patients we decided to perform two laparoscopic cholecystectomies rather than three or more.

On study days the OR team members may not switch between ORs but must work in their assigned OR. For instance prior to the study and on control days, nurses/doctors could be substituted during the course of the operative day in between the various ORs. Essentially, there was not strict team cohesiveness or team order. At the beginning of every study day, the team is formed and stays together for the scheduled procedures. On every study day for every selected procedure, the team composition is different. Ten minutes before the first case of the day starts, the OR team comes together and reviews the coming day and defines explicitly the roles of each individual team member. When the operating room is ready, the surgeon and nurse transport the patient into the operating room. The anesthesiologist provides the patient anesthesia. In the next step, the actual operation, all members of the team are present.

At the end of the operation, a nurse and the anesthetist transport the patient to the recovery room. The surgeon fills in a form containing qualitative questions about the procedure. This form gives information about whether the procedure has been performed uneventfully (Figure 1).

Date of surgery	
Patient number	
Has the medical procedure been performed as was expected prior to the procedure?	YES / NO If NO, specify
Where there other problems than medical which may have complicated the execution of the medical procedure	YES / NO If YES, specify

FIGURE 1. QUESTIONAIRE STUDY

After filling in the form, the surgeon goes to the preoperative area to welcome the next patient. At the end of the day, the forms are collected by an OR nurse and handed over to the researchers. At the end of the study days, we asked surgeons and nurses how they experienced working within a fixed team on similar cases during the day.

Study design on control days

On control days (every Friday) in the even weeks, the normal historical flow of patients occurred. Here, different OR teams performed two consecutive scheduled laparoscopic cholecystectomies under general anesthesia and two inguinal hernia repairs. The above mentioned three factors were not explicitly followed. On non-study days the members of the team perform common procedures for general surgery for example, Femur fracture, Patella fracture, Appendectomy (open and endoscopic), Circumcision, Acetabulum, Lumpectomy Mamma, Mastectomy, Recto-sigmoidresection (Hartmann Procedure, open), Sigmoid resection.

Scheduling patients

When patients have to undergo a surgical procedure like the inguinal hernia repair or a cholecystectomy, they are scheduled on regular OR days that suit the patient on a first come first served basis. This might be a Friday. If it is a Friday in the odd weeks than the patient is assigned to the study group. If it is a Friday in the even weeks than the patient is assigned to the control group. The resulting OR schedule is made by a staff member responsible for that job and who has no part or interest in the study. A day before surgery members of the OR team are selected randomly and assigned to a control group or study group.

Start –end study period

The study started June 6th 2008 and ended April 22th 2009

Statistical Considerations

Because of indications of log normality of surgical and procedure times [2,21,22,23] the recorded times are transformed to their natural logarithm. Consequently any calculations with case durations should be done with the log transformed OR time. For every group for every included procedure we calculate the mean times, standard deviation, 95% confidence intervals for the mean of the lognormal times, 5% lower and 95% upper Bayesian prediction bounds.

Confidence intervals were calculated to test if they are sufficiently narrow as to be managerially meaningful. We use a modified version of the Cox method [24,25] to calculate the 95%-confidence interval for the mean: For sample data with mean $\bar{Y}$ and variance S^2 the confidence level is calculated:

$$\exp\{\ \bar{Y} + \frac{S^2}{2} \pm\ t_{0.95} \sqrt{\frac{S^2}{n} + \frac{S^4}{2(n-1)}}\}$$

The 5% lower and 95% upper prediction bounds are

$$\exp\{\ \bar{Y} +\ t_{0.05,\ \ 0.95\ (n-1)} \sqrt{\frac{S^2}{n}(1+n)}\}$$

In our study we focus on four time intervals:

- preparation time: time between anesthesia-ready time and procedure start time
- surgical time: time between start incision and closing wound
- procedure time: time between patient entering and leaving the OR
- turnover time: time interval between previous patient leaving the OR and the following patient entering it.

Differences in mean OR time, surgical time and turnover time between the control group and study group are investigated with an independent samples *T*-test. We use Levene's test to test the null hypothesis that the population variance of the study group is equal to the control group. For Levene's test at $p < 0.05$, we accept the hypothesis that the two populations have unequal variances. For the *T*-test, at a *p*-value < 0.05, we conclude that there is a statistically significant difference in mean times between control group and the study group. The clinicians (W. Vrijland M.D., M de Quelerij M.D.) were explicitly excluded from participating in the practical part of the study. To prevent any possibility of conflict of interest they were not part of any OR team. The clinicians declared before the study that their one and only interest in the study was to contribute to science, whatever the outcome of the study. Statistical work was performed by Stepaniak and Heij. The results of this statistical work were presented to the clinicians.

Results

The first step in evaluating the results is to compare the three measured time intervals between the historical data (2005-May 2008) , the study group and the control group. There were no statistically significant differences between the time intervals measured for the historical data and control group for both procedures. In Table 1 we present the results of the questionnaire.

Inguinal hernia repair (according to the Lichtenstein technique), under spinal anesthesia

The results are presented in Table 2. For the hernia repair we have 68 patients on 17 study days (4 hernias per day on 1 OR, consecutively performed by the same team) and 68 patients on 17 control OR days (4 hernias per control day, in 2 ORs, per OR 2 consecutively performed NOT by the same team and 4 different teams). The number of different teams on control days are 68. Equal variances are not assumed for preparation time, surgery time and for procedure time ($p < 0.001$). The *T*-test results are on equal means of logtimes for preparation time (15.2 min- 6.8 min, $p < 0.001$) surgery time (51.2 min - 48.0 min, $p = 0.051$), procedure time (71.2 min -59.8 min, $p < 0.001$).

TABLE 1. RESULTS QUESTIONNAIRE

	Hernia Inguinalis	Lap. Cholestectomy
Performed as was expected prior	No = 0, Yes = 68	No = 2, Yes =24 Extra surgery needed
Any problems (other than medical	No = 67, Yes =1 Software data management system delayed start up (5 min)	No = 26, Yes =0
	Total 68/68	Total 26/26

Figure 2 shows a graph representing the preparation time of the 200 hernias before starting the study and the 68 during the study. The 95% confidence interval for the mean of procedure time in the study group is [56-64 minutes] versus [66-77 minutes] in the control group. For preparation time these intervals are respectively [6-7 minutes] in the study group and [14-17 minutes] in the control group. The mean turnover time for the study group is 7.6 (7.4 -7.8) minutes and for the control group 9.3 (9.1-9.5). Equal variances are not assumed ($p = 0.007$), *T*-test for difference in mean turn over time results in a *p* value of 0.001 .

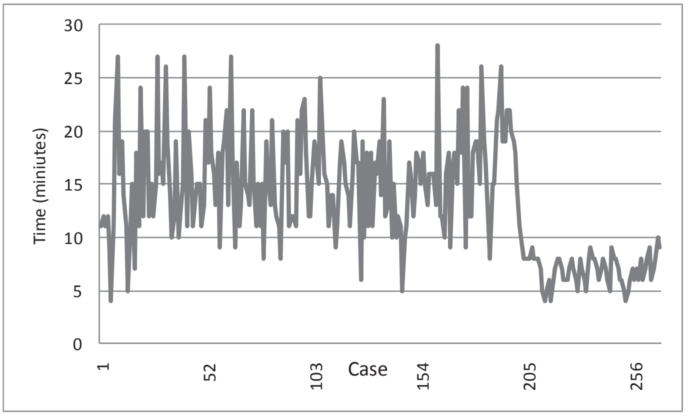


FIGURE 2. PREPARATION TIME OF 200 HERNIA INGUINIALIS PRIOR TO STARTING THE STUDY AND 68 DURING THE STUDY

Laparoscopic cholecystectomy, under general anesthesia

For the cholecystectomy we had 26 patients on 13 study days (2 consecutively performed by the same team) and 26 patients on 13 study days (in 1 OR , two consecutively laparoscopic cholecystectomy NOT performed by the same team). The number of different teams on control days are 26. In two cases on study days extra surgery had to be done: a Hernia Umbilicalis and a Naevus stomach (Table 1). These cases were included in the analyses since doing something unexpected differently in the OR is a reflection of reality. Comparing the study group and the control group equal variances may not be assumed for surgery time (p = 0.012) and for procedure time (p = 0.022).

The *T*-test results are on equal means of the log times for surgery time (82.8 min-88.1 min, p = 0.741) and procedure time (123.0 min - 117.7 min, p = 0.645). The assumption of equal variances is not rejected for preparation time (p = 0.925). The *T*-test shows a significant difference in preparation time (25.5 min - 22.2 min, p = 0.003). The mean turnover time for the study group is 11.3 (11.0 min - 11.5 min) minutes, for the control group the mean turnover time is 13.6 (13.3 min - 14.0 min) minutes. The difference in mean turn over time is significant. The assumption of equal variances is rejected (p = 0.016).

TABLE 2. RESULTS OF THE STUDY AND CONTROL GROUPS

Lap. cholestectomy	Mean time (ln)	SD (ln)	Mean	SD	Lower 5% CI	Upper 95% CI	5% Lower Prediction bound	95% Upper Prediction Bound
Procedure time								
control (n=26)	4.79	0.21	123.0	26.1	113	134	79	183
study (n=26)	4.76	0.13	117.7	15.4	111	125	88	154
historical data (736)	4.77	0.24	121.4	29.6	119	124	74	189
Surgical time								
control (n=26)	4.43	0.31	88.1	28.0	77	100	45	156
study (n=26)	4.40	0.18	82.8	15.0	77	90	57	117
historical data (736)	4.43	0.37	89.9	34.4	87	93	41	173
Preparation time								
control (n=26)	3.23	0.13	25.5	3.3	24	27	19	33
study (n=26)	3.09	0.14	22.2	3.1	21	24	17	29
historical data (736)	3.12	0.13	22.8	3.0	22	23	18	29
Hernia Inguinalis								
Procedure time								
control (n=68)	4.22	0.30	71.2	21.8	66	77	38	123
study (n=68)	4.06	0.25	59.8	15.2	56	64	35	95
historical data (n=704)	4.22	0.28	70.8	20.2	69	72	39	118
Surgical time								
control (n=68)	3.87	0.36	51.2	19.0	47	57	24	98
study (n=68)	3.81	0.35	48.0	17.3	44	53	23	90
historical data (n=704)	3.89	0.36	52.2	19.4	51	54	24	99
Preparation time								
control (n=68)	2.66	0.35	15.2	5.5	14	17	7	29
study (n=68)	1.90	0.21	6.8	1.5	6	7	4	10
historical data (n=704)	2.62	0.36	14.7	5.5	14	15	7	28
Historical data: 2005-May 2008								

Discussion

This study on the impact of similar consecutive cases on the turnover-, surgical-, and procedure time confirms the perception that repeating the same manual tasks may reduce the duration of these tasks. These results are well known for manufacturing processes and they form the basement of the lean manufacturing system. Based on our findings we affirm that organizations “learn by doing” [13,14]. By maintaining a fixed team for similar consecutive cases throughout the entire day we find a significant reduction in preparation time and turnover time for both studied procedures. Teams prepared the procedures in a more structured fashion in the study group. This explains the shorter preparation time in the study group as compared to the control group. Surgery time was not significantly different in the study group as compared to the control group. Surgeons do not work “faster or slower” when working on consecutive similar cases and surgeons do not compromise on quality of care to increase speed.

For the inguinal hernia repair we see a significantly shorter preparation,- and procedure times in the study group as compared to the control group. Also, the variation in the study group of the three time intervals is significant lower as compared to the control group. The average decrease of the procedure time in the study group (10 minutes per procedure) as compared to the control group has practical implications for planning purposes. A reason for the decreased operative time (because of the decrease in preparation time) may be the effect of the roles of each individual team member being explicitly defined before the start of the day. In the study group a significantly lower mean preparation time is found for the laparoscopic cholecystectomy. The mean procedure time for the laparoscopic cholecystectomy is not significantly lower in the study group. A possible technical explanation for the fact that no difference is found in the mean procedure time between the control group and study group is that in both study group and control group patients were included who experienced a cholecystis or an obstruction necessitating an endoscopic retrograde cholangiopancreatography. Both problems may involve a technically demanding operation that may require more dissection time.

Based on the results we may conclude that the consecutive concept helps to decrease the preparation time. Further, we assume that using consecutive planning with a fixed team has more effect on case durations of relatively less medically complex procedures than on more complex ones. The latter hypothesis should be further investigated in future studies. We asked surgeons and nurses on study days to describe their experiences during the study. In general they all experienced team spirit among the team and a smooth work-flow on study days when performing the inguinalis hernia repair procedure. The opposite is true for cholecystectomy. One explanation for this difference in experience is the duration and complexity of the cholecystectomy as compared to the hernia inguinalis. Team members identified some benefits of working a large part of the day together: because there is an extended briefing at the beginning of the day everybody in the team is made explicitly aware of their role. This effect of consecutive similar case planning on team spirit may be one of the factors causing reduced overall handling times in the OR.

Conclusion

Scheduling similar consecutive cases and performing with a fixed team results in lower turnover times and preparation times for the studied cases. The surgical time of the inguinal hernia repair decreases significantly and has practical scheduling implications. For more complex surgery like the laparoscopic cholecystectomy, there is no effect on surgery time.

Acknowledgement

We wish to thank the OR staff, surgeons and anaesthesiologists of the St Franciscus Hospital for their cooperation in this study.

References

1. Stepaniak PS, Mannaerts GHH, Quelerij M de, Vries G de. The effect of the operating room coordinatORs risk appreciation on operating room efficiency. *Anesth Analg* 2009;108:1249-56.
2. Stepaniak PS, Heij C, Mannaerts GHH, Quelerij M de, Vries G de. Modeling procedure and surgical times for CPT-anesthesia-surgeon combinations and evaluation in terms of case duration prediction and OR efficiency. *Anesth Analg* 2009;109:1232-45.
3. Cendan JC, Good M: Interdisciplinary work flow assessment and redesign decreases operating room turnover time and allows for additional caseload. *Arch Surg* 2006;141:65-9.
4. Harders M, Malangoni MA, Weight S, Sidhu T. Improving operating room efficiency through process redesign. *Surgery* 2006;140:509-16.
5. Seim AR, Andersen B, Sandberg WS. Statistical process control as a tool for monitoring non-operative time. *Anesthesiology* 2006;105:370-80.
6. Smith MP, Sandberg WS, Foss J, Massoli K, Kanda M, Barsoum W, Schubert A. High-throughput operating room system for joint arthroplasties durably outperforms routine processes. *Anesthesiology* 2008;109:25-35.
7. Sandberg WS, Daily B, Egan M, Stahl JE, Goldman JM, Wiklund RA, Rattner D. Deliberate perioperative systems design improves operating room throughput. *Anesthesiology* 2005;103: 406-18.
8. Hanss R, Buttgereit B, Tonner PH, Bein B, Schleppers A, Steinfath M, Scholz J, Bauer M. Overlapping induction of anesthesia: an analysis of benefits and costs. *Anesthesiology* 2005;103: 391-400.
9. Torkki PM, Marjamaa RA, Torkki MI, Kallio PE, Kirvela OA: Use of anesthesia induction rooms can increase the number of urgent orthopedic cases completed within 7 hours. *Anesthesiology* 2005; 103:401-5.
10. Mangum SR, Cutler K. Increased Efficiency through OR redesign and process simplification. *AORN* 2002;76:1041–46.
11. Sokolovic E, Biro P, Wyss P. Impact of the reduction of anesthesia turnover time on operating room efficiency. *Eur J Anesth* 2002;19:560-63.
12. Ramsay CR, Grant AM, Wallace SA, Garthwaite PH, Monk AF, Russell IT. Assessment of the learning curve in health technologies. A systematic review. *Int J Technol Assess* 2000:16:1095-108.
13. Pisano GP. Learning-before-doing in the development of new process technology. *Res Policy* 1996;25:1097-119.
14. Baloff N. Startup Management. *IEEE Transactions on EM* 1970:17: 132-41.
15. Dutton JM, Thomas A. Treating Progress Functions as a Managerial Opportunity. *Acad Manage Rev* 1984:9:235-47.
16. Argote LS, Beckman L, Epple D. The Persistence and Transfer of Learning in Industrial Settings. *Manage Sci*; 1990;36:140-54.
17. Argote LS. *Organizational learning: creating, retaining, and transferring knowledge*. Boston 1999, Kluwer Academic.
18. Hayes RH, Clark KB. Explaining Observed Productivity Differentials Between Plants: Implications for Operations Research. I*nterfaces* 1985:15:3-14.
19. Barley SR. Technology as an Occasion for Structuring: Evidence from Observations of CT Scanners and the Social Order of Radiology Departments. *Admin Sci Quart* 1986;31:78-108.
20. Orlikowski WJ. Case tools as organizational change: Investigating incremental and radical changes in systems development. *MIS Quarterly* 1993;17:309-40.
21. Strum DP, May JH, Vargas LG. Modeling the uncertainty of surgical procedure times: Comparison of log-normal and normal models. *Anesthesiology* 2000;92,1160-67.
22. Strum DP, May JH, Vargas LG. Lognormal distributions are a best fit model for scheduling surgical procedure times. *Anesth Analg* 1997;845-56.
23. Strum DP, Sampson AR, May JH, Vargas LG. Surgeon and type of anesthesia predict variability in surgical procedure times. *Anesthesiology* 2000;92:1454-66.
24. Zhou XH, Gao S. Confidence intervals for the log-normal mean. *Stat Med* 1997:16, 783-90.
25. Zhou XH, Gao S, Hui SL. Methods for comparing the means of two independent log-normal samples. *BIoMetrics* 1997;53:1129-35.

OE-LNM

5

The Effect of the OR Coordinators Risk Appreciation on Operating Room Efficiency

Evidence from a case study in a Dutch General Hospital

Pieter S. Stepaniak [ł] , Guido H.H. Mannaerts[±] , M.D., Ph.D., Marcel de Quelerij[¥] , M.D. , Guus de Vries [ł] , Ph.D.

Published in Anesth Analg 2009;108:1249-56

[ł] Institute of Health Policy and Management, Erasmus University Rotterdam
[±] Department of General Surgery, St. Franciscus Hospital Rotterdam
[¥] Department of Anesthesiology, St. Franciscus Hospital Rotterdam

Abstract

Background

The Operating Room Coordinator (ORC) is responsible for filling gaps in every operating room schedule. We have observed differences among the personalities of the four ORCs with regard to their willingness to accept taking on more risk concerning their daily planning. The hypothesis to be tested is that the relationship between the personality of each of the four ORCs and the risk an ORC is willing to take of cases running late affects OR efficiency.

Methods

In order to judge the personality of an ORC in relation to risk-taking in planning schedules, we apply the Zuckerman-Kuhlman Personality Questionnaire (ZKPQ) in our study. Seven anesthetists were asked to score every ORC on willingness to take risks in planning. To analyze which risk attitude creates more OR efficiency, the daily prognosis of the ORC compared with the actual OR program outcome was registered during a five-month period in 2006 and 2007. We analyze whether in the opinion of hospital management the costs of reserving too much OR time balances out with the costs of reserving too little OR time, and whether this result is consistent with the assignment of the management tasks of the ORC.

Results

Seven anesthetists classified the four ORCs into the risk-averse group (n=2) and the non risk-averse group (n=2). The ZKPQ results for risk-seeking indicate that there is a difference in risk appreciation between the different ORCs. The main finding in our study is that the non risk-averse ORC plans in more cases to fill the gaps in the OR program than the risk-averse ORC does. The number of extra cases performed by the non risk-averse ORC as compared to a risk-averse ORC is 188 in 2006 and 174 in 2007. The average end-of-program-time per OR/day for the non risk-averse ORC is 34 minutes (± 19 min, p = 0.0085) later than for the risk-averse ORC. We find that this hospital on average reserves more OR time for procedures than is actually required. The non risk-averse ORC takes more advantage of that extra OR time than the risk-averse ORC does by realizing extra cases during office hours. The success of the non risk-averse ORC can be linked to the fact that there is usually time available due to this over-reserving.

Conclusion

The conclusion of this study is that a non risk-averse ORC creates significantly less unused OR capacity without a great chance of running ORs after regular working hours or canceling elective cases scheduled for operation compared to a risk-averse ORC.

Introduction

Changes in the financing of the Dutch healthcare system have forced health organizations to focus more on the efficiency of their logistic processes. The operational risk of ORs is mainly related to elective surgical cases being completed outside regular working hours. A possible consequence of this extension past regular hours is that surgeons anticipate the availability of this extra OR time in their future planning. Having to work frequently beyond regularly scheduled hours can lead to both overtime costs and intangible costs, the latter resulting from dissatisfaction and reduced motivation on the part of the staff. Having to work overtime frequently in the ORs is one of the primary reasons [1] that nurses terminate their employment. Identified scheduling conflicts are a major cause of nursing staff turnover [2].

The OR is also an important financial production unit. The hospital management determines the OR capacity and assigns capacity to the different medical specialties. Increases in the efficiency of use of the ORs results in more production and therefore more revenue for the hospital. The Operating Room Coordinator (ORC) is a nurse anesthetist, selected for this specific job in this specific hospital. In our hospital there are four ORCs. Their responsibilities include rearranging case and staff assignments, as some OR cases take more or less time than originally planned, and unplanned acute patients require surgery. Their jobs involve frequent communication with the various stakeholders such as anesthetists, surgeons, and other OR staff. The responsibilities of the ORC in our study relate to the regularly scheduled work hours of 8 am to 4 pm.

Given the fixed OR capacity between 8 am and 4 pm, the ORC is assigned by hospital management to maximize OR efficiency by filling the gaps with as many cases as possible (planned and unplanned) under the constraint that ORs should close on average no later than 4 pm. Minimization of operating room inefficiency balances the additional costs of cases running late (i.e. overtime has to be paid out and staff morale dwindles) against the opportunity costs of paying idle staff. From an economic point of view, an ORC is constantly weighing the maximization of OR efficiency against minimizing reduced staff morale. As more cases are performed within the maximum margins of the available OR time without overstepping those boundaries frequently, maximum efficiency will eventually increase, and therefore also the contribution margin for the hospital. Every extra case performed in the OR provides a certain amount of Contribution Margin that goes toward the covering of fixed costs. The Total Contribution Margin (*TCM*) is Total Revenue (*TR*) minus Total Variable Cost (*TVC*): $TCM = TR - TVC$.

It is interesting to reveal how the ORC balances OR idle time with exceeding the scheduled time, given these constraints. One of the characteristics of the Dutch healthcare system is its strong supply-side controls (hospital, government). Since 2005, there has been a rapid transition to a demand-driven (patient) model, resulting in an increased competition among the hospitals. Additionally, the Dutch social system is founded on balancing work and private time, all in favor of private time. Today, on average, the ORs are open for business from 8 am to 4 pm. Because of the impending transition to a demand-driven model, this will lead to a need for hospital management to make different choices (i.e. opening hours). Longer opening hours will not decrease the need for an existing ORC since this very same ORC will have the task of filling gaps and maximizing OR efficiency.

We have observed differences between the personalities of the four ORCs, related to their willingness to take on more risk in their daily planning, resulting in a risk of cases running late. This was our motivation for analyzing the effect of the personality and risk aversity of an ORC on OR efficiency. The hypothesis to be tested was that the relationship between the personality of the ORC and the risk the ORC is willing to accept of cases running late affects OR efficiency. Specifically, we hypothesize that a risk-averse ORC causes more inefficiency for the OR.

Material and methods

A decision maker is said to be risk-averse if he prefers less risk to more risk, all else being equal. In the OR, a risk-averse decision maker will want all the ORs to be finished before 4 pm without any chance of running late. The opposite of risk aversion is risk-seeking. A risk-seeking decision maker will prefer more risk to less risk, and accepts the possibility of running late, all else being equal. There are numerous contributions to the conceptualization of subjective orientation toward risk [3,4,5]. Some studies analyze the interaction between personality feature variables which are not risk attitudes. These variables have been linked to decision-making on risky courses of action [6], impulsiveness [7] and decision-making style [8].

Zuckerman [9,10] developed the Zuckerman-Kuhlman Personality Questionnaire (ZKPQ) to assess personality along five dimensions. The results of the ZKPQ have been replicated across several studies. These results have shown for example that risk-taking is related to scores on the ZKPQ impulsive sensation seeking scale [9]. Zuckerman [6, 10,11] defines sensation seeking as a need for new and complex experiences and a willingness to take risk for one's own account. He has found that high sensation seekers tend to anticipate lower risk than low sensation seekers do, even for new activities. This finding indicates that a high sensation seeker is more likely to look for opportunities that provide the chance to take a risk, and that the will to take risks seems less threatening to this specific type of individual.

To assess personality versus risk-taking relationship of an ORC, we apply the ZKPQ in our study. The personal files of the ORCs indicated that their personalities were assessed by the ZKPQ test along five dimensions: Impulsive sensation seeking, neuroticism-anxiety, aggression-hostility, activity and sociability. This test was a standard procedure during the selection process of the ORCs. We used the scores of the impulsive sensation seeking dimension and used the explanatory table (Zuckerman) to rate the ORCs. In our study we grouped ZKPQ scores on impulsive sensation seeking as follows: the scores of very low and low were considered to be risk-averse, the average scores were considered risk-neutral, and the high and very high scores were considered to be non risk-averse.

Prior to the start of the study, seven anesthetists where asked to score every ORC on their risk appreciation. This risk appreciation could be: non risk-averse, risk-averse or risk-neutral. In 2006, prior to the start of the study, the ORCs were informed about this study whereas in 2007 they were not. In order to analyze which risk attitude creates maximum OR efficiency, the ORC's expectations with regard to how the OK program would materialize was registered during a five-month period in 2006 and 2007. This expectation, or prognosis, is proposed by the ORC and he informs the anesthetist on duty of this. When making the prognosis, the following aspects are estimated and noted by the ORC:

- which OR(s) need(s) time after business hours;
- which OR(s) are on schedule;
- the amount of available OR capacity for emergency surgery during the period from 2 pm until 4 pm. This capacity is designated for patients already on the waiting list and for emergency patients outside or inside the hospital who may possibly need emergency/acute surgery.
- the number of the planned elective patients that are to be rejected.

If at 4 pm, all the above-mentioned aspects have been accurately estimated, we say that the ORC's prognosis has materialized. In all other cases, the prognosis has not materialized. During the study period, we measured :

- Whether the prognosis of the ORC made at 2 pm coincides with the actual situation at 4 pm (% of all prognoses made).
- Accurate prognosis made at 2 pm that specific ORs would need extra time after regular working hours (% of all prognoses made).
- The average end time of all ORs.
- The average end time of all ORs after 4 pm.
- The average number of ORs in progress after 4 pm.
- The number of unnecessary rejections of planned elective patients.

We test for significance in the average end of program time among individual ORCs, and within the groups having a factorial ANOVA (significance level 0.05). The correlation between cases is considered to be independent but interchangeable between ORs.

As in Tessler [12], we analyze whether, in our hospital, limited hours serve to restrain the budget. This will help us to understand whether it is cost-effective for the OR management to proceed with a surgical case rather than to postpone it. Olivares [13] makes an estimation of cost parameters based on observed system behavior and assumed rational behavior in reserving OR time for an individual cardiac procedure. Based on Olivares' analysis, we analyze whether in the opinion of hospital management the costs of over-reserving a procedure are lower or higher than the costs of under-reserving a procedure, and whether this result is consistent with the assignment of the management tasks of the ORC. This result gives us insight into whether Olivares' analyses can be generalized to more than one procedure.

In this study no bias is present from seasonal influences or from various new specialized procedures. This study focuses on one of the many issues related to imperfect utilization of ORs. We quantify them and measure the effect of management decisions aimed at reducing imperfection. Due to fixed OR capacity in our hospital, the short-term objective in maximizing OR efficiency is to reduce under-utilized OR time [14]. This is because there are regular cancellations of patients due to medical reasons in the 24-hour period prior to OR.

In order to calculate the inefficiency related to the level of risk aversity, we use the following definition [15]. Operating room inefficiency is the sum of under-utilized OR time and over-utilized OR time, multiplied by the relative costs of overtime. Under-utilized time is hours of staffed operating time at straight time wages, but not used for surgery, set-up or clean-up of the OR. Over-utilized time is hours after operating room time, staffed at overtime. After finishing the first study, the ORC was asked to continue to register their prognosis of the progress of the OR program at 2 pm and actual outcome at 4 pm. The following data were excluded in order to compare the results: unexpected complications during an elective case after 2 pm (2006 n=2, resp. 2007 n=1), disruption of the elective program due to a patient who was brought in with an aneurysm (2007, n=1). Data were summarized using mean ± SD.

Results

The seven anesthetists, anonymously and independently of each other, classified two ORCs in the category of risk-averse, and two in the category of non risk-averse (n=2). Risk indifference was not scored. The results of the ZKPQ are shown in Table 1.

The expectations of the anesthetists as well as the results of the ZKPQ tests all indicate in the same direction, i.e. that there is a difference in risk appreciation between the ORCs.

Table 2 shows the quantitative results of the two groups.

TABLE 1.ZUCKERMAN-KUHLMAN PERSONALITY QUESTIONNAIRE
Scores of the Impulsive Sensation Seeking Dimension ZKPQ score per ORC

ORC	ZKPQ score
#1	81%
#2	92%
#3	25%
#4	32%

Explanation (11)
0 - 27% Very low impulsive sensation seeking
28 - 41% Low impulsive sensation seeking
42 - 70% Average impulsive sensation seeking
71 – 84% High impulsive sensation seeking
85 -100% Very high impulsive sensation seeking

TABLE 2. MAIN RESULTS PER TYPE ORC PER STUDY PERIOD

	Non risk-averse		Risk-averse	
	2006	2007	2006	2007
Working days	98	102	98	102
Prognosis of the ORC made at 2 pm matches the actual outcome at 4 pm (% of all prognoses made)	86%	88%	44%	53%
Accurate prognosis made at 2 pm that specific ORs will require extra time after regular working hours (% of all prognoses made)	86%	82%	18%	26%
Average end time of all ORs	3.53PM (SD =8min)	3.46 PM (SD=10min)	3.13 PM (SD=15min)	3.19 PM (SD=12min)
Average end time of all ORs still running after 4 pm	4.20PM (SD=18min)	4.18PM (SD=14min)	4.16PM (SD=17min)	4.19PM (SD=17min)
Average number of ORs in progress after 4 pm	13.8%	11.3%	8.8%	11.3%
Number of unnecessary rejections of planned elective patients (period of 5 months)	2	3	7	6

Non risk-averse group: In 87% of the cases, actual outcome at 4 pm matches the prognosis given at 2 pm over the two periods. In 84% of the cases, the expectation that the ORs would finish after working hours matched actual outcome. On average, the end time of the ORs is 3:50 pm (± 12 min). The average end times of ORs after 4:00 pm are 4:19 pm (± 17min).

The average percentage of ORs in progress after 4:00 pm is 12.6% (2.5%). The number of unnecessarily rejected planned elective patients in the study period of 5 months is 2 in 2006 and 3 in 2007.

Risk-averse group: In 22% of the cases, actual outcome matches the prognosis over the two periods. In 48% of the cases, the expectation that the ORs would finish after working hours matched actual outcome. On average, the end time of the ORs is at 3:16 pm (± 18 min). The average end time of elective ORs after 4:00 pm is 4:17 (± 17 min). The average percentage of ORs in progress after 4:00 pm is 10.1% (3.1%). The number of unnecessarily rejected planned elective patients in the study period of 5 months is 7 in 2006 and 6 in 2005. The difference in end time between the two ORC groups (risk-averse and non risk-averse) is 34 minutes (± 19 min) per OR per day ($p = 0.0085$).

Within these groups, we encountered some differences: The average time between the two risk-averse ORCs does not differ significantly ($p = 0.291$). The average end time of the two non risk-averse ORCs is significant (p=0.034). The comparison of the ZKPQ results between the ORCs within the non risk-averse group does not lead to any explanations for these differences. We can only conclude that within the risk-averse group, the two are each others' equivalent. Within the non risk-averse group, one ORC shows significantly better results than the other.

The number of extra cases performed by the non risk-averse ORC compared to a risk-averse ORC is 188 in 2006 and 174 in 2007. We can calculate the extra contribution margin for the hospital if for example in 2007 174 more Total Hip Replacements were performed (the patients for this surgery were actually available on a waiting list). The estimated Total Revenue (TR) for 174 cases is $2.1 mln. The estimated Total Variable Cost (TVC) is $ 0.9 mln. This results in an extra contribution margin of $1,2 mln/year. The contribution margin ratio is equal to (2.1 – 0.9) / 2.1) x 100% = 57%.

We analyzed ex-post how many cases the risk-averse group and non risk-averse group could have been planned in the time period from 2:00 pm-4:00 pm. For this analysis, we specifically used either the average time or the median of the case duration, whichever value was greater. For the risk-averse group, the numbers were 133 (2006) and 127 (2007). For the non risk-averse group, the results were 12 and 15. As mentioned before, these cases were actually available for filling gaps in the programs. The distribution of working days per ORC (Table 3) is uniformly distributed.

TABLE 3. DISTRIBUTION OF WORKING DAYS OF OPERATING ROOM COORDINATORS

	Non risk-averse ORC		Risk-averse ORC	
	#1	#2	#3	#4
2006				
Mon	9	10	10	10
Tue	10	10	10	10
Wed	10	10	9	10
Thurs	10	10	10	10
Fri	10	9	10	9
total	49	49	49	49
2007				
Mon	10	10	10	11
Tue	10	11	10	10
Wed	10	10	10	10
Thurs	10	10	11	10
Fri	10	11	10	10
total	50	52	51	51
TOTAL	99	101	100	100

The OR schedule is shown in Table 4.

TABLE 4. OR SCHEDULE

DAY/OR	1	2	3	4	5	6	7	8
Mon	Orto	Orto	Neuro	Gen Sur	Plast	Gyn	Gen Sur	ENT
Tue	Orto	Orto	ENT	Gen Sur	Gen Sur	Gyn	Gen Sur	Eye
Wed	Orto	Orto	Gen Sur	Jaw	Plast	Uro	Gen Sur	Gen Sur
Thurs	Orto	Orto	Gen Sur	Gen Sur	Plast	Neuro	Uro	Eye
Fri	Orto	Orto	Gen Sur	Gen Sur	Plast	GYN	Gen Sur	Gen Sur

We analyzed the demand for OR time during the historical period in time, and the study period. In the period 2005-2007, we looked at the average planned OR time and compared this with the average planned OR time during the study period. The average planned OR time is the sum of the planned OR time of all cases for a specific OR, including a standard turnover time of 11 minutes between cases divided by the number of days. There were no significant differences in the planned end time of the various OR suites (Table 5).

TABLE 5. AVERAGE PLANNED END TIME OR ROOMS

OR #	Historical	Study period
1	3:25 pm (21min)	3:21 pm (23min)
2	3:22 pm (32min)	3:19 pm (29min)
3	3:35 pm (29min)	3:28 pm (26min)
4	3:35 pm (18min)	3:29 pm (20min)
5	3:21 pm (27min)	3:22 pm (25min)
6	3:47 pm (17min)	3:45 pm (19min)
7	3:59 pm (14min)	3:58 pm (15min)
8	3:17 pm (34min)	3:19 pm (36min)

Furthermore, we studied the sample variance among OR-day combinations. For the study period we used Levene's test of homogeneity of variances. With $p = 0.903$ (2007) and $p = 0.189$ (2006), we can conclude that in both study periods we can accept the null hypotheses of equal variances. We performed the one-way ANOVA to compare means of case duration of the four ORCs. With a p value of 0.603, we accept the null hypotheses of equal means for the case duration for the four ORCs.

Finally, we conducted the factorial analysis with procedure time as dependent variable, and 'specialty' and 'day' as fixed factors.

Because we have a fixed OR schedule, we see that 'day' and 'specialty', and their interaction with each other are significant ($p < 0.001$). We also affirm in our study that a risk-averse person orders less than the normative benchmark and that a risk-seeking decision maker orders more than that very same normative benchmark [16].

We calculated the mean inefficiency per OR per day by considering each OR-day to be independent of all others. The relative cost of overtime in our study is 1.50. The cost per hour of over-utilized OR time includes: indirect costs, intangible costs, and retention and recruitment costs incurred on a long-term basis from staff working late. The mean inefficiency per OR per day for the risk-averse ORC is 0.87 (+/- 0.29), n=1,600. For the non risk-averse ORC, the mean inefficiency per OR per day is 0.46 (+/- 0.20), n= 1,600. This means that the non risk-averse ORC causes a lower OR inefficiency.

In 65.9% of all cases, the procedure was completed in as much time or less time than had been reserved by the OR. In 34.1% of all cases, more time was needed than was reserved. Comparing these results leads to the conclusion that case durations are over-estimated. The over-reserving leads to idle time. The non risk-averse ORC takes more advantage of this over-reserving than the risk-averse ORC does, by realizing extra cases during hours. The OR management assumes that from the cost perspective it is better to finish ORs during regular hours. Therefore we performed Tesslers' study [12] in our hospital (Table 6) to verify this assumption. As concluded by Tessler, we confirm that it is cost-effective to proceed with a surgery case after regular working hours rather than to postpone the case. This outcome helps OR management to improve the OR efficiency further.

TABLE 6. ZERO TOLERANCE FOR OVERTIME INCREASES SURGICAL PER CASE COSTS
Assumptions as in Tessler (13), calculated for our hospital

Average hourly wages (incl benefits)		Labor costs/hr/$
Post Anesthesia Care Unit Nurse		35.02
Operating Room Nurse		40.85
Surgical Ward Nurse		29.18
Nurse anesthetist		43.00
Administration		35.02
-Public Relations, Purchasing		
-Telecommunications, Garbage Disposal		
-Accounting, Human Resources		
Laundry/Housekeeping		17.51
Maintenance		19.84
Security		23.34
Pharmacy		37.35
Radiology		32.68
Laboratory		32.68
Central Supply Room		15.17
Physiotherapy		36.77
OR labour costs hourly		450
Marginal Tax Rates		
Marginal tax rates for individuals were derived from the Ministry of Finance		
The marginal tax rate varies between 15.75% and 52% depending on gross		
annual income		
Income	Tax Rate	
0 - $ 26,844	15.75%	
$ 26,844 - $ 48,239	23.50%	
$ 48,239 - $ 82,249	42.0%	
$ 82,249 and higher	52.0%	

COSTS CALCULATED IN THE SOCIETY PAYS MODEL (AFTER TAX VALUES)

US dollars			
		Proceed with case	Postpone case
OR Labour costs			
1.5	hr standard cost	377.58	
1.5	hr overtime cost	566.37	
3	hr standard cost		755.16
Nurse anesthetist costs			
1.5	hr standard cost	38.06	
1.5	hr overtime cost	57.08	
3	hr standard cost		76.11
OR supplies costs		292.95	292.95
Anesthesia supplies costs		39.00	39.00
Professional fees		310.25	310.25
Hospital costs per surgical bed/day			
	Labour	219.55	439.10
	Supplies	17.06	34.12
Hospital costs per patient bed/day			
	Administrative	8.75	17.51
	Technical		
	Laundry/ Housekeeping	2.92	5.84
	Maintenance	1.65	3.31
	Security	4.67	9.34
	Pharmacy	9.71	19.42
	Radiology	19.61	39.22
	Laboratory	14.71	29.41
	Central Supply Room	5.31	10.62
	Physiotherapy	5.15	10.29
Post Anesthesia Care Unit			
	Labour	109.25	218.49
	Supplies	7.34	14.68
Lost income for one day (after tax)			0.00
Professional fees saved			310.25
TOTAL		2106.96	2635.06

COSTS CALCULATED IN THE PATIENT PAYS MODEL

US dollars		
	Proceed with case	Postpone case
OR Labour costs		
3 hr standard cost	1348.50	1348.50
Nurse anesthetist costs		
3 hr standard cost	199.95	199.95
OR supplies costs	292.95	292.95
Anesthesia supplies costs	39.00	39.00
Professional fees	646.35	646.35
Hospital costs per surgical bed/day		
Labour	399.18	798.36
Supplies	17.06	34.12
Hospital costs per patient bed/day		
Administrative	13.89	13.89
Technical		0.00
Laundry/ Housekeeping	4.49	4.49
Maintenance	2.71	2.71
Security	7.07	7.07
Pharmacy	16.46	16.46
Radiology	33.23	33.23
Laboratory	24.51	24.51
Central Supply Room	7.93	7.93
Physiotherapy	8.58	8.58
Post Anesthesia Care Unit		0.00
Labour	182.08	182.08
Supplies	7.34	7.34
Lost income for one day (after tax)		0.00
TOTAL	3251.28	3667.52

COSTS CALCULATED IN THE HOSPITAL PAYS MODEL

US dollars			
		Proceed with case	Postpone case
OR Labour costs			
1.5	hr standard cost	674.25	
1.5	hr overtime cost	1011.38	
3	hr standard cost		1348.50
Nurse anesthetist costs			
1.5	hr standard cost	69.19	
1.5	hr overtime cost	181.77	
3	hr standard cost		199.95
OR supplies costs		292.95	292.95
Anesthesia supplies costs		39.00	39.00
Hospital costs per surgical bed/day			
	Labour	510.58	1021.15
	Supplies	17.06	17.06
Hospital costs per patient bed/day			
	Administrative	13.89	27.79
	Technical		
	Laundry/ Housekeeping	4.49	8.98
	Maintenance	2.71	5.42
	Security	7.07	14.15
	Pharmacy	16.46	32.92
	Radiology	33.23	66.47
	Laboratory	24.51	49.02
	Central Supply Room	7.93	15.85
	Physiotherapy	8.58	17.16
Post Anesthesia Care Unit			
	Labour	182.08	364.16
	Supplies	7.34	7.34
TOTAL		3104.47	3527.87

Discussion

In recent years, market forces have made their entry into the healthcare system in the Netherlands. As a result, the government and insurance companies wish increasingly to scrutinize the added value of care processes. Dealing with efficiency plays an important role in this. The more efficiently the processes can be organized, the more efficiently the various resources can be used. Many extensive studies on OR efficiency [17,18,19,20,21,22]. can be found in the literature. All these studies contribute to optimalizing the use of scarce and costly operating rooms, especially in the more so-called private labeled hospitals in a competitive environment. Up till now, the effect of the type of risk appreciation of an ORC in relation to the OR efficiency has not been described in any literature. The results of our study are in compliance with the findings in the literature we used: a high sensation seeker is likely to look for opportunities that provide the chance to take a risk, and this risk will seem less threatening to this kind of individual. Though there is a lot of evidence to support the link between personality and risk-taking, the literature shows that the exact nature is still unclear. The next step is to find out what happens in the mind of a risk-taker that is significantly different from what occurs in the mind of a non risk-taker.

Based on Olivares' preliminary analyses, we have to conclude that in the opinion of our hospital management, the cost of over-reserving a procedure is 48% lower than the cost of under-reserving. From that perspective we could conclude that frequently, the OR management prefers to reserve more time than actually needed, hence the conclusion that our hospital often prefers to be finished earlier rather than late. From an economic point of view, this is irrational behavior because the opportunity costs of idle OR time are considered to be lower than for utilized OR time. Hence, the conclusion of Olivares based on specific cardiac surgery cases cannot be generalized to a situation in which there is a heterogeneous mix of different operations and patients. When we mathematically conclude that an OR over-reserves, this fact does not mean that the OR management frequently prefers to reserve more time than actually needed. The fact is that for whatever reason, on average, this hospital over-reserves. On average, properly scheduled operating rooms will finish early two-thirds of the time and late one-third of the time [23]. In our hospital, that proportion is in accordance with these results.

Although the data from this study are statistically strong, there are some specific potential drawbacks that can be specified. The first factor is that the study population (4 ORCs) is relatively small. Despite being able to attain statistically significant data, it will be important to follow the trends as ORCs in the operating room system are introduced. The second factor is our decision to choose one axis: sensation seeking. But there are other axes, such as neuroticism-anxiety, aggression-hostility, activity and sociability that can be either important, necessary, or completely determinative for an ORC's success in planning the schedule. This has to be analyzed in future studies with a larger population of ORCs.

The third factor is the number of operating rooms. In our study we observed eight operating rooms. The effect of the difference on OR efficiency may be influenced by the number of ORs. Since two additional new operating rooms have been built, we recommend performing this study with ten ORs rather than eight, to examine the effect of a bigger span of control for the ORC on the OR efficiency. The ORC works in an environment of over-reserving. Hence the question arises whether the ORC will also be successful in the case of under-reserving. Risk aversity is a typically human attitude toward risk. In the case of under-reserving, the non risk-averse person will always explore the extremes of all the possibilities available. In such cases, this trait can lead to dissatisfaction among the OR staff, because the chance of having to work overtime structurally increases in combination with the risk that planned cases will be canceled. This effect leads to material and immaterial damage for both patients and hospital. In such a case it is quite easy to imagine that a risk-averse ORC will be more successful in his tasks at hand. This should be studied further in another setting.

One could take the point of view that even with accurate planning or deliberate over-booking, it would be best to have the non risk-averse ORC employed, because over-utilized time can be justified due to the fact that more cases will get done. The hospital then simply must take the strategic steps to bring the case volume up to the planned capacity of the OR suite, by adding rooms, expanding hours, etc. However, in the short term, a non risk-averse ORC who schedules cases into an already overbooked OR will create significant animosity among the staff.

Conclusion

The conclusion of this study is that a non risk-averse ORC creates significantly less unused OR capacity without a great chance of running ORs after regular working hours or canceling elective cases. Added to this, a non risk-averse ORC is cost-effective. This means that when recruiting an ORC, the risk-averse type must be one of the selection criteria. These findings will help management to further optimize OR efficiency and the results can be used in further research into a decision-support system to provide recommendations.

References

1. Stachota P, Normandin P. Reasons registered nurses leave or change employment status. *J Nurs Admin* 2003;33:111-18.
2. Thomson TP, Brown H. Turnover of licensed nurses in skilled nursing facilities. *Nurs Econ* 2002;20:66-69.
3. Sitkin SB, Pablo AL. Reconceptualizing the determinants of risk behaviour. *Acad Manage Rev* 1992;17:9-38.
4. Trimpop RM, Kerr JH, Kirkcaldy BD. Comparing personality constructs of risk-taking behaviour. *Pers Indiv Differ* 1999;26:237-54.
5. Weber EU, Hsee CK, Sokolowska J. What folkore tells about risk and risk taking? Cross-cultural comparisons of America, German and Chinese proverbs. *Organ Behav Hum Dec* 1998;75:170-86.
6. Zuckerman M. The psychophysiology of sensation seeking. *J Pers* 1990;59:313-45.
7. Eysenck SBG, Esenck HJ. The place of impulsiveness in a dimensional system of personality description. *Brit J Soc Clin Psyc* 1977;16:57-68.
8. Franken RE. Sensation seeking, decision making styles and preference for individual responsibility. *Pers Indi Diff* 1988;9:136-46.
9. Zuckerman M. An alternative five-factor model for personality. In "Halverson CF, Kohnstamm GA, Martin RP. The developing structure of temperament and personality from infacy to adulthood 1994:53-68. Hillsdale, NJ: Lawrence Erlbaum.
10. Zuckerman, M. Zuckerman-Kuhlman Personality Questionnaire (ZKPQ): An alternative five factorial. In "B. DeRaad (Ed.), Big Five Assessment". Seattle, WA: *Hogerefe and Huber Publishers* 2002.
11. Zuckerman M, Kuhlman M. Personality and risk taking: common biosocial factors. *J Pers* 2000;68:999-1029.
12. Tessler MJ, Kleiman SJ, Huberman MM. A zero tolerance for overtime increase surgical per case cost. *Can J Anesth* 1997;99:1036-41.
13. Olivares M, Terwiesch C. Reserving operating room capacity: a structural estimation using the newsvendor model. University of California 2004.
14. Dexter F, Traub R. How to schedule surgical cases into specific operating rooms to maximize the efficiency of use of operating room time. *Anesth Analg* 2002;94:933–42.
15. Dexter F, Macario A. When to Allocated operating room time to increase operating room efficiency. *Anesth Analg* 2004;98:758-62.
16. Eeckhoudt L, Gollieb C. Schneider T. Risk-aversion, prudence and temperance: A unified approach. Catholic Faculties of Mons, Belgium and Lille, France 1995.
17. Dexter F, Epstein RH, Marsh HM. A statistical analysis of weekday operating room anesthesia group staffing costs at nine independently managed surgical suites. *Anesth Analg* 2001;92:1493-8.
18. Dexter F, Macario A. Forecasting surgical groups total hours of elective cases for allocation of block time: Application of time series analysis to operating room management. *Anesthesiology* 1999;91:1501-8.
19. Dexter F, Epstein RH, Traub R. Making management decisions on the day of surgery based on operating room efficiency and patient waiting times. *Anesthesiology* 2004;101:1444-53.
20. Dexter F, Epstein RH. Operating room efficiency and scheduling, Anesthesiology 2005;18:195-8.
21. Friedman DM. Sokal SM. Increasing operating room efficiency through parallel processing, *Ann Surg* 2006;243:10-4.
22. Kanconck WM, Walter PF. Operating room scheduling procedure times: comparison of log-normal and normal models. *Anesthesiology* 2000;92:1160-7.
23. Strum DP, Vargas LG, May JH. Surgical subspecialties block utilization and capacity planning. A minimal cost analysis model. *Anesthesiology* 1999;90:1176–85.

6

Quality improvement: balancing the risks of overtime and cancellation of scheduled cases

Pieter S. Stepaniak[ł], Ronald A.C. van der Velden[¤], Albert P.M. Wagelmans[¤], Ph.D., Joris van de Klundert[ł], Ph.D.

Submitted for publication

[ł] Institute of Health Policy and Management, Erasmus University Rotterdam
[¤] Econometric Institute, Erasmus School of Economics, Erasmus University Rotterdam

Abstract

In its influential 'Crossing the quality chasm', the Institute of Medicine [1] identifies six quality dimensions of health care, among which are efficiency and timeliness. The six dimensions together, make quality improvement a complex matter, as interventions which yield improvement regarding one dimension may have a negative effect regarding another: the quality dimensions form conflicting objectives. In this research we simultaneously address efficiency and timeliness of care in the operating theatre.

We formally model the real time surgery scheduling to minimize a weighted sum of cancellation of scheduled cases, overtime cost, moving scheduled cases from the day to the service operating room and scheduling emergency/acute cases after an imposed time limit. Stepaniak et al. [2] show how risks attitudes of OR planners influence the quality of scheduling. We formally model heuristics which are based on different risk attitudes and analyze their mutual performance. More generally, we analyze Monte Carlo based optimization methods and use recent actual data from the St. Franciscus Gasthuis, Rotterdam, The Netherlands.

Introduction

Operating theatres daily present dynamic situations which result from unanticipated developments of scheduled cases, arrival of emergency cases, and the scheduling decisions taken during the day by the operating room coordinator (ORC). The task of the ORC is to ensure that operating rooms (ORs) finish on time and that all scheduled cases as well as the emergency cases are completed. At the end of the dynamic days however, ORs often finish late and scheduled cases have experienced delays or been canceled.

Delays or cancelation add to the patient's inherent anxiety associated with surgery and engenders anger and frustration. They have been shown to be an important determinant of patient dissatisfaction across the continuum of preoperative-operative-postoperative care [3]. Delays in scheduled surgical cases affect patient satisfaction even more than the intraoperative anesthesia experience [4]. The delays and cancelations may have a significant impact on quality of care for other cases as well. They generate time pressure and overtime work, which is known to be one of the primary reasons [5] for churn of OR nurses. Likewise, identified scheduling conflicts are a major cause of nursing staff turnover [6].

The obvious solution to reduce the daily number of scheduled cases may indeed serve to reduce the operational difficulties encountered, but reduces productivity in the long run. As a result, waiting times increase and hospitals loose revenue. The ability of an ORC to handle the dynamic developments that occur during an OR day is therefore crucial to keep help care accessible and affordable, without endangering employee satisfaction, and putting customer satisfaction and quality of care at stake. To this purpose, the ORC may decide to rearrange case and staff assignments when cases take more or less time than planned, and/or emergency cases arrive. The results of Chapter 2 and 3 provide the ORC with more reliable estimates of the duration of surgical cases. Chapter 4 addresses how the risk attitude of an ORC affects the quality of the scheduling decision making. It measured risk attitude of an ORC using the sensation seeking dimension of the Zuckerman-Kuhlman Personality Questionnaire (ZKPQ).

Chapter 5 reveals that a risk-seeking ORC cancels fewer cases with lower OR inefficiency as compared with the non risk seeking ORC. In this research we formalize risk attitudes in heuristics developed to solve the real time scheduling problems ORCs face during the day. We formulate the optimization problem from a quality perspective, where we focus on effectiveness and timeliness of the surgery services. The heuristics consider a variety of scenarios for how the day might develop and select a decision based on the outcomes in the various scenarios. Risk attitudes are modeled by assigning weights to the scenarios. Risk averseness is modeled by assigning much weight to bad case scenarios, and risk seeking behavior is modeled by assigning much weight to favorable case scenarios. We evaluate the heuristics using data from the year 2009 of the St. Franciscus Gasthuis (SFG) in Rotterdam and use Monte Carlo optimization. The simulation environment we build to that purpose allows us to simulate and analyze the effects of different risk attitudes of ORCs on e.g. the number of rejected patients, overtime costs, OR inefficiency.

We now briefly review previous work on closely related problems. Dexter [7] uses computer-based, hypothetical OR suites, to test different OR scheduling strategies to develop an OR scheduling strategy aimed at maximizing OR utilization. OR utilization depends greatly on and increases as, the average length of time patients wait for surgery increases. Charnetski [8] uses simulation to study the problem of assigning time blocks to surgeons on a first-come, first-served basis when the goal is to balance the waiting cost of the surgeon and the idle cost of the facilities and operation room personnel.

The proposed heuristic recognizes that different types of procedures have different service time distributions and sets case allowances based on the mean and the standard deviation of the individual procedure times. Murray 's [9] challenge was to design a schedule that would use every surgical bed in the hospital seven days a week and not use one bed more or less. Developing such a schedule would require relating OR procedures to bed use. Dexter [10] uses computer simulation to evaluate ten scheduling algorithms described in the management sciences literature to determine their relative performance at scheduling as many hours of add-on elective cases as possible into open OR time.

The reader may note that the aforementioned references aim to optimize resource use, rather than quality of care. Before considering solution techniques, and testing them in a simulation environment, we therefore proceed with a formal problem analysis in the next section. The position we take is that resource utilization is not a goal in itself, and less so for the real time scheduling decisions made by an ORC. The ORC starts with a given schedule and deals with the turn of events as it materializes while performing scheduled cases and emergency cases as they newly arrive. The ORC may cancel scheduled cases, or defer them to the service OR. All other cases have to be performed, potentially yielding overtime work. The task of the ORC is therefore to balance the costs of working overtime with the effects cancelations have on patient satisfaction and patient health.

Problem definition

Formally, the scheduling problem an ORC faces during a day at the OR is known as a parallel machine scheduling problem. More specifically, the problem is a stochastic scheduling problem, as the arrival process of emergency cases is a stochastic process and the durations of the cases are stochastic variables. The distributions of the stochastic variables are however assumed to be a priori known [11]. Closely related stochastic parallel machines scheduling problems have been widely studied in the context of manufacturing and computer-communication systems [12,13,14,15,16,17].

The field of stochastic scheduling is motivated by the design and operational problems arising in systems where service resources must be allocated over time to cases with random features vying for their attention [18]. Obviously, operating rooms and their crews can be considered as such service resources. The daily task of each operating room and the crew working in the room is to process a list of scheduled cases. Moreover, some of the rooms and crews can be assigned emergency cases which arrive as the day proceeds. It is common to distinguish two types of emergency cases: acute and emergency. Acute cases must be scheduled without further delay, typically within 30 minutes. Emergency cases be scheduled on the day of arrival, typically within four hours. Thus we have a scheduling problem with a stochastic arrival process, stochastic case durations, and parallel machines.

In the SFG one operating room and crew typically remain in operation at the end of the day and during the following night to service newly arriving emergency cases. This service capacity is oftentimes also used to process a number of cases which have been eliminated from the daily programme, or have arrived during the day and have not been processed yet. Thus, in the problem under consideration, the input consists of:

- A set O of n operating rooms
- A set of case types CT. Each case type $t \in CT$ has a stochastic duration $d(t)$.
- A set of emergency case types EC, $EC \subseteq CT$ For each emergency case type $t \in EC$, the arrival process is denoted by $A(t)$.
- Sets SC and EC of scheduled and emergency cases, respectively. We let $C \equiv SC \cup EC$ denote the total set of cases. Notice that the emergency cases of EC are not part of the problem initial input, they are implicitly defined by the arrival processes $A(t)$.
- Linearly ordered subsets $\{S_1, S_2,..S_n\}$ that form a partition of SC. For i = 1..n linearly ordered subset S_i corresponds to the list $\{C_1, C_2,..\ C_m\}$ of scheduled cases which are planned to be processed in order in operating room i.
- A subset $OE \subseteq O$ to which emergency cases can be assigned. An emergency case e_j scheduled in room $i \in OE$ can be scheduled between any two subsequent cases of S_i.
- Opening time span T during which all ORs are opened. We assume they open at time 0, meaning that they are opened until time T.
- Service operating room $SO \in O$ which serves as the service operating room.
- Service opening time span N, which is the time span during which cases that are

not served during the day can be scheduled in the service operating room *SO.* We assume it lasts from time *T* till time *T+N*, implying that a case running in *SO* before *T* can continue after *T* without incurring overtime.

Postponing the definition of the objective function, we define by Real Time Surgical Scheduling (*RTSS*), the problem of scheduling all cases so as to minimize total costs. The classical objective functions of parallel machine scheduling problems are makespan minimization and minimization of the sum of the completion times. Neither of these apply to the problem at hand, as will be shown in the analysis below. The problem at hand is related to scheduling problems in which the objective is a weighted functioned of the makespan and penalties for rejected cases. Such scheduling problems with rejection have been studied by various authors [19,20,21,22,23]. The prime objective we state for operating room scheduling is to minimize the loss of quality incurred by canceling cases. Obviously, cancelation is not equally harmful or undesirable for each of the cases. Cancelations is not only undesirable from a patient perspective, it is also undesirable from a hospital perspective, as cancelation means lost revenue (at least for the day of cancelation).

Moreover, if the patient remains in the hospital until surgery takes place, cancelation induces hospital costs without extra revenues (e.g. when revenue is Diagnoses Related Group based). As there are no realistic data available to express the patient and provider effects on canceling a case, we assume that they depend on the surgery type. For all practical purposes, we therefore assume in the remainder that they are proportional to the expected duration of the surgery, with parameter α. (The analysis or solution methods however, do not rely on this assumption.) For any solution to the *RTSS*, let *CC* the set of canceled cases, and let C_i be the completion time of room *i*, that is the time at which the last case assigned to room *i* is finished. Notice that we only know set of canceled cases after all cancelation decisions have been made, i.e. at the end of the day. As already explained in the introduction, prohibiting cancelations, especially when there are relatively many emergency cases, and/or many cases which take longer than expected, is not desirable. Overtime is costly, and high work pressure and frequent overtime lead to loss of quality of care and employee dissatisfaction. Hence the sum of the overtime is also considered in our objective, with a penalty of β per time unit.

As operating room SO stays open after a working day ORCs may decide that cases which cause overtime are not canceled but referred to *SO* be processed after time *T* and before time *T +N*. Although doing so will typically not lead to problems it is considered to be undesirable, if only because it may cause delays for surgery to emergency cases which newly arrive. Thus, we penalize reassigning cases to *SO* to be processed between *T* and *T+*

N with penalty γ. Let us denote by *CN* cases assigned to be executed in *SO* between *T* and *T* +*N*. As is common in practice a time limit is imposed during emergency and acute cases should be executed after arrival (240 minutes for emergency cases and 30 minutes for acute cases. If emergency and acute cases are executed later than these time limits then they are penalized. Let us denote by *EA* set emergency and acute cases not executed within time limits with penalty δ: $e_j = 1$, if not executed within time limits
Then we consider in the remainder the following objective function:

$$Minimize\colon\ E\ \{\alpha \sum\nolimits_{c\ \in CA} d\big(t(c)\big) + \beta \sum\nolimits_{i=1..n} (\max(0, C_i - T)) + \gamma \sum\nolimits_{c\ \in CN} d\big(t(c)\big) +\ \delta\ \sum\nolimits_{c \in EA} e_j\ \}$$

For the ORC, the stochasticy plays a role during the decision making process, as emergency cases arrive, and durations of cases become known only after scheduling decisions have been made, yet additional decision making is required. Such scheduling problems have become known as online decision making meaning that at a point in the time axe we cannot see future cases, whereas we only know the cases before or at this point of time. Simultaneously decisions have to be made real time at the same point of time about future cases. The designation "on-line, real-time" means that in the OR decision making is needed to respond to stimuli (arrival of emergency cases, expected duration of cases) at numerous ORs. Hence, we can define *RTSS* can be characterized as a on-line, real-time problem.

While making decisions in real-time, it is useful to evaluate the expected value of the objective function as it is influenced by the scheduling decisions. The first component is easy to evaluate as a scheduling only change it when the decision is to cancel a scheduled case. The change then equals γ times the expected duration of the canceled case. The effect scheduling decisions, e.g. canceling a case, or inserting an emergency case, have on expected overtime is however harder to assess. Let (*i*) denote the expected overtime of room *i* and let, $EC(i) = \Sigma_{c \in Sj} E(d(t))$ that is the sum of the expected durations of the cases *j* scheduled in room *i*. Then, contrary to $EC(i) = \Sigma_{c \in Sj} E(d(t))$, $E(i) \neq \max(0, EC(i))$, i.e. the expected overtime is not equal to the expected surplus duration over the opening time span *T*.

Indeed one easily envisions examples where the expected completion time $EC(i) < T$ while there is a positive probability that $C_i > T$. Thus, in order to evaluate the effect of scheduling decisions on the objective function, it is required to know the joint distribution function of all $C \in S_i$. As we have shown in Chapters 3 and 4 that three-parameter lognormal distributions better approximate durations than alternatives studied in the literature, we proceed by exploring the joint distribution function of a sum of independent three-parameter lognormally distributed variables.

The probability density function of the three-parameter lognormal distribution is as follows:

$$f_X(x;\ \mu,\sigma,\theta) = \frac{1}{(x-\theta)\sigma\sqrt{2\pi}} e^{-\frac{(\ln(x-\theta)-\mu)^2}{2\sigma^2}}$$

Here θ is the shift parameter (the case where θ = 0, is called the two-parameter lognormal model).

By the central limit theorem the distribution of the sum of independent lognormal distributed random variables tends to a normal distributions as *n* tends to infinity. As the number of cases per operating room seldom exceeds ten, the central limit theorem does not apply. However, a closed-form expression does not exist for the sum distribution and it is difficult to numerically calculate the distribution.

The closed form distribution has been approached by other methods. Fenton [24] approximated the distribution by a lognormal distribution which has the same moments as the exact sum distribution. Barakat [25] used the inverse Fourier transform of the characteristic function. Analytical work on statistical properties is reported in [26,27,28,29,30,31,32]. Jingxian [27] uses the moment generating function as a tool in the approximation and does so without the extremely precise numerical computations at a large number of points that were required by the other proposed methods in the literature. The methods by Schwartz-Yeh [28], and Beaulieu-Xie [29] approximate the lognormal sum by a single lognormal random variable, and provide different recipes for determining the parameters of the lognormal probability density function. The methods by Slimane [30] and Schleher [31] instead compute a compound distribution or specify it implicitly. Beaulieu et al. [32] have studied in detail the accuracy of several of the above methods, and show that each method has its own advantages and disadvantages; none is unquestionably better than the others.

We conclude for the moment that it is not known how to analytically determine the sum distribution of lognormal cases durations and hence expected overtime. More informally this means that risks of overtime are hard to assess for ORCs. As the *RTSS* not only has stochasticity in the case durations but also in the arrival processes of emergency cases, our solution methods and overtime estimates rely on Monte Carlo simulation, rather than on analytical approximations.

Having specified the input and the objective function, we now turn to the constraints of the problem, thus defining the solution space available to the ORC. First of all, we assume that the assignment of scheduled cases is given, as is the linear order of the cases per operating room. Thus the order of the cases cannot be modified, except for the insertion of emergency cases. Emergency cases can only be scheduled in dedicated operating rooms, which typically have slack time to accommodate emergency cases. Cases that have already started cannot be interrupted (preempted) for emergency cases. Further, emergency cases cannot be canceled. When a case for operating room *i* is canceled, it is the last scheduled case in the linear order of cases assigned to room *i*. As an alternative to being cancelled, the last scheduled case can be referred to operating room *NO* to be scheduled between *T* and *T+N.* Cancelation and referral decisions cannot be undone. The ORC does not have information on regarding future arrivals of emergencies or durations of cases other than the information described in the problem input.

Solution Methods

To start the analysis of *RTSS* and subsequently derive solution methods, let us first consider the offline deterministic version of the problem i.e. the version in which all durations and all emergency cases are known a priori. The problem is therefore to select an operating room for each of the emergency cases, and to choose scheduled cases to delay to the night shift or to cancel. If there is only one operating room to which all emergency cases are assigned, this problem reduces to deciding which of the scheduled cases are delayed and which ones are canceled. This problem can be seen to be Non-deterministic Polynomial time Complete (NP-complete). Intuitively, it means that a solution to any search problem can be found and verified in polynomial time by a special (and quite unrealistic) sort of algorithm, called a non-deterministic algorithm. Such an algorithm has the power of guessing correctly at every step.

Suppose that there is only one operating room available during the daytime and that $T = N$. Furthermore, let the set of cases consist entirely of emergency cases, i.e. $C = EC$ and assume that all these cases arrive at the beginning of the day. Finally, assume that these cases are all of the type that needs to be scheduled on the day of arrival and that the sum of the durations of the cases is $T + N = 2T$. We let $\gamma = 0$. Then there exists a solution of cost 0 if and only if *EC* can be partitioned into two subsets of cases, where the sum of the durations of the cases of either of these subsets equals *T*. Thus, this special case of *RTSS* is equivalent to the partitioning problem which is NP-Complete.

If there are two or more operation rooms to which emergency cases can be assigned, *RTSS* can easily be shown to be strongly NP-Complete, as it then contains 3 partitioning as a special case. Thus we conclude that *RTSS* is already hard to solve without considering the stochasticity, or the fact that it is an online problem. At present, it cannot be ruled out that the offline version of *RTSS* is intractable as it is not known how to express the joint distribution function of the sum of three-parameter lognormally distributed random variables. For this reason, worst case analysis and analysis of the *RTSS* with stochastic durations are highly non-obvious.

We conclude that for all practical purposes, the solution of *RTSS* will rely on heuristic approaches. To the best of our knowledge, no practical applications are known where *RTSS* is solved using scheduling software. ORCs make scheduling decisions, based on real-time information on the execution of the scheduled cases and emergency cases. ORCs typically work without formal rules and/or procedures, and the quality of the scheduling decisions therefore depends on the attitudes, estimations, and problem solving skills of the ORCs. In Chapter 5, we measure risk attitudes of ORCs using the sensation seeking dimension of the Zuckerman-Kuhlman Personality Questionnaire (ZKPQ) , and subsequently show that risk seeking ORCs are better schedulers than risk averse ORCs. We now propose a formalization of these risk attitudes in the context of *RTSS*, and analyze heuristics in which risk attitude is modeled by means of a parameter.

We consider heuristics which produce scheduling decisions during the day, at predefined moments. These moments will be called time instants. Thus, at time instant *t*, a heuristic uses all information about the turn of events until *t* – including scheduling decisions already made, as well as information regarding scheduled and expected cases which are not completed at time *t*, including their expected durations. The heuristics we propose consider all feasible decisions that take effect at *t*, and choose one based on risk attitude.

To evaluate a feasible decision in our heuristic approach, we sample a fixed number of scenarios, each of which completely specifies all arrivals of emergency cases after t, and the durations of all cases to be completed after *t* according to the scheduling decisions made. We define the cost of a scenario by the cost of the optimal solution for the offline problem instance of *RTSS* as specified by a scenario. Since we want to evaluate a feasible solution at time instant t, we in fact consider the conditional cost of a scenario, i.e., the cost of an optimal solution for the scenario, under the condition that the decision under consideration is indeed taken at time *t*. We subsequently define risk attitude on the basis of the scenarios that are taken into account when evaluating decisions. Risk averse ORCs are modeled by considering only a subset of scenarios with high conditional cost for the decision under consideration, whereas risk seeking ORCs are modeled by considering only a subset of scenarios that have low conditional cost for the decision under consideration. In the end, both types of ORCs choose that decision that they evaluate as best.

To formalize this idea, consider the outcomes of a decision for a set of M scenarios. To evaluate the decision, a family of functions is used. Each of these functions sorts the costs under the different scenarios and then takes the average of a subset of these sorted costs. Family members differs in the subset that is used and different subsets represent different risk attitudes. The subsets depend on parameters $\phi \in [0,1]$, $\omega \in (0,1)$ as follows. Let x be the vector of sorted outcomes with x_i an element of this vector. We assume x_1 is the smallest cost (best case) and x_m is the largest cost (worst case).

For given ϕ and ω we define a function $f_{\phi,\omega}(x)$ on the vector of sorted outcomes as follows:

$$f_{\varphi,\omega}(x) = \frac{1}{\omega M}\sum\nolimits_{i=1+\varphi(1-\omega)M}^{\omega M+\varphi(1-\omega)M} x_i$$

which is the average of the outcomes with indices between the boundaries 1 + ϕ(1-ω)M and ωM + ϕ(1-ω)M , which is an interval containing ωM outcomes.
We have three special cases:
For ϕ = 0 we have

$$f_{0,\omega}\ (x) = \frac{1}{\omega M}\sum_{i=1}^{\omega M} x_i$$

which corresponds to the average of the first ωM elements in vector x.
For ϕ = 1 we have

$$f_{1,\omega}(x) = \frac{1}{\omega M}\sum_{i=1+(1-\omega)M}^{M} x_i$$

which corresponds to the average of the last ωM *e*lements in vector x.
For ϕ = 0.5 we have

$$f_{0.5,\omega}(x) = \frac{1}{\omega M}\sum_{i=1+0.5(1-\omega)M}^{\omega M+0.5(1-\omega)M} x_i = \frac{1}{\omega N}\sum_{i=1+(0.5M-\omega M)}^{0.5M+0.5\omega M} x_i$$

which corresponds to the average of the middle ωM elements of vector x.

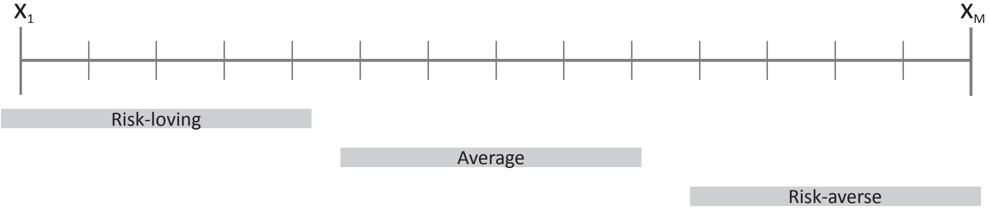
We can view these cases in a more practical, human way:

- A person with ϕ = 0, would be the ultimate optimist (risk seeking), who only takes the best possible outcomes into account and does not care about any scenario that would result in a worse outcome.

- A person with ϕ = 1 would be the ultimate pessimist or risk-averse person, whose decisions are guided by worst things that may possibly happen
- A person with ϕ = 0.5 bases his or her decision on the more usual outcomes, ignoring the real extreme cases (good or bad) cases.

This is illustrated in the figure below, where we assume $\omega = 0.3$ and $M = 15$. Note that the three person types all take the average of $\omega M = 5$ observations[1]. However, the non risk-averse ORC averages the six best outcomes while the risk-averse person averages the six worst outcomes. The average person takes some observations in between while ignoring the extreme outcomes on both sides.

FIGURE 1. Averaging outcomes



The calculation of the cost of a scenario requires the determination of the optimal schedule for the corresponding deterministic offline instance. As the number of emergency cases is typically small, as is the number of ORs accepting emergency cases, the instances in our simulation model allow us to find this optimal solution quickly by enumeration.

Simulation model

General description

We simulate the execution and modification of the OR planning at SFG. By execution, we mean the starting and ending of cases while by modification, we mean adaptation of the schedule in response to newly available information and in view of the predefined objective function. Decisions about modifications are made with the heuristic described before. First, we will outline the general structure of our simulation and describe the most important concepts. Then we will turn to the process of making decisions and changing the schedule.

[1] In the preceding text, we have assumed that all values are integral. In our implementation we first calculate β*M and round this down; also, the limits for the summation are rounded down

Next, we will perform a sensitivity analysis of the most important parameters using test data from the SFG. Finally, we use the simulation environment to evaluate the impact of the "personality of the ORC".

Description of the model

We simulate separate, independent working days using discrete event simulation: the system is modeled by means of a chronologically ordered discrete set of events. As these events are processed one at a time, the state of the system changes and new events may be generated. The simulation starts at 8:00 AM and ends when all regular ORs have completed their final case. Because we compare the simulation results with real life day per day data from the SFG, we have chosen not to consider interdependencies between working days, e.g. by rescheduling canceled cases the next day. The simulation is based on a 3-month period in the year 2009. Total number of surgical cases in this period amounts to 3,027 of which 301 emergency cases and 39 acute cases. The number of ORs is ten. For every surgical case we know the scheduled and actual case duration, scheduled and actual start- and end time, whether the case is elective, urgent or acute and the scheduled and actual operating room where the case is performed. The numbers of ORs and durations of workday are optimal for what we are scheduling. Holidays and weekends are excluded from the data. Based on the data, all relevant events on the days of surgery and the adjustments can be simulated and the outcomes can be compared to the historical outcomes.

For each treatment, we have estimated the parameters of the lognormal distribution that can be used to estimate the case duration. All electives cases were known at 8 AM, the beginning of the working day. For emergency arrivals, we do not exactly know the time at which they arrived. We will assume the following about their arrival:

- Around 50% of the emergency cases arrive between the end of the previous day and 08.00 AM (SFG, 2010). These emergency cases are considered at the start of the day. The remaining emergency cases arrive at a random time between 8 AM and 4 PM (see below for details).
- The simulation uses historical urgent and acute cases.
- Urgent and acute cases have to be started within 240 and 30 minutes after arrival, respectively.
- The subset *OE* of operating rooms to which emergency cases can be assigned may vary per day.
- The number and hours of emergency cases were built into the staffing planned.

To generate random urgent and acute case arrivals between 08.00 and 16.00 for the scenarios, we have collected data about the arrivals of emergency cases in 2008. We assume that the time of day arrivals occur according to a non-homogeneous Poisson process with a piecewise constant arrival rate. The arrival rates are estimated using the mean number of arrivals per 30 minutes time interval. For each random arrival, we sample a random emergency case from the historical dataset. The state of the system at a certain time of the day consists of the status of the planning: the starting and ending times of all cases that have been completed, the starting times and expected duration of the cases that are being performed at that moment, the ordered lists of cases scheduled for future execution in each of the rooms, the list of cases that will be performed in the service OR and finally the list of cases that have been canceled and the cases that have not yet been assigned to any room.

In each room, we start the first case at 8:00 AM. When a case starts, the corresponding 'finish event' is generated using the historic duration of the case (so that we can compare our outcomes with historic data). Of course place heuristics don't use this generated duration, but work with the parameters of the distribution of the duration of cases of that type. After a case has finished, 9 minutes are scheduled for cleaning time. After cleaning, the next case assigned to the room starts as soon as possible (if there is one). Cases cannot start more than 60 minutes earlier than scheduled.

Decision making

For reasons of computation times, we have limited the frequency by which rescheduling is considered. A first rescheduling occurrence is at 8 AM when the newly arrived cases are considered, possibly leading to modifications of the original schedule. During the day we consider rescheduling whenever a case finishes with a ending time that differs 15 minutes or more from the scheduled ending time. Rescheduling is also considered when a new emergency case arrives, and at 16.00, the scheduled closing times of the ORs. Finally, rescheduling is considered at least every 60 minutes.

Rescheduling must take the following rules into account:

- The sequence of elective cases within an operating room is fixed and cannot be changed during the day.
- When an emergency/acute case arrives, it is placed in the series 'non-scheduled'. There is no room assigned to this specific case.
- If before 4 PM there is OR capacity available in a room then the next scheduled elective case or urgent/acute case is started.
- Scheduled cases can be moved from the originally assigned room to the service OR or can be canceled.
- Cases that are not yet assigned to any room, can be assigned to a room or to the service (so that they are performed after 4 PM).

- Canceled cases or cases moved to the service OR cannot be scheduled again in the day schedule (before 4PM).
- Cases cannot be paused or stopped once they have started

The rescheduling heuristic uses Monte Carlo optimization as follows. It starts by generating a set of scenarios. A scenario consists of a random realization for the duration of each of the remaining cases including a set of randomly generated emergency cases still to arrive. For each scenario all assignments of future arrivals to ORs are enumerated. These assignments decisions are complemented by optimal decisions regarding cancelation of elective cases and rescheduling of elective cases in the service OR. Optimality refers is regarded here with respect to the aforementioned cost function which serves as objective function. The cost of a scenario are set to be equal to the minimum costs - over all assignments for the emergency cases generated in the scenario, - of the thus created optimal schedule per assignment. The rationale behind using the cost of minimum cost schedules for optimal assignments, is that this coincides with the scheduling objectives taken into account during the day.

We now turn to specifying the coefficients of the objective function. As SFG aims to avoid cancelation of cases at all costs, we set the corresponding parameter at plus infinity, i.e. α = 1,000,000. In order to find suitable values for the weights β and γ, we have presented actual ORCs with several dilemmas in which there is a choice between an amount of overtime and another amount of service time (see appendix). Based on the choices made by the ORCs, we set $\beta = 1$ and $\gamma = 2$ i.e. one minute of overtime is twice as costly as one minute of work in the service room. Finally, we have set δ = 1,000,000 (plus infinity) to value late emergencies as the response times are legal obligations.

Experiments and results

The simulation model described in the previous section has been implemented in Microsoft Visual C++ (version 9.0, 2008 Express Edition). A number of parameter values can be set by the user before starting the actual simulation. In this section we will describe the experiments that we have done and the results we have found. Unless indicated otherwise, we will assume the default settings of Table 1.

TABLE 1. DEFAULT SETTINGS

Service period	From 4 PM until 8 PM
ORC policy	$\phi = 0.5$, $\omega = 0.4$
Weights	$\alpha = 1{,}000{,}000$ $\beta = 2$ $\gamma = 1$ $\delta = 1{,}000{,}000$
Scenarios	30

In our experiments, we have considered 30 scenarios while evaluating each possible decision. The choice of 30 scenarios is based on the fact that in real life a rational choice takes into account the cognitive limitations of both knowledge and cognitive capacity of the human being [34]. It is interesting to find out the effects between the different risk attitudes when we assume that human capacity will have a hard time analyzing a large number of scenarios. Therefore in our comparison of simulation results with the historic outcomes, we will use a simulation with 50 scenarios. We now first present the results (*based on 50 scenarios*) in comparison to historical data.

TABLE 2. COMPARISON BETWEEN RESULTS SIMULATOR AND HISTORICAL DATA

	Non risk-averse policy	Mean policy	Risk-averse policy	Historical results
Rejected cases	24	27	30	25
Overtime (minutes)	5,238	4,060	4,745	2,291
Service time (minutes)	9,121	10,964	11,269	12,871
Value objective function	24,019,597	27,019,084	30,020,759	25,017,453

The last column gives the historical results. The three preceding columns give the results for various choices of the risk aversion parameter ϕ. The first column are the result for $\phi = 0$, the most risk seeking variant. The next columns use $\phi = 0.5$, and $\phi = 1$, the most risk averse variant. The simulation results show that the process of cancelation works realistically. At the same time, it reveals that the preferences regarding overtime versus referring to service OR may work differently in practice, than stated by the ORCs in the presented dilemmas. In the remainder we continue nevertheless on the basis of the stated preferences, and further analyze the role of risk aversion and other parameter settings.

Influence of ORC risk aversion on the planning

The modeling of risk aversion is especially interesting as it models the effect of variations of risk attitude between ORCs. As was discussed in the first chapter of this thesis, the psychological profile of the ORC will influence the adjustments made and thus the actual planning performance. Figure 2 compares the results of a risk minded heuristic ($\phi = 0$) with a risk averse heuristic ($\phi = 1$). The risk minded heuristics results in less service time, less cancelations, and a better objective function value. It does however generate more overtime.

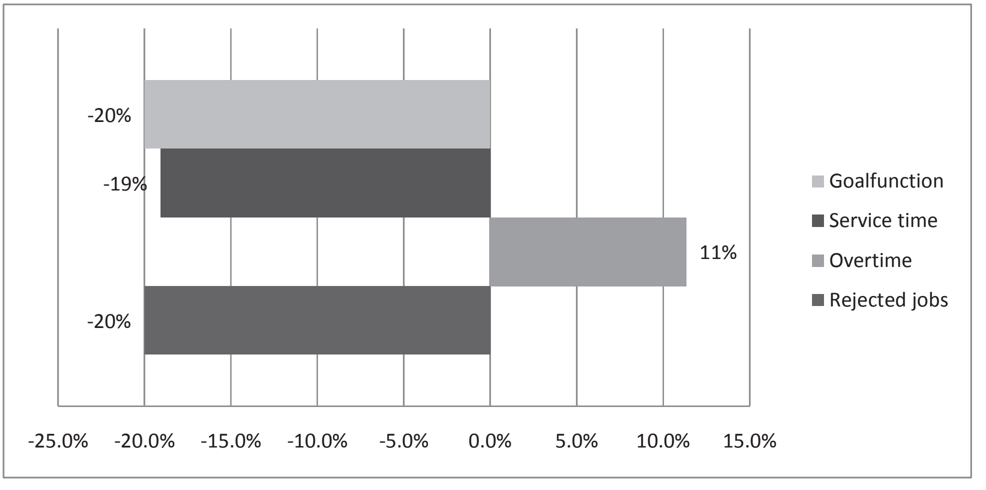


FIGURE 2. EFFECT OF NON RISK AVERSE POLICY AS COMPARED TO RISK AVERSE POLICY, BASED ON 50 SCENARIOS

Figure 3 to 6 more generally analyze how each of the 4 objective function components varies in value with ϕ.

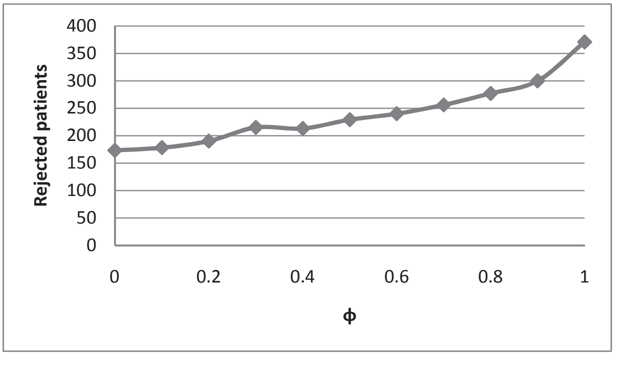


FIGURE 3. RISK AVERSION VERSUS REJECTED PATIENTS

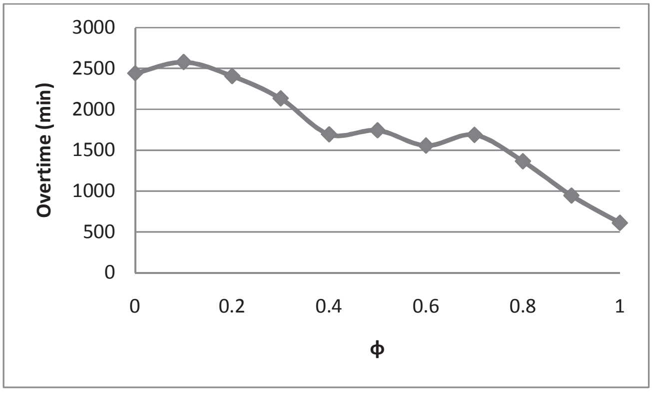


FIGURE 4. RISK AVERSION AND OVERTIME

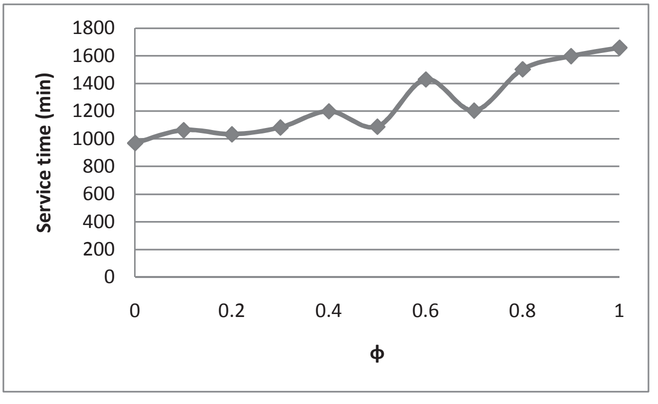


FIGURE 5. RISK AVERSION AND SERVICE TIME

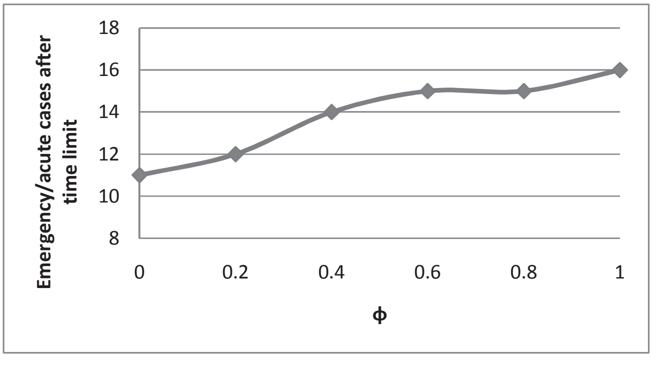


FIGURE 6. RISK AVERSION AND NUMBER OF EMERGENCY CASES AFTER TIME LIMIT

We clearly see that risk aversion leads to an increase in the number of cancelations, increase in service time and decrease in overtime. A risk-averse person focuses on the worst scenarios (which may include a larger number of emergency case arrivals or long case durations). Since service time is limited, the presumption of an increased workload will lead to more cancellations. The number of emergency/acute cases performed after 240/30 minutes respectively after arrival increases slightly as risk aversion increases. In the following presented outcomes we find that the number of performed emergency/acute cases after the imposed time limit varies between 11-16 cases. Because of this relatively low variation these results are omitted in the following tables.

END Service time

Next we consider the effect of postponing the end time of the service OR and analyze the effect on overtime, service time and rejected patients. Note that in reality the task of the service staff is primarily to operate urgent/acute cases in the period after 4PM till the next day (8AM). It is therefore undesirable to defer a set of elective cases to the service OR which together yield a high work load, as this effectively blocks the capacity for the primary purpose of treating emergency cases. As the length of the service time interval to which cases can be deferred increases, there are less cancellations, and less overtime for regular ORs. (see Figures 7,8, 9).

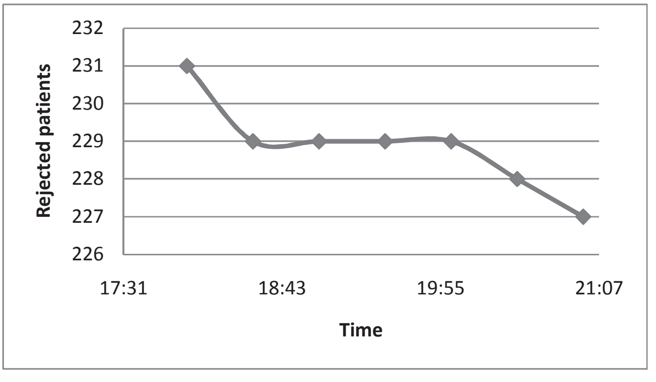


FIGURE 7. END SERVICE TIME VERSUS REJECTED PATIENTS

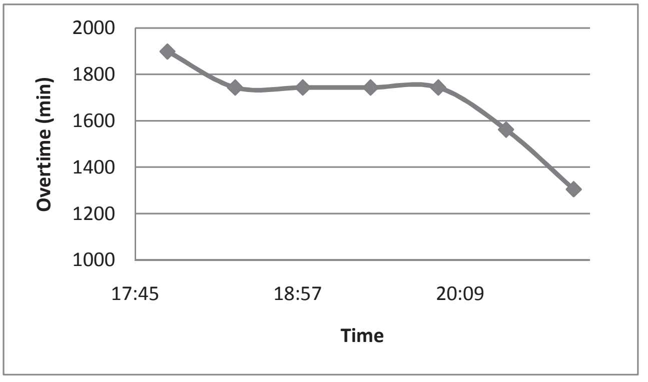


FIGURE 8. END SERVICE TIME VERSUS OVERTIME

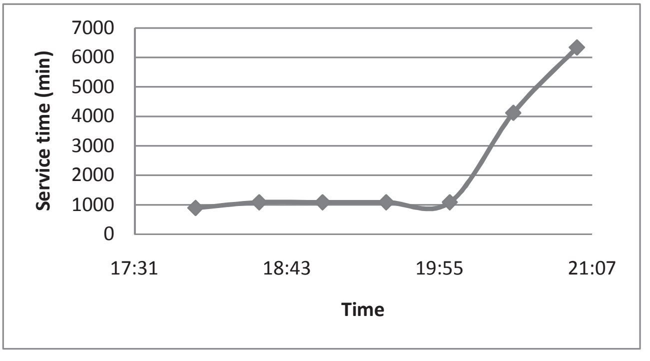


FIGURE 9. END SERVICE TIME VERSUS SERVICE TIME

Weight factors

We now turn to the sensitivity of the outcomes of the heuristics to the values of the objective function parameters (based on 30 scenarios). If we lower α, the cost of cancelation, while keeping all other parameters at their default values, we find the results shown in Table 3. We find that a cancelation cost of 1,000 suffices to ensure that overtime and service time are always preferred above case cancelation. As we lower α further, we see an increase of rejected cases and decrease in overtime and service time.

TABLE 3: EFFECT OF DECREASING COSTS REJECTION PATIENTS

α	Rejected patients	Overtime (minutes)	Service time (minutes)
1,000,000	32	4,460	10,574
100,000	38	4,590	10,347
10,000	31	4,652	10,797
1,000	31	4,652	10,797
100	77	2,810	7,010
10	350	728	1,154

Further, if we increase β, the cost of overtime, and keep all other weights at their default values then we see first more rejected cases. The second effect is that overtime decreases while service time increases.

TABLE 3: Effect of increasing costs of overtime

β	Rejected cases	Overtime (minutes)	Service time (minutes)
1	23	7,195	5,958
2	32	4,460	10,574
3	48	4,286	12,537
4	48	3,701	13,390
5	56	3,864	13,626
10	54	3,796	15,158
20	65	4,275	15,350
50	60	4,434	15,146
100	65	4,154	16,208

Number of scenarios

Finally, we test how many scenarios should be considered by the heuristic when evaluating possible planning adjustments. This can be controlled in two ways: First by changing the fraction ω of these scenarios that will be used to calculate an average score for each action. Second, by changing the total number of scenario's that should be evaluated for each action. First we have adjusted factor ω while keeping the number of evaluated scenarios constant at 30 and using a value of ϕ = 0.5. In Figure 10, we see an effect on the number of cancelations: as we consider a broader set of outcomes (up to ω = 0.7), the number of rejected cases decreases and then increases. Considering more information, i.e. taking more scenarios into account, apparently leads to better scheduling decisions. In Figure 13 we see the effect of ω on the objective function. As ω increases to a value of 0.7 the value of the objective function first decreases and subsequently increases.

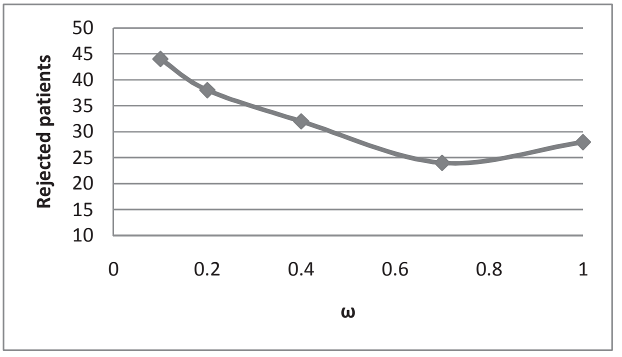


FIGURE 10. RELATION OMEGA AND REJECTED PATIENTS

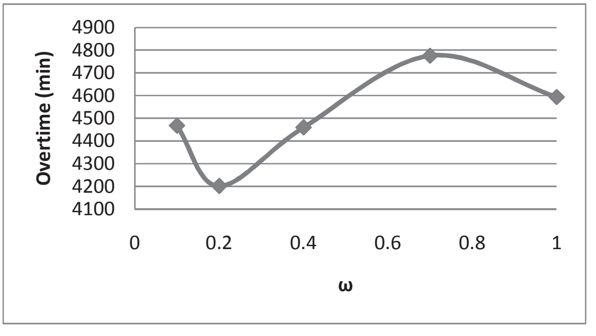


FIGURE 11. EFFECT OMEGA ON OVERTIME

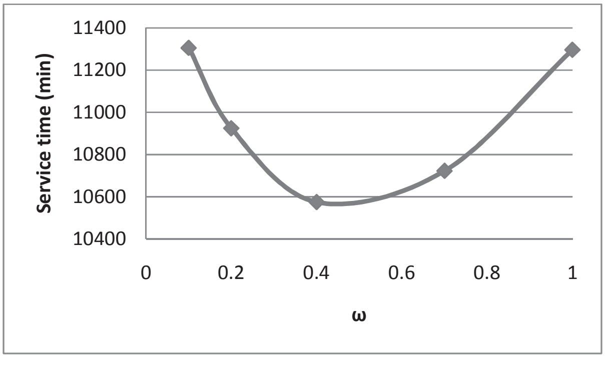


FIGURE 12: EFFECT OMEGA ON SERVICE TIME

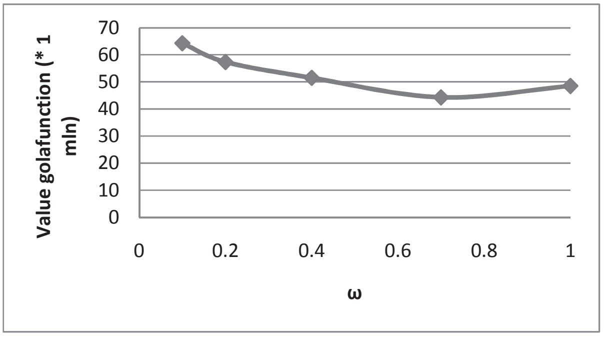


FIGURE 13. EFFECT OF OMEGA ON OBJECTIVE FUNCTION VALUE

Also, we have kept ω fixed while changing the total number of scenarios. We have done this for three types of ORCs, having values ϕ = {0.0, 0.5, 1.0}, (a non risk-averse, mean and risk-averse planner, respectively). Referring to the objective function (Figure 14) we see that the various risk averseness levels generate schedules with converging costs, as the number of scenarios considered by the heuristic increases. In particular we notice that the quality of the solutions of the risk averse variant improves considerably.

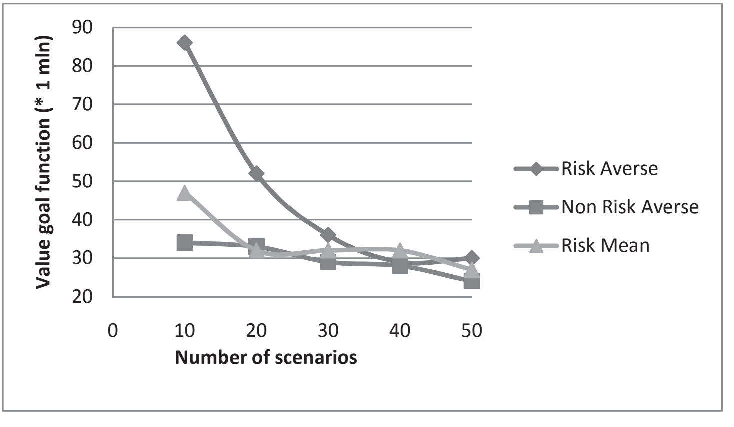


FIGURE 14. EFFECT OF NUMBER SCENARIOS ON OBJECTIVE FUNCTION VALUE

Conclusion

In this study we modeled the dynamics the ORC faces daily and especially how risk averseness influences the quality of the scheduling decisions. This study confirms earlier results in chapter 5 that a non risk averse ORC creates lower costs and less rejected patients as compared to a risk-averse ORC and an higher utilization during working days. The results of this study supports ORCs and management in their daily scheduling decision making without the usual psychological effects in decision making involved. We may also conclude that the simulations results fit the historical performance quite well, but not perfectly. As the results of the risk minded heuristic are better than the historical results, we have indications that the heuristics may yield better results than presently obtained in practice.

Further research

In view of the potential effects in quality and cost of care, risk attitudes of ORCs need explicit attention in operations room management. Moreover, the fact that the risk minded heuristics may outperform existing practice, strongly suggests to consider their potential in more detail and in other settings. Indeed, the fact that our study population (1 hospital, 10 ORs, four ORCs) is relatively small is a first limitation. We strongly suggest repeating the study in other hospitals and further improvement of the heuristics in the process. A second limitation is to consider only the sensation seeking axis of psychological characteristics. Including the other axes, neuroticism-anxiety, aggression-hostility, activity and sociability can be expected to generate further valuable insight in the performance of ORCs and be valuable for improving scheduling heuristics. More generally, improvement of the heuristics is an interesting direction for further research. This will certainly benefit from theoretical work on fundamental properties of the scheduling models and solution methods.

References

1. Institute of Medicine & Committee on Quality health care in America. Crossing the Quality Chasm: A new health system for the 21st Century. Washington DC, *National Academic Press*, 2001.
2. Stepaniak PS, Mannaerts GHH, Quelerij M de, Vries G de. The effect of the operating room coordinators risk appreciation on operating room efficiency. *Anesth Analg* 2009;108:1249-56.
3. Jiju A, Frenie Jiju A, Kumar M, Byung Rae Cho. Six sigma in service organisations: Benefits, challenges and difficulties, common myths, empirical observations and success factors. *Int J Qual Rel Manage* 2007;24:294-311.
4. Tarazi EM, Philip BK. Friendliness of or staff is top determinant of patient satisfaction with outpatient Surgery. *Am J Anesth* 1998;25:154-157.
5. Stachota P, Normandin P. Reasons registered nurses leave or change employment status. *J Nurs Admin* 2003;33:111-18.
6. Thomson TP, Brown H. Turnover of licensed nurses in skilled nursing facilities. Nurs Econ 2002;20:66-9.
7. Dexter F, Macario A, Traub RD, Hopwood M and Lubarsky DA. An operating room scheduling strategy to maximize the use of operating room block time: Computer simulation of patient scheduling and survey of patients' preferences for surgical waiting time. *Anesth Analg* 1999;89:7-20.
8. Charnetski JR. Scheduling operating room surgical procedures with early and late completion penalty costs. *J Oper Manag* 1984;1:91-102.
9. Murray CV. Creating an optimal operating room schedule. *AORN Journal*, March, 2005.
10. Dexter F, Macario A, Traub R. Which algorithm for scheduling add-on elective cases maximizes operating room utilization? Use of bin packing algorithms and fuzzy constraints in operating room management. *Anesthesiology* 1999;5:1491-500.
11. Megow N, Uetz M, Vredeveld T. Models and algorithms for stochastic online scheduling. *Math Oper Res* 2006;31:513-25.
12. Möhring RH, Radermacher FJ, Weiss G. Stochastic scheduling problems: General strategies. *Z Oper Res* 1984;28:193-260
13. Möhring RH, Schulz AS, Uetz M. Approximation in stochastic scheduling: the power of LP-based priority policies. *J ACM* 1999;46:924-42.
14. Fleischer R, Wahl M. Online scheduling revisited. *J Sched* 2000;3:343-53.
15. Rudin JF. Improved bounds for the on-line scheduling problem. Ph.D thesis, The University of Texas at Dallas, 2001.
16. Sgall J. On-line scheduling: a survey. Online Algorithms: The State of the Art. *Comp Sci* 1998;1442:196–231.
17. Karger DR, Phillips SJ, Torng E. A better algorithm for an ancient scheduling problem. *J Algorithm* 1996;20:400–30.
18. Encyclopedia of Optimization. Panos M. Pardalos Kluwer. Academic Publishers Group 2001.
19. Cao Z, Wang Z, Zhang Y. and Liu S. On several scheduling problems with rejection or discretely compressible processing times. Lecture Notes in Computer Science, Springer-Verlag, Beijing, China, 2006:90-98.
20. Cao Z, Zhang Y. Scheduling with rejection and non-identical job arrivals. *J Syst Science Complex* 2007;20:529–35.
21. Epstein L, Noga J, Woeginger GJ. On-line scheduling of unit time cases with rejection: Minimizing the total completion time. *Oper Res Let* 2002;30:415-420.

22. Cao Z. Xiaoguang Yang A. PTAS for parallel batch scheduling with rejection and dynamic job arrivals. *Theor Comput Sci* 2009;410:27-9.
23. Engels DW, Karger DR, Kolliopoulos SG, Sengupta S, Uma RN, Wein J. Techniques for scheduling with rejection. *J Algorithm* 2003;49:175-91.
24. Fenton LF. The sum of log-normal probability distributions in scatter transmission systems, *IRE Trans Comm Syst* 1960;8;57-67.
25. Barakat R. Sums of independent lognormally distributed random variables, *J Opt Soc Am* 1976;66:211-16.
26. Crow EL, Shimizu K. Lognormal distributions: Theory and application, Dekker, New York, 1988.
27. Jingxian Wu, Neelesh Mehta, Jin Zhang. A flexible lognormal sum approximation method. *Mitsubishi electric research laboratories*, Inc 2005.
28. Schwartz S, Yeh Y. On the distribution function and moments of power sums with lognormal components. *Bell System Technical Journal* 1982;61:1441-62.
29. Beaulieu NC, Xie Q. An optimal lognormal approximation to lognormal sum distributions, *IEEE Vehicular Technology* 2004;53:479–89.
30. Slimane SB. Bounds on the distribution of a sum of independent lognormal random variables, *IEEE Transactions on Communications* 2001;49:975–78.
31. Schleher DC. Generalized gram-charlier series with application to the sum of log-normal variates, *IEEE Transactions on Information Theory* 1977:275–80.
32. Stüber GL. Principles of mobile communications. *Kluwer Academic Publishers*, 1996.
33. Beaulieu NC, Abu-Dayya A, Mclance P. Estimating the distribution of a sum of independent lognormal random variables. *IEEE Transactions on information theory* 1995;43:2869-73.
34. Simon Herbert. Bounded Rationality and Organizational Learning. *Organ Sci* 1991;2:125-34.

Appendix

Questionnaire ORCs:

- If there are at 3.30 PM two cases to perform, of which one has a scheduled duration of 45 minutes and the other a scheduled duration of 60 minutes, which one (or both) of these cases are moved to the service OR?
- Would you rather start a very important case with a scheduled duration of 140 minutes in the scheduled OR at 2.50 PM or at 4 PM in the service OR? And would you start the same case at 2.20 PM in the scheduled OR or at 4 PM in the service OR?
- Would you prefer to perform a case with a scheduled duration of 90 minutes in the service OR, or would you rather schedule this very same case in a operating room with only 60 minutes of capacity left? And what if only 45 minutes of capacity would be left?

7

Summary, general conclusions and future perspectives

Introduction

When looking at an OR in an era in which both cost-containment and quality of health care are considered of prime importance, hospitals have to utilize ORs effectively and efficiently. As an OR manager at St. Franciscus Hospital, Rotterdam, I fully realized what a valuable resource OR capacity is, particularly when subject to high demand for care. My experiences and impressions motivated me to start studying how to control the enormous variation in activities in the OR. I started by looking at the variations in case durations, surgical processes, and scheduling processes. As this thesis will demonstrate, a fundamental understanding of the variation and proper control in the operating room makes it possible to improve its efficiency and effectiveness, and therefore also improve the quality of care provided to the patients. The results of the studies described in this thesis prove that it is possible to run the OR more efficiently, effectively and in a more patient-centered way.

In its influential 'Crossing the quality chasm', the Institute of Medicine [1] identifies six quality dimensions of health care, among which are efficiency, effective and timeliness. The six dimensions together, make quality improvement a complex matter.

Donabedian's structure-process-outcome model [2] is generally used as the basis for much of the work addressing quality and outcomes. Donabedian framed the concept of quality assurance in terms of three types of measures: structure (what do we need to have to be able to achieve quality), process (what do we need to do to achieve quality), and outcomes (what do we need to achieve). Donabedian suggests that each dimension can be judged independently or in conjunction. Furthermore, he argues that the outcome will be positive if structure and process are adequate.

Surgical delay has been shown to be an important determinant of patient satisfaction across the continuum of preoperative-operative-postoperative care [3]. Delays in scheduled surgical cases affect patient satisfaction even more than the intraoperative anesthesia experience [4]. Delays in surgery resulting from cancellations, bumping of cases and poor scheduling can have a significant impact on quality of care for scheduled cases as well [5]. Delays only add to the patient's inherent anxiety associated with surgery and engenders anger and frustration. The operating room, by its very nature, is an extremely stressful, uncertain, dynamic, and demanding environment where staff members need to manage multiple highly technical tasks, often simultaneously [6,7]. The Joint Commission on the Accreditation of Healthcare Organizations has identified time pressures to start or complete the procedure as one of four contributing factors to increased wrong site surgery [8]. Similar to other professions, the undue pressures of time that result from falling behind create stress that can lead to cutting corners or inadvertent error. Relative to other hospital settings, errors in the operating room can be catastrophic (i.e. wrong site surgery, retained foreign body, unchecked blood transfusions). In some cases these errors can result in high-profile consequences for the patient, surgeon or hospital [9]. In other words, poor scheduling and the subsequent induced variation in processes reduces outcome. Evidence indicates that case scheduling in practice often is performed poorly [10,11]. Additionally, methods which improve the reliable estimate of surgical cases naturally lead to improved timeliness, efficiency, and effectiveness of OR processes [12,13,14,15,16].

Reasoning along these lines, W. Edwards Deming concluded that the real enemy of quality is variation in processes. A main objective in operations management is therefore to identify sources of variation [17]. Though variation exists in every process and always will, controlling the identified variation enables to improve the health service delivery processes [18]. This thesis presents various results regarding variation at the OR. Some of the results are of a descriptive nature, and provide a better understanding of the nature of the variation, thus facilitating better control. Other results regard OR scheduling, that is one of the processes by which OR managers control variation. This thesis shows how risk attitudes are related to effective control of variation, and present mathematical optimization models exploring these findings, to provide decisions support systems to optimally control variation.

Conclusions

The main findings of the thesis are the following

In *Chapter 2* the use of accurate statistical models to predict surgical and procedure times is studied. The study is based on a complete set of surgical cases of two large European teaching hospitals in the period 2005-2008, involving 85,312 cases and 92,099 hours in total. The conclusion of this study is that the percentage of cases fitting the normal, two- and three-parameter lognormal models is higher for surgical time than for total procedure time. The evidence supports the idea that type of surgery is the most important single source of variability amongst surgeries. Using the bisection method and applying the three-parameter lognormal model fits procedure time and surgical time better than the two-parameter lognormal model without shift parameter. The posterior distribution is suitable for predicting case durations as helpful in managing variation of cases while making scheduling decisions. More specifically we find that compared to the standard way of case scheduling encountered in the teaching hospitals under consideration, use of the mean of the three-parameter lognormal distribution for case scheduling reduces the mean over-reserving OR time per case up to 53.1% and the under-reserving OR time up to with 55.6%. Using the three parameter lognormal model for case scheduling causes a lower mean over-utilized OR time up to 20.0 (19.7-20.3) minutes per OR per day as compared to the standard method and 11.6 (11.3-12.0) minutes per OR per day as compared to the bias-corrected scheduled OR time.

In *Chapter 3* the possible dependence of procedure times on surgeon factors like age, experience, gender, and team composition is investigated. The effect of these factors is estimated for over 30 different types of medical operations in two hospitals, by means of ANOVA models for logarithmic case durations. Depending on the type of operation (CPT) and on the hospital, procedure times may depend on several surgeon factors. In particular, for complex operations, factors like relevant work rate experience of the surgeon and composition of the surgical team may have large effects. Team composition explains up to 20% of variation, and when combined with work rate, even 30%. Other relevant factors are age of the surgeon and time of the day. Gender has nearly never any effect, and the only effect that is significant (at the 5% level) is found for cataract, where female surgeons work 8% faster than male surgeons. A predictive out-of-sample analysis for case durations in 2009 shows that surgeon factors help in predicting case durations. As compared to the methodology currently employed in both hospitals under study, mean absolute prediction errors are reduced by up to 18 minutes and up to 18% of the median procedure time.

The most significant gains are obtained for relatively complex CPTs, especially those involving endoscopic and laparoscopic procedures. As the complexity of surgical procedures shows an ever increasing trend, surgeon factors may become even more important in the future. The practical implementation of (ANOVA or other) prediction models is done best after consultation of surgeons, OR management, and other staff involved in the operation room activities. As hospitals differ widely in aspects like surgical experience with different specializations, organizational structure, OR protocols and OR logistics, the effect of surgeon factors will differ among hospitals. As a more general conclusion, we also note that surgeons are a major source of variation, which is presently often uncontrolled, but not necessarily entirely uncontrollable. Standardization of skills and processes, and recruitment policies might aid to controlling it.

In *Chapter 4,* we study the impact of scheduling cases of a same type consecutively on the turnover-, surgical-and procedure time. We find that maintaining a fixed team for similar consecutive cases throughout the day yields a significant reduction in preparation time and turnover time. Teams prepared the procedures in a more structured fashion in the study group. This explains the shorter preparation time in the study group as compared to the control group. Surgery time was not significantly different in the study group as compared to the control group. Surgeons do not work “faster or slower” when working on consecutive similar cases and surgeons do not compromise on quality of care to increase speed. For the inguinal hernia repair we see a significantly shorter preparation,- and procedure times in the study group as compared to the control group. Also, the variation in the study group of the three time intervals is significantly lower in comparison to the control group. The average procedure time is 10 minutes less which has practical implications for planning purposes. A reason for the decreased operative time (because of the decrease in preparation time) may be the effect of the roles of each individual team member being explicitly defined before the start of the day.

In the study group a significantly lower mean preparation time is found for the laparoscopic cholecystectomy The mean procedure time for the laparoscopic cholecystectomy is not significantly lower in the study group. A possible technical explanation for the fact that no difference is found in the mean procedure time between the control group and study group is that in both study group and control group patients were included who experienced a cholecystis or an obstruction necessitating an endoscopic retrograde cholangiopancreatography. Both problems may involve a technically demanding operation that may require more dissection time. Based on the results we conclude that scheduling cases of a same type decreases preparation time. We didn't find significant effects on procedure time, which might be due to the fact that we considered medically complex procedures. Whether procedure time is reduced for less complex procedures is a topic worthy of future research.

Chapter 5 focuses on the Operating Room Coordinator (ORC). The ORC is responsible for adjusting operating room schedules in response to the deviations from the planning that occur during the day. We have observed differences among the personalities of the four ORCs with regard to their willingness to take on more risk concerning their daily planning. We investigated the relationship between the risk attitude and the acceptance/cancellation of cases. Zuckerman developed the Zuckerman-Kuhlman Personality Questionnaire (ZKPQ) to assess personality along five dimensions. Zuckerman defines sensation seeking as a need for new and complex experiences and a willingness to take risk for one's own account. He has found that high sensation seekers tend to anticipate lower risk than low sensation seekers do, even for new activities.

This finding indicates that a high sensation seeker is more likely to look for opportunities that provide the chance to take a risk, and that the willingness to take risks seems less threatening to this specific type of individual. The ZKPQ results for risk-seeking indicate that there is a difference in risk appreciation between the different ORCs. Risk-averse ORCs plans in less cases than non risk-averse ORCs. The number of extra cases performed by the non risk-averse ORC as compared to a risk-averse ORC is 188 in 2006 and 174 in 2007. The average end-of-program-time per OR/day for the non risk-averse ORC is 34 minutes (± 19 min, $p = 0.0085$) later than for the risk-averse ORC. In our study a non risk-averse ORC creates significantly less unused OR capacity without a great chance of running

ORs after regular working hours or canceling elective cases. Added to this, a non risk-averse ORC is cost-effective.

In *Chapter 6* we simultaneously address efficiency and timeliness of care in the operating theatre. In this study we modeled the dynamics the ORC faces daily and especially how risk averseness influences the quality of the scheduling decisions. This study confirms earlier results in chapter 5 that a non risk averse ORC creates lower costs and less rejected patients as compared to a risk-averse ORC and an higher utilization during working days. The results of this study supports ORCs and management in their daily scheduling decision making without the usual psychological effects in decision making involved. We may also conclude that the simulations results fit the historical performance quite well, but not perfectly. As the results of the risk minded heuristic are better than the historical results, we have indications that the heuristics may yield better results than presently obtained in practice.

Future perspectives

Practical perspectives

The results of this study can be straightforwardly implemented in any OR decision support system. Based on my experience the main condition to implement successfully is a coherent view of OR staff, surgeons, anesthetists towards helping patients in a timely, efficient and effective manner. It enables to make scheduling decisions in a patient centered manner, rather than to let the perspectives of OR staff, surgeons or anesthetists et cetera prevail.

Future research

The results presented in the thesis are significant to help reduce uncontrolled variation as enemy of quality [17]. Based on our findings we propose the following directions for future research to improve OR scheduling:

In the simulation for case duration prediction and efficiency gains (Chapter 2), we omitted cases with a frequency less than ten. As a result, the real efficiency gains may be over-estimated. Further study is called for regarding the estimation of low frequency cases and the resulting effect on scheduling decisions. In the study regarding the consecutive scheduling of cases of a same type, we considered relatively complex surgical procedures (Chapter 4). For these cases, the duration time of the procedure was not significantly effect. We hypothesize that the effect on case durations of relatively less complex procedures is significant and propose to test this hypothesis.

The study which addressed the risk attitudes of ORC's (Chapter 5) only takes the sensation seeking dimension out of five dimensions of ZKPQ into account. The dimensions, neuroticism-anxiety, aggression-hostility, activity and sociability, might significantly influence scheduling decision making as well. We encourage further research in this novel, multidisciplinary, direction.

Although the results from the simulation (chapter 6) are strong, there are some potential drawbacks. The first factor is that the study population (1 hospital, 10 ORs and four ORCs) is relatively small. It is therefore important to repeat this study in other hospitals and to further improve the methodology based on the new situations and data. When patients are rejected then we assume that they are not scheduled in the future. It is interesting to analyze the effect on our results when rejected patients are moved to a waiting list or the next day. We used a relative simple questionnaire to estimate the cost coefficients. A more realistic approach is using actual costs and use them in the simulation model.

Finally, our heuristics bring about a number of scheduling questions regarding the performance of heuristics, and performance bounds that are attainable. This strongly suggests more fundamental research on the complexity and nature of the real time scheduling problems considered in our study. More generally, improvement of the heuristics is an interesting direction for further research. This will certainly benefit from theoretical work on fundamental properties of the scheduling models and solution methods.

References

1. Institute of Medicine & Committee on Quality health care in America. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington DC. *National Academic Press*, 2001.
2. Donabedian A. The quality of care: how it can it be assessed? *JAMA* 1988;260:1743-48.
3. Tarazi EM, Philip BK. Friendliness of or staff is top determinant of patient satisfaction with outpatient Surgery. *Am J Anesth* 1998;25:154-57.
4. Brown DL, Warner ME, Schroeder DR and Offord KP. Effect of intraoperative anesthetic events on postoperative patient satisfaction. *Mayo Clinic Proceeding* 1997;72:20-5.
5. Litvak E. Optimizing patient flow by managing its variability. In Berman S. (ed.): Front Office to Front Line: Essential Issues for Health Care Leaders. Oakbrook Terrace, IL: *Joint Commission Resources* 2005:91-111.
6. Reason J. "Safety in the operating theatre part 2: Human error and organizational failure." *Qual Saf Health Care* 2005;14:56-60.
7. Silen-Lipponen M, Tossavainen K, Turunen H and Smith A. Potential errors and their prevention in operating room team work as experienced by Finnish, British and American nurses. *Int J Nurs Pract* 2005;11:21-32.
8. ACOG Committee Opinion: Patient safety in the surgical environment. *Obstet Gynecol* 2006;107:429-33.
9. Makary MA, Sexton BJ, Freischlag JA, Millman EA et al. Patient safety in surgery. *Ann Surg* 2006;243:628-35.
10. Litvak E, Long MC. Cost and quality under managed care: Irreconcilable differences? *Am J Manag Care* 2000;6:305-12.
11. McManus ML, Long MC, Cooper A, Mandell J, Berwick DM, Pagano M, Litvak E. Variability in surgical caseload and access to intensive care services. *Anesthesiology* 2003;98:1491-96.
12. Wickizer TM. Effect of hospital utilization review on medical expenditures in selected diagnostic areas: An exploratory study. *Am J Public Health* 1991;81:482-84.
13. Dexter F, Epstein RH, Marsh HM. A statistical analysis of weekday operating room anesthesia group staffing costs at nine independently managed surgical suites. *Anesth Analg* 2001;92:1493-98.
14. Dexter F, Traub RD. Sequencing cases in the operating room: Predicting whether one surgical case will last longer than another. *Anesth Analg* 2000;90:975-79.
15. Dexter F, Traub RD, Macario A. How to release allocated operating room time to increase efficiency: Predicting which surgical service will have the most underutilized operating room time. *Anesth Analg* 2003;96:507-12.
16. Lapierre SD, Batson C, McCaskey S. Improving on-time performance in health care organisations: case study. *Health Care Manage Sci* 1999;2:27-34.
17. Tannat G. Design for six sigma: launching new products and services without failure Gower *Publishing Limited*, 2002.
18. McLaughlin CP, Kaluzny AD. Continuous Quality Improvement in Health Care, Third Edition Center for Health Services Research, *University of North Carolina at Chapel Hill*, Chapel Hill, North Carolina, 2006.

8

Summary in Dutch

Introductie

Na een aantal jaren in het bedrijfsleven te hebben gewerkt, heb ik in 2004 mijn carrière voortgezet als manager van het operatiekamercomplex (OK) in het St. Franciscus Ziekenhuis, te Rotterdam. In het bedrijfsleven heb ik geleerd om op een efficiënte en effectieve wijze processen te organiseren vanuit het perspectief waarin de klant centraal staat. Met deze ervaring begon ik op 1 januari 2004 op de OK. Een omgeving die gekenmerkt wordt door veel kostbare apparatuur en verschillende hoog opgeleide professionals die klaar staan om patiënten te helpen. Kenmerkend voor de OK is, dat het een belangrijke schakel in de zorgketen is die economische toegevoegde waarde creëert voor het ziekenhuis.

Omdat de OK een arbeidsintensieve omgeving is waar in dit geval jaarlijks meer dan 11.000 patiënten (2004) werden geopereerd verwachtte ik dat processen zo efficiënt en effectief mogelijk waren ingericht. Hierbij ondersteund door geavanceerde planningssystemen met als doel patiënten tijdig en klantgericht te helpen. Mijn verwachting bleek niet overeen te komen met de praktijk. Hierbij een aantal voorbeelden om dit te illustreren. Omdat OK programma's niet goed waren gepland, was er veelvuldig sprake van lege OKs.

Als gevolg van de minder goede planning kwam het relatief vaak voor dat operateurs, operatie- en anesthesie medewerkers tijdens werktijd niet aan het werk waren en/of na reguliere werktijd moesten doorwerken om de operaties af te maken. De realisatie van operatietijden week soms fors af van de oorspronkelijke planning. Omdat de planning niet altijd werd gehaald ging dit vaak ten koste van reeds geplande electieve patiënten.

Het maken van een goede planning en vervolgens op het geplande tijdstip opereren van patiënten was bijna elke dag een uitdaging. Het resultaat van de niet optimale planning leidde soms tot flinke frustraties en stress bij OK personeel, operateurs, management en niet in de laatste plaats de patiënt omdat hij/zij van het programma werden gehaald. Als de OK wordt bezien in een tijd dat de kosten van de gezondheidszorg stijgen, bestuurders van ziekenhuizen stellen dat de patiënt centraal staat, de kwaliteit van zorg en de veiligheid onder een vergrootglas ligt bij de inspectie, is het niet meer dan logisch dat ziekenhuizen hun operatiekamers op een effectieve en efficiënte wijze moeten inrichten. Eind 2006 kwam ik op basis van mijn ervaringen en indrukken tot de conclusie dat de omvangrijke variaties die inherent zijn aan de planning van operaties beter konden worden gemanaged. Ik begon in 2007 met het bestuderen van de variaties die gerelateerd zijn aan operatietijden, processen rondom een operatie en de planningsprocessen.

In dit proefschrift wordt aangetoond dat een beter begrip van de variatie van operatietijden en processen en een juiste beheersbaarheid van de variatie leidt tot een verbetering van de efficiëntie, effectiviteit en tijdigheid van zorg op de OK. Daarmee verbetert de kwaliteit van zorg aan patiënten. W. Edwards Deming heeft eerder geconcludeerd dat de echte vijand van kwaliteit de variatie in processen is. Een belangrijke opdracht voor management is de factoren die variatie veroorzaken te identificeren [1]. Hoewel variatie in elk proces aanwezig is en zal blijven, leidt beheersing van geïdentificeerde variatie tot verbetering van de kwaliteit van zorgprocessen [2].

In dit proefschrift worden verschillende resultaten gepresenteerd betreffende het beheersen van variatie op de OK. Sommige resultaten zijn beschrijvend van aard, andere bieden diepgaand inzicht in de aard van variatie en wat bijdraagt om deze variatie te beheersen. Andere resultaten hebben betrekking op het plannen van operaties. In dit proefschrift wordt aangetoond hoe verschillende soorten risicohoudingen ten opzichte van elkaar staan in relatie tot het op een effectieve wijze beheersen van variatie. Wiskundige optimalisatiemodellen worden gebruikt om deze aspecten te onderzoeken met als doel een beslissingsondersteunend systeem te ontwikkelen teneinde op een optimale wijze variatie te beheersen.

Theorie

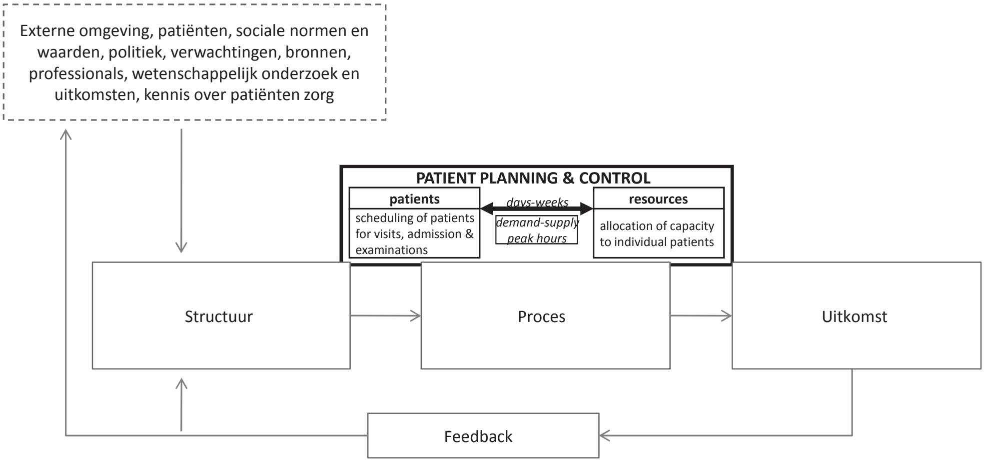
Het Institute of Medicine (IoM) onderscheidt zes doelstellingen voor kwaliteitvolle zorg. Deze doelstellingen zijn[3]:

- Veilig: vermijdbare letsels minimaliseren.
- Effectief: zorg gebaseerd op wetenschappelijke kennis aan al wie daar baat bij heeft.
- Efficiënt: vermijden van verspilling (materiaal, ideeën, energie).
- Patiënt centraal: respectvol voor en responsief aan individuele voorkeur, nood, waarden.
- Tijdig: zonder wachttijden en schadelijke vertragingen voor patiënt en zorgverlener.
- Gelijkwaardig: kwaliteitsvolle zorg los van geslacht, etniciteit, locatie, sociaal-economische status.

Donabedian[4] ontwikkelde het structuur-proces-uitkomst model waarmee vooruitgang in kwaliteit kan worden gemeten. *Structuur* is gedefinieerd als wat nodig is om in staat te zijn kwaliteit te leveren. *Proces* houdt in wat we moeten doen om kwaliteit te bereiken en *uitkomst* is wat aan kwaliteit bereikt moet worden. Het onderzoek in dit proefschrift heeft ten aanzien van IoM betrekking op de effectiviteit, efficiëntie en tijdigheid van zorg.

Voor wat betreft de organisatie en afstemming van aanbod en vraag zijn er verschillen tussen de industrie en de zorg. Vissers et al. [5] heeft voor de zorg een besturingsraamwerk ontwikkeld van vijf niveaus. Dit model wordt toegepast op de OK. In het onderzoek worden de bovenste drie niveaus als exogeen beschouwd. Op het vierde en vijfde niveau zijn de patiëntgroepen en patiënten op de OK het resultaat van een samenspel van keuzes en krachten in de bovenste drie niveaus van het besturingsmodel. Het management op de OK zal gegeven deze patiëntgroepen en patiënten de logistieke processen zodanig moeten inrichten dat er evenwicht wordt gevonden tussen flexibel georganiseerd zijn, zodat verstoringen van het proces (spoed, uitloop) kunnen worden opgevangen en de reguliere vraag naar zorg. Hierbij spelen planningregels een belangrijke rol.

Personeel op de OK voert haar werkzaamheden op soms lange dagen en onder tijdsdruk uit. De *Joint Commission on the Accreditation of Healthcare Organizations* heeft na onderzoek vastgesteld dat de tijdsdruk om te starten of een verrichting te beëindigen één van de vier factoren is die bijdraagt aan een verhoogde kans op het maken van fouten [6]. Vergelijkbaar met andere beroepen (o.a. in de luchtvaart) veroorzaken tijdsdruk en het risico niet op schema te lopen stress bij OK personeel , dit kan tot fouten resulteren. In sommige gevallen kunnen deze fouten vergaande consequenties hebben voor patiënt, chirurg of het ziekenhuis [7]. Tijdsdruk kan ontstaan als de geplande operatietijden niet juist zijn en als de onzekerheid in het niet op tijd realiseren van een OK programma de overhand krijgt. De onzekerheden in de OK schema's die ontstaat als gevolg van niet juist plannen heeft dus invloed op de kwaliteit van zorg. Een beter begrip van de variatie kan resulteren in betere schatting van OK tijden. Een betere schatting draagt vervolgens bij aan een grotere kwaliteit van zorg in termen van veiligheid, effectiviteit, efficiëntie en tijdigheid. Omdat de OK een *leading source* is [5], heeft beheersen van variatie van OK roosters en processen als tweede positief effect dat het de kwaliteit in de gehele zorgketen verbeterd. Dit kan worden bereikt door het verbeteren van de planning en control cyclus van de OK.



FIGUUR 1. SYNTHESE VAN HET VIJFDE NIVEAU UIT HET LOGISTIEK FRAMEWORK 3 EN DONABEDIAN'S MODEL VOOR KWALITEIT

Conclusies van het onderzoek

In het tweede hoofdstuk wordt beschreven hoe operatietijden beter kunnen worden geschat. Door gebruik te maken van een model dat OK-tijden beter voorspelt, is het directe resultaat een verbetering van de OK-efficiëntie en het op het afgesproken tijdstip opereren van patiënten. OK tijden worden over het algemeen gekenmerkt door een scheve verdeling en hebben een minimale tijd operatietijd. In hoofdstuk twee wordt aangetoond dat OK-tijden beter worden gemodelleerd door een 3-parameter lognormale verdeling. Daartoe is in twee Europese ziekenhuizen retrospectief onderzoek gedaan naar alle operaties (*n*= 85.312) in de periode 2005-2008. Voor verrichtingen die niet vaak voorkomen is met behulp van een combinatie van historische tijden en inschatting van operateurs gekomen tot een betere schatting van de OK-tijd. Tenslotte wordt aangetoond dat gebruik maken van deze modellen leidt tot minder over- en onderschatting van OK tijden vergeleken met een planning waar gebruik wordt gemaakt van het gemiddelde van de laatste tien waarnemingen. Tenslotte wordt aangetoond dat de OK-efficiëntie toeneemt bij gebruik van het 3-parameter lognormale model.

In het derde hoofdstuk is onderzocht of er specifieke factoren van invloed zijn op de variatie van OK-tijden. De onderzochte factoren zijn leeftijd, ervaring, en geslacht van de chirurg en teamsamenstelling. De factoren zijn geschat door gebruik te maken van variantie-analyse (ANOVA, 'Analysis of Variance') voor logaritmisch verdeelde OK-tijden. De praktische relevantie van de uitkomsten van de analyse is getoetst door over de periode

2009 de geplande en werkelijke OK-tijden met elkaar te vergelijken in 2009. In het geval van met name endoscopische verrichtingen blijken deze factoren significant als de operateur minder dan 1 keer per 3 weken opereert. De factoren die significant van invloed zijn op OK-tijden zijn: team samenstelling, ervaring en het moment op de dag dat een operatie wordt uitgevoerd. Door rekening te houden met deze significante effecten in de planning neemt de *out of sample* voorspelling van OK tijden van 1.250 operaties in 2009 toe met meer dan 15% vergeleken met de huidige wijze van plannen (laatste 10 gemiddeld). De meest significante effecten worden behaald bij relatief complexe operaties, in het bijzonder de endoscopische en laparoscopische operaties. Omdat er in de trend van complexe operaties een toename is te zien, zal voor de voorspelbaarheid van deze operatietijden het belang van de onderzochte factoren toenemen.

In het vierde hoofdstuk wordt het effect onderzocht van opereren in straatjes op de operatie- en wisseltijden. In deze opzet verrichten vaste OK-teams repeterend eenzelfde type operatie. Gekozen is voor een relatief laag complexe verrichting (Hernia Inguinalis volgens Lichtenstein) en een relatief hoog complexe verrichting (laparoscopisch cholestectomie). De veronderstelling is dat als gedurende de dag een vast OK-team in straatjesopzet dezelfde operatie uitvoert de operatietijd, snijtijd, voorbereidingstijd en wisseltijd minder is.

Het onderzoek bevestigt dat het repeteren van dezelfde taak de tijdsduur van deze taak reduceert. Het onderzoek toont aan dat bij de laag complexe verrichting de operatietijd minder is in de studiegroep dan in de controlegroep. Voor zowel de laag als hoog complexe verrichting is de voorbereidingstijd in de studie groep significant lager dan in de controle groep. Een verklaring hiervoor is dat OK-teams in een straatjesopzet meer gestructureerd de operatie voorbereiden. Voor beide operaties is geen verschil in snijtijd gevonden tussen de studie- en controlegroep. Operateurs werken niet sneller of langzamer in straatjes.
In hoofdstuk 5 wordt onderzocht wat de relatie is tussen de risico-aversiteit van een OK-coördinator en de OK-efficiëntie. De OK-coördinator is verantwoordelijk voor het vullen van gaten in een operatieprogramma zodat OKs zo veel mogelijk worden benut en minimaal uitlopen. Dit zonder dat geplande patiënten van de OK-lijst worden gehaald. Er zijn in het ziekenhuis van studie vier OK coördinatoren. Tussen deze coördinatoren zijn verschillen waargenomen in de wijze waarop zij bereid zijn om risico te nemen om geplande operaties uit te laten voeren en/of extra operaties in te plannen. De getoetste hypothese is dat er een relatie is tussen de risico-aversiteit van een OK-coördinator en de OK-efficiëntie. Het onderzoek toont aan dat een niet risico-averse OK-coördinator een hogere OK-efficiëntie creëert zonder dat dit leidt tot significant meer overwerk vergeleken met een risico-averse OK-coördinator. Tevens zegt de niet risico-averse OK-coördinator minder electief geplande patiënten af. Tenslotte wordt aangetoond dat het vanuit kosten overwegingen beter is om

een operatie waar de kans op overwerk aanwezig door te laten gaan in plaats van af te zeggen.

In hoofdstuk 6 wordt een onderzoek gepresenteerd waar de kennis uit eerdere hoofdstukken wordt samengevoegd. Het doel in dit hoofdstuk is om tegelijkertijd efficiëntie en tijdigheid van zorg in de OK te verbeteren door middel van een simulatieprogramma. Dit programma werkt op realtime basis en heeft een doelstellingsfunctie ingebouwd om de gewogen kostensom van afgezegde operaties, overwerk kosten, verschuivingen naar de dienst OK en het later dan afgesproken uitvoeren van spoedpatiënten te minimaliseren. Eerder in dit proefschrift is aangetoond dat de risicohouding van een OK-coördinator van invloed is op de kwaliteit van de OK-planning. Door gebruik te maken van heuristiek die gebaseerd is op verschillende risico houdingen kan de uitkomst van een OK-planning worden geanalyseerd. Hierbij wordt gebruik gemaakt van Monte Carlo optimalisatiemethoden die getest worden op recente (2009) gerealiseerde operaties in het St. Franciscus Gasthuis, te Rotterdam. De resultaten van de simulatie tonen aan dat een niet risico-averse houding van de OK-coördinator leidt tot minder afzeggingen van operaties, minder werk in de dienst OK, en een betere benutting overdag. De prijs van deze risico- houding is meer overwerk.
In hoofdstuk 7 volgt de samenvatting van dit proefschrift en worden de belangrijkste conclusies genoemd. Tevens wordt beschreven wat aspecten zijn die in de toekomst nader onderzoek behoeven.

Literatuur

1. Tannat G. Design for six sigma: launching new products and services without failure Gower *Publishing Limited*, 2002.
2. McLaughlin CP, Kaluzny AD. Continuous quality improvement in health care, third edition Center for health services research, *University of North Carolina at Chapel Hill*, Chapel Hill, North Carolina, 2006.
3. Institute of Medicine & Committee on Quality health care in America. Crossing the quality chasm: A new health system for the 21st Century. Washington DC. *National Academic Press*, 2001.
4. Donabedian A. Explorations in quality assessment and monitoring: The definition of quality and approaches to its assessment (Vol 1). Ann arbor, MI: *Health Administration Press*, 1980.
5. Vissers JMH, Bertrand JWM, De Vries G. A framework for production control in health care organisations. *Prod Plan Control* 2001;12:591-604.
6. ACOG Committee Opinion: Patient safety in the surgical environment. *Obstet Gynecol* 2006;107:429-33.
7. Makary MA, Sexton BJ, Freischlag JA, Millman EA et al. Patient safety in surgery. *Ann Surg* 2006;243:628-35.

TEST
GS412 TAS451
139°/39
HDG 145 MAG
15
12
-09
40
ARPT
WX+T
0
TFC
+10
TGO
EDDS
-20
SUNEG
FMC
AD2002-24-51 : fuel usage restrictions
p/n HVM8-009

Appendices

Dankwoord

Dit proefschrift heeft mede zijn inhoud en vorm gekregen mede dankzij de steun van anderen die ik graag wil bedanken.

Allereerst mijn twee promotoren Prof.dr. ir. G. de Vries en Prof.dr. J. van de Klundert.

Beste Guus. In ons eerste gesprek waarin ik aangaf te willen promoveren heb je veel wijze woorden gezegd. Zelfs na een uur op mij inpraten bleef ik vasthouden aan mijn doel om te promoveren. Dat 'vasthouden' was tijdens het traject al snel een begrip. Dank voor je betrokkenheid en tijd zowel op het inhoudelijke vlak als voor alle andere gesprekken die we hebben gevoerd. Het is een eer om je eerste promovendus te zijn.

Beste Joris. Jij bent onlosmakelijk verbonden aan mijn traject om te promoveren. Mede door jouw bijdrage heeft dit proefschrift aan diepgang en scherpte gewonnen. No guts, no glory. Dank daarvoor.

Dr. C. Heij. Beste Christiaan, wij hebben elkaar voor het eerst ontmoet tijdens de revisie van het tweede artikel. Dankzij jouw scherpe opmerkingen en enthousiasme heb ik in relatief korte tijd flink tempo in mijn onderzoek kunnen maken. De snelheid waarmee je de stukken van uitgebreide, zeer bruikbaar commentaar voorzag, overtrof op een gegeven moment (even) de snelheid waarmee ik de zoveelste versie kon maken. Dank voor al onze gesprekken en de samenwerking.

Prof.dr. Albert Wagelmans. Beste Albert. Ook zonder jou had ik niet de slag kunnen maken om te komen tot een volwaardig proefschrift. Dank voor je inzichten, tijd en humor. Als je een aanbevelingsbrief nodig hebt, dan ben ik uiteraard bereid om deze voor je te schrijven.

Ik wil de leden van de leescommissie Prof.dr. Franklin Dexter, Prof.dr. Ger Koole, en Prof.dr. Albert Wagelmans bedanken voor hun tijd en betrokkenheid bij mijn promotie.

R. van der Velden. Beste Ronald. Ik ben blij dat we samen 'ORSim®' hebben ontwikkeld. Ik schrijf 'we', maar de duizenden codes zijn door jou geschreven. Door snel inzicht te krijgen in de werkelijke OK wereld heb je een ingenieuze vertaalstag kunnen maken naar een mooi planningsprogramma. Om in jouw woorden te spreken: het is 'Super de luxe'.

Dr. G.H.H. Mannaerts. Beste Guido. Jij gaf mij in 2006 het laatste duwtje om een promotietraject in te gaan. Dank voor je support en dat je vandaag mijn paranimf bent.

M. de Quelerij. Beste Marcel. Dank voor al je inspanningen, relativeringen en opmerkingen en ook dat je vandaag mijn paranimf bent.

M. van Stolk. Beste Margaret. Wij waren en destijds het voorbeeld van succesvol duo-management in de zorg. Dank voor je vertrouwen en steun.

Fabbers van het SFG, in het bijzonder Marco Kuipers en Rien van den Oever. Bedankt voor het aanleveren van alle data op de meest onmogelijke momenten en deadlines. Marco, binnenkort samen de golffrequentie maar flink opvoeren!

Dr. P. Batenburg. Beste Piet. Ik wil je bedanken voor de kans die je mij hebt gegeven om op de OK van het SFG te werken en van je te leren. Nu zijn we al een tijd op de goede weg in het mooie Catharina Ziekenhuis.

Prof.dr. A. van Zundert. Beste André, dank dat je tijd hebt willen besteden om de vele kleine puntjes op de spreekwoordelijke *i* te zetten.

J. Maat. Beste Jeroen. Jouw kennis van Excel is ongekend. Dank voor al je hulp.

K. Jamdagni. Beste Kumar. Dank voor al je editing werk. Het is dus mogelijk om naast te promoveren ook verdieping te krijgen in een vreemde taal.

Naast de inhoudelijke ondersteuning wil ik mijn 'vlieg'-vrienden bedanken voor de check and balances Has van Helvoort, Marco Bouman, Joop Peek en Willem de Kleijnen. Een promotietraject kan worden vertaald naar een APPROACH checklist:

' We will be making the ILS approach to runway 24 in EHRD. Localizer frequency is 109.1 the course is 243°. The decision height is 251 feet, which we set in the altitude warning system.
In case of a missed approach we will climb to 2000 feet and then proceed to ROT.
I fly the approach it's my decision to Go Around or land'

J. Ouwehand en R. Jonkman. Beste Joke en Ronald. Dank voor het feit dat ik als manager met jullie heb kunnen en mogen samenwerken. Ik heb het vaker gezegd: jullie zijn de Top OK teamleiders van Nederland.

Beste Art en Louise. Bedankt voor de fysieke momenten in de sportschool. De kracht inspanningen en gezelligheid heeft vaak tot creativiteit of oplossing van problemen geleid.

Lieve Mariëlle, bedankt voor alles en op naar de volgende vink.

Curriculum vitae

Pieter Szymon Stepaniak was born on April 19th, 1967 in Breda, the Netherlands. He followed the MAVO (Maassluis), HAVO (one year) and VWO (Vlaardingen). In 1987 he entered the Erasmus University in Rotterdam studying Econometrics, Economics and received his master degree (MSc) in 1992 (Free doctoral exam). After serving his military duty he worked for ErasmusMC, PricewaterhouseCoopers and KPN. In 2004 he started as manager operating rooms in the St. Franciscus Hospital Rotterdam and continued (2007) his career in healthcare as director in the Tweesteden Hospital, Tilburg. In January 2007 he started his Ph.D studies under supervision of Prof.dr.ir. Guus de Vries, joined by Prof.dr. Joris van de Klundert in 2009. From June 2010 onwards he works as manager operating rooms in the Catharina Hospital, Eindhoven. In his private time he is a commercial pilot and private day trader at the financial stock market.

List of publications

Pieter S. Stepaniak, Christiaan Heij, Guido H. H. Mannaerts, Marcel de Quelerij, and Guus de Vries. Modeling Procedure and Surgical Times for Current Procedural Terminology-Anesthesia-Surgeon Combinations and Evaluation in Terms of Case-Duration Prediction and Operating Room Efficiency: A Multicenter Study. Anesth Analg 2009;106:1232-45.

Pieter S. Stepaniak, Christiaan Heij and Guus de Vries. Modeling and prediction of surgical procedure times. Stat Neerl 2010;64:1-18.

Pieter S. Stepaniak, Christiaan Heij, Wietske Vrijland, Marcel de Quelerij and Guus de Vries. Working with a fixed OR team through the day on consecutive similar cases and the effect on OR case duration and turnover time: By random assigning patients to study and control days. Accepted for publication, Arch Surg (24/09/2009).

Pieter S. Stepaniak, Guido H. H. Mannaerts, Marcel de Quelerij, and Guus de Vries. The Effect of the Operating Room Coordinators Risk Appreciation on Operating Room Efficiency. Anesth Analg 2009;108: 1249-56.

Pieter S. Stepaniak, Ronald van der Velden, Albert Wagelmans and Joris van de Klundert. Quality improvement: balancing the risks of overtime and cancellation of scheduled cases. Submitted for publication.


Thomson
QATAR