HEALTHCARE LOGISTICS: THE ART OF BALANCE

MARK VAN HOUDENHOVEN

Healthcare Logistics:
The Art of Balance

Zorglogistiek: De kunst van het balanceren

Proefschrift

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

Prof.dr. S.W.J. Lamberts

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

woensdag 19 december 2007 om 11:45 uur
door
Mark Van Houdenhoven
geboren te Wageningen

ERASMUS UNIVERSITEIT ROTTERDAM

Promotiecommissie

Promotor
Prof.dr. E.W. Steyerberg
Prof.dr. H.W. Tilanus

Overige leden
Prof.dr. J. Bakker
Prof.dr. J. Klein
Prof.dr. ir. J.M.H. Vissers

Copromotor
Dr. G. Kazemier

Content

Introduction to the topic and outline of this thesis

Introduction

Healthcare management is a very complex and demanding business. The processes involved – operational, tactical and strategic – are extremely divers, sophisticated, and we see medical-technological advancements following on each other's heels at breathtaking speed. And then there is the constant great pressure exerted from many sides: ever-increasing needs and demands from patients and society, thinking about organizations, growing competition, necessity to incorporate these rapidly succeeding medical-technological advancements into the organization, strict cost containment, growing demand for healthcare, and a constant tightening of budgets. These developments force healthcare managers in the individual organizations to find a balance between said developments, the feasibilities of organization in question, and the desired healthcare outcomes in an ever-changing world.

The search for individual organizational balances requires that the world of professional competencies, i.e. the clinicians, and the world of healthcare managers should speak the same language when weighing the various developments and translating the outcomes into organizational choices. For the clinicians to make the right choices they must be facilitated to appraise the effects of their choices on organizational outcomes. Likewise, the healthcare managers' decision-making process should include the effects on the medical policies pursued by the individual clinicians in the own organization.

This thesis places a focus on developing methods for allocation of hospital resources within a framework that enables clinicians and healthcare managers to balance the developments on the various levels, thus providing a basis for policy-making. The framework recognizes four hierarchical levels: strategic level, tactical level, operational off-line level, and operational on-line level. In addition, it covers four areas of planning and control: medical planning, resource capacity planning, material coordination, and financial planning. This thesis includes several examples elaborating the development of tools that within the total framework facilitate the search for an individual organizational balance.

The framework as such is a conceptual structure in which the individual actors and interested parties in a healthcare organization and society engage in dialogue, with the ensuing policy-making based on a trade-off between the various developments and interests. It thus provides for a common language that facilitates the search for the balance, steers this in the right direction, and prevents that actors will find themselves at cross-purposes and thus eventually would arrive at a suboptimal policy for the organization in question.

The field of operations research is an interdisciplinary science which deploys scientific methods like mathematical modeling, statistics, and algorithms to decision making in complex real world problems which are concerned with coordination and execution of the operations. OR offers many mechanisms and techniques for a trade-off between the interests at different levels and the risks perceived by the interested actors, such as medical specialists, managers, and staff. In this thesis OR techniques are developed and made suitable for problem solving in hospitals concerning the use and allocation of capacity for operating rooms and intensive care units. The use of these techniques in healthcare is still in its infancy. The care-driven boundary conditions and goals should first be adequately translated into the operations research techniques, before these techniques can be of any use for weighing up the interests at the various levels and areas in the framework. This weighing up may then serve as a basis for decision-making and consequently the acceptance of risks concerning, for example, overtime, revenues, reputation, lack of capacity for emergency cases, refusal of patients, flexibility for the specialists, utilization, pricing, negotiations with health insurers, size of the teams on call in the night, et cetera. What's more, the use of these techniques can only be successful if the healthcare-driven boundary conditions, in the primary process, are also adequately translated in the search for available and/or novel, suitable mathematical methods. Vice versa it would seem essential to select and apply the appropriate techniques for the different trade-off problems. This can only be achieved when managers, clinicians and mathematicians take a real interest in one another's processes and learn to use each other's languages. By way of example, the **appendix** presents three mathematical articles elaborating on selection of the appropriate techniques, and which prove that these techniques are applicable in hospitals at strategic, tactical, and operational levels. C**hapter 4**, **chapter 7**, and **chapter 9** show how the use of these techniques impacts the weighing up of the outcomes in the primary process.

Outline of this thesis

In chapter 2 a framework for planning and control – a conceptual structure – has been developed that facilitates the actors at the various levels in making a trade-off between the different interests, given the qualities and characteristics of the individual organization.

- *Is it possible to typify care, the case mix, in such a way that this typing may serve as a basis for setting up the operational, tactical and strategic processes in the organization?*
- *Does a hierarchical conceptual structure exist that provides for the actors in a healthcare organization to jointly translate the effects of external and internal developments into collective policy in the different fields of planning and control?*

Utilization by different specialties of scarce and shared capacity such as operating rooms, IC beds, and radiology, is considered a measure of efficiency. In practice, insurers, governmental agencies, clinicians and healthcare managers, for example, often take utilization as an absolute measure, departing from the implicit fundamental idea that 100% utilization should be the ultimate goal for every organization. **Chapter 3** provides an answer to the question whether 100% utilization is feasible and thus worth striving after.

- *Is there any relation between operating room utilization, the hospital's case mix, staff interest to minimize the risk of overtime, and managerial considerations to minimize costs and numbers of cancelled patients?*
- *Does a method exist that allows to establish in advance the maximally attainable utilization of a hospital, i.e. the norm utilization, given the characteristics of this hospital?*
- *How can the norm utilization be used in the deliberations and trading off between health insurers, clinicians and healthcare management?*

Being private institutions, hospitals are themselves responsible for generating revenues. The importance of adequate patient care in a region cuts straight through the hospitals' independent responsibility. Quality of care, for example the need to limit the transports of trauma patients, calls for synchronization and transparency between the hospitals in that region so as to achieve they can admit all regional trauma patients as well as their own elective patients without the need to increase capacity. In **appendix 1** a model is developed in which several hospitals in a region jointly reserve a small number of beds for regional emergency patients. A mathematical method is presented for computing the number of regional beds for any given acceptance rate. The analytic approach was inspired by overflow models in telecommunication systems with multiple streams of telephone calls. In **chapter 4** the management implications and boundaries for the use of the model are given.

- *What are the strategic consequences of sharing IC capacity for individual hospital?*
- *What risks are involved in utilization of regional IC capacity?*

Every society has a need for availability of healthcare provision for unplanned care, or: emergency care. The organization of healthcare during evenings, nights and weekends is a strategic availability problem, in contrast with the organization of care during daytime hours, which represents an efficiency problem. **Chapter 5** deals with effectiveness aspects of this availability function for an operating room department during the night.

- *Is it possible to develop a model for computing optimum size of a nighttime emergency team in an operating room department that appraises the medical component, the cost component and the effects on staff?*
- *Can this model be used prospectively for each and every healthcare organization?*
- *Does the model offer opportunities for developing local, regional or national policy?*

Little or no evidence is available for methods handling unplanned care within regular working hours. While hospitals typically keep an emergency team on standby during regular working hours, the effectiveness of this approach in terms of waiting times, utilization and overtime has not yet been investigated so far. In **chapter 6** a model was developed to reserve time for emergency operations in an operating room department.

- *What parameters are decisive for reserving time for emergency operations during the regular working hours?*
- *What techniques for the handling of emergencies are available in other business sectors?*
- *What is the optimum method to handle emergencies through reserving time in an operating room department, and what are the key relevant considerations?*

A patient's care process may involve multiple consecutive steps using different facilities, such as outpatient departments, radiology units, clinical wards, operating rooms, et cetera. At an operational level each individual clinician will plan patients for his or her own involvement in the total care process. By definition these individual choices will lead to suboptimal use of the available resources if the effects on the next steps in the individual patient's care process, or the effects on the next steps for other patients, are not known or not being accounted for. The effect of individual medical policy-making on the likelihoods of resources being used by other patients is the subject of research in **chapter 7**.

- *What problems may arise if clinicians independently of one another are planning the use of consecutive capacities, like OR and ICU, for different patients within the same group or for different patient groups?*
- *Does cyclic case scheduling allow solving the perceived problems, and if so, what are the associated limiting conditions?*
- *What outcome measures determine the successfulness of a new planning method or planning support system for the clinicians, patients, and healthcare managers?*
- *How can the individual medical responsibility of clinicians in the planning of individual patients be guaranteed in a more centralized planning method?*

Major vascular surgery is associated with long length of in-hospital stay. Cardiac risk factors serve to identify patients at increased risk. Recent studies found that statin, aspirin and beta-blocker therapy are associated with improved postoperative outcome. Still, the effects of all these factors on length of stay have not yet been defined.

In **chapter 8** the relation is established between medical planning and its effect on resource capacity planning in terms of predictability of clinical capacity utilization, expressed in Length of Stay.

- *The aim of this study is to determine the effect of cardiac risk factors and (preventive) statin, aspirin and beta-blocker therapy on Length of Stay, and to deduce from these factors a model that predicts Length of Stay.*

The various specialties and specialists in fact mutually compete for operating room capacity, in that each aims at using the scarce and expensive resources maximally and on their own discretion, preferably on the moments that suit them best. It follows that the various specialties and specialists fully independently of one another plan surgery of their individual patients in the block periods made available to them.

In **chapter 9** the following hypothesis is tested: Lowering of organizational barriers between surgical departments combined with mathematical techniques to generate OR schedules can significantly improve OR efficiency. To this aim the findings from the research reported in **appendix 2** were used.

- *What problems may arise if clinicians independently of one another plan utilization of the OR?*
- *Given existing mathematical techniques, such as bin packing and portfolio techniques, would it be possible to improve the efficiency, and what effects would this have on flexibility of the specialties and specialists?*
- *What efficiency gain could be achieved using such techniques, and for which hospital types would these techniques be suitable?*

Up till now clinicians generally agree that Length of Stay is unpredictable for individual patients. We may expect that experienced intensivists are able to come up with a prediction, with level of accuracy dependent on the individual intensivist's experience. Nevertheless, we speculate that comprehensive evaluation of the association between preoperative, intraoperative, and postoperative prognostic variables on the one hand, and Length of Stay on the other hand, might be translated into a mathematical model that with some precision would predict Length of Stay. In **chapter 10** we first developed three such models based on extensive retrospective data on patients undergoing esophagectomy. We opted for this group of patients as they proportionally have great impact on use of the ICU capacity and show a wide range in Length of Stay.

- *Would it be possible to develop a prospective tool, on the basis of individual patient characteristics, peroperative-, postoperative- and intra-ICU data, that enables the planner to make individual predictions for total Length of Stay on the ICU?*
- *What are the effects of this prediction on the utilization of this expensive ICU resource, numbers of operations in the available facilities, and cancellations?*

Variability of operation times leads to sub-optimal use of operating room capacity. This then results in higher healthcare costs and cancellation of planned procedures. Reliable prediction of operative time is therefore mandatory, but subjective predictions made by surgeons or historical mean durations have only limited predictive value for individual patients. In **chapter 11** we aimed to devise an individual prediction model taking into account the surgeons' estimate and characteristics of the surgical team, the operation and the patient.

- *Is it possible to design a prospective model that is usable to predict duration of the procedure for an individual patient?*
- *Are the present registration methods in the operating room department and the data they yield suitable to develop predictive models and thus, to solve operational logistic problems in healthcare?*
- *To what extent will better individual predictions for operating times contribute to optimizing utilization of operating room capacity?*

To facilitate deliberations between mathematicians, clinicians and managers, so as to benefit from each other's skills, techniques and knowledge, we need not only create a uniform language, but also apply uniform and measurable concepts. Relevant in this case are measurable indictors of, for example, efficiency and overtime. In addition it would seem important to translate the primary process into benchmarks that are generally recognizable and measurable, and into relevant performance indicators, such as start of operation, start of inducing anesthesia, surgical period, end of surgical period, et cetera. **Chapter 12** provides an example of a uniform time registration model for surgical suites. With this we have created a conceptual framework for general use in the Netherlands, and comparable with international standards. Using the newly developed definition- and time registration method enables to gain insight into the effectiveness of processes in OR suites.

- *Would it be possible to develop a comprehensive and relatively simple definition- and time registration method for registration in surgical suites which provides the possibility to apply operations research methods in surgical department?*

Benchmarking compares performances of organizations in view of achieving lasting improvement. Still too often, however, comparison is 'between apples and oranges', with less favorably performing organizations unnecessarily falling prey to naming and shaming. In **chapter 13** a benchmarking method was developed and successfully applied to the operating room departments of the eight university medical centers in the Netherlands, without experiencing the negative side effects of the current methods.

- *Is it possible to compare operating room departments of the university medical centers without experiencing the negative side effects of benchmarking as a result of 'naming and shaming' and the mistaken feeling of euphoria arising when one has reached No. 1 position without really comprehending?*
- *Would this model also be applicable to other sectors of healthcare and society?*

Chapter 14 presents a general discussion.

Chapter 15 presents a summary for the chapters in this thesis.

Chapter 16 presents a Dutch summary for the chapters in this thesis.

CHAPTER 2

A framework for Hospital Planning and Control

Houdenhoven, M. Van, Wullink, G.,
Hans, E.W., Kazemier, G.

Introduction: "management by logistical buzzwords"

During the last 20 years, the costs of health care in the U.S. have risen from approximately 10 percent to 15 percent of the GDP [1]. In Europe a similar trend can be observed; in many countries the average annual increase of the total health care expenditure is approximately 5 percent. A major portion of the health care expenditure is spent in hospitals [1]. Hospital efficiency and management is, hence, receiving an increasing amount of attention in practice and in the literature.

The pursuit for making hospitals more efficient has led to the application of logistical concepts from other industries, for which it is believed that they can be successful in hospitals as well. The reason is obvious: while hospitals have historically been devoid from a focus on efficiency, fierce competition in industry has led to many success stories and a vast amount of logistical concepts that focus on efficiency [2]. Concepts like *Just-In-Time*, *Focused Factory* [3], *Pull versus Push* [4], *Material Requirements Planning* [5], *Manufacturing Resource Planning / MRP* and *Agile* [6] or *Lean Manufacturing* [7] were swiftly added to the vocabulary of the health care managers and management consultants. Companies like *Toyota* [8], or *Wall Mart* [9], where some of these concepts were developed, are often used as an example to demonstrate the success of these logistic concepts.

Not all implementations of the aforementioned concepts, however, have been successful [2]. There are two main pitfalls for hospitals that should be avoided when copying concepts from other industries. The first pitfall is:

> *Management and efficient use of hospital resources require a concept that facilitates a dialogue and cooperation between all stakeholders.*

To be successful, a concept should address the (generally conflicting) objectives of all stakeholders. Glouberman and Mintzberg [10;11] nicely illustrate the complexity of hospital organization; the four parties involved (nurses, doctors, administrators, and trustees) form four coalitions that have their own, generally conflicting, objectives. If there is a lack of a common language, communication between these coalitions is often poor. As a result, when improving the hospital's efficiency, it is extremely difficult to bring about organizational changes.

A planning and control concept supports addressing all stakeholders' objectives, and provides a common language, if it concerns *all* areas of interest concerning hospital management. Many logistical concepts, however, just focus on one area of interest. For example, MRP is primarily material oriented, and mostly ignores resource capacity planning. In fact, it even plans against infinite capacity, which is why the majority of the MRP implementations have been a failure [12 2]. Hierarchical Production Planning [13] is primarily resource capacity planning oriented, but ignores material coordination. Other authors [14;15] recognize the importance of other areas of interest, and introduce planning and control frameworks that focus on multiple areas: technological planning, resource capacity planning, and material coordination. In hospitals, however, the management areas are: medical planning, resource capacity planning, material coordination, and financial planning.

Logistical concepts mostly focus on productivity. Health care managers have traditionally focused on providing the best possible service [16]. Therefore, in hospitals, productivity cannot be seen separately from objectives like quality of labor and quality of care. Moreover, clinicians possess a considerable amount of *professional autonomy* induced by their expertise [17]. They typically have their own *personal objectives*, which are often related to their academic career or their salary. In addition, as a result of their *professional ethics* such as the Hippocratic oath, doctors will always choose what is best for an individual patient, often without considering the interest of a patient group.

The second pitfall of copying concepts from other industries is:

> *The system characteristics of the hospital should match the system characteristics where the concept was designed for.*

Hospitals do not form a homogeneous business. Therefore, there is no single solution for all hospitals types. More importantly, hospitals have unique properties that make them hard to compare to production environments. For example, providing health care involves a high degree of variability and uncertainty. Yet, some logistical concepts (e.g. MRP and JIT) require a stable production system, with a predictable demand, and assume the product structure is fully known. Many companies ignored this while trying to copy the successes of Toyota, and hence failed because their production environment differed too much from Toyota's. Unlike products/parts in manufacturing, health care processes can generally not be interrupted or stored. Finally, the financial (reward-based) system of the health care system differs from the, typically market-oriented, financial system in industry. Hospitals are paid for what they do, not for the success of a treatment.

To avoid both pitfalls we demonstrate how a planning and control framework can be filled with techniques for hospital management, given the typology of the case mix. The combination of the framework and the typology facilitates a dialogue between *planning and control experts*, *clinicians* and *managers* to select the planning and control mechanisms to consistently materialize the hospital objectives on all areas of interest.

The paper is organized as follows. Section 2 outlines the literature on frameworks for planning and control. Section 3 presents the generic framework for hospital planning and control. Section 4 proposes the typology for hospitals. Section 5 demonstrates illustrates the ideas discussed in this paper using an example from the Erasmus MC (Rotterdam, the Netherlands). Finally, in Section 6 we draw conclusions.

Literature on hierarchical frameworks for planning and control

As argued by Vissers et al. [18], the importance of hospital planning and control is obvious: while their budgets are tightened, hospitals are faced with a growing demand for care, and higher expectations for service quality. Hospital planning and control comprises the coordination between hospital resources, patient and material flows, medical policy, and the financial system such that the hospital's objectives are realized.

Various researchers have proposed frameworks for (hierarchical) planning and control in hospitals. Vissers et al. [18] and De Vries et al. [19] propose a framework for production control in hospitals. It distinguishes five hierarchical levels: strategic planning, patient volumes planning & control, resources planning & control, patient group planning, and patient planning and control. The approach is based on the idea that a hospital is organized in relatively independent business units. It focuses on resource capacity planning, and does not explicitly consider areas of interest like medical planning, financial planning, and material coordination. Moreover, it does not consider the online operational level, where, for instance, emergency coordination is addressed.

Roth and Van Dierdonck (1995) [20] propose an MRP-based framework for hospital planning. It uses DRGs to define patient groups that can be planned using MRP techniques, and stochastic resource requirements to deal with variability of health care. An obvious drawback of this MRP based approach is the rigidity of the MRP approach itself. Merode et al. [21] discuss the implementation of Enterprise Resource Planning (ERP) in hospitals. They acknowledge that ERP systems have trouble with dealing with variability and stochasticity. As an alternative they suggest to use Knowledge Based Systems in combination with Advanced Planning Systems. Blake and Carter [22] propose a framework for OR planning. They propose a hierarchical decomposition of strategic, administrative, and operational planning. They argue that besides efficiency, planning also serves a tool for communication between different departments.

In manufacturing, planning and control has a rich tradition. The well-known frameworks typically organize planning and control functions hierarchically, with a strategic, tactical and operational level. Some specifically focus on *resource capacity planning* [13;23], some on other areas of interest like *technological planning* or *material coordination*. The frameworks are designed for manufacturing planning and control in complex organizations, which are characterized by an unpredictable demand [14;24]. Since these frameworks address many of the aforementioned areas of interest and all hierarchical levels of control, they offer a sound basis for our framework for hospital planning and control. To use it as a framework for hospitals, however, requires significant modifications.

A generic framework for hospital planning and control

Hospital planning and control should address the following areas of interests:

- *Medical planning* While manufacturing engineers do technological planning, in hospitals, doctors perform this role. We refer to this area of interest as medical planning. It comprises the planning of the medical activities. Medical decisions made by clinicians regarding, e.g., diagnoses and treatments interact with other areas of interest, such as, financial control and material logistics. The more complex the health care process, the more important the medical planning becomes. Typical, performance indicators in this area of interest are quality of care or research output.
- *Resource capacity planning* Deals with efficiently using the hospital's scarce resources. These resources include people, tools, operating rooms, CT/MRI scanners, etc. Typical performance indicators are utilization, overtime, and underutilization.
- *Material coordination* Deals with the distribution of materials, prostheses, blood, (sterile) instruments, etc., to support the primary process. It encompasses functions like inventory control and purchasing. In this area of interest, performance indicators such as service rate and response time are crucial.
- *Financial planning* Comprises all functions regarding hospital finances, like financial planning and control, cost-price calculation or investment planning. Major performance indicators in this area are profitability, liquidity, or solvability. Because of the unique financial system of health care a framework must explicitly address this managerial area.

A framework should emphasize the interaction between the areas of interest and the various levels of control. Moreover, it should support the insight in the trade-off between performance indicators in the four areas of interest. In accordance to many frameworks for manufacturing planning and control we choose a hierarchical decomposition [23] into a *strategic*, *tactical*, and *operational* level. Since variability is inherent to the medical process, which implies dealing with unplanned events such as emergencies, we specifically discern between *offline* and *online* operational planning. We briefly explain the differences between these four levels:

- *Strategic planning* Addresses the formulation of long-term objectives, or mission statements of an organization, and the determination of the investments needed to achieve these. These organizational objectives should be decomposed into consistent and concrete strategic objectives on all four areas of interest.
- *Tactical planning* Translates strategic objectives or choices into medium-term objectives. As an example we mention resource allocation decisions by middle management (e.g. department managers). While strategic planning uses patient forecasts and/or historical information, tactical planning, like operational planning, deals with actual / expected patients. As opposed to the operational planning, at the tactical level the longer planning horizon creates more flexibility in the dimensioning of the involved resources. While in operational planning the resource capacity is typically given, in tactical planning resource capacity can be temporarily expanded (e.g. overtime, temporary extra staff).
- *Operational offline planning* Deals with the in-advance day-to-day control of expected activities. It comprises the detailed coordination of the resources and materials that were made available at the previous planning stage, to achieve the desired service levels. The adjective “offline” refers to the fact that operational offline planning concerns operational planning *in advance*. Operational offline decisions are typically delegated to lower management or clinicians. Examples are diagnosing, department scheduling, inventory replenishment ordering.
- *Operational online planning* Involves all control mechanisms that deal with monitoring the process and reacting to unforeseen or unanticipated events. Examples are: treatment planning in case of an emergency, patient rescheduling due to temporary resource unavailability, rush ordering policies regarding sterilizing instruments for surgery. Objectives are for instance waiting time for an emergency patient or waiting time for instruments for a surgery in case of a rush order.

Figure 1 displays the 4-by-4 framework that results from our horizontal and vertical decomposition. We do not explicitly give the decision horizon length for the planning levels, since these depend entirely on the specific application. The dimensions of the framework are generic, the contents is not. We filled in some example planning and control functions for a general hospital. These certainly do not cover all possible functions. If applied to e.g. a trauma center or an operating theatre department, the contents of the framework will change.

Techniques or mechanisms – varying from straightforward to advanced or from reactive to proactive – should be properly geared to each other to guarantee optimal interaction. Horizontal interaction allows trade-offs to be made between the hospital’s (generally conflicting) objectives. Vertical interaction (downwards) within a managerial area allows formulating goals and restrictions on lower levels that comply with the hospital’s strategic objectives, and (upwards) gives feedback about the realization of these objectives (e.g. the required resource capacity).

An important factor in determining hospital planning and control policies is the way the health care system in a country is organized. Insurance systems, governmental health care policies, and demographic factors determine the way a hospital interacts with its environment. We explicitly mention this, but do not incorporate this in our framework, since it is not part of the internal hospital organization.

The framework serves as a common language for managers, clinicians and experts on planning and control to formulate objectives in terms of performance indicators on all organizational levels and in all areas of interest.

	Medical planning	Resource capacity planning	Material coordination	Financial planning
Strategic	Research and treatment methods	Case mix planning, layout planning, capacity dimensioning	Supply chain and warehouse design	Agreements with insurance companies, investment plans
Tactical	Definition of medical protocols	Allocation of time and resources to specialties, rostering	Supplier selection, tendering	Determining and allocating budgets, annual plans
Operational offline	Diagnosis and planning of an individual treatment	Patient scheduling, workforce planning	Purchasing, determining order sizes	RNG billing
Operational online	Diagnosing emergencies and complications	Monitoring, emergency coordination	Rush ordering,	Billing complications,

Hierarchical decomposition

Managerial areas

Figure 1 *Framework for hospital planning and control*

A hospital typology framework

One of the most important strategic choices for hospital management is which case mix to serve. It affects all lower level objectives, and thus the way the hospital is managed. The design and application of planning and control instruments strongly depend on three case mix characteristics: (1) the extent in which patient trajectories can be predicted, which we call *intra clinical variability*, (2) the variability in the duration or length of stay of patients, which we call *inter clinical variability*, and (3) the *volume* of patient categories. The fields OR/MS and economics offer a vast amount of techniques to deal with different degrees of variability. For this reason, we propose a hospital typology based on variability for the selection of planning and control instruments. In the subsequent sections we elaborate on these determinants, and present the hospital typology. Like the framework, the typology can be applied to any hospital, or hospital department.

Inter clinical variability Inter clinical variability of a patient group concerns the variability of the duration of a treatment or length of stay (LOS). Depending on the type of treatment and/or patient condition, the duration/LOS may be fairly certain or completely unpredictable.

A high degree of inter clinical variability requires *robust planning* approaches – proactive or reactive – that account for the uncertain durations or LOSs. *Proactive* planning is an offline approach that tries to use the available flexibility in the plan to prevent that corrective decisions have to be made when uncertainties materialize. This flexibility is usually obtained by deploying reserve or slack capacity. As a result, managers should realize that if the inter clinical variability is high, the maximum resource utilization that can be attained decreases [norm paper]. *Reactive* planning is an online planning approach that deals with reacting to unforeseen events, for example, emergency surgery planning.

Intra clinical variability Intra clinical variability is the variability in the pathway that is followed by an individual patient. A clinician must make complex decisions regarding the care pathway of a patient. Therefore, if the intra clinical variability is high, instruments for planning and control should focus on supporting clinicians in their planning decisions by, for example, offering them planning alternatives or at least making the consequences of their decisions transparent. If, however, the intra clinical variability is low, clinicians need to interfere less with planning and control. Clinicians can work according to protocols and predetermined plans that are geared to optimal material and resource capacity usage.

While inter clinical variability concerns planning of a single resource or department, intra clinical variability concerns inter-departmental planning. Instruments for planning and control in hospitals with a high degree of intra clinical variability should therefore always take into account the impact of planning decisions on other hospital departments. Figure 2 illustrates the two determinants inter clinical variability and intra clinical variability.

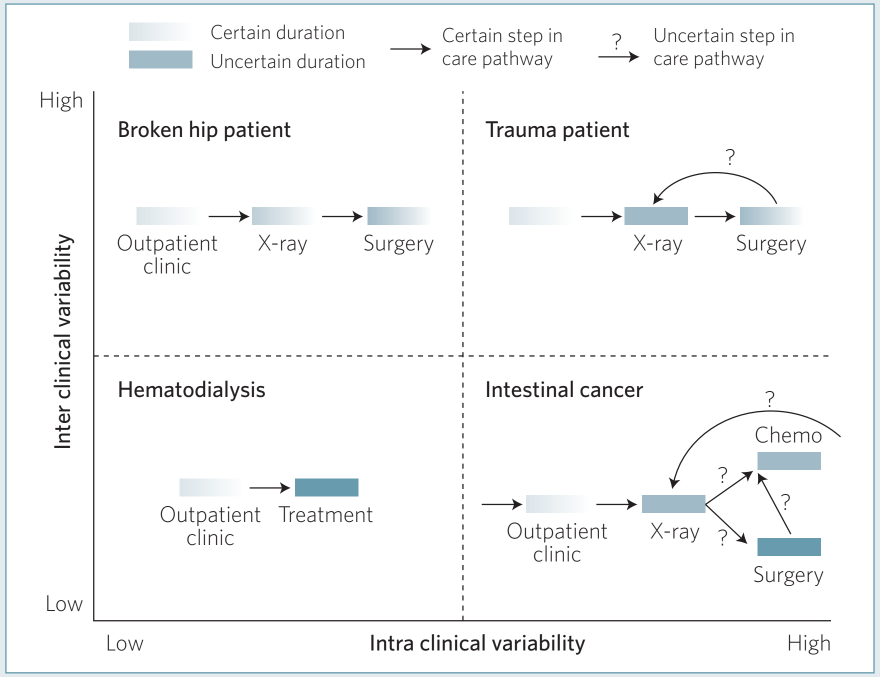


Figure 2 *Typology of a case mix*

Volume The volume of a patient category is largely determined by strategic choices and demographic factors. An advantage of large volumes is that it offers possibilities – using the portfolio effect – to dampen the effects of uncertainty. If their volume is large, case mixes with a high inter/intra clinical variability can theoretically be served better. Also, if the volume of a particular patient category is large, a hospital can decide to separate it from other patients in specialized departments.

We now have three case mix characteristics based on which we can typify hospitals: intra and inter clinical variability, and volume. As an example figure 3 depicts four hospital types: a cataract, cardiac, academic, and cancer hospital.

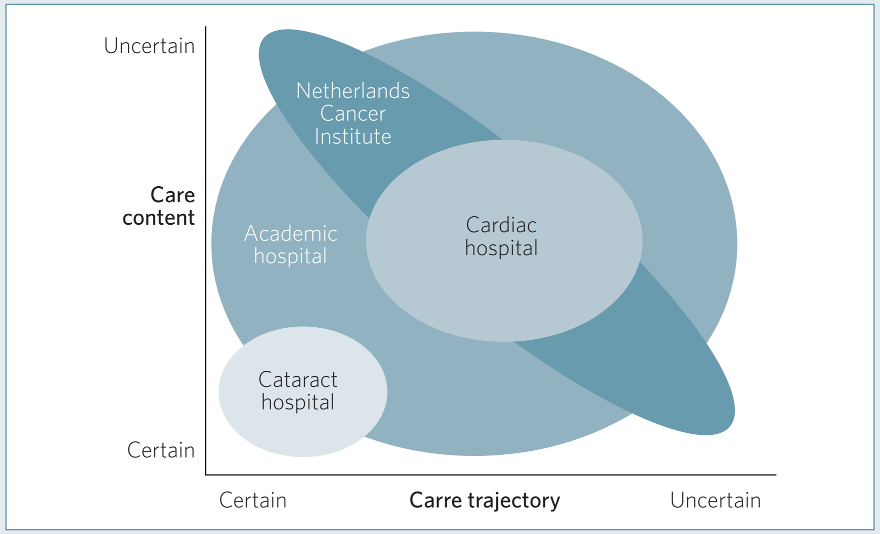


Figure 3 *Hospital types*

The outpatient OR of the Erasmus MC

This section describes a case of the outpatient clinic of the Erasmus MC (Rotterdam, the Netherlands) to which the ideas in this paper apply.

How it started In March 2003 the Board of directors of the Erasmus MC (Rotterdam, the Netherlands) decided to build an outpatient OR. The aim was to create a low-threshold, recognizable, patient friendly, and accessible OR department, with optimized logistics and no influence of disturbing emergency cases. The case mix of the department would consist patients on which low or medium complex surgery would be performed and that would not be admitted to the hospital. The business case of the outpatient OR was that by only planning elective, low complex surgery inventory costs could be kept to a minimum and OR capacity could be used optimally, since there was no disturbance of emergency cases. The intended location of the new OR department was on the first and second floor of the south wing of the main building of the Erasmus MC.

Design phase To determine the size of the case mix all surgical departments were asked to determine the number of patients they thought suitable to operate upon in such an environment. This investigation phase resulted in a case mix that required three operating rooms, a recovery room with six places, several offices, dressing rooms, etc. It appeared, however, that the physical space of the intended location had just about the required size and little room for inventory was left.

The team that was assigned to execute the project decided that the limited space for inventory should be no problem, since only elective low complex surgery would be done at the new OR department. This allowed for low inventory levels combined with a JIT inventory policy. During contract negotiations with suppliers so-called procedure trays were developed the supplier would deliver periodically.

Success! On the first of November in 2003 the outpatient OR department was opened. It appeared an enormous success and surgeons, personnel, and patients were very content with the department. Gradually, however, it appeared that the estimations of the surgical departments concerning the size of the case mix had been too optimistic. After several months the utilization rates of the outpatient OR department appeared to be disappointingly low and ORs were empty during parts of the day. There were two main reasons for the underutilization of the department. First, the surgical departments had been too optimistic in their estimations of the new patients that could be operated. Second, a surgery that does not involve a hospital admission is financially less attractive than a surgery with admission. Therefore, the new patient categories that surgical departments had planned to operate upon, did not come to the new OR department. The size of the OR department, the staffing, and the facilities, however, had already been set up. The resulting low utilization rates, however, contributed even more to the success of the department.

What went wrong Surgeons and personnel liked to work in the new department and since there was always an OR available gradually more and more other, more complex surgery was executed in the new ORs. Then things started to go wrong. The case mix of the new OR, which was originally indented to be in the lower left part of the typology (Figure 4) moved to the upper right corner of the typology.

More complex surgery, more complications, and more variability in the session time were the inevitable consequences. This required more sophisticated operating materials than available in the procedure trays supplied by the contractor. Despite the little space at the department, inventories were built up to deal with the more complex case mix. Particularly, the last aspect, growing inventories, diminished the aimed benefit of the business case of the outpatient OR department.

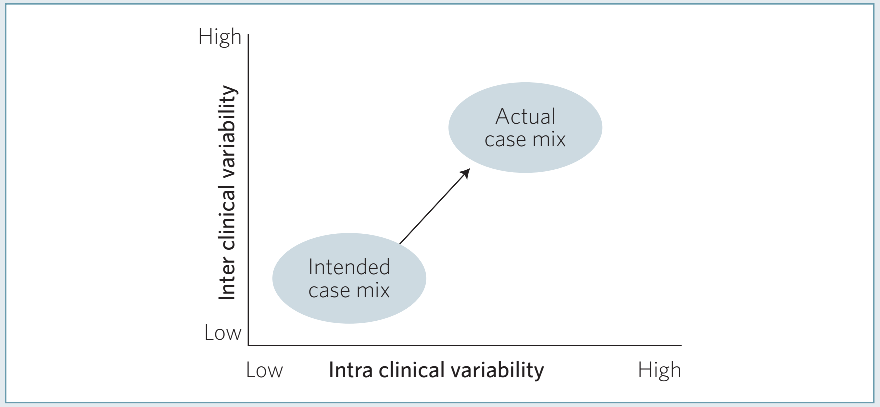


Figure 4 *Typology of a case mix*

Lessons to be learned This case illustrates the consequences of the use of the framework and the typology. First concerning the framework. The inventory policy at the tactical and strategic level has to match the resource capacity constraints determined in the resource capacity planning column, i.e., the size of the department and the ORs. Moreover, the inventory policy has to match the case mix that, i.e., the medical planning.

Second, the case illustrates that the position of the case mix in the typology determines the logistic policies regarding inventory, but also regarding capacity planning. The patient planning method on the new OR department was intended to be pretty straightforward. Patients were planned early and little free capacity was planned to deal with inter clinical variability. The introduction of more complex surgery made this planning method unsuitable and too rigid. The same argument holds for the selected inventory policy. JIT was suitable for low complex elective surgery

Conclusions

As a result of the increasing costs of health care and the introduction of (managed) competitive health care in western countries there is a great need for new and adequate approaches to hospital management. As in traditional manufacturing, OR/MS can fulfill an important role. Many managers and consultants that work in health care have recognized this development. This, however, did not yet result in a structured approach to hospital management. Efforts to adopt hyped concepts from manufacturing frequently resulted in failures and misunderstandings between managers and professionals in health care. The main reasons are that the concepts focus on a part of the areas of interest, and were developed for a system that is entirely different. They do not account for that hospitals can be very different, and have several (generally conflicting) objectives.

In this paper we propose a reference framework for hospital planning and control. It hierarchically structures all planning and control functions of a hospital in all areas of interest. This offers a common language for all stakeholders that are involved in hospital management: clinicians, managers, and experts on planning and control. Any research that focuses on hospital process optimization can use this framework to position problem areas, analyze the control functions that are involved, and analyze the relations between adjacent and related control functions. Also, new techniques from for instance the area of OR/MS or economics can be applied in a structured way. The second contribution of this paper is a typology for hospital types. This typology enables the formulation of different objectives for different types of hospitals, and accordingly the selection of different instruments for planning and control.

The strength of the approach introduced in this paper is that it can not only be applied to hospitals, but also to specific hospital departments, such as the operating room department. With the combination of the reference framework for planning and control and the typology for hospitals we believe that hospital managers and clinicians are better suited for managing the competitive hospitals of the future.

Reference List

1 OECD Health Data 2005. 2005. IRDES. Ref Type: Report
2 Hopp WJ, Spearman ML. Factory Physics – Foundations of Manufacturing Management. IRWIN, Boston, 1996.
3 Skinner W. The focussed factory. *Harvard Busines Review* 1974;**May-June:**113-21.
4 Lee YH,.Lee B. Push-pull production planning of the re-entrant process. *International Journal of advanced Manufacturing Technology* 2003;**22:**922-31.
5 Orlicky J. Material Requirements Planning. McGraw-Hill, London, 1975.
6 Murray LW. Agile manufacturing: Forging new frontiers – Kidd,PT. *Journal of Product Innovation Management* 1996;**13:**181-2.
7 Lang JD,.Hugge PB. Lean Manufacturing. *Aerospace America* 1995;**33:**28-&.
8 Ohno T. Toyota production system: beyond large-scale production. Productivity Press, 1988.
9 Daft RL. Management. Fort Worth: The Dryden Press, 2000.
10 Glouberman S,.Mintzberg H. Managing the care of health and the cure of disease –

Part II: Integration. *Health Care Manage.Rev.* 2001;**26**:70-84.
11 Glouberman S,.Mintzberg H. Managing the care of health and the cure of disease – Part I: Differentiation. *Health Care Manage.Rev.* 2001;**26**:56-69.
12 Darlington J, Moar C. MRP Rest in Peace. Management Accounting (UK), 1996.
13 Bitran GR, Tirupati D. Hierarchical Production Planning. North-Holland, Amsterdam, 1993.
14 Zijm WHM. Towards Intelligent Manufacturing Planning and Control Systems. *OR Spektrum* 2000;**22**:313-45.
15 Hans EW, Herroelen W, Leus R, Wullink G. A hierarchical approach to multi-project planning under uncertainty. *In press, corrected proof in Omega* 2006.
16 Carter M. Diagnosis: Mismanagement of Resources. *OR/MS today* 2005; **April 2002**.
17 Fitzgerald, J. A., Lum, M., and Dadich, A. How can human technologies improve decision-making practices in the operating theatre? 8-25-2005. Aalborg, Denmark, Paper presented at the 4th Health Care Technology Management Conference. Ref Type: Conference Proceeding
18 Vissers JMH, Bertrand JWM, De Vries G. A framework for production control in health care organizations. *Production Planning & Control* 2001;**12**:591-604.
19 De Vries G, Bertrand JWM, Vissers JMH. Design requirements for health care production control systems. *Production Planning & Control* 1999;**10**:559-69.
20 Roth AV,.Van Dierdonck R. Hospital Resource Planning. *Production and Operations Management* 1995;**4**:2-29.
21 Merode GG, Groothuis S, Hasman A. Enterprise Resource Planning for Hospitals. *International Journal of Medical Informatics* 2004;**73**:493-501.
22 Blake JT,.Carter MW. Surgical process scheduling: a structured review. *J.Soc.Health Syst.* 1997;**5**:17-30.
23 Hax AC, Meal HC. Hierarchical Integration of Production Planning and Scheduling. North Holland TIMS, 1975.
24 Hans EW, Herroelen W, Leus R, Wullink G. A hierarchical approach to multi-project planning under uncertainty . *In press, corrected proof in Omega* 2006.

CHAPTER 3

A Norm Utilisation for Scarce Hospital Resources: Evidence from Operating Rooms in a Dutch University Hospital

Mark Van Houdenhoven, Erwin W. Hans, Jan Klein, Gerhard Wullink, Geert Kazemier

Journal of Medical Systems, Volume 31, Number 4/August 2007, p. 231–236

Introduction

The utilisation rate is commonly used as an indicator to measure the efficiency of the use of hospital resources such as the operating rooms (OR). Researchers often argue that high utilisation is cost-effectiveness and goes hand in hand with high quality of health care [1–6]. In these studies utilisation is typically considered as an absolute measure that can be used to resolve the perceived efficiency problem in health care [7–11]. In accordance with this trend, governmental institutions, such as the audit commission of the National Health-care Service, offer programmes to improve utilisation of operating room departments [12;13]. Practice learns, however, that utilisation rates of 100% are rarely achieved.

While trying to maximise the utilisation of the OR, managers are confronted with the complexity of the case mix and high costs for overtime of the OR staff. Typically, in hospitals with a complex case mix with a high variability in the duration of surgical cases, utilisation rates are significantly lower than 100%. For example, 100% utilisation for a cancer medical centre is utopian, since a considerable number of the surgical cases cannot be completed within the planned duration because of the unpredictability of the disease. Typically in a cancer centre the duration of surgical cases is unpredictable and they can last much longer than expected. This causes frequent overtime in operating rooms of a cancer centre.

The aim of this study was to address the association between operating room utilisation, the case mix, and the accepted risk of overtime, using straightforward statistical analyses. For this purpose, we propose a method to establish this association by calculating a norm utilisation for the OR. We compared the sensitivity of the method with respect to the two main input parameters, namely the case mix and the accepted risk of overtime. The study uses prospectively collected data of the Erasmus Medical Centre (Erasmus MC) in Rotterdam, the Netherlands.

Methods

The records of 180,000 surgical cases performed at Erasmus MC's main location from January 1, 1994 through December 31, 2005, were used. The following surgical departments have been investigated: General Surgery, Gynaecology, Oral Surgery, Ear-Nose-Throat Surgery, Neurosurgery, Ophthalmology, Orthopaedic Surgery, Plastic Surgery, and Urologic Surgery. All data were entered electronically by the nursing staff in the Hospital Information System and validated by the responsible surgeon and anaesthetist. Based on this extensive database we computed several case mix characteristics such as the standard deviation of the duration of elective and emergency cases and the number of elective and emergency cases per specialty. For this study, we assumed that the number of cases performed, the average duration, and the standard deviation for both elective and emergency cases describe the patient mix. Figure 1 depicts all measurement moments that were used in this study.

We assumed that cases were scheduled using the block planning approach [14]. In this case scheduling approach, a surgical department schedules its elective surgical cases in the blocks of operating time that are available for that surgical department. Scheduling surgical cases is done using average case durations. To reduce the risk of overtime induced by emergency cases and the variability of the duration of elective and emergency cases the surgical department has to plan reserve capacity in the OR block [15;16]. The amount of reserve capacity is calculated on basis of the variability of the duration of elective and emergency surgical cases.

Reserve capacity to operate upon emergency surgical cases is assigned to all elective operating rooms. Hence, no emergency operating room exists. Upon arrival the decision is made in which operating room the emergency case will take place, the emergency patient has to wait until the preceding cases in this operating room is ended. The total amount of reserved capacity for emergency cases is based on the average duration and the expected number of emergency surgical cases. Hence, the reserved OR capacity is used to deal with the variability of elective cases, emergency cases, and to deal with the variability of emergency cases. Figure 2 shows this schematically.

There are various ways to compute the utilisation rate. In this study the utilisation of OR capacity is defined as the time an operating room is occupied to perform elective and emergency surgical cases, expressed as a percentage of the length of time an operating room is available and staffed during a certain period (see figure 1) [14;17].

The method requires some mathematical notation (see Table 1). Let s denotes a surgical department. The average number of elective surgical cases executed in an OR block by surgical department s is n_s. μ_s is the average duration and σ_s the standard deviation of a surgical case of department s. Emergency surgery for surgical department s is characterised by the average duration of the emergency surgical case (μ_s^e), the standard deviation of the duration (σ_s^e), and the average number of emergency cases per OR block (n_s^e). The accepted risk of overtime for surgical department s is denoted by r_s.

The following steps were taken to calculate the amount of OR time that is required for operating all elective and emergency cases given the variability

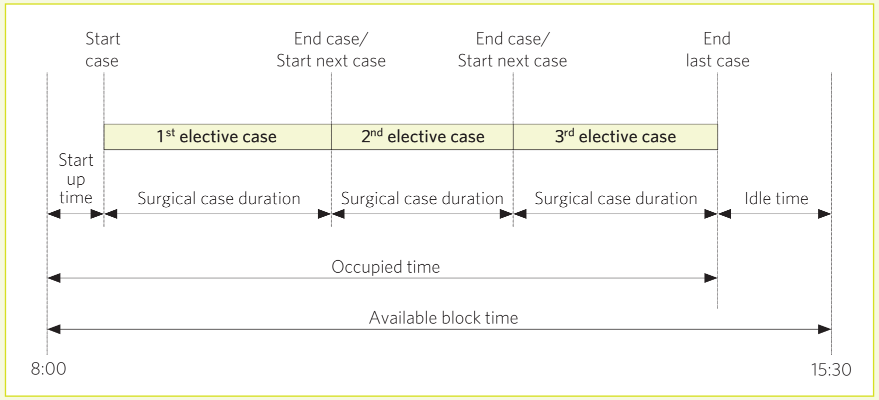


Figure 1 *Representation of the measurement moment at a single operating room. The ratio between the Occupied time and the Available block time defines the utilisation of the operating room. Anaesthetic time, cleaning time, and changeover times are included in the case durations.*

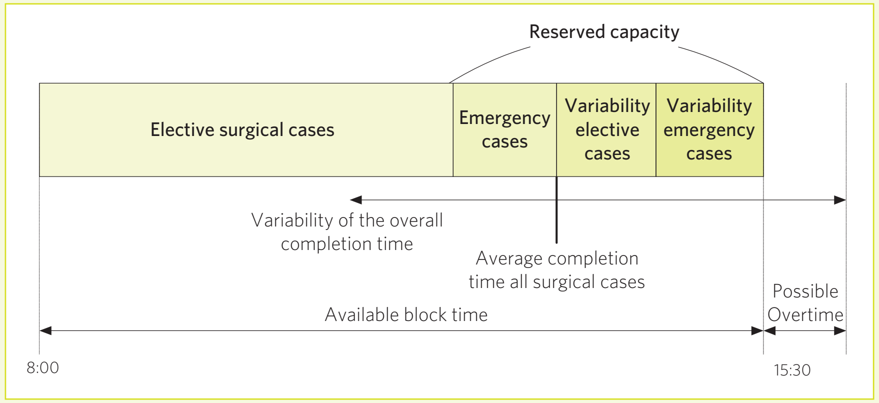


Figure 2A *A schematic representation of a surgical case plan of an operating room. Reserved capacity is included to perform emergency surgical cases and to deal with the variability of the overall completion time.*

and a predetermined accepted risk of overtime. The total expected duration of all elective cases in one OR block is $n_s \cdot \mu_s$. The standard deviation of the total duration of these cases equals $\sqrt{n_s \cdot \sigma_s^2}$. The total capacity to operate all elective cases of surgical department s is thus $n_s \cdot \sigma_s + \alpha(r_s)$, where $\alpha(r_s)\sqrt{n_s \cdot \sigma_s^2}$, where $\alpha(r_s)$ is a function that gives a factor that yields a probability r_s of working in overtime. The outcome of this function depends on the distribution of the total duration of cases. For simplicity we shall assume here that this is a normal distribution. In this case, suppose $r_s = 50\%$, then it follows from a normal distribution that $\alpha(r_s) = 0$. The chosen total capacity would thus be $n_s \cdot \mu_s$. Or, suppose $r_s = 31\%$, then $\alpha(r_s) = 0{,}5$, and the chosen total capacity would be $n_s \cdot \mu_s + 0.5\sqrt{n_s \cdot \sigma_s^2}$. The term $\alpha(r_s)\sqrt{n_s \cdot \sigma_s^2}$ is thus the total required reserve capacity. This approach does not depend on the chosen distribution; if another distribution is used (e.g. the sum of lognormal distributions), only the outcome of $\alpha(r_s)$ changes. The capacity to execute the expected number of emergency cases is determined analogously: $n_s^e \cdot \mu_s^e + \alpha(r_s)\sqrt{n_s^e \cdot (\sigma_s^e)^2}$. If elective and emergency cases are scheduled together in one OR, the total reserve capacity that guarantees an overtime risk r_s for surgical department s is: $\alpha(r_s) \cdot \sqrt{n_s \cdot (\sigma_s)^2 + n_s^e \cdot (\sigma_s^e)^2}$. The utilisation rate for specialty s is then:

Symbol	Meaning
s	A surgical department
μ_s	Average elective case duration of surgical department s
σ_s	Standard deviation of the duration of elective cases of surgical department s
n_s	Average number of elective cases per OR block of surgical department s
μ_s^e	Average emergency case duration of surgical department s
σ_s^e	Standard deviation of the duration of emergency cases of surgical department s
n_s^e	Average number of emergency cases per OR block of surgical department s
r_s	Accepted risk of overtime for surgical department s
μ	Overall average elective case duration
μ^e	Overall average emergency case duration
σ	Overall standard deviation of elective cases
σ^e	Overall standard deviation of emergency cases

Table 1 *Overview of mathematical notation used in the research.*

$$\frac{n_s \cdot \mu_s + n_s^e \cdot \mu_s^e}{n_s \cdot \mu_s + n_s^e \cdot \mu_s^e + \alpha(r_s)\sqrt{n_s \cdot (\sigma_s)^2 + n_s^e \cdot (\sigma_s^e)^2}}$$

To investigate the sensitivity of the outcome of the formula we have varied the standard deviation of the case duration to evaluate the effect of different case mixes on utilisation rates. We also investigated the sensitivity of the utilisation with respect to various values of r, i.e., the predetermined accepted risk of overtime.

Results

Table 2 summarises the Erasmus MC data, which has been the input of all experiments. The accepted risk of overtime in Erasmus MC is 31%.

Applied to the case of General Surgery in the Erasmus MC the proposed formulas result in the following computations:

- The time required to perform elective and emergency cases:
  $= n_S \cdot \mu_S + n_S^e \cdot \mu_S^e = 6.77 \times 172.39 + 0.94 \times 131.75 = 1167$ minutes;
- The factor for calculating the required reserved capacity is $\alpha(31\%) = 0.5$;
- Required reserved capacity:
  $\sqrt{n_s \cdot (\sigma_s)^2 + n_s^e \cdot (\sigma_s^e)^2} = \alpha(31\%)\sqrt{6.77 \cdot (106.32)^2 + 0.94 \cdot (102.78)^2}$
  $= 147$ minutes;
- Utilisation of operating room time: $1167/(1167 + 147) \times 100\% = 90\%$

	SUR	GYN	ORAL	ENT	NEC	TRAU	OPT	ORT	PLC	URO
Total available time per day	1,350	450	450	450	450	450	450	450	900	450
Average number of cases per day n_S	6.8	3.7	2.5	2.7	1.3	3.2	6.0	2.6	4.8	2.9
Average case duration (minutes) μ_S	172	104	150	119	256	108	67	136	154	125
Average case standard deviation (minutes) σ^e	106	56	88	135	157	60	34	77	133	92
Number emergency surgical cases per day n_S^e	0.9	0.1	0.0	0.1	0.2	0.3	0.0	0.2	0.2	0.1
Average emergency case duration (minutes) μ_S^e	132	76	134	87	141	138	107	118	174	93
Average emergency standard deviation (minutes) σ_S^e	103	47	75	69	87	89	62	75	166	51

Table 2 *Input values based on data of Erasmus MC.*

	SUR	GYN	ORAL	ENT	NEC	TRAU	OPT	ORT	PLC	URO
Reserved capacity per OR block (minutes)	147	55	70	112	92	59	43	65	150	78
Norm utilisation (%)	90	88	84	75	80	87	91	86	84	83

Table 3 *Results, Erasmus MC with an accepted risk of overtime of 31%.*

Table 3 shows that, given an accepted risk of overtime of 31%, the utilisation varies between 75% (for Ear, Nose, and Throat surgery, including head and neck surgery) and 91% (for Ophthalmology).

The association between the patient mix, represented by the variability of the case durations, and utilisation and the association between the accepted risk of overtime and utilisation is shown for three surgical departments: General Surgery, Ear-Nose-Throat Surgery, and Ophthalmology. Figure 3 addresses the correlation between the variability in the duration of cases and utilisation. As may be expected, if there is no variability in the duration of the surgery, the utilisation is 100%. Utilisation decreases, however, as variability increases, since the required reserved capacity depends on the variability. Utilisation increases from 89% to 95% if the standard deviation is varied from the standard deviation of General Surgery (Erasmus MC) to half of this value. Figure 4 shows that the utilisation increases when the accepted risk of overtime grows.

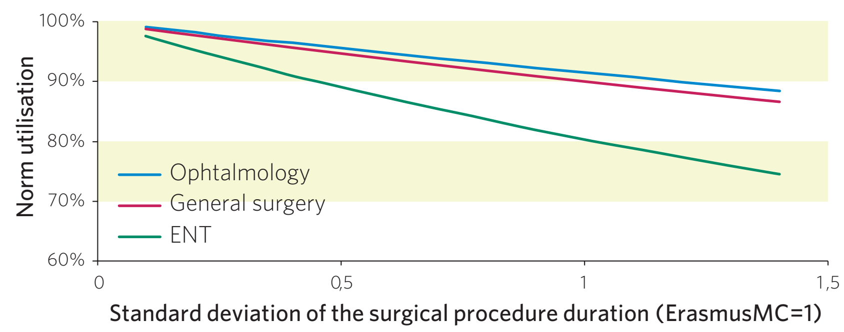


Figure 3 *Utilisation versus variability in the standard deviation of the case duration.*

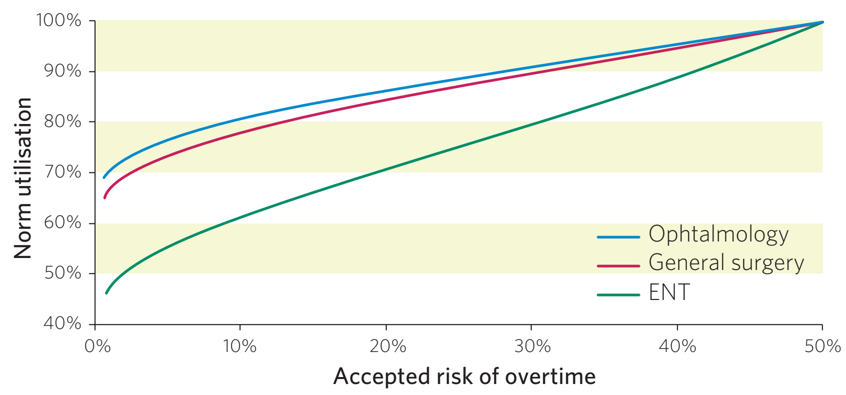


Figure 4 *Utilisation versus the accepted risk of overtime.*

Discussion

The relation between the utilisation of operating room time, patient mix, and the accepted risk of overtime was addressed. With relatively straightforward statistical techniques, the association between the required reserved capacity and the acceptable risk of overtime and the variability of the case mix are established. This results in a method to determine a realistic norm utilisation rate for a scarce hospital resource such as an operating room considering the complexity of the case mix and managerial choices such as working in overtime.

Little research has been performed to the association between an operating room patient mix, willingness to accept risk of overtime, and operating room performance in terms of capacity utilisation. Some work on utilisation rates acknowledges the limitations of utilisation as a performance measure [18]. Strum et al. focus on cost efficiency of the OR and use a minimal cost analysis model to make the explicit trade-off between the costs for over and underutilisation [19]. Their approach serves to determine the best OR time allocation to specialties on a daily basis. As far as we know, a generic method that establishes the association between patient mix characteristics and managerial choices, such as working in overtime, lacks. The calculations presented in this paper do not capture the entire complex reality of health care. The assumptions and simplifications, however, do not violate the general idea of the approach. For instance, following the central limit theorem of statistics, we assumed a normal distribution for the total duration of the surgical cases. A more accurate distribution (e.g., a distribution based on the sum of lognormal distributions) can be used if sufficient historical data are available. The idea remains applicable but the complexity of the calculations increase.

Furthermore, we assumed an identical average and standard deviation for the duration of all the cases of a given surgical department. This assumption leads to a slight overestimation of the required reserve capacity and hence to a conservative norm utilisation. We did not account for cancelled cases. If a hospital has information about the average number of cancelled cases however, this can be incorporated in the formula for the norm utilisation. Finally, the amount of emergency cases may differ significantly over the days of the week. All of these aspects can be dealt with, without harming the general idea of the approach. There are several definitions of utilisation, which may yield different results, but rely on the same basic assumptions. Another definition to compute utilisation may yield slightly other rates, but the idea of the method proposed in the paper remains the same. Also, the reserve capacity that is computed with the proposed method can be allocated to one or more operating rooms. The computations and the resulting norm utilisation, however, remains the same regardless the way the reserve capacity is allocated.

The method proposed in this paper was developed to determine an utilisation rate for each surgical department individually. OR management can also choose to determine a single utilisation rate for the total department by not distinguishing between surgical departments. Omitting the s index in the proposed formula and using the case of all departments together can achieve this.

The operating room of the Erasmus MC was used as an example. The proposed method, however, can be used for many other scarce hospital resources to obtain realistic utilisation rates, such as the radiology department, wards or the intensive care.

The utilisation rate computed in this paper can be perceived as the theoretical maximum benchmark utilisation rate. It can be realised only if one assumes that the operating room department is making use of optimal planning techniques, has adequate management, makes the best use of its personnel, and has adequate equipment available to it, which is in practice exceptionally. Contrarily, aiming at an utilisation rate higher than the calculated utilisation is practically impossible given a risk of working in overtime and the variability of the case mix.

In operating room benchmarking studies, utilisation rates are typically compared without considering patient mix characteristics and managerial choices. These widely used comparisons typically overlook the key factors that actually determine utilisation. We conclude, therefore, that the use of utilisation as an absolute measure is an unfair comparison of hospitals departments and probably leads to the wrong conclusions. The method developed for computing the maximum achievable norm utilisation should be the basis for a new approach to operating room and hospital benchmarking and internal performance measurements for hospital boards.

In the various health-care systems, hospital funding is based on the number of surgical cases actually performed. If norm utilisation is not incorporated in the price agreed upon, a hospital will probably lack the resources to finance the required reserve capacity required to deal with the variability of the case mix and managerial choices. To ensure cost-efficient health care given the complexity of different types of health care these factors should be used into negotiations between financers and providers of health care.

This paper showed that reserving capacity results in a lower utilisation rate. Hospital boards under pressure to increase their utilisation may hence decide to achieve higher targets by refusing complex and emergency care that involves a high variance. From a societal perspective, such an eventuality is highly undesirable. To prevent this calculating behaviour, hospitals should be judged on their utilisation with respect to their own norm utilisation.

We proposed a method to determine a norm utilisation rate the operating room. This method accounts for the complexity of health care processes and the differences between hospitals. The method proposed can be used for managerial issues to evaluate and steer their own performance, but also externally to support contract negotiations and benchmarking of hospitals.

References

1 Gordon T, Paul S, Lyles A, Fountain J. Surgical unit time utilization review: resource utilization and management implications. J Med Syst 1988; 12(3):169–179.

2 OECD Health Data 2005. 2005. IRDES.

3 Young A. The empty theatre. BMJ 1990; 300(6735):1288–1289.

4 Dowdall JM. The use of operating theatres in the Northern Ireland Health and Personal Social Services. Belfast: Northern Ireland Audit Office. 2003.

5 Audit Commission. Operating Theatres, Review of national findings. 2003. 2003.

6 Bevan Report. The Management and utilisation of operating departments. 1989.

7 Overdyk FJ, Harvey SC, Fishman RL, Shippey F. Successful strategies for improving operating room efficiency at academic institutions. Anesth Analg 1998; 86(4):896–906.

8 Dexter F, Macario A, Lubarsky DA, Burns DD. Statistical method to evaluate management strategies to decrease variability in operating room utilization: application of linear statistical modeling and Monte Carlo simulation to operating room management. Anesthesiology 1999; 91(1):262–274.

9 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(5):1491–1500.

10 Kuo PC, Schroeder RA, Mahaffey S, Bollinger RR. Optimization of operating room allocation using linear programming techniques. J Am Coll Surg 2003; 197(6):889–895.

11 Eaton L. Trusts cancel 10% of operating theatre sessions. BMJ 2002; 324(7347):1174.

12 Agency NM. Theatre Programme, Step Guide to improving Operating Theatre Performance. 2003.

13 Healthcare CF. Inspecting Informing Improving, Acute hospital portfolio review, Day surgery. 2005.

14 Viapiano J, Ward DS. Operating room utilization: the need for data. Int Anesthesiol Clin 2000; 38(4):127–140.

15 Gallivan S, Utley M, Treasure T, Valencia O. Booked inpatient admissions and hospital capacity: mathematical modelling study. BMJ 2002; 324(7332):280–282.

16 Hans EW, Wullink G, Houdenhoven M, Kazemier G. Robust Surgery Loading. European Journal of Operational Research 2006; In press.

17 Mazzei WJ. Maximizing operating room utilization: a landmark study. Anesth Analg 1999; 89(1):1–2.

18 Dexter F, Macario A, Traub RD, Lubarsky DA. Operating room utilization alone is not an accurate metric for the allocation of operating room block time to individual surgeons with low caseloads. Anesthesiology 2003; 98(5):1243–1249.

19 Strum DP, Vargas LG, May JH. Surgical subspecialty block utilization and capacity planning: a minimal cost analysis model. Anesthesiology 1999; 90(4):1176–1185.

CHAPTER 4

Regional synchronization Intensive Care Capacity

Richard J. Boucherie, Mark Van Houdenhoven,
Nelly Litvak, Marleen van Rijsbergen.

Translated from Dutch,
STAtOR, june 2005; 1–2, p. 23–27

Introduction

'Hundreds of patients a year die unnecessarily' was the premise put forward in the 6 November 2001 edition of NOVA, a Dutch television news show, in which the shortage of Intensive Care (IC) capacity was discussed (www.novatv.nl). Having recognized that there are capacity problems indeed, The Dutch Minister of Health, Welfare and Sport has initiated a number of studies. These studies have shown that problems around admission and discharge of patients have great impact on issues affecting efficacy of IC units (ICUs).

In 2001, ICUs in the Netherlands refused roughly 10% of trauma patients on the ground of capacity problems; 4% were accepted in spite of lack of available capacity, and 3% were discharged early and transferred to general wards (Hautvast et al. 2001). A major cause of these problems is the shortage of IC nurses. An IC bed can only be made available if enough nurses can be deployed.

Another factor that reinforces ICU capacity problems is a hospital's complex supply chain logistics (synchronization between the departments). Non-availability of a bed for a planned operation (elective patient) is very costly for the hospital, seeing that operating (room) capacity is wasted as well. For this reason hospitals will refuse regional trauma patients (for example as a result of an accident) as well when beds are still available, yet needed for elective patients at a later time.

Regional intake

The Netherlands has adopted a region-wide approach for the intake of trauma patients. In view of providing good quality of care it would be very undesirable to move patients to ICUs outside of the region. Patients will only be referred to a hospital outside of the region if all IC beds within the own region are occupied. In the present system of decentralized use of IC capacity this may also happen as a result of simultaneous reservation of capacity for elective patients in some ICUs, and other ICUs having actually reached full capacity. Region-wide intake requires insight into numbers of available IC beds in each hospital within the region, as well as clearly delineated rules for admission of regional trauma patients.

Optimal intake

The aim of this study is to determine to what extent efficiency (proportion of refused patients) of the ICUs of all hospitals in the region will improve when a small portion of IC capacity is allocated to regional trauma patients. The eventual goal is to define the structure required to maximally accommodate for the demand of care within the region at a minimum IC capacity.

Inventory taking of IC capacity

Inventory taking of region-wide IC capacity requires a model that accounts for both the unpredictability of patient flow and the uncertainty of duration of stay. In addition, adequate modeling of differences in patient flows is a prerequisite for reaching a clear conclusion. To this aim we made an inventory of the various patient flows within ICUs.

Figure 1 depicts the patient flows for two ICUs. The ICU forms an important link in a hospital's patient care, and admits broadly three patient flows. An internal trauma patient (flow 3), for instance an emergency case within the hospital, is always admitted to the hospital's own ICU. If no bed is available here, the patient is assigned to a so-called overbed, which implies that ICU

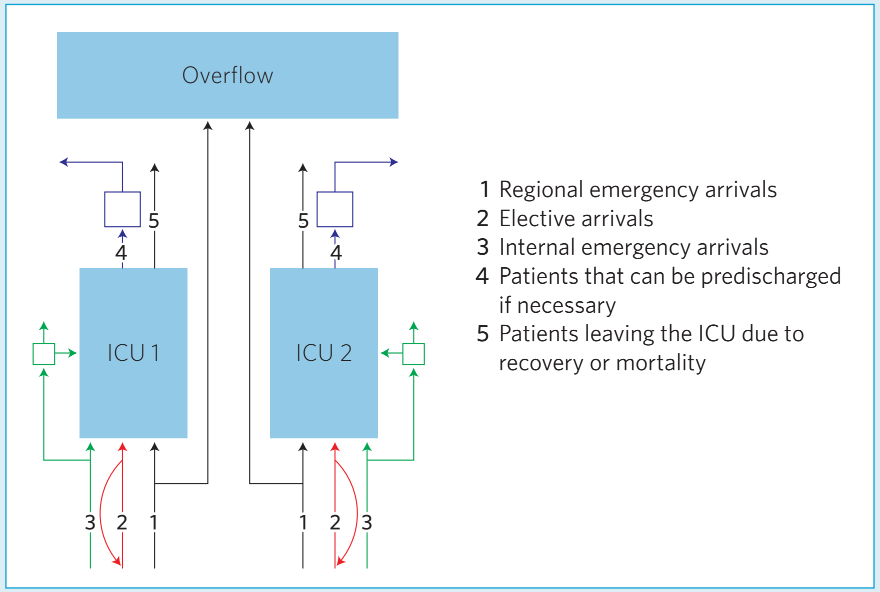


Figure 1 *Overview of the ERM including all patient streams.*

staff temporarily will have to manage more beds. The overbed will be withdrawn as soon as a patient is discharged from the ICU. A planned ICU patient (flow 2) is one who, for example, has undergone an elective operation. An elective operation cannot go underway until an IC bed needed after the operation is available. Elective patients (operations) therefore are cancelled if ICU capacity is fully used. Regional trauma patients (flow 1), for example as a result of an accident, are refused admission if IC capacity is fully used, and are referred to the Overflow. Leaving aside region-wide capacity, these patients will be referred to an ICU outside the region, and this is this group represented by Overflow. Taking region-wide capacity into account, however, Overflow will represent the region-wide capacity. Patients will leave the ICU as a result of early discharge (flow 4), or as a result of improved condition or death (flow 5). Flow 4 is not further pursued here.

Modeling with the use of queuing theory

Behavior of ICUs in a region strongly resembles that of circuit switched telephone systems. A stochastic process generates telephone calls in these systems, and a call will engage a telephone line for the duration of the call. The number of telephone lines for each switchboard is restricted. Switchboards share a joint overflow – of limited capacity – to handle calls blocked at the switchboards. The queuing theory proposes a very accurate approach for the determination of fraction of blocked calls (local and overflow), the so-called Equivalent Random Method (ERM), see Wilkinson 1956.

Substituting circuit for bed and call for patient makes this method suitable to analyze IC capacity. Internal trauma patients, who must be admitted to the ICU, cannot be modeled, however, using the ERM standard version. This is why a generalization of ERM was developed. The procedure is as follows. For a single ICU with region-wide beds (modeled as overflow with finite capacity) the fraction of refused regional trauma patients can be determined. It is not possible, however, to determine an analytic expression for the fraction of refused trauma patients per ICU for a region with multiple ICUs and regional beds. Here the ERM is a perfect godsend. For each ICU, the mean and variance of number of patients in the overflow with unlimited capacity are determined. The mean and variance of total number of patients in the overflow then are calculated by summation of the values for the separate ICUs. Using the mean and variance thus determined we construe a single ICU that yields the same values for mean and variance of the number of patients in the overflow, see Figure 2 in which we have replaced two ICUs by a single Equivalent Random ICU. Next, regarding the Equivalent Random ICU thus construed as a regional ICU with the intended capacity, we can determine the fraction of refused regional trauma patients. From this fraction we can also deduce the fractions of refused regional trauma patients per ICU.

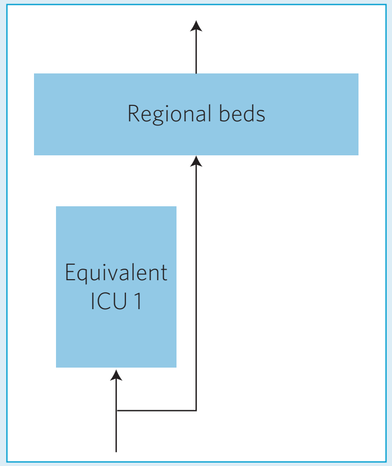


Figure 2 *Equivalent Random IC with regional emergency patients.*

We performed a simulation study intended to evaluate accuracy of the ERM approach. Accuracy was found to be high: key indicators such as mean duration of ICU stay and fraction of refused patients are approximated at a precision within 10% for realistic patient flows. It would seem justified therefore to use the analytic results obtained with the use of ERM for a study into the distribution of capacity between ICU and region-wide ICU.

Rijnmond region

The Rijnmond region was taken as starting point for a detailed study. In this region, Erasmus University Medical Center (Erasmus MC) functions as trauma center. This role puts extra pressure on the available IC capacity in Erasmus MC, resulting in refusal of planned patients. Insight into the benefit of a region-wide approach is therefore of great interest to Erasmus MC. Detailed data on patient flows and fractions of refused patients were obtained for Erasmus MC (EMC). Aggregated data were used for three hospitals in Rijnmond region: Albert Schweizer (AS), Dirksland (D), en Sint Franciscus (SF). IC capacity of these four hospitals is as follows: EMC: 36; AS: 13; D: 5; SF 11. We based our study on a region including these four hospitals. Patient flows in other hospitals in the region were left out of consideration.

A first study concerns assigning the capacity that will become available as regional ICU. In this set-up, capacity of the existing ICUs remains unchanged, as a result of which the fraction of refused planned patients and number of ‘overbeds’ remain unchanged. We set out to determine the required number of extra beds for regional trauma patients in two situations. One, considering patient flows at each ICU separately, and two, the ICUs allowing each other full access to the available regional IC capacity, assuming that regional trauma patients can be assigned unrestrictedly to the available regional beds. Aiming at a refusal rate not exceeding 1%, we find that in the first situation – no cooperation – the total required number of extra beds is 16 (EMC: 9; AS: 3; D: 0; SF: 4). In the second situation – full cooperation – the total required number of extra beds to accomplish the same fraction of refused patients is 11. The distribution of these beds over the hospitals is not dictated by the approach. Synchronization therefore results in a reduction of 5 beds (31%). Other scenarios for patient flows and choice of maximal fractions of refused patients provide similar pictures.

A second optimization problem is the problem of allocating existing IC beds for a regional function. It assumes that each IC allocates part of its capacity to be used as regional IC capacity. Clearly, the regional trauma patients will benefit most from a large region-wide capacity, for only these patients can make use of this. On the other hand, small region-wide capacity would be optimal to planned patients in particular. The analytical model enables us to perform a sensitivity analysis taking into account the goals for refused fractions set on a managerial level.

For that matter, a simulation study would be an alternative approach to sensitivity analysis. The analytical ERM approach has a great advantage over simulation, however, in that it enables to rapidly compute many scenarios for the distribution of beds for different patient flows. For this purpose ERM does not require detailed data on patient flows. The analytical overflow model in addition provides fundamental insight into the nature of the overflow problem.

Conclusions

This newly developed mathematical model facilitates the trade-off between local and region-wide IC capacity. It is comparable to models developed for telecommunication systems. This parallel shows the strength of mathematical modeling and analysis: mathematical models typically can be deployed on a broad scale, even beyond the context in which they have been developed.

From the perspective of the logistics of a single hospital the allocation of capacity to region-wide capacity for trauma patients would not always be advantageous, particularly with regard to acceptance of planned patients. Nevertheless, provided that each ICU in de region allocates beds for regional use in a transparent way, this collaborative effort allows for reaching the set goals for fractions of refused trauma and planned patients at a smaller number of IC beds. Indeed, the patient perspective propagated by TPG/Bakker 2004 calls for the deployment of shared region-wide capacity for trauma patients.

References

1 *Het kan écht. Betere zorg voor minder geld*, TPG/Bakker, 2004.

2 Hautvast, J.L.A., et al. (2001). *Plaats in de herberg, een studie naar determinanten van opname- en ontslagproblemen in* IC-*afdelingen in Nederland.* Julius Centrum voor Huisartsgeneeskunde en Patiëntgebonden Onderzoek, Utrecht.

3 Wilkinson, R.I. (1956). Theories for toll traffic engineering in the USA. *The Bell Systems Technical Journal*, 35, 421–514.

CHAPTER 5

A Simulation Model for Determining the Optimal Size of an Emergency Team on Call in the Operating Room at Night

Jeroen M. van Oostrum, Mark Van Houdenhoven,
Manon M.J. Vrielink, Jan Klein, Erwin W. Hans, Markus Klimek,
Gerhard Wullink, Ewout W. Steyerberg, Geert Kazemie.

Anesthesia Analgesia in Revision

Introduction

Relative to daytime surgical schedules, the nighttime schedules account for much smaller numbers of surgical cases. Irregular hours payments and work-sleep regulations for operating room (OR) staff also contribute to higher costs during the night. Facing OR staff shortages as well, OR department managers must therefore critically appraise nighttime workforce deployment (1).

Appropriate size of the emergency team, with acceptable frequency of calling team members from home, should ensure sound treatment for all patients. Previous studies show that analytical methods can help to decide on numbers of operating and anesthesia nurses needed (2–7). These studies, however, implicitly assume that all patients are operated upon almost immediately, without considering the option to postponing procedures if possible. For surgical procedures to be successful, they should start within a specific time interval, dependent on the nature of the emergency patient. For instance, a patient with abdominal aortic aneurysm must be operated on within 30 minutes after arrival, reimplantation of an amputated finger must take place within 90 minutes after arrival, and perforated gastric ulcer requires intervention within 3 hours after arrival. While studies have defined safety intervals for decision-making during the day (8), the option of postponing operations by a safe time interval during the night shift has not yet been addressed.

This study was designed to determine the optimal OR staff on call at nights by explicit modeling of patients' safety intervals and by discrete-event simulation modeling. The simulation model provided insight in the trade-off between the main outcome measures surgery on time and numbers of times team members are called from home. A case study was performed for the main OR department of Erasmus MC, Rotterdam, the Netherlands.

Data and Methods

Erasmus MC is a tertiary referral centre and has maintained a database with information on all surgical procedures since 1994. The information includes duration of the various procedures, the surgeon and surgical department involved, exact nature of the procedure, patient arrival time, and composition of the team present during the procedure. Anesthesia and surgery nurses prospectively approved these data immediately after a surgical procedure, and surgeons retrospectively approved all data.

In this study we propose a discrete simulation model for determining the optimal size of emergency teams (i.e., anesthesia and surgery nurses) on call at night. The model involves several issues already addressed by others: sequencing of emergency patients (9) and determination of staff requirements (2,3,10). In addition, however, we model medically sound safety intervals for emergency patients. This would provide for postponing an emergency case such that fewer nurses need to come in from home.

To operate on emergency patients, anesthesia and surgery nurses are on call either in the hospital or at home. For this study the hours from 11:00 P.M. through 7:30 A.M. were defined as the night shift. We included the six surgical departments that yearly performed at least eight procedures during the night shift. These are listed in Table 1, each with relevant data on surgical procedures and ICU requirements.

We used simulation as a tool for analysis because of its flexibility to incorporate uncertain operating times and "what if" or scenario analyses (11–14). Several earlier studies have used simulation successfully to assess effects of staff reduction on patients waiting times or staff requirements (2,3,9,10,15). The model was built in eM-Plant (Tecnomatix, Plano, USA) and comprised the following parameters: (a) holding, (b) operating rooms, (c) recovery room, (d) anesthesia nurses (either at home or in the OR department) (e) surgery nurses (either at home or in the OR department), and (f) patients.

Waiting time for emergency patients and the frequency of calling operating and anesthesia nurses from home are the primary outcome measures in this study. These outcome measures combined with the number of nurses in the hospital provide OR management insight in the costs of night shifts and the corresponding waiting time of emergency patients at their department.

Surgical Department	Proportion of all surgical procedures	Duration of surgical procedure Mean	Duration of surgical procedure Variance	Proportion of ICU patients
General surgery[a]	47.1%	156.16	118.45	15.6%
Traumatology	15.9%	146.14	82.10	2.7%
Neurosurgery	15.5%	126.24	72.32	28.6%
Plastic surgery	9.9%	200.32	142.45	10.6%
Gynecology	7.2%	74.05	41.02	3.7%
ENT Surgery	4.4%	90.21	54.88	10%

Table 1 *Data per surgical department over all night shifts in the period 1994–2004 at the main OR department of Erasmus MC. ([a] Including vascular and transplant surgery)*

Modeling

The model started at the beginning of the night shift with an empty recovery room and no patients waiting for emergency surgery (i.e., an empty holding). Recovery room capacity is unlikely to be a bottleneck in the process, since patients recovering from earlier evening shift procedures are typically taken care of by evening shift nurses or recovery nurses. Hence, the assumption of an empty recovery room was valid. The model allowed for the possibility that evening shift procedures (i.e., before 11:00 P.M.) were continuing after start of the night shift. We modeled this by assuming at start a single OR occupied with a probability of 0.40 and with a double OR occupied with a probability of 0.18. Remaining times of the surgical cases running into the night shift were drawn from a lognormal distribution based on the case mix data. The above occupancy probabilities were based on this case mix as well.

We assumed that emergency patients will arrive according to a Poisson distribution, which was modeled time-dependent. Table 2 shows the assumed inter-arrival times for each of the night shift hours, also expressed as mean number of patients arriving in a particular hour. Furthermore we assumed that each patient was instantly available for surgery (i.e., essential tests or scans already having been performed).

Hour of the night shift	Inter-arrival times in minutes	Expressed in mean number of patients per hour
11:00 p.m. – 0.00 a.m.	175	0.34
0.01 a.m.- 1.00 a.m.	204	0.29
1.01 a.m. – 2.00 a.m.	520	0.12
2.01 a.m. – 3.00 a.m.	656	0.09
3.01 a.m. – 4.00 a.m.	1386	0.04
4.01 a.m. – 5.00 a.m.	1782	0.03
5.01 a.m.- 6.00 a.m.	1386	0.04
6.01 a.m.- 7.00 a.m.	891	0.07
7.01 a.m. – 8.00 a.m.	1040	0.06
Mean number of patients per night	–	1.1

Table 2 *Mean inter-arrival time of emergency patients during the night shift*

Type of surgical procedure determined composition of the team required to be present. Complex procedures, such as liver transplantation or unstable polytrauma patients require a large team of two anesthesia nurses and three surgery nurses. Standard procedures require one anesthesia nurse and two surgery nurses. Table 3 shows for each surgical department the proportion of procedures requiring a large team.

Upon arrival of a patient, the model checked availability of an OR and enough members of the emergency team before the end of the safety interval related to the procedure. If so, the patient was operated on, either immediately or when an OR and a team became available. If too few emergency team members were available within the safety interval, the additionally required members were called in from home. We assumed that once assigned to a procedure, a nurse would be occupied for its duration.

Procedure durations were drawn from lognormal distribution s(16) for the surgical departments involved, based on the data set of the case under consideration (see Table 1). After completion of the surgical procedure, team members called in from home were assumed to leave. Patients at this point are assigned to the ICU or the recovery room, given the probability in Table 1. Transport time to the ICU or recovery, or time for the nurses to return the OR, was taken to be 30 minutes. One anesthesia nurse assisting in the procedure also transports the patient to the recovery room. Here, at least two anesthesia nurses are needed to watch patients through the night. If only one anesthesia nurse was available, the second was called in from home. The recovery duration was drawn from a lognormal distribution using a historical mean of 70.2 minutes and a variance of 37.0 minutes, again based on the data of the OR department under consideration. The surgery nurses were assumed to clean the OR and restocked materials after the surgical procedure. Figure 1 schematically depicts the simulation model.

Surgical department	Percentage
General Surgery	30.0 %
Plastic Surgery	35.4 %
Neurosurgery	0.0 %
Traumatology	20.0 %
Gynaecology	17.0 %
ENT surgery	0.0 %

Table 3 *Proportions of surgical procedures requiring a large emergency team*

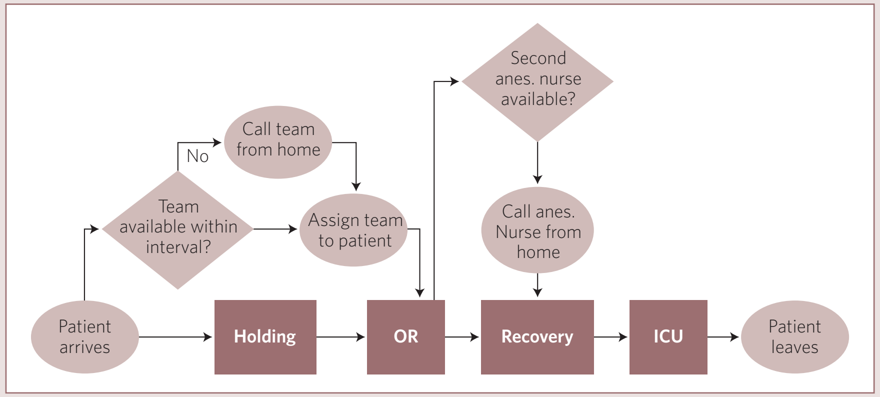


Figure 1 *Concept Process Diagram of the Simulation Model*

We determined four safety intervals based on clinical experience of the surgeons and OR staff in Erasmus MC. Then, based on a surgical department's patient mix and types of the procedures we determined proportions of patients to be assigned to each of the four safety intervals. Table 4 shows these safety intervals for the six Erasmus MC surgical departments involved.

Safety Intervals	General Surgery	Plastic Surgery	Neuro-Surgery	Trauma-tology	Gynae-cology	E.N.T. Surgery
< 30 minutes	15	0	74	15	26	33
< 90 minutes	25	32	14	17	18	6
< 3 hours	20	18	10	24	21	29
< 8 hours	40	50	2	44	35	32

Table 4 *Proportions of emergency patients per surgical department assigned to the four safety intervals*

Scenarios

To evaluate compositions of emergency teams, we defined nine scenarios. Current practice in Erasmus MC (Scenario 1, Table 5) was assumed "safe", seeing that in the past 10 years no emergency patients have been in severe danger because of OR staff shortage or lateness. This scenario was used as the reference scenario against which to evaluate the other eight scenarios. In each subsequent

Scenario	1	2	3	4	5	6	7	8	9
Number of anesthesia nurses (in-house + on call at home)	4+1	3+1	3+1	2+2	3+1	2+2	2+2	2+1	2+1
Number of surgery nurses (in-house + on call at home)	5+1	5+1	4+1	4+1	3+2	3+2	2+2	2+2	2+1

Table 5 *Scenarios of emergency team compositions*

scenario, one nurse was excluded from the night shift or placed on call at home instead of being present at the hospital. Scenarios with fewer staff than available in Scenario 9 were not considered since these would result in excessive waiting time for emergency patients.

We performed sensitivity analyses on the safety intervals. This allows comparison of our discrete-event simulation model with existing methods that do not deploy patient safety intervals, such as the one described by Tucker *et al.* (6). Four alternatives were analyzed; each was constructed by excluding one or more safety intervals. The proportion of patients previously assigned to these intervals was distributed among the remaining safety intervals according to the original ratios. The following alternatives were defined:

1 Excluding safety intervals of eight hours;
2 Excluding safety intervals of three and eight hours;
3 Excluding safety intervals of 90 minutes, three, and eight hours;
4 Excluding safety intervals of 30 minutes;
5 Excluding safety intervals of 30 minutes and 90 minutes.

Note that the third alternative corresponds with Anesthesia Billing as explained in Tucker et al. (6). Further sensitivity analyses were performed on arrival intensity of patients during the night and the likelihood of occupied ORs at 11:00 P.M. This holds the following alternatives:

6 – 10% arrival intensity
7 – 20% arrival intensity
8 – 30% arrival intensity
9 + 10% arrival intensity
10 + 20% arrival intensity
11 + 30% arrival intensity
12 + 25% likelihood of occupied ORs at 11.00 P.M.
13 + 50% likelihood of occupied ORs at 11.00 P.M.
14 – 25% likelihood of occupied ORs at 11.00 P.M.
15 – 50% likelihood of occupied ORs at 11.00 P.M.

Based upon preliminary experiments we tested the alternatives for Scenarios 1 and 6. Before conducting the experiments, the model was validated by comparing the output of scenario 1 with actual practice. The key validation measure was number of times anesthesia or surgery nurses were called from home. Validation was provided by this number in the model being the same as in practice.

The number of runs required to obtain reliable results was determined by the following equation:

$$n_r^*(\gamma) = \min\left\{ i \geq n : \frac{t_{i-1,1-\alpha/2}\sqrt{S^2(n)/i}}{|\bar{X}(n)|} \leq \gamma' \right\}$$

Equation 1 Determination of the number of runs (19)

Where $n_r^*(\gamma)$ is the minimum number of runs for obtaining a relative margin of error of $\bar{X}(n)$, given an average value of. The value $S^2(n)$ represents the variance of $\bar{X}(n)$ and γ is the probability distribution of t, which is set at 0.05. A relative error of 0.1, which is a common value in simulation studies, yields a total of 10,300 days (17). To measure patients' safety, we categorized amounts of time exceeding the safety interval in four categories: 0–10, 11–20, 21–30, and more than 30 minutes after the safety interval.

Results

Table 6 presents proportions of patients treated too late during the night shift. Figures show a steady increase in total percentage from Scenario 1 (current situation) to 6. Scenarios 7, 8, and 9 show a substantial increase of patients treated more than 30 minutes late. Proportions of patients treated too late in Scenario 1, which reflects the current staffing, correspond to the real situation in Erasmus MC, which is an indication for the validity of the approach. Reducing the numbers of anesthesia and surgery nurses following Scenarios 1 to 6 only slightly increases proportions of patients treated too late. For instance, in Scenario 6 the percentage of patients treated 30 minutes after their safety intervals has increased by no more than 2.5 percent points relative to Scenario 1 (1.4% vs. 3.9%). Correspondingly, total percentage of patients treated too late has increased by only 2.3 percent points in Scenario 6 (10.6% vs. 12.9%).

Figure 2 shows percentages of nights the first anesthesia nurse and the surgery nurse is called in from home in the different scenarios, and Figure 3 likewise

Safety interval	SC1	SC2	SC3	SC4	SC5	SC6	SC7	SC8	SC9
Total too late	10.6	11.2	12.7	12.6	12.9	12.9	15.5	17.0	23.1
between 0 min. and 10 min. too late	2.9	2.8	2.8	2.7	2.8	2.8	2.6	2.3	2.2
between 10 min. and 20 min. too late	3.2	3.0	3.1	3.1	3.2	3.2	3.0	2.6	2.3
between 20 min. and 30 min. too late	3.1	2.9	2.9	2.9	3.0	3.0	2.9	2.4	2.2
more than 30 min. too late	1.4	2.4	3.9	4.0	3.9	3.9	7.1	9.7	16.4

Table 6 *Proportions of emergency patients treated too late (SC: Scenario; in procent)*

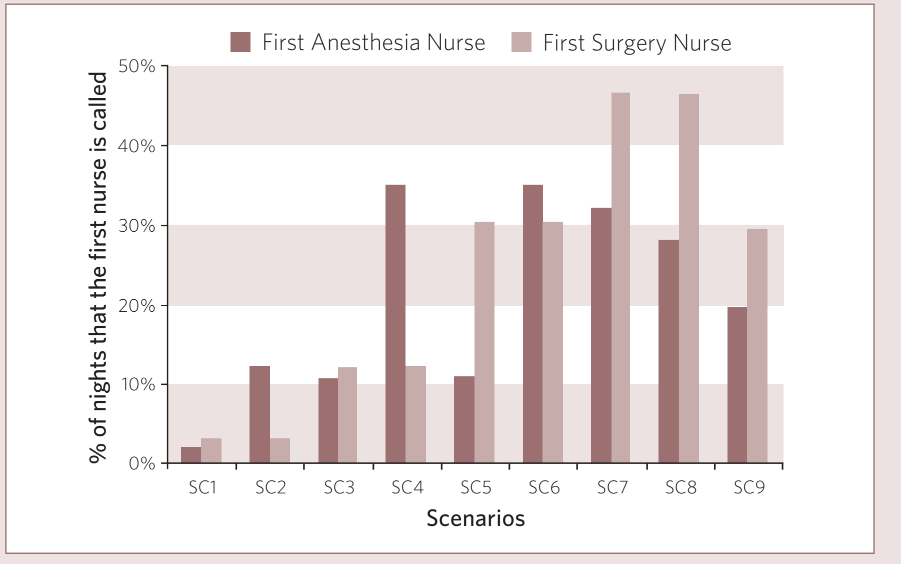


Figure 2 *Proportions of nights in which the first nurses are called in from home*

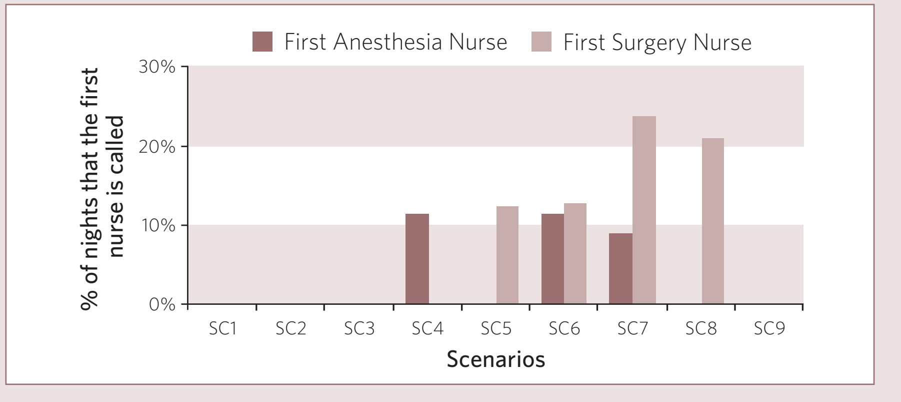


Figure 3 *Proportions of nights in which the second nurses are called in from home*

Reference scenario	Safety interval	SA: Varying safety intervals				
		SA1	SA2	SA3	SA4	SA5
SC1	Total too late	12.0	12.8	23.3	7.8	3.4
SC1	between 0 min. and 10 min. too late	3.2	3.5	4.8	2.2	1.1
SC1	between 10 min. and 20 min. too late	3.5	3.8	5.8	2.8	1.1
SC1	between 20 min. and 30 min. too late	3.5	3.5	6.5	2.1	1.1
SC1	more than 30 min. too late	1.8	2.0	6.1	0.7	0.1
SC6	Total too late	14.6	15.5	27.4	9.6	4.5
SC6	between 0 min. and 10 min. too late	3.1	3.3	3.9	2.3	1.3
SC6	between 10 min. and 20 min. too late	3.5	3.8	5.0	2.7	1.3
SC6	between 20 min. and 30 min. too late	3.4	3.5	5.8	2.4	1.5
SC6	More than 30 min. too late	4.6	4.9	12.7	2.1	0.4

Table 7 *Proportions of emergency patients treated too late given various sensitivity analysis scenarios (SC: Scenario; SA: Sensitivity Analysis; in procent)*

Reference	Safety interval	SA: varying arrival intensity during the night						SA: likelihood of occupied ORs at 23.00 hours			
		SA6	SA7	SA8	SA9	SA10	SA11	SA12	SA13	SA14	SA15
SC1	Total too late	10.6	10.4	10.4	11.9	11.7	11.7	9.7	13.2	7.5	15.1
SC1	between 0 min. and 10 min. too late	3.0	2.7	2.8	3.2	3.0	2.9	2.4	3.6	1.9	3.9
SC1	between 10 min. and 20 min. too late	3.3	3.4	3.2	3.3	3.5	3.2	2.9	3.8	2.3	4.3
SC1	between 20 min. and 30 min. too late	3.1	3.2	3.1	3.4	3.4	3.3	2.9	3.9	2.0	4.6
SC1	more than 30 min. too late	1.2	1.1	1.3	2.0	1.9	2.3	1.6	1.9	1.3	2.2
SC6	Total too late	12.7	12.4	12.2	14.3	14.6	14.8	11.7	15.6	9.4	17.8
SC6	between 0 min. and 10 min. too late	3.0	2.6	2.8	3.0	3.0	3.0	2.3	3.6	1.9	3.8
SC6	between 10 min. and 20 min. too late	3.3	3.3	3.1	3.4	3.6	3.2	2.8	3.8	2.5	4.3
SC6	between 20 min. and 30 min. too late	3.3	3.2	3.3	3.4	3.3	3.4	2.9	3.8	2.2	4.7
SC6	More than 30 min. too late	3.2	3.2	3.0	4.4	4.6	5.2	3.6	4.4	3.0	5.0

Table 8 (con'd of Table 7) *Proportions of emergency patients treated too late given various sensitivity analysis scenarios*

shows this for the second nurses. As from Scenario 4, frequencies of calling team members in from home increase significantly, then sharply decline for Scenario 9. In this scenario significantly more patients are postponed to the day team.

Tables 7 and 8 show the results of the sensitivity analyses. The sensitivity analysis for Scenario 3 (SA3) shows that setting all safety intervals to 30 minutes leads to a substantial increase of patients treated too late. In addition, SA1 to SA5 show that results are sensitive to the use of safety intervals. Outcomes are insensitive to variation in arrival intensity (SA6 to SA11), but results are sensitive to the number of occupied ORs at 23.00 hours (SA12 to SA15).

Discussion

A simulation model was presented to determine optimal size of the emergency team on call during the night, i.e., from 11:00 P.M. through 7:30A.M., using safety intervals of emergency patients. The main contribution of this study as compared to other studies in this field is that it combines aspects of patient safety, uncertainty of the procedure duration, and nocturnal OR staffing in a simulation approach (2,7,18). While this is a single center study, variation of the input parameters showed that the approach can be generalized for use in other centers. To do so, hospitals need to obtain data regarding patient arrival rates and safety intervals. Retrospectively, frequencies per safety interval can be computed per surgical department.

The case study indicated that cost reductions can be realized by reducing the size of the emergency team during the night without jeopardizing patients' safety. This is best illustrated in Scenario 6, with reduction by two surgery nurses and two anesthesia nurses as compared to Scenario 1. The consecutive Scenarios 7 to 9, with even greater reduction, are associated, however, with substantial increase of patients being treated too late. The choice for Scenario 6 potentially makes two surgery and two anesthesia nurses available for the daytime surgical schedules. Overall this would increase productivity of the OR department. Historically, the main OR department in Erasmus MC deployed four anesthesia nurses and five surgery nurses during the night shift, forming two emergency teams permanently present in the OR. Statistics over the past four years, however, indicate a structural overcapacity of these teams. In 45% of the night shifts, no new patients were admitted for surgery after 11 P.M.; on average, 1.1 patients per night were operated on; and in one out of every seven nights, two teams had to work simultaneously to perform all emergency surgeries on time. Changing from scenario 1 to scenario 6 in Erasmus MC would reduce night-shift costs by approximately 27%, corresponding to an annual cost reduction of 275,000 euro.

Recent studies show that labor costs of emergency teams during regular hours, second shifts, and weekends can be significantly reduced, but most authors exclude the night shift, (7,19,20) or focus on single specialty OR departments (2–4). Safety intervals to allow safe postponing of specific emergency cases are not being considered in these studies.

Dexter and O'Neill (7) propose a statistical method to determine the weekend staffing requirement of the OR department. It assumes an expected workload and computes the staffing requirements based on this workload. For all that, it does not incorporate safety intervals that might flatten the workload and hence reduce the number of staff required to be called in from home.

Tucker and colleagues proposed a queuing approach to determine OR staff requirements (6). This approach as well does not address the issue of treating patients on time, and the authors do not incorporate detailed characteristics of surgical departments and the OR department. This approach, therefore, typically overestimates the probability of multiple cases performed at the same time since no delaying of cases within their safety intervals was considered.

Several studies showed that safety intervals or medical triage systems for emergency patients are hard to establish (4,21). In this paper we used safety intervals determined by surgeons of Erasmus MC. We do not claim that these intervals are valid in general, but we show that using such intervals facilitates medical decision-making regarding the treatment of emergency patients. Sensitivity analyses showed the impact of safety intervals. Hence, we recommend all hospitals to consider safety intervals when deciding upon the required staff during night shifts, and not to rely solely on Anesthesia Billing since the latter does not account for safety intervals.

Research has proven that a significant part of the procedures performed during the night shift can be postponed to the day shift (22–24). We think that the approach of using safety intervals helps to pinpoint the procedures that really cannot be postponed. In future research, we would like to investigate the performance of the model with more gradual safety intervals. Such gradual intervals allow the researcher to model the benefits of early treatment in terms of mortality and risks of complications for certain patient categories.

In the model we assumed transport or travel times between the OR and the ICU or the wards to be 30 minutes. Shortening of these times is likely to improve the performance in all scenarios, which in the end allows a further reduction of number of nurses required to be on call in the OR during the night.

In conclusion, this study shows that a discrete simulation model is of use in determining the best size and composition of an emergency team, taking into account the patients' safety. Its flexibility provides for varying the input variables, such as safety intervals frequencies, which indicated the sensitivity of the outcome measures to the safety intervals. Moreover, the approach allows evaluating different scenarios as a means to support complex managerial decision-making. Any hospital that reconsiders its staffing during night shifts should carefully consider the safety intervals of the hospital's patient mix. Using safety intervals and adopting simulation modeling will enable them to reduce staffing during the night without negatively affecting quality of care.

References

1 Simoens S, Villeneuve M, Hurst J. Tackling Nurse Shortages in OECD Countries: OECD 2005.
2 Lucas CE, Dombi GW, Crilly RJ. Neurosurgical trauma call: use of a mathematical simulation program to define manpower needs. J.Trauma 1997;42:818–24.
3 Lucas CE, Middleton JD, Coscia RL. Simulation program for optimal orthopedic call: a modeling system for orthopedic surgical trauma call. J.Trauma 1998;44:687–90.
4 Charalambous CP, Tryfonidis M, Paschalides C et al. Operating out of hours in acute orthopaedics: variations amongst surgeons and regions in the United Kingdom. Injury 2005;36:1306–10.
5 Dexter F, Macario A, Qian F, Traub RD. 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.
6 Tucker JB, Barone JE, Cecere J et al. Using queueing theory to determine operating room staffing needs. J.Trauma 1999;46:71–9.
7 Dexter F, O'Neill L. Weekend operating room on call staffing requirements. AORN Journal 2001;74:664–71.
8 Dexter F, Epstein RH, Traub RD, Xiao Y. Making Management Decisions on the Day of Surgery Based on Operating Room Efficiency and Patient Wating Times. Anesthesiology 2004;101:1444–53.
9 Badri MA, Hollingsworth J. A simulation model for scheduling in the emergency room. IJOPM 1992;13:13–24.
10 Lucas CE, Buechter KJ, Coscia RL. Mathematical modeling to define optimum operating room staffing needs for trauma centers. Journal of American College of Surgeons 2001;192:559–65.
11 De Angelis V, Felici G, Impelluso P. Integrating simulation and optimisation in health care centre management. Eur J Oper Res 2003;150:101–14.
12 Jun JB, Jacobson SH, Swisher JR. Application of discrete-event simulation in health care clinics: A survey. J Oper Res Soc 1999;50:109–23.
13 Dexter F, Macario A, Traub RD. Optimal sequencing of urgent surgical cases. J Clin Monit Comput 1999;15:9.
14 Costa AX, Ridley SA, Shahani AK. Mathematical modelling and simulation for

planning critical care capacity. Anesth Analg 2003;58:320–7.
15 Lowery JC, Davis JA. Determination of operating room requirements using simulation. In: Farrington PA, Nembhard HB, Sturrock DT, Evans GW eds. Proceedings of the 1999 Winter Simulation Conference. Arizona, USA, 1999 1568–72.
16 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.
17 Law AM, Kelton WD. Simulation modeling and analysis Singapore: McGraw-Hill Inc., 2000.
18 Dexter F, Epstein RH, Marsh HM. Costs and risks of weekend anesthesia staffing at 6 independently managed surgical suites. AANA Journal 2002;70:377–81.
19 Epstein RH, Dexter F. Statistical power analysis to estimate how many months of data are required to identify operating room staffing solutions to reduce labor costs and increase productivity. Anesth Analg 2002;94:640–3.
20 Dexter F. A Strategy to Decide Whether to Move the Last Case of the Day in an Operating Room to Another Empty Operating Room to Decrease Overtime Labor Costs. Anesth Analg 2000;91 925–8.
21 Fitzgerald J, Lum M, Dadich A. Scheduling unplanned surgery: a tool for improving dialogue about queue position on emergency theatre lists. Aus Health Rev 2006;30: 219–31.
22 Sherlock DJ, Randle J, Playforth M. Can nocturnal emergency surgery be reduced? BMJ 1984;289:170–1.
23 McKee M, Priest P. Which general surgical operations must be done at night? Ann R Coll Surg Engl. 1991;73:295–302.
24 Yardeni D, Hirschl RB, Drongowski RA. Delayed versus immediate surgery in acute appendicitis: do we need to operate during the night? J Pediatr Surg 2004;39:464–9.

CHAPTER 6

Closing emergency operating rooms improves efficiency

Gerhard Wullink, Mark Van Houdenhoven,
Erwin W. Hans, Jeroen M. van Oostrum,
Marieke van der Lans, Geert Kazemier.

Journal of Medical Systems, accepted, 21 augustus 2007

Introduction

Postponing emergency surgery may increase a patient's risk of postoperative complications and morbidity. Waiting times depend on the speed at which an operating room (OR) can organize its resources to operate upon an emergency patient. A common approach to deal with emergency procedures is to reserve OR capacity; this is believed to increase responsiveness to the arrival of an emergency patient [1;2].

There are two basic policies for reserving OR capacity for emergency patients: in dedicated emergency ORs or in all elective ORs. The first policy, reserving capacity in dedicated emergency ORs, would combine short waiting times with low utilisation of expensive OR capacity. Hence, it is an expensive option, since one or more entire ORs cannot be used for elective surgery. Emergency patients arriving at a hospital that has adopted the first policy will be operated immediately if the dedicated OR is empty and will have to queue otherwise, whereas patients arriving at a hospital that has adopted the second policy can be operated once one of the ongoing elective cases has ended. Other planned cases will then be postponed to allow the emergency operation. Thus, besides influencing waiting times of emergency patients, the choice of either policy will have impact on the amount of overtime and OR utilisation.

Little evidence is available on the performance in terms of waiting times, OR utilisation, and overtime for the policy of reserving capacity for emergency patients in all elective ORs. In this study we determined the best policy to reserve time for emergency patients. We assessed the policies using a discrete-event simulation model for this purpose.

Data and Methods.

Erasmus MC with 1,300 beds is the largest teaching hospital and tertiary referral centre in the Netherlands. It provides for the complete spectrum of surgical procedures, including transplantation and trauma surgery. Of the 34,500 admissions per year, some 20,000 involve a surgical procedure. Data on more than 180,000 surgical procedures have prospectively been collected since 1994, including procedure duration, the procedure name, the procedure type (elective or emergency), and surgical specialty involved. Data had been approved immediately after the surgical procedure by the surgery or anaesthesia nurse. The duration of surgical procedures, both emergency and elective, is assumed to be lognormal [3]. Table 1 shows the descriptive statistics of the central OR department of the Erasmus MC.

Number of different surgical procedure types	328
Mean number of elective cases per day	32
Mean case duration (minutes)	142
Standard deviation of the case duration (minutes)	45
Mean number of emergency cases per day	5
Mean emergency case duration (minutes)	126
Standard deviation of emergency case duration (minutes)	91

Table 1 *Descriptive statistics of the OR in Erasmus MC*

A block planning approach to schedule the elective procedures was assumed [4;5]. We assumed that on average 12 ORs per day, five days per week were staffed and available. The availability of the staffed Ors was limited to 450 minutes per day. Moreover, all ORs were assumed to be multi-functional, i.e., all procedures types can be performed in all ORs.

We developed a discrete event simulation model [6;7], using the simulation software tool eM-Plant (Plano, USA). This simulation model was a representation of the Erasmus MC 12 OR set-up. We simulated days independently of each other. In the first emergency policy, with emergency capacity allocated to one dedicated emergency OR, the remaining free OR time is allocated to exclusively elective ORs. In the second policy, with emergency time allocated to each elective OR, the reserved OR time is distributed evenly over all elective ORs. Figure 1 illustrates these policies.

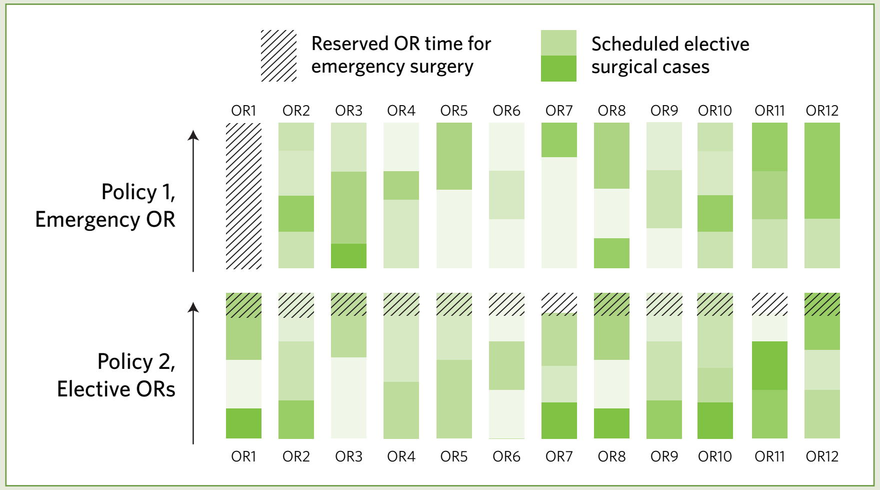


Figure 1 *Visualization of the two studied policies for allocating reserved OR time.*

A given elective OR program forms the starting point for the comparison. Emergency patients arrive according to a Poisson process (with mean inter-arrival time of 1/5 day): inter-arrival times are mutually independent and exponentially distributed. Emergency operation is on a first-come-first-served basis and is performed either after the first completion of an elective operation or at the emergency OR, depending on the policy adopted. Elective procedures planned in an OR are postponed until after the emergency operation and might be executed in overtime. A schedule with elective surgical cases is the input for the simulation model. These schedules are constructed by applying a first-fit algorithm [8]. The first-fit algorithm subsequently assigns for each surgical department separately surgical cases to the first available OR. The resulting surgical case schedule specifies therefore for each OR the elective surgical cases to be performed.

Overtime is defined as the time used for surgical procedures after the regular block time has ended. Efficiency of OR utilisation is calculated as the ratio between the total used operating time for elective procedures and the available time. The sequential procedure [9] to determine the run length of the simulation with a maximum deviation 10% and a reliability of 90% yielded a run length of 780 days, which includes approximately 4000 emergency patients.

Results

Waiting times are plotted cumulatively in Figure 2. In policy 1, with use of a dedicated emergency OR, all 4000 emergency patients were operated on within 7

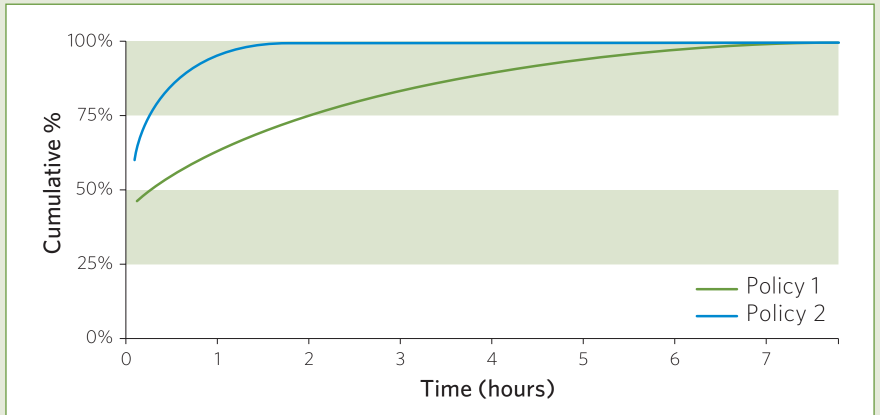


Figure 2 *Cumulative percentage of emergency patients in the two studied policies, treated within a certain (waiting) interval*

hours. The mean waiting time was 74 (± 4.4) minutes. In policy 2, with capacity for emergency surgery allocated to all elective ORs, all 4000 emergency patients were operated upon within 80 minutes. The mean waiting time was 8 (± 0.5) minutes.

Table 2 shows values for the other two performance indicators broken down for type of policy. Efficiency of OR utilisation computed for all ORs in the first policy is 74%; for the second policy it is 77%. Overall, the second policy, with emergency capacity allocated to all elective ORs, substantially outperforms the first policy, with a dedicated emergency OR, on all outcome measures.

Emergency policy	Policy 1	Policy 2
Total Overtime per day (hours)	10.6	8.4
Mean number of ORs with overtime per day	3.6	3.8
Mean emergency patients' waiting time (minutes)	74 (4 4.4)	8 (4 0.5)
OR utilisation* (%)	74	77

Table 2 *Overview results of the outcome measures (* The OR utilisation is the ratio of elective surgery hours performed and the available capacity)*

Discussion

This study showed that reserving capacity for emergency surgery in elective ORs performs better than the policy of a dedicated OR for emergency procedures in a large teaching hospital, based on a discrete-event simulation study with the three performance indicators: waiting time, overtime, and cost effectiveness of the OR.

The policy of allocating OR capacity for emergency surgery to elective ORs requires the OR department to be flexible. Upon arrival of an emergency patient, one of the ORs will have to fit the emergency operation into the elective OR schedule. The patients originally planned will have to be operated on either in another OR or at a later time. This requires flexibility of OR staff and surgeons in dealing with and accepting frequent changes to the original elective surgical case schedule. Also it requires OR to be equipped for all kinds of emergency surgery. Although OR departments that have physical overcapacity, i.e. OR departments where in general some of the ORs are unused, do not face this problem as they may allocate the emergency patient to an empty room that is sufficiently equipped. This way the OR staff have to move to this room, but not all rooms need to be fully equipped for all emergency surgery.

Besides reserving OR capacity for emergency patients, ORs generally need to reserve capacity to cope with the variability in the session durations. In the elective policy, reservation might be shared to increase the flexibility for dealing with unexpected long case duration and emergency surgery, whereas the dedicated policy does not offer the opportunity to use this overflow principle.

In OR departments that have dedicated emergency ORs it is common practice to re-assign staff to elective ORs to deal with temporary staff shortages. Hence, upon arrival of an emergency patient, the team may be incomplete, which implies the patient must wait until the team is complete again, typically when one of the ongoing elective cases ends. This practice considerably reduces the advantage of a dedicated OR.

A dedicated emergency OR may cause queuing of emergency patients, confronting OR management and surgeons with the question which patient should be operated on first. Since such decisions are typically based on medical urgency, trauma procedures or a ruptured abdominal aneurysm will often be given preference over, for instance, fracture surgery. Hence, surgeries of specialties with less acute cases are more likely to be postponed. This would be less so if capacity for emergency surgery were to be allocated to all elective ORs, providing for various emergency patients to be operated on simultaneously.

Implementation of the policy by which emergency capacity is reserved in all elective ORs, requires all stakeholders on the OR to strictly adhere to the policy. In fact, the surgical departments that use a single OR face the so-called prisoner's dilemma. A single surgical department may benefit from not reserving capacity for emergency surgery, whereas this is disadvantageous for all surgical departments together. If one or more surgical departments do not reserve free OR capacity on their own ORs and hence must use reserved capacity of other specialties, the latter face the prisoner's dilemma. Successful implementation, therefore, would require dedication of all surgical departments.

In conclusion, we have compared two policies to reserve OR capacity for emergency surgery. Results obtained from a discrete-event simulation study show that distribution of free OR capacity evenly over all elective ORs performs better than dedicated ORs on measures reflecting quality of patient care, staff satisfaction, and cost-effectiveness. The policy of reserving free capacity can be successfully implemented on ORs only if all stakeholders were to participate. Moreover, besides the quantitative benefits as shown in this paper, it offers several, more soft advantages to improve ways of dealing with the variability that is inherent to medical processes.

References

1 Gallivan S, Utley M, Treasure T, Valencia O. Booked inpatient admissions and hospital capacity: mathematical modelling study. BMJ 2002; 324(7332):280–282.

2 Hopp WJ, Spearman ML. Factory Physics – Foundations of Manufacturing Management. Boston: IRWIN, 1996.

3 Strum DP, May JH, Vargas LG. Modeling the uncertainty of surgical procedure times: comparison of log-normal and normal models. Anesthesiology 2000; 92(4):1160–1167.

4 Strum DP, Vargas LG, May JH. Surgical subspecialty block utilization and capacity planning: a minimal cost analysis model. Anesthesiology 1999; 90(4):1176–1185.

5 Ozkarahan I. Allocation of Surgical Procedures to Operating Rooms. Journal of Medical Systems 1995; 19(4):333–352.

6 Jun JB, Jacobson SH, Swisher JR. Application of discrete-event simulation in health care clinics: A survey. Journal of the Operational Research Society 1999; 50(2):109–123.

7 Costa AX, Ridley SA, Shahani AK. Mathematical modelling and simulation for planning critical care capacity. Anaesthesia 2003; 58:320–327.

8 Hans EW, Wullink G, Houdenhoven M, Kazemier G. Robust Surgery Loading. European Journal of Operational Research 2006; In press.

9 Law AM, Kelton WD. Simulation modeling and analysis. Singapore: McGraw-Hill Inc., 2000.

CHAPTER 7

Fewer ICU refusals and a higher capacity utilization by using a cyclic surgical case schedule

Mark Van Houdenhoven, Jeroen M. van Oostrum,
Gerhard Wullink, Erwin Hans, Johann L. Hurink,
Jan Bakker, Geert Kazemier.

Journal of Critical Care, Accepted, Article in Press,
accepted 25 july 2007, Published online first

Introduction

Mounting health care costs force hospital managers to maximize utilization of scarce resources and simultaneously improve access to hospital services. Efforts are therefore directed at developing planning methods that may deal with these seemingly conflicting objectives (1).

Typically, Dutch hospitals use a block planning approach for surgical scheduling (2). In this approach surgeons of various departments plan their patients in blocks of OR time assigned to their specific department. The method of planning largely determines the utilization of the available OR capacity, and thus the efficiency of the OR department. Implicitly a substantial part of the surgical schedules is basic and performed in a cyclic manner. In addition, the surgical schedule determines the daily number of patients flowing from the OR to the Intensive Care Unit (ICU) postoperatively and hence influences surgical and non-surgical patients' access to the ICU. Scheduling surgical cases without taking into account the inherent ICU or ward occupancy will result in peak demands on these hospital resources. Such peak demands may lead to bed shortages and thus to cancellation of surgical cases or refused ICU admissions for other indications (3). Moreover, the uncertain duration of operations and ICU stay, as well as the unforeseeable emergency cases are complicating factors in surgical scheduling.

Faced with similar challenges regarding availability of services, peak demands, and capacity utilization, industry has developed methods to deal with these problems. One of these methods is to explicit create and use master schedules, which are repetitively used, for subsequent production steps. In such a master schedule repetitive jobs are scheduled leading to improved utilization of scarce resources and coordination in the supply chain (4,5).

Based on this experience the aim of the paper is to assess, by means of computational experiments, the benefit of a comprehensive cyclic case schedule for a university hospital and a virtual hospital with a different case mix.

Materials and methods

Erasmus MC's main OR department consists of sixteen ORs. Planning data have been electronically collected since 1994. From this extensive database we obtained data on frequencies and durations of specific surgical cases, and on the standard deviation of duration of all surgical cases. We also obtained data on related length of stay in the ICU if applicable. Erasmus MC has a tertiary referral case mix. Its mean utilization rate was 85.5%.

The Erasmus MC case mix differed from case mixes of community hospitals. Therefore besides the experiments using Erasmus MC data, experiments are performed using a case mix of a virtual hospital. The procedure for constructing a dataset of the virtual hospital was as follows. The surgical cases from the Erasmus MC dataset were put in descending order of frequency. We then selected surgical cases from the ordered list until half of the total surgery volume of the Erasmus MC was accounted for. Subsequently the frequency was doubled to obtain a case mix with the same volume as the case mix of the Erasmus MC. Table 1 depicts the data for Erasmus MC and the virtual hospital.

We used the block planning method that is currently used in the Erasmus MC as starting point for the analysis (2,6). In the Erasmus MC's block planning method, months in advance blocks are assigned to surgical department that subsequently plan their patients in the available OR time according to strict rules. One of these rules is to plan reserved OR time for emergency patients and the reduction of overtime (7–10). The amount reserved for the latter depends upon a chosen probability, which is in the Erasmus MC set at 31%.

The use of an MSS implies the following three stages in the case scheduling process. First the length of a cycle period is determined and an MSS is constructed for that period. Thereafter surgical departments will assign actual patients to the surgical case types incorporated in the MSS. Patients who require a surgical case that is not incorporated may be scheduled in one of the OR blocks that are kept free. At this stage all patients are assigned to a specific day, for which the clinicians can make the appointments with the patient for surgery. Stage three finally provides for the admission of emergency cases and possible replanning of elective cases.

	Erasmus MC	Virtual Hospital
Data		
Total annual case volume (hours)	18,549	18,861
Mean case duration (minutes)	142	104
Standard deviation (Minutes)	36	30
Mean number of required ICU beds per day	6	5
Assumptions		
Mean OR utilization	85.5%	85.5%
Accepted risk on overtime	31%	31%

Table 1 *Descriptive characteristics of the datasets of Erasmus MC and the virtual hospital*

The focus of this paper is on the first stage, i.e., determining the optional cycle period and the construction of an MSS for such a period. The choice for a particular cycle period was of importance since it determined the number of surgical cases that could be incorporated in the schedule. A longer cycle period leaded to a larger set of surgical cases that is on average performed at least once. Given the cycle length and, consequently the number of case types per cycle, the optimal case schedule was constructed by mathematical optimization techniques. Its ultimate aim was twofold: optimizing the use of OR time and minimizing the peak demands of required ICU beds for elective surgical patients (8). We applied the method of Van Oostrum et al. by which first the OR utilization is maximized by reducing the unused OR capacity, and subsequently the ICU bed requirements are levelled (8). Maximization of OR capacity was accomplished by generating sets of case types that fitted in one OR, such a set is referred to as Operating Room Day Schedule (ORDS). A column-generation approach generated and selected an optimal set of ORDSs (11).

Such an approach starts with an initial set of ORDSs that is generated by a longest processing time heuristic (12). A longest processing time heuristics applied to surgical case scheduling orders at first all case types based on their expected duration. Then the first case in line is selected and scheduled in an empty ORDS, followed by the next ones in line unless the sum of durations exceeds the available OR time of the ORDS under consideration. Upon this moment a new ORDS is created and the heuristic continues adding surgical cases until the capacity limit is exceeded. After all case types are scheduled this way, an initial set of ORDSs is created that covers all case types to be scheduled in the MSS.

Subsequently the unused time (slack) in the ORDSs is calculated and applying Linear Programming (LP) techniques a new ORDS is constructed that may reduce the total capacity needed. This ORDS is added to the available set of ORDSs. A selection of ORDS is made using LP-techniques that covers all-surgical case types; and again the slack in the all ORDS is calculated (8). Using the renewed slack calculation a new ORDS that may reduce the required OR capacity is constructed and added to the existing set of ORDS. These steps are repeated until no ORDS can be constructed that possibly benefit the amount of required OR capacity.

Hence, each ORDS consists of case type that causes a certain bed requirement profile. To reduce peak ICU peak demands, the selected ORDSs were assigned to a specific OR and particular day during the cycle. Bases on an LP formulation of this problem (8) all possible combination of assignment of ORDS to a specific OR and a particular day were considered using computer modelling. The ICU bed demands of the resulting case schedule were calculated based on the ICU requirements of surgical cases performed in the previous cycles and surgical cases performed in current cycle. For this purpose we used the mean ICU length of stay for each case type. See figure 1 for an example of how an MSS might be constructed. The computer-modelling package AIMMS (Paragon decision technology B.V., Haarlem, the Netherlands) was used to construct the MSS for both Erasmus MC and the virtual hospital.

Any surgical case that was not incorporated in an MSS was scheduled following the current Erasmus MC scheduling method, resulting in similar performance measures. The value of different MSSs was assessed by two outcome measures: the increase in OR utilization and the levelling of the number of ICU beds occupied by elective surgical patients. For both Erasmus MC and the virtual hospital cycle periods of 1, 2, and 4 week were examined.

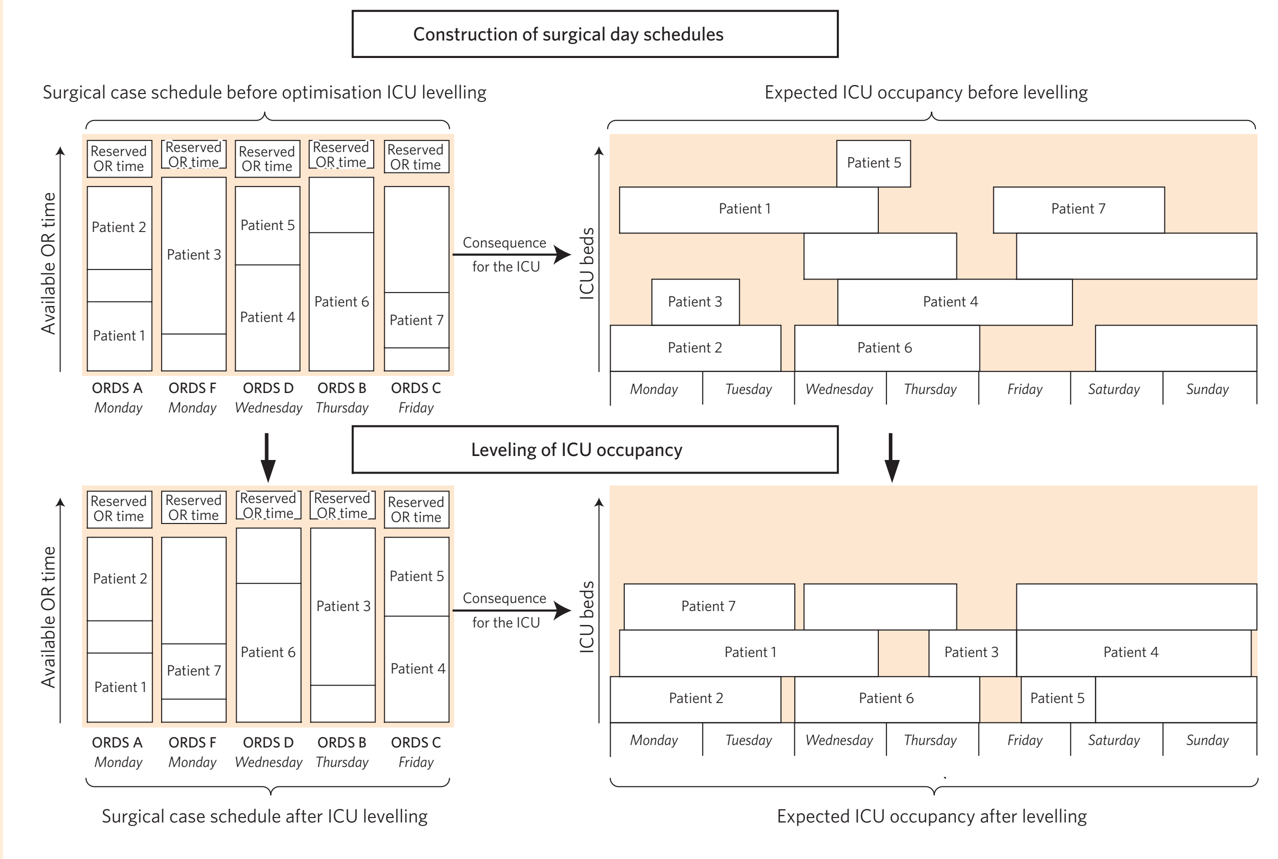


Figure 1 *Overview methodology for construction of an MSS*

	Cycle period			
	1 year	4 weeks	2 weeks	1 week
Proportion of surgical cases incorporated in an MSS				
Erasmus MC	100%	53%	42%	27%
Virtual Hospital	100%	80%	75%	62%
Proportion of the ICU demand of surgical patients determined by an MSS				
Erasmus MC	100%	45%	38%	17%
Virtual Hospital	100%	74%	69%	68%

Table 2 *Description of the influence of length of the cycle period on the proportion of surgical cases incorporated in an MSS and on the proportion of the ICU demand of surgical patients*

Results

Use of an MSS can improve OR utilization considerably by up-to 6.3 percent point. Simultaneously the ICU workload from such an MSS can be optimal levelled, resulting in less surgery cancellation and fewer ICU refusals. The length of a single MSS cycle has a strong influence on the obtainable improvement of OR capacity. Also the virtual hospital potentially has benefits more than the Erasmus MC does (see Figure 2).

Figure 3 presents for the Erasmus MC a comparison of the ICU demand when it used an MSS, with a two-week cycle period, compared to a situation when no MSS is used. Comparable results hold for other cycle periods of the Erasmus MC and the virtual hospital.

Data analysis yields that the cycle period is important for the proportion of surgical cases incorporated in the MSS as well as that of total related ICU workload. The MSSs for Erasmus MC incorporated fewer cases than that for the virtual hospital. Also shorter cycle periods resulted in smaller proportions.

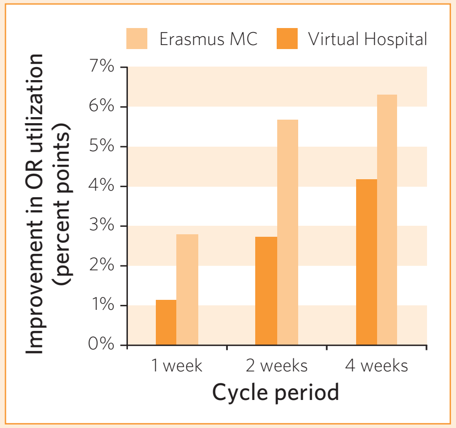


Figure 2 *OR utilization for both hospitals and various cycle periods*

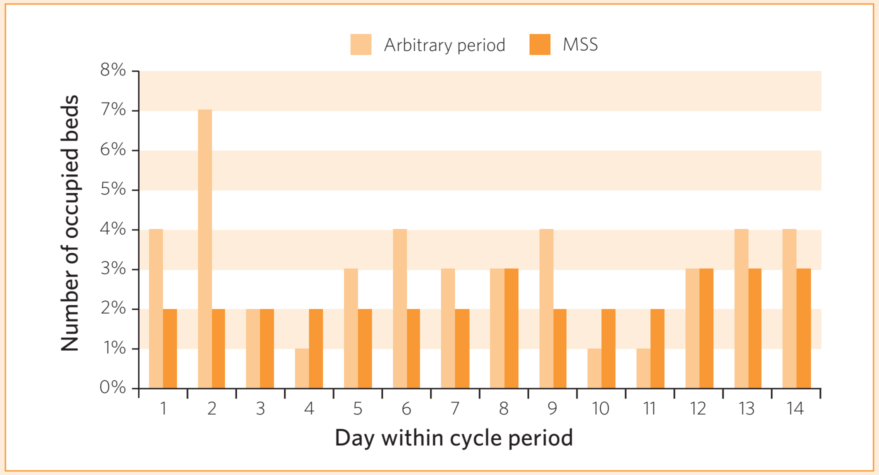


Figure 3 *Number of occupied ICU beds by patients for which a surgical case was incorporated in an MSS with a cycle period of two weeks compared to an arbitrary chosen two-week period (03/06/2004 until 16/06/2004) in the Erasmus MC. Note that approximately 70% of the ICU demand of elective surgical patients is incorporated in the MSS (see Table 2).*

Discussion

The aim of this paper was to determine the benefits of a Master Surgical Schedule, in terms of improved OR utilization and levelling of ICU workload. Computational experiments showed for the Erasmus MC and a virtual hospital that a cyclic case schedule is indeed able to reduce peak demands on the ICU while at the same time it increased OR utilization. Apparently the seemingly conflicting goals of efficiency and access to hospital services can be optimised simultaneously.

Existing literature on surgical case and ICU scheduling is mostly concerned with scheduling of add-on cases, emergency cases, and allocation of OR and ICU time to departments (13–15). Only a few authors have investigated the use of cyclic surgical case scheduling approaches (16–18). None of them, however, proposes, a case scheduling method that actually schedules individual surgical case types, accounts for uncertain case durations, and levels the associated workload on ICUs. Hence, MSSs described in this paper enriches the available literature and available case scheduling methods.

We assumed that the Erasmus MC block planning method was used. This implies that OR time is reserved to deal with emergencies and to lower the risk on overtime. Hence, 100% utilization is not obtainable. A higher accepted risk on overtime results in a higher norm utilization. In combination with the assumption that surgical cases that were not incorporated in an MSS were scheduled following the current Erasmus MC practice, the potential improvement may therefore differ for other hospitals depending on their choice to accept overtime and their current OR scheduling practice.

Like the durations of surgical cases, length of stay on the ICU and surgical wards may be highly unpredictable, particular in a tertiary referral environment. A system that guarantees no cancellation of surgical cases needs a considerable amount of reserve capacity (10). Unless this capacity is available, levelling of bed requirements by taking into account mean length of stay reduces the probability of peak demands. This helps to reduce the number of case cancellation. An adequate registration system is therefore indispensable to predict surgical duration and bed usage. Note that levelling of ICU bed requirements only concerns the proportion of surgical cases incorporated in an MSS and that therefore the obtained benefits strongly depend on the proportion of ICU bed requirements incorporated in an MSS. The remaining part of the ICU bed requirements might be levelled according to other principles such as the method of Kim and Horowitz (19).

When a single surgical department schedules its patients independently from other departments, the result is a sub-optimal schedule in terms of ICU demands and OR utilization. A more flexible hospital organisation and cooperation between different surgical departments may further improve the surgical schedules in terms of OR utilization. An MSS as described in this paper offers the opportunity to integrate such flexibility in the care pathway and hence optimize OR utilization and level ICU demand.

The use of ORs by various surgical departments on the same day has large organisational implications such as the requirements for specialized equipment, multi employable personnel in all ORs, and possibly longer changeover times. Moreover, the daily activities of clinicians are influenced by the unpredictable durations of surgical cases of other surgical departments. OR utilization is higher when surgical cases of multiple surgical departments can be scheduled in the same OR, on the same day (9). A hospital should make a trade-off between OR utilization and the flexibility to schedule surgical cases from multiple specialties in the same OR on the same day. Nevertheless, a cyclic planning approach that includes the use of an MSS is also applicable to a single surgical department.

Clinicians are responsible for the patient scheduling, which is a requirement for implementation. In addition, most clinicians already have a repetitive schedule. The same type of patients is every week operated on the same day. An MSS offers the opportunity to optimize OR utilization and level ICU bed requirement for all clinicians together. Therefore it functions as communication tool between planners, clinicians, and other services within hospitals for which an MSS structures for example material coordination. Consequently, the week-to-week case scheduling requires less effort and the administrative burden on medical staff is lowered since an MSS provides a substantial part of the final surgical schedule.

Our findings show that the proposed cyclic OR planning policy results in a leveled outflow of patients towards the ICU. While in this study we have focussed on reduction of surgical case cancellation due to ICU bed shortages, levelling of other resource requirements might be beneficial for other aspects of a hospital's organization too, for example required intra-operative navigations systems, numbers of fluoroscopy equipment, availability of beds on the ward, and fluctuations in the required number of postoperative CT scans.

References

1 Dexter F, Epstein RH. Operating Room Efficiency and Scheduling. Current Opinion in Anaesthesiology 2005;18:195–8.
2 Magerlein JM, Martin BM. Surgical Demand Scheduling: A Review. Health services research 1978;13:418–33.
3 McManus ML, Long MC, Copper A, et al. Variability in Surgical Caseload and Access to Intensive Care Services. Anesthesiology 2003;98:1491–6.
4 Schmidt E, Dada M, Ward J, Adams D. Using cyclic planning to manage capacity at ALCOA. Interfaces 2001;31:16–27.
5 Tayur S. Improving Operations and Quoting Accurate Lead Times in a Laminate Plant. Interfaces 2000;30:1–15.
6 Ozkarahan I. Allocation of Surgeries to Operating Rooms by Goal Programming. Journal of Medical Systems 2000;24:339–78.
7 Hopp WJ, Spearman ML. Factory Physics – Foundations of Manufacturing Management. Boston: IRWIN, 1996.
8 Van Oostrum J, van Houdenhoven M, Hurink J, et al. A model for generating Master Surgical Schedules to allow cyclic scheduling in operating room departments. OR Spectrum 2006; In press.
9 Hans EW, Wullink G, Houdenhoven M, Kazemier G. Robust Surgery Loading. European Journal of Operational Research 2006;In press.
10 Gallivan S, Utley M, Treasure T, Valencia O. Booked inpatient admissions and hospital capacity: mathematical modelling study. BMJ 2002;324:280–2.
11 Barnhart C, Johnson EL, Nemhauser GL, et al. Branch-and-price: Column generation for solving huge integer programs. Operations Research 1998;46:316–29.
12 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.
13 Zhou J, Dexter F. Methods 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;89:1228–32.
14 Gerchak Y, Gupta D, Henig M. Reservation Planning for Elective Surgery Under Uncertain Demand for Emergency Surgery. Management Science 1996;42:321–34.
15 Ozkaraham I. Allocation of Surgeries to Operating Rooms by Goal Programming. Journal of Medical Systems 2000;24:339–78.
16 Beliën J, Demeulemeester E. Building cyclic master surgical schedules with leveled resulting bed occupancy. European Journal of Operational Reseach 2006;In press.
17 Blake JT, Donald J. Mount Sinai Hospital Uses Integer Programming to Allocate Operating Room Time. Interfaces 2002;32:63–73.
18 Vissers JMH, Adan IJBF, Bekkers JA. Patient mix optimisation in cardiothoracic surgery planning: a case study. IMA Journal of Management Mathematics 2005;16: 281–304.
19 Kim SC, Horowitz I. Scheduling hospital services: the efficacy of elective-surgery quotas. Omega 2002;30:335–46.

CHAPTER 8

Influence of Cardiac Risk Factors and Medication on Length of Hospitalization in Patients Undergoing Major Vascular Surgery

Marleen A. van de Pol, Mark Van Houdenhoven, Erwin W. Hans, Eric Boersma, Jeroen J. Bax, Harm H.H. Feringa, Olaf Schouten, Marc R.H.M. van Sambeek, Don Poldermans.

The American Journal of Cardiology, Volume 97, 15 May 2006, Pages 1423–1426

Because health care costs are increasing, hospitals are forced to use their scarce resources as efficient as possible. A shorter in-hospital length of stay (LOS) increases the average income for each bed per day as LOS accounts for 31 % of hospital costs in surgical inpatients [1]. Major vascular surgery is commonly performed and associated with a long and variable LOS, due to postoperative events, such as heart failure, myocardial infarction, and stroke. Research has been primarily focused on the identification of risk factors that are associated with a increased perioperative cardiac event rate, in order to improve postoperative outcome by treatment of underlying risk factors. Recently, a number of patients characteristics was identified that are associated with an increased risk of prolonged LOS due to perioperative events [2–7]. Given the fact that recent studies suggested that cardioprotective medical therapy (i.e. statin, aspirin, beta-blockers) was associated with a reduced risk of postoperative complications [8–16], we hypothesized that a model, based on cardiac risk factors and cardiopreventive medical therapy, can predict LOS more accurately, so a more efficient hospital planning can be made.

The Erasmus Medical MC is a metropolitan university hospital that acts as a tertiary referral centre for approximately 30 affiliated hospitals. Between 1990 and 2004, 2.374 noncardiac vascular surgical procedures, including aortic aneurysm repair, carotid endarectomy and peripheral vascular surgery, were performed in patients above the age of 15 years in our centre. We excluded 484 (20,4%) procedures that resulted in a LOS exceeding 90 days. LOS was defined as days from date of admission to the hospital to date of discharge from the hospital. Patients were included as many times as they were admitted for vascular surgery, on the understanding that these procedures were >30 days apart. Thus, the operation (not the patient) was the unit of analysis, which is consistent with clinical practice.

Based on hospital records and personal interviews at the time of surgery, a medical history was recorded. Each patient's medical history is classified according to the ninth International Classification of Diseases (ICD–9). For this study, we used medical conditions that have been associated with an increased risk of perioperative cardiovascular complications, including diabetes mellitus (ICD–9 250), myocardial infarction (ICD–9 410, 411, and 412), angina pectoris (ICD–9 413 and 414), prior heart failure (ICD–9 428), cerebrovascular accident (ICD–9 430), renal disease (ICD–9 580), hypertension (ICD–9 401), and Chronic Obstructive Pulmonary Disease (COPD) (ICD–9 490–496). Medical therapy (i.e. statins, beta-blockers and aspirin) were noted in all patients.

Data analysis was performed using SPSS 11.5 software. Continuous data are described as median values and corresponding 25th and 75th percentiles, and dichotomous data are described as numbers and percentages. Patient characteristics are described in Table 1. Length of hospital stay was not normally dis-

Factor	n	Median Length of In-Hospital Stay (quartiles)	P Value
Male gender	1665 (88%)	13 (7.25-22)	0.183
Age >=70 years	928 (49%)	15 (8-23)	<0.001
Prior myocardial infarction	612 (32%)	13 (7-21)	<0.001
Angina Pectoris	370 (20%)	14 (8-21)	0.156
Prior heart failure	119 (6%)	17 (10-29)	<0.001
Prior stroke	350 (19%)	9 (6-18)	<0.001
Chronic Obstructive Pulmonary Disease	326 (17%)	17 (12-26.25)	<0.001
Renal failure	125 (7%)	16 (11-27.5)	<0.001
Diabetes mellitus	205 (11%)	14 (7.5-24)	0.321
Hypertension	908 (48%)	14 (8-21)	0.037
Hypercholesterolemia*	262 (14%)	12.5 (8-17)	0.129
Prior Coronary Artery Bypass Graft	257 (14%)	15 (8-22)	0.070
Percutaneous Coronary Intervention	66 (3%)	12 (6-21.25)	0.544
Statins	473 (25%)	12 (7-17)	<0.001
Diuretics	300 (16%)	15 (9-24)	<0.001
Angiotensin Converting Enzyme inhibitor	547 (29%)	14 (8-22)	0.053
Calcium antagonists	636 (34%)	14 (9-21)	<0.001
Nitrates	321 (17%)	15 (9-24.5)	<0.001
Beta-blockers	618 (33%)	14 (8-20)	0.219
Digoxin	44 (2%)	13.5 (7.25-22.75)	0.775
Aspirin	789 (42%)	8 (6-16)	<0.001
Warfarin derivatives	292 (15%)	15 (9-25)	<0.001
Urgent surgery	169 (9%)	14(9-22)	0.527
Aortic surgery	1021 (54%)	16(12-23)	<0.001
Carotid surgery	534 (28%)	7(5-8)	<0.001
Peripheral surgery	1057 (56%)	14(9-23)	<0.001

Table 1 *Patient characteristics (* Is removed from analysis due to high correlation with statins)*

tributed. Therefore, the relation between clinical characteristics, cardiovascular medication and LOS was first evaluated using Mann-Whitney univariate tests. For these tests, a P-value <0.05 was considered statistically significant.

All variables entered the multivariate state, irrespective of results of univariate analysis. We used backward linear regression to construct a multivariate model that predicted LOS. All variables entered the equation and were subsequently removed, until all remaining variables had a P-value<0.10.

A total of 2.374 patients, excluding 484 patients who were admitted >90 days to the hospital, mostly because they were on a waiting list for a nursing home facility or rehabilitation centre, were enrolled in the study. From these patients, 1021 underwent aortic surgery, 1057 underwent peripheral vascular surgery, and 534 patients underwent carotid endarereictomy. In-hospital mortality was 7.2% (137 of 1.890 patients).

Univariate analysis showed that elderly age, prior myocardial infarction, congestive heart failure, prior cerebrovascular accident, hypertension, renal dysfunction, and COPD were associated with a prolonged LOS. Importantly, statin and aspirin use were associated with a reduced LOS. All variables were then entered in multivariable analysis. The model that was developed from these factors explained 14.1% of variance in LOS. Outcomes of multivariate analysis are presented in Table 2 and graphically represented in Figure 1.

The main finding of this study was an association between LOS and clinical risk factors such as age, heart failure and COPD in patients undergoing major vascular surgery. Importantly, medical therapy, such as aspirin and statin use, commonly associated with a reduced post-operative morbidity, was also associated with a reduced LOS. An effect of beta-blockers on LOS could only be observed in high-risk patients with proven coronary artery disease. This is in line with the cardioprotective effect of beta-blockers on post-operative cardiac events, which also demonstrated a beneficial effect only for those patients with multiple cardiac risk factors or a positive dobutamine stress echocardiography as a marker of significant coronary artery disease [12]. The combined information of risk factors and medical therapy enables the treating physician to plan LOS of individual vascular surgery patients more efficiently. Patients >= 70 years of age, with a previous myocardial infarction and angina pectoris had an extended stay, compared to those without risk factors. However, if aspirin, statins and beta-blockers were prescribed the LOS could be reduced.

The relation between the risk factors and LOS is complex. Elderly age is a well known risk factor associated with an increased LOS as comorbid conditions are more prevalent in the elderly, and importantly early discharge depends not only on the adequate support systems at home, but also on the availability of skilled nursing facilities. These results are confirmed in a Veteran Affair study evaluating more than 8.000 patients showing a prolonged LOS in patients of

Factor	Effect on LOS (days)	Significance	95% CI	
Constant	*9.756*	*<0.001*	*5.678*	*13.833*
Male gender	-2.282	<0.001	-3.582	-0.982
Prior heart failure	2.877	0.031	0.263	5.490
Chronic Obstructive Pulmonary Disease	1.871	0.026	0.222	3.519
Hypertension	-1.212	0.041	-2.376	-0.047
Diabetes mellitus	2.341	0.021	2.313	0.021
Renal failure	2.349	0.069	-0.181	4.879
Age (years)	0.141	<0.001	0.088	0.194
Statins	-3.153	<0.001	-4.581	-1.725
Aspirin	-1.171	0.085	-2.501	0.160
Aortic surgery	2.823	<0.001	1.254	4.392
Carotid surgery	-7.078	<0.001	-9.108	-5.047
Peripheral surgery	2.273	0.004	0.736	3.810

Table 2 *Results of multivariate analysis*

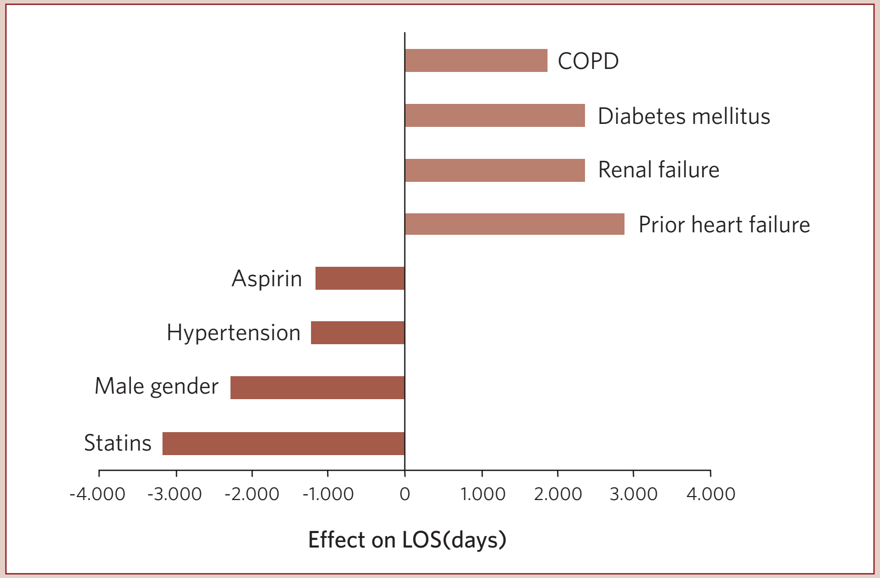


Figure 1 *Factors that influence LOS*

80 years or older after major vascular surgery [4]. Another important risk factor is COPD. Patients with COPD, based on a preoperative pulmonary function test (Forced Expiratory Volume in 1 second < 70%), had a prolonged LOS. Such pulmonary complications may result in a prolonged mechanical ventilation and stay at the intensive care. Whether preoperative evaluation and perioperative medical therapy such as steroids will shorten LOS, has yet to be determined, although the recent introduction of less invasive endovascular procedures may have a potential beneficial effect on pulmonary function in those patients with COPD. Cardiac risk factors such as heart failure are a well known risk factor for postoperative morbidity and prolonged LOS [4,5,7]. Importantly, cardiovascular medical therapy such as statins, aspirin, and beta-blockers in high-risk patients was associated with a reduced LOS. The effect of beta-blockers on perioperative morbidity in high-risk patients is well known. As shown by the study of Boersma et al, beta-blockers were associated with an improved postoperative outcome, in high risk patients, those with two or more risk factors and especially those with proven coronary artery disease [13]. Also a recent study of Powell et al showed that beta-blockers decreased the time from surgery to discharge (LOS) in patients undergoing infrarenal vascular surgery [16]. The potential effect of statins and aspirin need to be confirmed in large prospective studies, evaluating potential side effects such as an increased bleeding tendency or myopathy in relation with a reduced LOS.

A limitation of the study is that diabetes mellitus was noted as a dichotomous variable, while studies in patients undergoing cardiac surgery have shown that regulation of diabetes mellitus in associated with postoperative outcome [17]. Therefore we were unable to assess the influence of diabetes regulation (tight vs. non-tight) on LOS[17,18].

In conclusion, the model we developed explains 14,1% of variance in LOS and enables the treating physician to plan surgical procedures more efficiently. Other factors that might explain variance in LOS are intraoperative factors (i.e. blood loss and surgical complications), and postoperative care (i.e. availability of a nursing home). The effect of these intra- and postoperative factors is very large, which explains the relatively low explained variance. However, all variables included in the multivariate stage showed to have a significant influence on LOS. Therefore this information should be used in the preoperative planning stage. In order to improve planning for patients undergoing major vascular surgery all risk factors (pre-, intra- and postoperative) have to be included in a prognostic model which will be further refined on an individual patient base during hospital stay.

References

1 Macario A, Vitez TS, Dunn B, McDonald T. Where are the costs in perioperative care? Analysis of hospital costs and charges for inpatient surgical care. *Anesthesiology* 1995; 83:1138–1144.

2 Chang JK, Calligaro KD, Lombardi JP, Dougherty MJ. Factors that predict prolonged length of stay after aortic surgery. *Journal of Vascular Surgery* 2003; 38:335–339.

3 Collins TC, Daley J, Henderson WH, Khuri SF. Risk factors for prolonged length of stay after major elective surgery. *Ann Surg* 1999; 230: 251–259

4 Kazmers A, Jacobs L, Perkins A. The impact of complications after vascular surgery in veterans affairs medical centers. *J Surg Research* 1997;67:62–66.

5 Kertai MD, Klein J, van Urk H, Bax JJ, Poldermans D. Cardiac complications after major vascular surgery. *Acta Anaesthesiol Scand* 2003; 47:643–654.

6 Bartholomew BA, Harjai KJ, Grines CL, Boura JA, Grines LL, Stone GW, Cox DA, Brodie BR, O'Neill WW. Variation in Hospital length of stay in patients with acute myocardial infarction undergoing primary angioplasty and the need to change the diagnostic related group system. *Am J Cardiol* 2003; 92:830–833.

7 Mangano DT, Goldman L. Preoperative risk assessment of patients with known or suspected coronary disease. *N Engl J Med* 1995; 333: 1750–1758.

8 Kertai MD, Boersma E, Westerhout CM, Klein J, van Urk H, Bax JJ, Roelandt JRTC, Poldermans D. A combination of statins and beta-blockers is indepentdently associated with a reduction in the incidence of perioperative mortality and nonfatal myocardial infarction in patients undergoing abdominal aortic aneurysm surgery. *Eur J Vasc Endovasc Surg* 2004; 28:343–352.

9 Kertai MD, Boersma E, Klein J, van Sambeek M, Schouten O, van Urk H, Poldermans D. Optimizing the prediction of perioperative mortality in vascular surgery by using a customized probability model. *Arch Intern Med* 2005; 165:898–904.

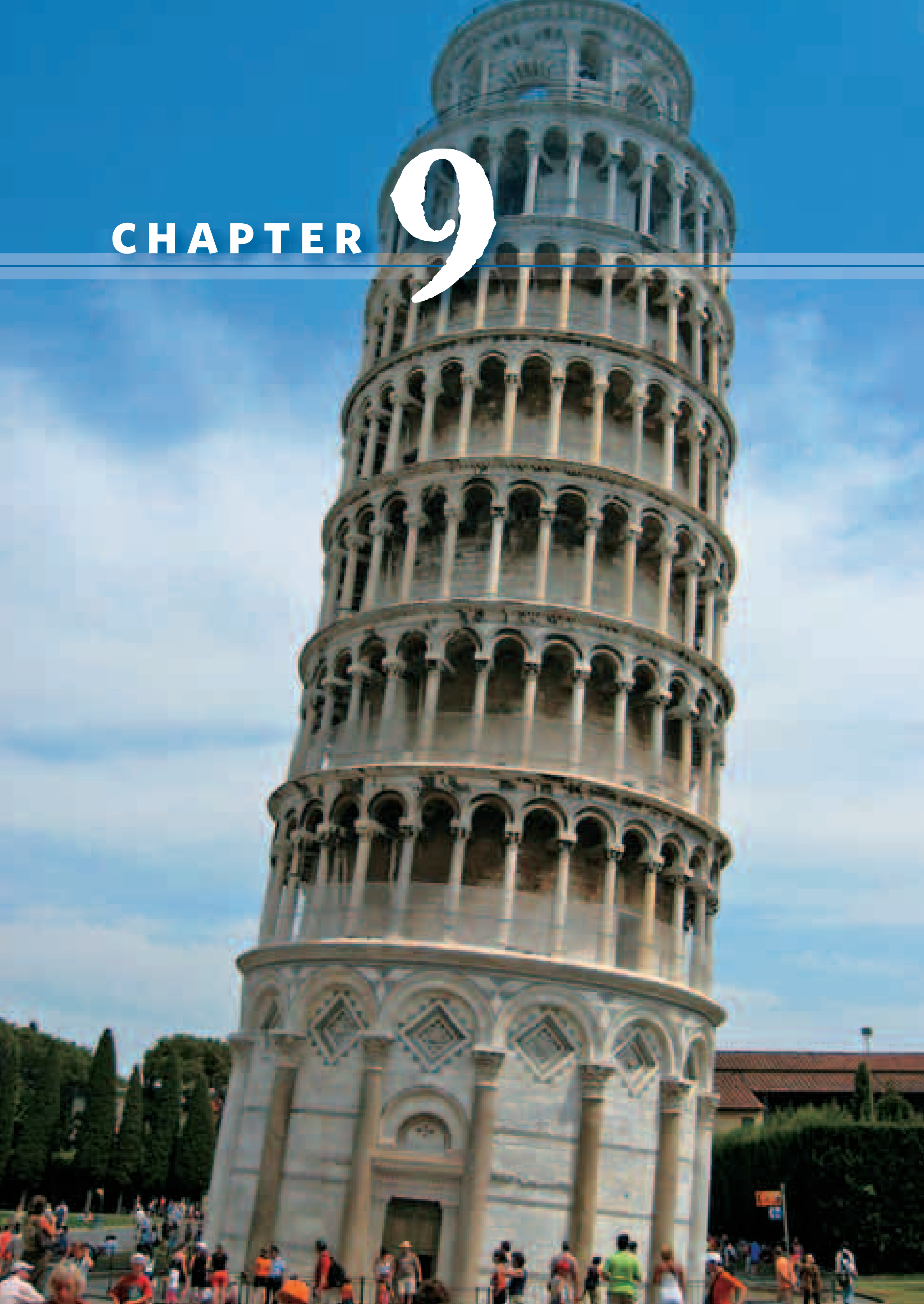
10 Poldermans D, Bax JJ, Kertai MD, Krenning B, Westerhout CM, Schinkel AF, Thomson IR, Lansberg PJ, Fleisher LA, Klein J, van Urk H, Roelandt JR, Boersma E. Statins are associated with a reduced incidence of perioperative mortality in patients undergoing major noncardiac surgery. *Circulation* 2003; 107: 1848–1851.

11 O'Neil-Callahan K, Katsimaglis G, Tepper MR, Ryan J, Mosby C, Ioannidis JPA, Danias PG. Statins decrease perioperative cardiac complications in patients undergoing noncardiac vacular surgery. *J Am Coll Cardiol* 2005; 45:335–342.

12 Durazzo AES, Machado FS, Ikeoka DT, De Bernoche C, Monachini MC, Puech-Leão P, Caramelli B. Reduction in cardiovascular events after vascular surgery with atorvastatin: a randomised trial. *J Vasc Surg* 2004; 39:967–976.

13 Poldermans D, Boersma E, Bax JJ, Thomson IR, van de Ven LL, Blankensteijn JD, Baars HF, Yo TI, Trocino G, Vigna C, Roelandt JR, van Urk H. The effect of biso-

prolol on perioperative mortality and myocardial infarction in high-risk patients undergoing vascular surgery. *N Eng J Med* 1999; 341:1789–1794.
14 Mangano DT, Layug EL, Wallace A, Tateo I. Effect of atenolol on mortality and cardiovascular morbidity after noncardiac surgery. *N Engl J Med* 1996; 335: 1713–1720.
15 Auerbach AD, Goldman L. Beta-blockers and reduction of cardiac events in noncardiac surgery. *JAMA* 2002; 287:1435–1444.
16 POBBLE Trial Investigators. Perioperatibe Beta-blockade (POBBLE) for patients unidergoing infrarenal vascular surgery: Results of a randomized double-blind controlled trial. *J Vasc Surg* 2005;41:602–609.
17 Van den Berghe G, Wouters P, Weekers F, Verwaest C, Bruyninckx F, Schets M, Vlasselaers D, Ferdinande P, Lauwers P, Bouillon R. Intensive insulin therapy in critically ill patients. *N Engl J Med* 2001;19:1359–1367.
18 Van den Berghe G. How does blood glucose control with insulin save lives in intensive care? *J Clin Invest* 2004;114:1187–1195.



CHAPTER 9

Improved efficiency by applying bin packing and portfolio techniques to surgical case scheduling

Mark Van Houdenhoven, Jeroen M. van Oostrum,
Erwin W. Hans, Gerhard Wullink, Geert Kazemier.

Anesthesia Analgesia, September 2007, Volume 105, No.3, p. 707–714

Introduction

Optimal use of scarce and expensive facilities such as operating rooms (ORs) requires efficient planning. The Erasmus University Medical Center (Erasmus MC), Rotterdam, the Netherlands, to this aim developed an OR business model based on controlled surgical case scheduling and management contracts. OR department managers nevertheless still explore new ways to improve OR efficiency.

The main inpatient OR department in Erasmus MC is run as a facilitating department that provides staffed and fully equipped ORs for the various surgical departments. A block planning approach has been adopted, in which blocks of OR time are made available to surgical departments in advance (1,2). Departments may only assign patients to OR blocks that were made available to them. These organizational barriers regrettably result in suboptimal use of OR time. The OR business model furthermore incorporates the annual management contracts specifying the yearly amounts of OR time available for each surgical department. Capacity for emergency cases and uncertainty of case durations is accounted for by determining target OR utilizations for each surgical department independently. Any surgical case schedule therefore must include free OR time, or: planned slack. Since target utilizations differ, planned slack also differs between surgical departments.

In summary, for the planning of surgical cases the surgical departments must adhere to the following rules:

1. Submit elective case schedules two weeks in advance;
2. Maximize use of OR time and not exceed block times;
3. Plan elective cases using historical mean case durations;
4. Include planned slack to deal with emergency cases and variability of case durations.

Provided these rules are adhered to, the OR department "guarantees" that all scheduled surgical cases and emergency cases will be executed, whatever happens during the day. Moreover, applying these rules consequently helps surgical departments in their yearly contract negotiations about OR time with the hospital board.

The hypothesis to be tested was: combining advanced mathematical algorithms with lowering of organizational barriers between surgical departments improves OR efficiency. Several methods to improve efficiency have been proposed in the literature. Strum *et al.* (3) reported benefit of approaching the OR planning problem as a newsvendor problem. Dexter *et al.* (4) recently showed the benefits of various approaches to surgical case scheduling. A broad overview of relevant literature is presented by McIntosh *et al.* (5). Mathematical algorithms to optimize surgical case schedules is a widely researched topic (3,6–8). Several studies addressed the application of bin-packing techniques such as the Best Fit Descending heuristic (9,10) or Regret Based Random Sampling (11), yet within single departments. Finally, there is evidence that approaching the OR scheduling problem as a portfolio problem (12) may deal with the unpredictability of case durations and improve efficiency this way (11). Similar portfolio techniques are already in use for case mix management problems (13,14).

Given the business model employed by the main OR department of Erasmus MC, efforts are still focused on improving the current OR utilization. The aforementioned mathematical methods were examined. In addition we report a computer simulation study assessing promising methods for creating efficient surgical schedules within scenarios that represent various degrees of lowering organizational barriers.

Data and Methods

Data Erasmus MC is a university hospital and tertiary referral centre in Rotterdam, the Netherlands. Erasmus MC has 1,237 beds and admits 34,500 patients per year, 60% to 70% of whom undergo operation. The main inpatient OR suite consists of 16 operating rooms, providing for the complete spectrum of surgical cases, including transplantation and trauma surgery. Organizationally, the Erasmus MC inpatient OR department is subdivided into four units, each serving a set of specialties (Table 1). Prospective data, approved immediately after the surgical procedure, are available for more than 180,000 surgical cases since 1994. Data on expected and real case durations and variations in durations for the ten largest surgical departments were retrieved. Based on frequency, mean duration, and standard deviation of case duration,

Unit 1	Unit 2	Unit 3	Unit 4
ENT surgery	General surgery	Oral surgery	Gynaecological surgery
Neurosurgery		Trauma	Urology
Ophthalmology		Orthopaedic surgery	
		Plastic surgery	

Table 1 *Clustering of ten Erasmus MC surgical departments into four units*

data were classified into four to eight homogeneous categories per surgical department (Table 2). Table 3 shows the OR suite fixed weekly block plan. All OR blocks in this study consisted of 450 minutes.

	General surgery		Gynaecological surgery		Plastic surgery		ENT surgery		Orthopaedic surgery		Urology	
Cat.	Mean (SD)	Freq.	Mean (SD)	Freq.	Mean (SD)	Freq.	Mean (SD)	Freq.	Mean (SD)	Freq.	Mean (SD)	Freq.
1	150 (89)	8%	80 (65)	2%	119 (107)	5%	102 (125)	4%	107 (58)	9%	121 (68)	3%
2	67 (31)	3%	52 (19)	14%	63 (22)	14%	40 (17)	33%	61 (23)	10%	59 (30)	5%
3	100 (44)	12%	73 (33)	19%	82 (28)	17%	65 (24)	19%	83 (30)	18%	74 (26)	30%
4	135 (52)	19%	98 (32)	25%	112 (36)	21%	102 (35)	12%	109 (38)	21%	102 (49)	15%
5	171 (63)	20%	125 (43)	32%	139 (39)	22%	127 (32)	14%	160 (43)	21%	152 (49)	17%
6	213 (89)	3%	156 (41)	2%	187 (57)	11%	182 (65)	8%	199 (45)	16%	230 (68)	21%
7	262 (87)	25%	213 (82)	6%	432 (181)	10%	254 (75)	5%	291 (102)	5%	385 (123)	8%
8	351 (124)	9%					549 (203)	6%				
	Trauma		Ophthalmology		Neurosurgery		Oral surgery					
Cat.	Mean (SD)	Freq.	Mean (SD)	Freq.	Mean (SD)	Freq.	Mean (SD)	Freq.				
1	100 (68)	7%	83 (46)	1%	192 (165)	8%	97 (37)	1%				
2	62 (23)	22%	46 (14)	35%	113 (41)	17%	87 (29)	44%				
3	81 (30)	32%	60 (22)	42%	171 (62)	14%	130 (43)	44%				
4	122 (38)	20%	95 (30)	17%	255 (62)	28%	238 (87)	11%				
5	176 (92)	19%	127 (34)	5%	324 (73)	12%						
6					492 (177)	21%						

Table 2 *Characteristics of the ten main surgical departments in Erasmus MC. Eachcategory represents the patient mix for a department. (Abbreviations: Cat. Category, SD Standard Deviation, Freq. Frequency. Sample sizes: General Surgery 31,209, Gynaecological Surgery 10,163, Plastic Surgery 14,318, ENT surgery 17,103, Orthopaedic Surgery 11,859, Urology 11,876, Trauma 8,385, Ophthamology Surgery 9,801, Neurosurgery 10,370, and Oral Surgery 2,608. Surgical cases were classified based on their expected duration. Surgical cases for which no prediction of the case duration was available are grouped in Category 1. Mean and standard deviation (SD) are given in minutes.)*

Specialty	Number of operating rooms per day of the week				
	Mon	Tue	Wed	Thu	Fri
General surgery	3	3	3	3	3
Gynaecological surgery	1	1	1	1	1
Oral surgery	1	1	1	1	1
ENT surgery	2	2	2	1	2
Neurosurgery	2	2	2	2	2
Trauma	1	1	0	1	1
Ophthalmology	1	1	1	1	1
Orthopaedic surgery	1	1	2	1	2
Plastic surgery	2	2	2	2	1
Urology	2	2	2	2	2

Table 3 *Fixed weekly block plan for the inpatient OR department of Erasmus MC with 16 operating rooms*

Mathematical representation of Erasmus MC's surgical case scheduling

Surgical case scheduling is like finding the combination of surgical cases that makes optimal use of available OR time. In the field of applied mathematics, this problem is known as the bin-packing problem. Currently, surgical departments schedule their surgical cases using a First-Fit approach (15). Searching from the beginning, patients are selected from a waiting list and scheduled in the first available OR in a particular week.

In our study waiting lists were generated based on different surgical case categories representing each department's case mix (Table 2). Subsequently a First-Fit algorithm simultaneously selected and scheduled surgical cases for the period of one week, which in practice is done approximately two weeks before the date of surgery. This algorithm scheduled next cases only if the previous surgical case had been scheduled and if the algorithm concluded that it was impossible to fit the previous case in any of the available OR blocks. If the case did not fit in any of the available blocks, it was placed back on the waiting list. The algorithm terminates once it reached the end of the waiting list. Note that for scheduling of cases the mean duration was used and that no planned overtime was allowed, as prescribed by the Erasmus MC rules. The resulting surgical case schedules comprised surgical cases, planned slack, and unused OR capacity (Figure 1).

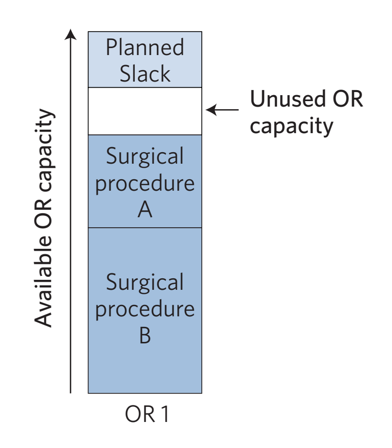


Figure 1 *Graphical representation of a surgical case schedule, which typically includes various surgical cases, planned slack, and unused* OR *time.*

Planned slack and the portfolio effect The financial world deals with uncertainty by using the portfolio effect. This ensures that the expected return of a stock portfolio is less vulnerable to fluctuations on the stock market. The term 'portfolio effect' then indicates that portfolio risk falls with increasing diversity, as measured by the absence of correlation (covariance) between portfolio components (16). We earlier found application of the portfolio effect to surgical case scheduling to be successful in increasing OR efficiency since it reduces the required amount of planned slack, given an accepted risk of overtime (11). The approach clustered surgical cases with similar variability in the same OR block, assuming these to be uncorrelated.

We illustrate the portfolio effect applied to surgical case scheduling by the following example: consider two OR blocks, for both of which two surgical cases are scheduled. One case with (*mean*, *standard deviation*) = (100,10) and one case with (*mean*, *standard deviation*) = (100,50), see Figure 2 (all values are given in minutes). We assumed that case durations are described by a normal distribution function. In this example we now compared this situation (the left-hand side of Figure 2) with the situation in which surgical cases with similar variance are clustered. In the first situation, the standard deviation of total duration is the same for both OR blocks: $\sqrt{(50^2+10^2)} = 51.0$ minutes. The total planned slack for the two blocks is thus $102.0 \cdot b$ minutes, where b is a risk factor to deal with risk of overtime. Since the sum of the durations follows a normal distribution the following holds: P (*mean* + *ß·standard deviation*) ≤ accepted risk of overtime, such that given a certain accepted risk of overtime the risk factor can be calculated. In the second situation the total planned slack is: $(\sqrt{(50^2+50^2)} + \sqrt{(10^2+10^2)}) \cdot \beta = 84.9 \cdot \beta$ minutes. This means a $17.1 \cdot \beta$ minutes reduction in the total required planned slack time, and thus an equal increase in available capacity. This portfolio profit will increase with higher variability of the cases concerned. This example illustrates that rescheduling a surgical case can reduce the extent of planned slack.

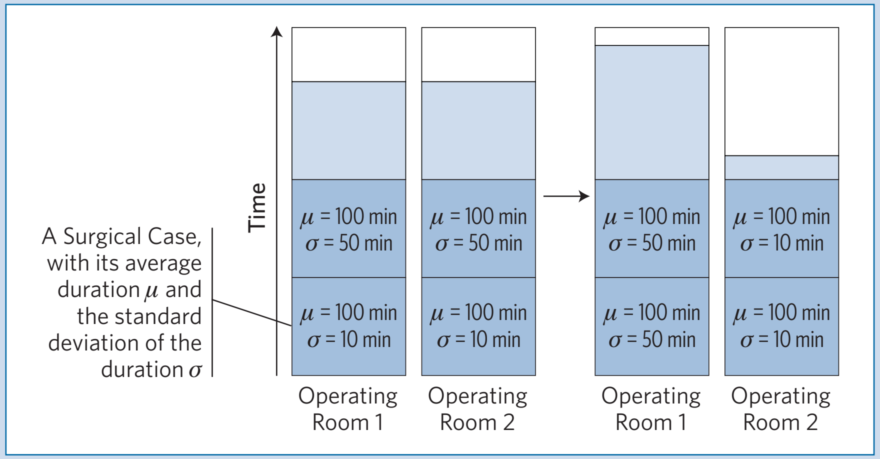


Figure 2 *Example of planned slack reduction as a result of the portfolio effect. The sum of the shaded areas in the two* OR *blocks on the left exceeds that in the two* OR *blocks on the right.*

Organizational barriers We constructed three scenarios to investigate the impact of lowering organizational barriers imposed by block planning (Table 4). The scenarios are graded as to interdepartmental flexibility, i.e. scheduling cases of different departments in the same OR on one day, and flexibility of rescheduling surgical cases between days of the week compared to the current situation. Rescheduling of surgical cases throughout the week does not affect patients since they have not been scheduled yet.

Scenario	Interdepartmental flexibility	Flexibility over the week
1	OR block consists of surgical cases of a single department	Rescheduling on the same day
2	OR block consists of surgical cases of a single department	Rescheduling within the same week
3	OR block consists of surgical cases of a department within one unit.	Rescheduling on the same day

Table 4 *Description of scenarios representing various flexibility levels (See Table 1 for the clustering of surgical departments in organizational units. The flexibility is applied to the construction of surgical schedules two weeks in advance.)*

In this study we assumed application of the scenarios directly after the construction of the surgical case schedules, approximately two weeks before the actual execution of the schedule (Figure 3). This enables OR departments to take necessary steps to ensure feasibility, for example regarding material logistics, ranging from specific surgical material to complete navigation system for complex craniotomy surgery. Surgical departments are responsible for the scheduling of semi-urgent or add-on elective patients who need operation on a day for which a surgical case schedule is already set. For this purpose, departments may schedule cases without assigning a patient to it, or by canceling one or more of the elective cases. Emergency patients are operated on within the reserved OR time as described earlier.

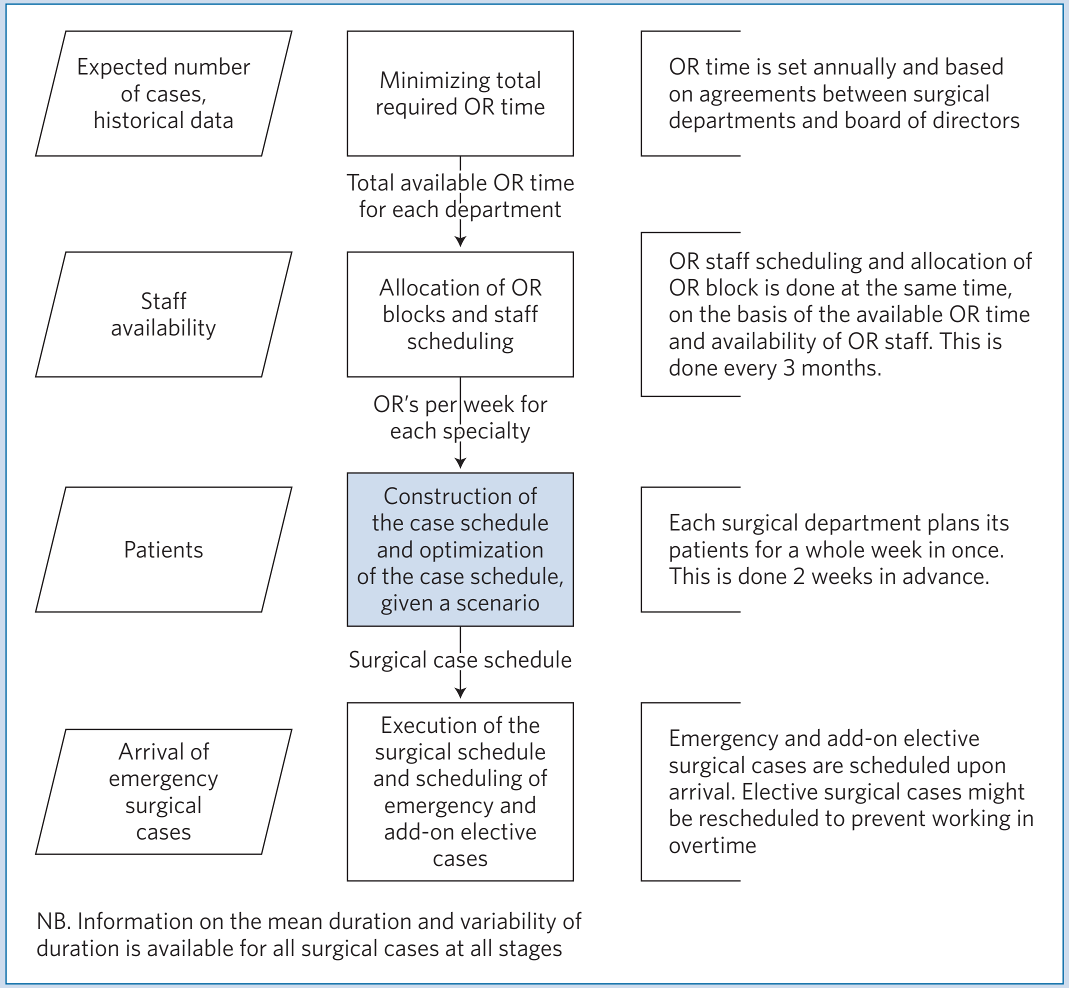


Figure 3 *Positioning of the OR scheduling process. The focus of this paper is on scheduling surgical cases approximately two weeks in advance, methodology for scheduling add-on and elective cases is beyond the scope of this paper and therefore not explicitly described in the figure.*

Advanced mathematical algorithms Application of the different scenarios to a surgical case schedule implied re-scheduling of surgical cases according to the organizational flexibility of the scenario under consideration. A bin-packing algorithm, based on work of Hans *et al.* (11), who used regret-based random sampling (RBRS), did the rescheduling of the surgical schedules given the scenarios. Figure 4 shows how rescheduling surgical cases saves OR time. The objective of the algorithm is to minimize planned slack, by exploiting the portfolio effect and the required number of OR blocks. RBRS procedures start with removing all cases of the existing surgical schedule to a list. Then, RBRS iteratively schedules a random surgical case from the list until all cases are scheduled. The drawing probability of each of the cases is based on the case's Best Fit suitability. This randomized procedure gives a new solution (a "surgical case") every time it is executed. We stopped the algorithm after generating

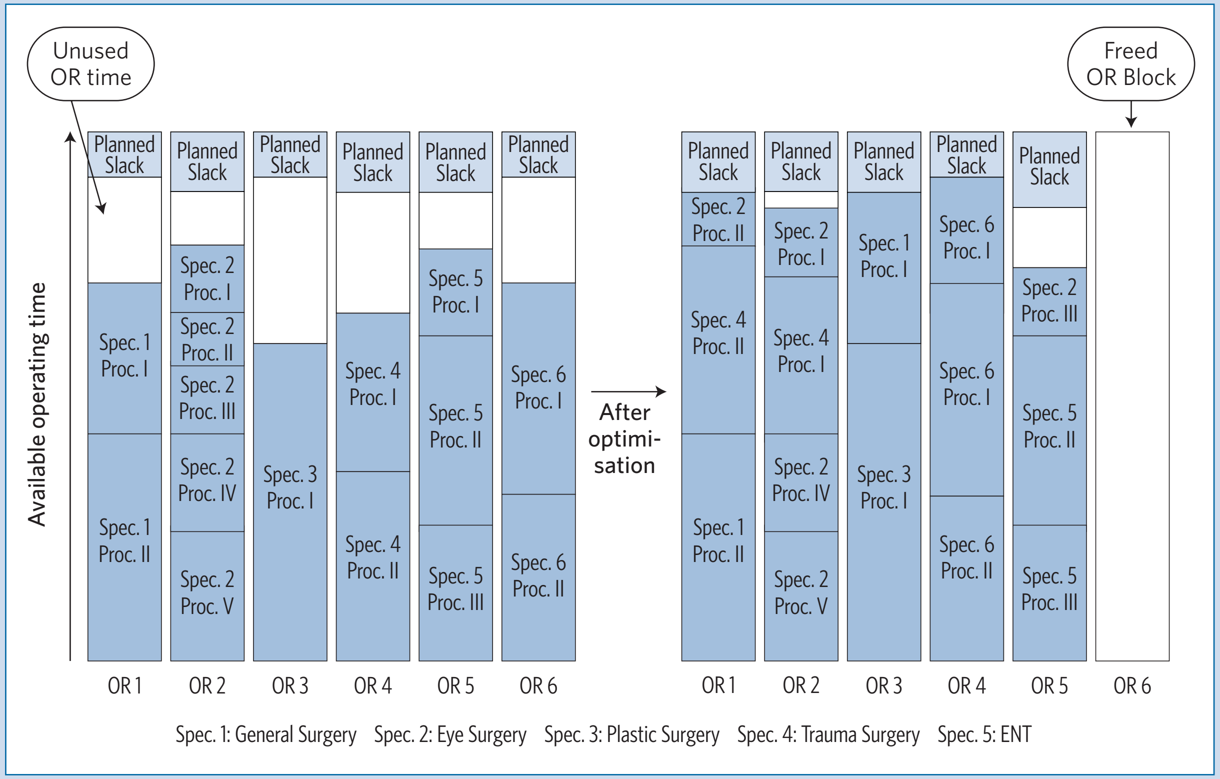


Figure 4 *Example of creating a free OR block by reallocating surgical cases.*

a preset number of 1,500 surgical case schedules. The generated schedules were evaluated on the objective criterion (amount of free OR capacity) and the best schedule was saved (11). The algorithm was coded in the Borland Delphi computer language (Cupertino, USA).

Experimental design The Erasmus MC's main inpatient department considered using the newsvendor approach of Strum *et al.* (3). Subsequently, we investigated the benefits of the RBRS that exploited the portfolio effect and relaxation of the organizational constraints. To this aim the surgical case scheduled created by the RBRS algorithm were compared to the surgical schedules constructed by the First-Fit approach. The RBRS algorithm was compared with the Best Fit algorithm (7) to assess the performance of advanced mathematical algorithms over available and simpler heuristic techniques.

We performed a robustness analysis on the influence of unpredictability of case duration on OR utilization, wherein the unpredictability was represented by the standard deviation of case duration. The influence of number of ORs within an OR department on OR utilization was investigated as well. Both analyses were carried out for each of the three flexibility scenarios. The outcome measures of this study are OR utilization and the number of freed OR blocks, so-called freed ORs. OR utilization was defined as the ratio between the total duration of elective surgical cases and the total staffed OR capacity per week. Hence it is similar to what is known in the literature as "raw OR utilization" (17).

Results

Applying the newsvendor approach of Strum *et al.* (3) did not lead to improved efficiency. With staffing costs determined by the allocated capacity and overtime by a relative cost ratio of 1.5 and increasing the block times with 15 minutes, it even decreased efficiency (See Table 5). Therefore new ways to increase OR efficiency were explored, as described in the previous section.

	Mean (minutes)	Standard deviation (minutes)	Proportion
Underutilization	59	68	52%
Overutilizaton	40	94	47%

Table 5 *OR performance (Measures are based on 30 consecutive months from 01–01–2004 onwards.)*

	Current situation	Scenario 1	Scenario 2	Scenario 3
	Mean 4 SE (%)	Mean 4 SE (%)	Mean 4 SE (%)	Mean 4 SE (%)
First Fit case schedule	77.4 4 0.2			
RBRS algorithm		77.5 4 0.2	81.9 4 0.2	78.8 4 0.2
Best Fit descending		77.4 4 0.2	81.2 4 0.2	78.2 4 0.2
		Mean (95% CI)(%)	Mean (95% CI)(%)	Mean (95% CI)(%)
RBRS algorithm versus Best Fit Descending		0.1 (- 0.5 - 0.6)	0.7 (0.2 - 1.2)	0.6 (0.1 - 1.0)

Table 6 *OR utilization rates (Utilization is defined as the ratio between the total amount of elective surgical cases and the total allocated OR capacity. SE standard error. Mean and 95% confidence interval (CI) of differences between RBRS algorithm and Best Fit Descending in the three scenarios are examined by a paired t-distribution.)*

Increased flexibility in the three scenarios increased the number of freed OR blocks (see Table 6). Eventually this resulted in an improved utilization rate of 4.5 percent points (95% confidence interval 4.0% – 5.0%). Both the Best Fit Descending heuristic and the RBRS algorithm improved utilization. The latter, more advanced, algorithm significantly outperformed the first heuristic by 0.7 percentage point in Scenario 2 (95% confidence interval 0.2% – 1.2%). Applying either the RBRS algorithm or the Best Fit Descending did not significantly improve the initial surgical schedules when combined with scenario 1, i.e. blocks consists of surgical cases of a single department and cases are rescheduled on the day. No significant difference was measure between the Best Fit Descending heuristic and the RBRS algorithm in Scenario 1 (Table 6).

Number of freed OR blocks and hence OR utilization increased relatively to the standard deviation of case duration within one department (Figure 5). The RBRS algorithm and the portfolio effect did not significantly improve the original schedule in scenario 1, regardless the standard deviation in the patient mix. Furthermore, in scenarios 2 and 3, the benefits of the RBRS algorithm increased with the standard deviation of case duration.

Figure 6 shows the association between number of ORs and OR utilization rate expressed in number of freed OR days for the three scenarios. The findings shows that if more flexibility would be achievable, benefits progressively increase with the number of cases performed daily relatively to the available hours provided.

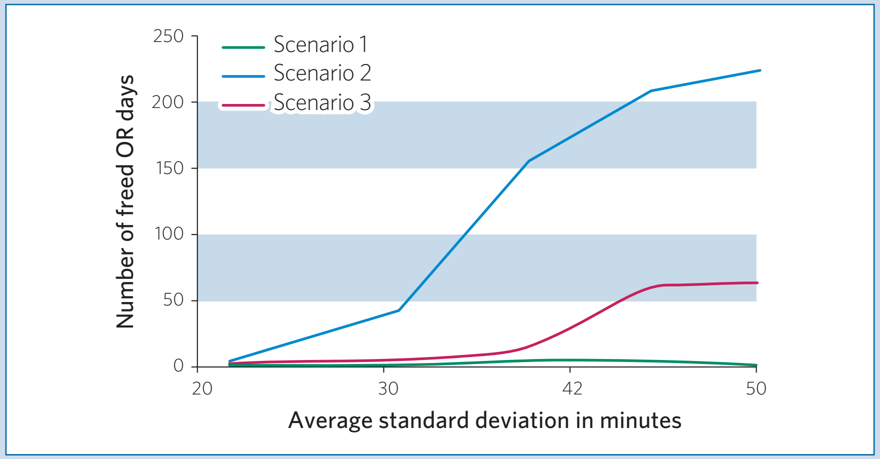


Figure 5 *Graphic representation of the number of freed* OR *blocks in Erasmus* MC *when the RBRS algorithm is applied in combination with the three scenarios in which the standard deviation of case duration is varied.*

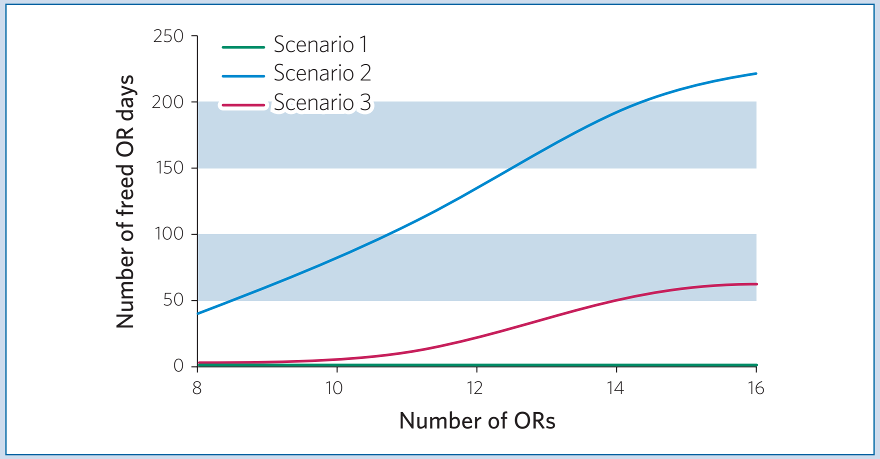


Figure 5 *Graphic representation of the number of freed* OR *blocks when the RBRS algorithm is used in combination with the three scenarios for* OR *departments of various sizes.*

Discussion

The study showed how to improve OR efficiency by combining advanced mathematical and financial techniques with the lowering of organizational barriers. The combination facilitates OR departments to improve OR efficiency when current methods will no longer benefit (3,7). The method is applicable in hospitals that set their surgical case schedules approximately two weeks in advance and potentially improves OR utilization by 4.5%. Improved efficiency implies that more operations can be performed at the same OR capacity or that less OR capacity is needed for the same number of operations. We also showed that potential benefits vary for different OR departments, depending on the uncertainty in case duration and number of ORs within one OR department. Absolute measures of this study are hard to compare with results from other studies, because Erasmus MC employs a specific method of reserving OR time in surgical schedules.

The algorithms used aimed to free OR blocks, because capacity that was previously allocated in these blocks is not accounted for while calculating the utilization rate. This is true for all OR departments that have sufficient flexibility in their staff scheduling to allow changes approximately two weeks in advance.

We assumed in the analysis that surgical case durations show normal distribution. Other studies have shown that a lognormal distribution is a better approximation of the real duration (18). Calculation of planned slack, which is required to simulate the portfolio effect, requires a closed form probability distribution. This is not the case for a lognormal distribution, and this is why we have opted for a normal distribution, which may modestly influence the outcomes. Since the amount of planned slack is similarly calculated for the RBRS algorithm compared to the Best Fit heuristic, we do not expect that the assumption influences the calculated outcomes.

Many hospital use information technology (IT) systems to actual schedule their surgical cases in the available blocks. The mathematical techniques presented in this paper can easily be incorporated in such IT-systems, permitting planners to actually use the mathematical algorithms. Using the techniques addressed in this paper, and given a flexibility scenario agreed upon beforehand, the set of cases planned by the different departments in their blocks is collectively optimized after surgeons have set their patients' surgery dates. After optimization, each department can match its surgeon and bed planning with the new, more efficient, case schedule.

Lowering organizational barriers might have some negative effects and will require a more flexible attitude of surgical departments and individual surgeons. First, allowing various surgical departments to use the same OR may result in longer waiting times for surgeons. Second, surgeons may be planned in various ORs on the same day. Third, having surgeons operate on different days in the week requires adjustment of their other tasks, especially in hospitals where surgeons are highly specialized and where cases cannot be interchanged between surgeons. All these issues should be carefully addressed and weighed against the efficiency increase. The essential consideration, we believe, is that the drawbacks for a surgical department can be compensated for by the huge amount of extra OR capacity, which can be used to shorten the waiting list and earn more money.

Another aspect of implementation of the techniques is the required additional flexibility of the ORs. Each OR has to be uniformly equipped so that all surgical departments may operate in it. The efficiency increase achieved by the proposed method would justify the investment to equip all ORs generically.

Each hospital can choose a flexibility scenario that matches its requirements. Even more scenarios can be made to show benefits of even lower organizational barriers. The potential benefits can be calculated by comparing the current case scheduling strategy, in this paper represented by a first-fit algorithm, and the future situation in which the portfolio effect and bin-packing techniques have been applied and organizational constraints have been relieved. This paper provides a tool for any hospital type to make their own trade-off between flexibility and higher utilization of OR capacity.

References

1 Ozkarahan I. Allocation of Surgeries to Operating Rooms by Goal Programming. J Med Syst 2000;24:339–78.

2 Magerlein JM, Martin BM. Surgical Demand Scheduling: A Review. Health serv res 1978;13:418–33.

3 Strum DP, Vargas LG, May JH. Surgical subspecialty block utilization and capacity planning: a minimal cost analysis model. Anesthesiology 1999;90:1176–85.

4 Dexter F, Macario A, Cowen DS. Staffing and case scheduling for anesthesia in geographically dispersed locations outside of operating rooms. Curr Opin Anaesthesiol 2006;19:453–8.

5 McIntosh C, Dexter F, Epstein RH. The impact of service-specific staffing, case scheduling, turnovers, and first-case starts on anesthesia group and operating room productivity: a tutorial using data from an Australian hospital. Anesth Analg 2006; 103:1499–516.

6 Ogulata SN, Erol R. A Hierarchical Multiple Criteria Mathematical Programming Approach for Scheduling General Surgery Operations in Large Hospitals. J Med Syst 2003;27:259–70.

7 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.

8 Marcon E, Kharraja S, Simonnet G. The operating theatre planning by the follow-up of the risk of no realization. Int J Prod Econ 2003;85:83–90.

9 Dexter D, Traub RD, Macario A. How to release allocated operating room time to increase efficiency – Predicting which surgical service will have the most under-utilized operating room time. Anesth Analg 2003;96:507–12.

10 Dexter F, Traub RD. How to schedule elective surgical cases into specific operating rooms to maximize the efficiency of use of operating room time. Anesth Analg 2002;94:933–42.

11 Hans EW, Wullink G, Houdenhoven M, Kazemier G. Robust Surgery Loading. European Journal of Operational Research 2006; Available online.

12 Markowitz HM. Portfolio selection. 2nd ed. Cambrigde: Blackwell Publishers Ltd., 1991.

13 Dexter F, Ledolter J, Wachtel RE. Tactical decision making for selective expansion of operating room resources incorporating financial criteria and uncertainty in subspecialties' future workloads. Anesth Analg 2005;100:1425–32.

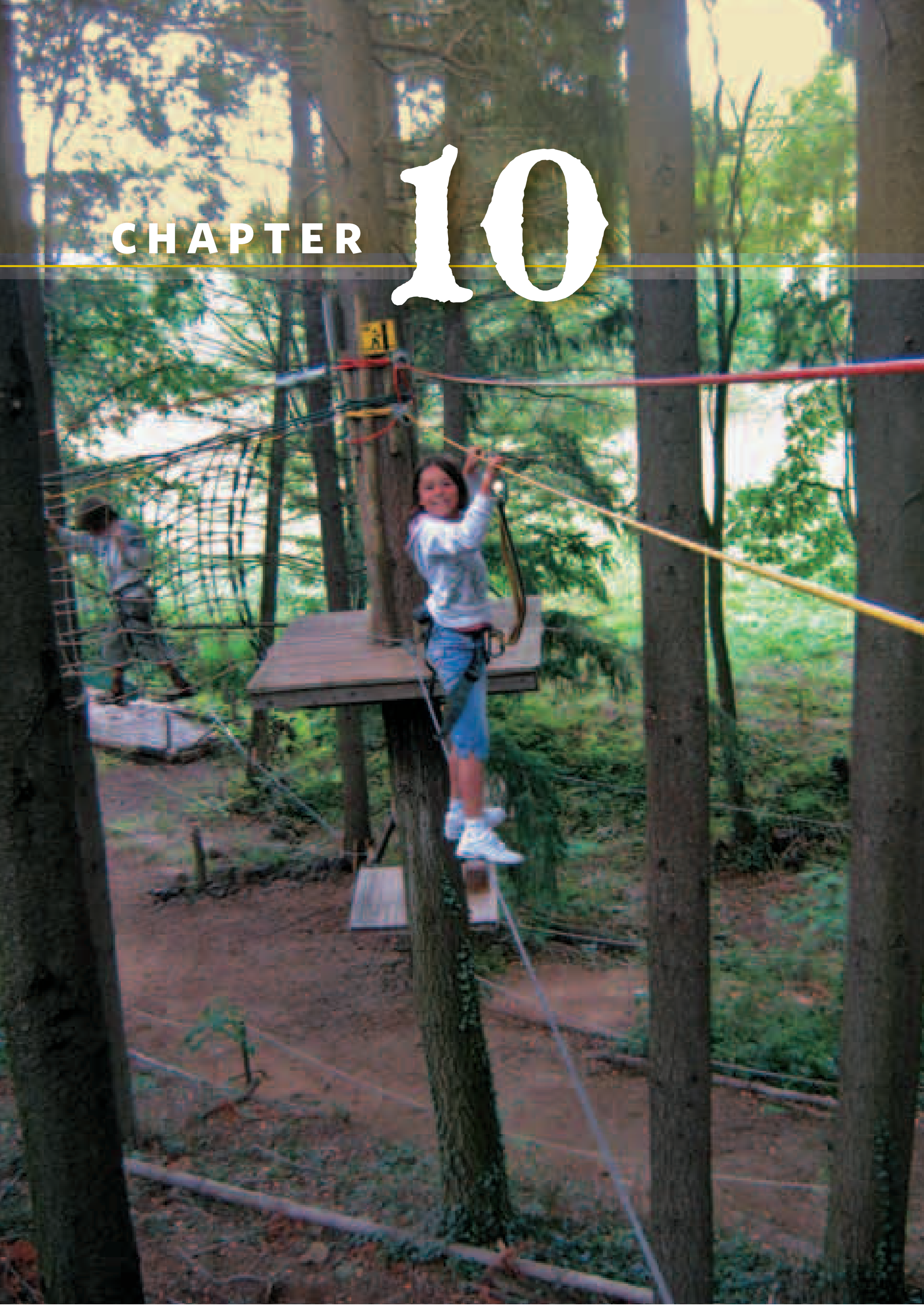
14 Dexter F, Ledolter J. Managing risk and expected financial return from selective expansion of operating room capacity: mean-variance analysis of a hospital's portfolio of surgeons. Anesth Analg 2003;97:190–5.

15 Dexter F, Macario AMD, Traub RD, *et al*. An Operating Room Scheduling Strategy to Maximize the Use of Operating Room Block Time: Computer Simulation of Patient Scheduling and Survey of Patient's Preferences for Surgical Waiting Time. Anesth Analg 1999;89:7–20.

16 Markowitz H. Portfolio Selection. The Journal of Finance 1952;7:77–91.

17 Dexter F, Macario A, Traub RD, Lubarsky DA. Operating room utilization alone is not an accurate metric for the allocation of operating room block time to individual surgeons with low caseloads. Anesthesiology 2003;98:1243–9.

18 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.



CHAPTER 10

Optimizing Intensive Care capacity using individual length-of-stay prediction models

Mark Van Houdenhoven, Duy-Tien Nguyen,
Marinus J. Eijkemans, Ewout W. Steyerberg, Hugo W. Tilanus,
Diederik Gommers, Gerhard Wullink, Jan Bakker,
Geert Kazemier

Critical Care, 2007; 11 (2), published online 27 March 2007

Introduction

Intensive care units (ICUs) consume a considerable portion of hospital budgets. Moreover, costs are predicted to rise with the emergence of new treatment methods. Problems with ICU capacity are nevertheless common, and studies conducted in ICUs have documented high rates of refusal to admit because of lack of empty beds [1,2]. In addition, the need to serve the 'greying' population is likely to increase demand for ICU beds further, exacerbating the current strain on ICU capacity. Consequently, hospitals will face an increase in numbers of cancelled surgical procedures that necessitate postoperative intensive care, and higher rates of refusal to admit other critically ill patients [2,3]. The only way to remedy these problems is apparently to improve the efficiency with which the available ICU and operating room capacity is used, in other words to optimize patient planning.

Patient planning depends importantly on reliable and adequate management information. Key elements in the ICU setting are the patient's expected length of stay (LOS) in the ICU at admission and possible changes in expected LOS resulting from later treatment. Starting from the admission date and expected LOS, the planner will be able to pinpoint the anticipated date at which an ICU bed will once again become available. This information, along with subsequent changes in a patient's expected LOS, is needed to schedule the next operating room patient who requires postoperative intensive care or to reserve emergency patient capacity on the ICU. In addition, information on expected LOS preoperatively facilitates scheduling of individual surgical procedures on specific dates. This information can be used to predict ICU admission dates and LOSs. Information that emerges during the surgical procedure and the postoperative stay in the ICU can influence the LOS predicted by the preoperative model. This so-called online patient planning can help to improve OR and ICU programmes.

Clinicians generally assume that LOS of individual patients is unpredictable. Intensivists are expected to be able predict LOS roughly, but the accuracy of this prediction depends largely on the intensivist's experience. We speculate that if comprehensive evaluations of the association between preoperative, intraoperative and postoperative prognostic variables on the one hand, and LOS on the other are translated into a mathematical model, then this model might be able to predict LOS with greater accuracy.

The main goal of this study was to develop a model that will provide planners with a tool to predict the LOS of individual patients in the ICU. Data on a cohort of consecutive patients undergoing an elective oesophagectomy were used to create and validate such models. Predictive power was assessed to determine the best performing model. In a second cohort of patients, the LOSs of individual patients were predicted prospectively to determine the potential gain of this best model on a day-to-day basis.

Methods

Data Data from 518 consecutive patients who underwent elective oesophagectomy with reconstruction for carcinoma at the Erasmus University Medical Centre, Rotterdam, The Netherlands, between January 1997 and April 2005, were retrieved from the hospital information system. These data were combined with detailed data from a prospective database held at the Department of Surgery. The Erasmus University Medical Centre includes a total of 1,212 beds on several locations. It is a trauma centre for a catchment area that includes 5.2 million people. The main site includes 32 ICU beds and 19 operating rooms.

The outcome variable of the present study was LOS, defined as the time in days between admission and discharge from the ICU. Admission to and discharge from the ICU were based on the national protocol [4]. For patients discharged to the ward and readmitted to the ICU within 48 hours, the intervening stay on the ward was included in the LOS. Definitions of these variables and supporting references [5-7] are given in Table 1; reports that provide evidence supporting the use of these variables in the model are also referenced [8-12].

Model construction Only those variables that were present in more than 1% of the patients were included, in order to avoid unstable estimates. On clinical grounds, two expert surgeons (HWT and GK) and two expert anaesthetists (DG and JB) formed a preselection of factors from the potentially prognostic variables in order to prevent overfitting [13-15]. Only these selected variables were used to build the three models. In those patients with missing values, data were completed using multiple imputation methods. This was done under the assumption that the distribution of the missing date and the complete data were the same [16].

The imputed model included both the independent potentially prognostic variables and the outcome variable LOS. Given the inherently skewed distribution of LOS, a natural log transformation was used [17].

Univariate linear regression analysis was used to test which of the variables contributed to LOS with $P \leq 0.20$. The mean and standard deviation are reported for those variables that are normally distributed. The median and interquartile ranges are given for non-normal distributions.

Significant variables in the univariate analyses were entered as potentially

	Construction sample (n = 518)	Application sample (n = 65)	Reference
Patient characteristics			
Age (years)	63 (55-70)	60 (56-68)	[8-12]
Male sex	407 (79)	48 (74)	[9-12]
BMI (kg/m2)	25 (22-28)	26 (23-29)	[8]
ASA 1, 2	89 (17)	28 (43)	[5]
Hypertension	192 (37)	35 (54)	[6]
Previous stomach operation	132 (25)	19 (29)	[6]
Preoperative serum haemoglobin (mmol Fe/l)	8.4 (7.6-9.2)	8.7 (7.4-9.4)	[11,12]
Preoperative serum creatinin (mmol/l)	78 (68-89)	78 (68-90)	[11,12]
Preoperative FEV1 (l)	2.9 (2.4-3.5)	3.2 (2.4-3.7)	
Preoperative chemotherapy	170 (33)	17 (26)	[8,10]
Preoperative radiotherapy	55 (11)	8 (12)	[8,10]
Aetiology			
Gastroesophageal reflux disease	63 (12)	9 (14)	[6]
Barrett's esophagus	43 (8)	9 (14)	[6]
Other	66 (13)	13 (20)	[6]
Comorbidities			
Cardiac	134 (26)	24 (37)	[6]
Respiratory	91 (17)	7 (11)	[6]
Vascular	65 (13)	6 (9)	[6]
Neurological	33 (6)	7 (11)	[6]
Diabetes mellitus	51 (10)	7 (11)	[6]
Other carcinoma	53 (10)	4 (6)	[6]
Other	40 (8)	2 (3)	[6]
Tumour characteristic			
Adenocarcinoma	340 (66)	51 (79)	
pTNM stage 0	27 (5)	7(11)	[7]
I	64 (12)	5 (8)	[7]
IIa	120 (23)	13 (20)	[7]
IIb	46 (9)	5 (8)	[7]
III	193 (37)	21 (32)	[7]
IV	68 (13)	14 (22)	[7]
Radicality (R0)	400 (77)	54 (83)	
Session variables			
Expected duration of the procedure (min)	240 (180-270)	266 (262-314)	
Duration of the procedure (min)	301 (254-359)	333 (290-368)	[11]
Total age of the two head surgeons (years)	83 (72-88)	84 (74 – 94)	
Transthoracic approach	114 (22)	14 (22)	[8,10-12]

	Construction sample (n = 518)	Application sample (n = 65)	Reference
Session variables (sequel)			
Reconstruction using colon	24 (5)	3 (5)	[10-12]
Oesophagus and cardia resection	506 (98)	65 (100)	[6]
Splenectomy during surgical procedure	15 (3)	2 (3)	
Absolute crystalloid administration (l)	6.0 (4.5-7.0)	4.0 (2.3-5.5)	[8]
Absolute colloid administration (l)	1.5 (1.5-2.0)	1.5 (1.5-2.0)	[8]
Erythrocyte concentrate transfusion	276 (53)	22 (33)	[8,12]
Fresh frozen plasma transfusion	36 (7)	6 (10)	[8,12]
Absolute blood loss (l)	1.5 (1.0-2.2)	1.1 (0.7-1.5)	[8,12]
Absolute urine production (l)	0.7 (0.4-1.3)	0.4 (0.3-0.7)	
Epidural analgesia during procedure	467 (90)	57 (88)	
Vasopressor administration	214 (41)	63 (97)	
Duration of vasopressor therapy (hours)	0 (0-1.5)	270 (206-337)	
Minute volume (l)	7.8 (7.2-8.8)	7.8 (7.0-8.4)	
Positive end-expiratory pressure (cmH2O)	5 (4-7)	6 (5-7)	
Serum oxygen saturation (%)	98 (96-100)	100 (98-100)	
End temperature (aC)	35.8 (35.2-36.4)	36.3 (36.0-36.9)	
Lactate (mmol/l)	1.7 (1.2-2.2)	1.4 (1.0-2.0)	
Postoperative variables			
Duration of mechanical ventilation	0.63 (0.13-5.58)	0.54 (0.12-4.12)	
Surgical complications			
Postoperative bleeding	19 (4)	1 (2)	[6]
Chylothorax	20 (4)	3 (5)	[6]
Leakage of anastomosis	38 (7)	9 (13)	[6]
Necrosis of anastomosis	18 (4)	2 (3)	[6]
Other	42 (8)	11 (17)	[6]
Nonsurgical complications			
Pulmonary: pneumonia, atelectasis, or ARDS	198 (38)	24 (37)	[6]
Infection: urinary tract, sepsis	31 (6)	1 (2)	[6]
Thrombosis, embolism	20 (4)	3 (5)	[6]
Other	135 (26)	26 (40)	[6]
Length of stay in the ICU (days)	4.0 (2.0-7.9)	4.2 (2.9-7.9	

Table 1 *Characteristics for both cohorts of patients who underwent oesophagectomy with reconstruction for cancer (Values are expressed as number (%) or, for continuous variables, as median (25^{th} to 75^{th} percentile). ARDS, acute respiratory distress syndrome; ASA, American Society of Anaesthesiology Physical Status Score; BMI, body mass index; FEV1, forced expiratory volume in 1 s; ICU, intensive care unit.)*

prognostic variables into a backward, stepwise selection procedure to construct a multivariable linear model that provides a natural logarithm transformed prediction of LOS (ln[LOS]). Because LOS can be predicted based on expanding sets of available information at three stages, three multivariable linear models were constructed. First, preoperative data were used to build a preoperative prediction model. Then, intraoperative data were incorporated to construct a postoperative model. To construct an intra-ICU model, which was used after three days on the ICU, all selected data were used. This final model was constructed to improve accuracy based on new information from the last three ICU days. The criterion for retention of variables in the model was $P <$ 0.20, which ensured high power for inclusion of variables with somewhat weaker predictive effects [14]. Interactions between variables and nonlinear relationships were explored. A smearing factor to correct the 'back transformation' bias was needed to obtain the estimated LOS, because a natural logarithmic transformation on LOS was used [18]. Goodness-of-fit was assessed graphically by plotting observed LOS against predicted LOS in a calibration plot. The predictive power of the model was expressed as the percentage explained variation (multiple r^2) on the logarithmic scale.

Internal validity was assessed with bootstrap sampling to obtain estimates of the optimism of the multiple r^2 [14,19,20]. This optimism indicates the expected decrease in model performance when it is applied in future patients [21]. Bootstrap samples were drawn with replacement and with the same size as the original sample. Regression models were constructed in each bootstrap sample and tested on the original sample. This was repeated 200 times to obtain stable estimates of the optimism of the model [21].

Analyses were performed using SPSS version 11 (SPSS Inc, Chicago, IL, USA) and S-Plus version 6 (Insightful Inc, Seattle, WA, USA).

Model application{Level 2 heading}

After internal validation of the models, the gain in terms of usage of ICU capacity with the model exhibiting the highest r^2 was assessed in routine clinical practice. Prospective data were collected for consecutive patients who underwent elective oesophagectomy with reconstruction for carcinoma. The data were collected during the period from May 2005 to April 2006, which were the 12 months after construction of the model (Table 1). The prediction model was assessed by comparing the total overestimation and underestimation of the required ICU days if the mean LOS was used (the old situation) with the total overestimation and underestimation of the required ICU days if the prediction model was used (the new situation). The overestimation and underestimation in the old situation were calculated by subtracting the observed LOS from the mean LOS. The overestimation and underestimation in the new situation were calculated by subtracting the observed LOS from the predicted LOS. The mean LOS was used rather than the median LOS, because use of the median will favour the prediction model because the LOS is skewed. Therefore, use of the mean LOS will result in a more conservative gain in comparison with the median LOS.

Both the old and new situations have three possible outcomes: negative, indicating that the ICU bed was reserved for too long and that the number of ICU days was overestimated; zero, indicating perfect prediction; or positive, indicating that the ICU bed was reserved for an insufficient period and that the number of ICU days was underestimated. The total overestimation and the underestimation were calculated for both the mean LOS approach and the LOS prediction model for both the old and the new situations.

Results

Retrospective population The mean LOS was 8.76 days and the median LOS was 4.0 days (interquartile range 2.0 to 7.9 days). Overall, 6.8% of the patients were discharged from the ICU within 1 day after their surgical procedure, 37% within 3 days, 56% within 5 days and 69% within 7 days (Figure 1). Thirty-eight patients (7.3%) were readmitted to the ICU after a stay shorter than 48 hours on the ward. Table 1 lists the retrieved data for variables that were thought to be potentially prognostic, broken down into patient characteristics, tumour and session characteristics, and postoperative complications within the first 72 postoperative hours. ICU mortality was 2.5% and total in-hospital mortality was 4.1%.

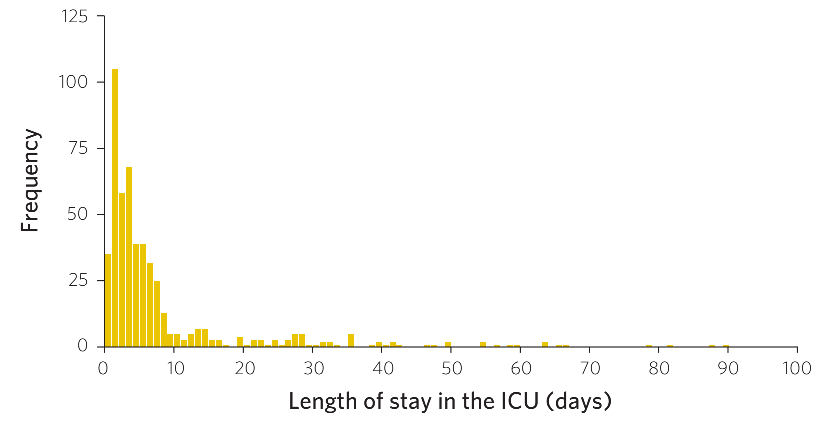


Figure 1 *Distribution of length of stay in the ICU. ICU, intensive care unit.*

Univariate analysis The following preoperative variables (Table 1) were associated with longer LOS: older age ($P < 0.001$), American Society of Anesthesiology's Physical Status 3 or 4 ($P = 0.001$), presence of five out of seven comorbidities ($P < 0.001$ to 0.14), squamous cell carcinoma ($P = 0.003$), transthoracic approach instead of transhiatal ($P < 0.001$), reconstruction using colon instead of stomach ($P = 0.02$), previous chemotherapy ($P = 0.003$) and lower forced expiratory volume in 1 s during preoperative screening ($P < 0.001$). Intraoperative variables associated with longer LOS were higher absolute amount of colloids administered ($P = 0.01$), greater absolute blood loss ($P = 0.04$), longer duration of vasopressor administration ($P = 0.03$), higher respiratory minute volume ($P = 0.005$) and lower arterial oxygen saturation ($P < 0.001$). Patients with any complication occurring within 72 hours after surgery also had significantly longer LOS ($P < 0.001$).

Preoperative, postoperative and intra-ICU multivariable models The multiple r^2 for the preoperative model was 21% and the optimism was 6%; hence, the r^2 after validation was 15%. The preoperative model had a 95% confidence interval (CI) with relative bounds between 0.5 and 2.5. This implies that LOS may be from 50% shorter to 254% longer than the mean LOS. Patient age ($P = 0.001$), presence of gastroesophageal reflux disease ($P < 0.001$), neurological comorbidity ($P < 0.001$) and a transthoracic instead of transhiatal approach ($P < 0.001$) were the variables that contributed most to the increase in LOS for the preoperative model.

For the postoperative model, the multiple r^2 was 25% and the optimism was 9%; the r^2 after validation was 17%. The 95% CI with relative bounds was comparable to that of the preoperative model. Apart from the variables included in the preoperative model, higher absolute amount of colloids administered ($P = 0.03$) and a maximum respiratory minute volume during the surgical procedure ($P < 0.001$) were the variables found to contribute to LOS in the postoperative model.

The multiple r^2 of the intra-ICU model was 56% and the optimism was 11%, resulting in an r^2 of 45% after validation. The intra-ICU model had a 95% CI with relative bounds between 0.3 and 3.4, implying that LOS may be from 70% shorter to 340% longer than the mean LOS. Complications occurring within 72 hours in the ICU (five complications had $P < 0.001$ and two complications had $P < 0.06$) were the variables found to contribute to LOS in the intra-ICU model. Results are shown in Table 2 and formulas to calculate the LOS of the preoperative, postoperative, and intra-ICU models can be found in Additional files 1, 2, and 3, respectively.

The goodness-of-fit of the three models is shown in Figure 2, which reveals considerable variation. Preoperative and postoperative LOS predictions exhibit variation and are not symmetrically distributed around the regression line.

	Preoperative model	Postoperative model	Intra-ICU model
R(Constant)	1.26	0.44	1.82
Expected session time (min)	1.10 (1.01-1.21)		
Patient age (per decade)	1.16 (1.06-1.28)	1.20 (1.09-1.31)	1.09 (1.02-1.17)
FEV1 (l)	0.91 (0.81-1.03)	0.85 (0.75-0.96)	
Gastroesophageal reflux disease (yes/no)	1.46 (1.14-1.89)	1.53 (1.19-1.96)	
Vascular comorbidity (yes/no)	1.29 (1.01-1.66)	1.32 (1.03-1.69)	
Neurological comorbidity (yes/no)	1.74 (1.24-2.43)	1.82 (1.31-2.53)	
Previous chemotherapy (yes/no)	0.81 (0.68-0.97)		
Previous radiotherapy (yes/no)			0.78 (0.63-0.97)
Transthoracic approach (yes/no)	2.13 (1.74-2.62)	1.79 (1.44-2.24)	1.21 (1.05-1.40)
Reconstruction using colon (yes/no)	1.56 (1.05-2.30)	1.52 (1.03-2.23)	
Observed session time (min)		1.07 (0.98-1.16)	
Volume administration of colloids (liter)		1.14 (1.01-1.29)	
Absolute intraoperative blood loss (l)		0.94 (0.87-1.02)	
Absolute intraoperative urine production (l)		1.12 (0.99-1.25)	
Epidural analgesia administration (yes/no)		0.83 (0.69-1.01)	
Respiratory minute volume (l)		1.09 (1.04-1.15)	1.05 (1.01-1.09)
Positive end-expiratory pressure (cmH_2O)		1.03 (0.99-1.07)	
Chylothorax surgical complication (yes/no)			1.31 (0.96-1.79)
Anastomosis leakage complication (yes/no)			1.83 (1.47-2.28)
Other complication (yes/no)			1.71 (1.38-2.10)
Pulmonary nonsurgical complication (yes/no)			1.97 (1.72-2.26)
Myocardial infarction (yes/no)			1.54 (0.93-2.56)
Infection (yes/no)			1.61 (1.25-2.07)
Other nonsurgical complication (yes/no)			1.41 (1.22-1.62)
Multiple r^2	21%	25%	56%
Optimism	6%	9%	11%
Optimism corrected r^2	15%	17%	45%

Table 2 *Multivariable preoperative, postoperative and intra-ICU linear LOS analyses (Unless stated otherwise, values are expressed as coefficient (95% confidence interval). FEV1, forced expiratory volume in 1 s; ICU, intensive care unit; LOS, length of stay.)*

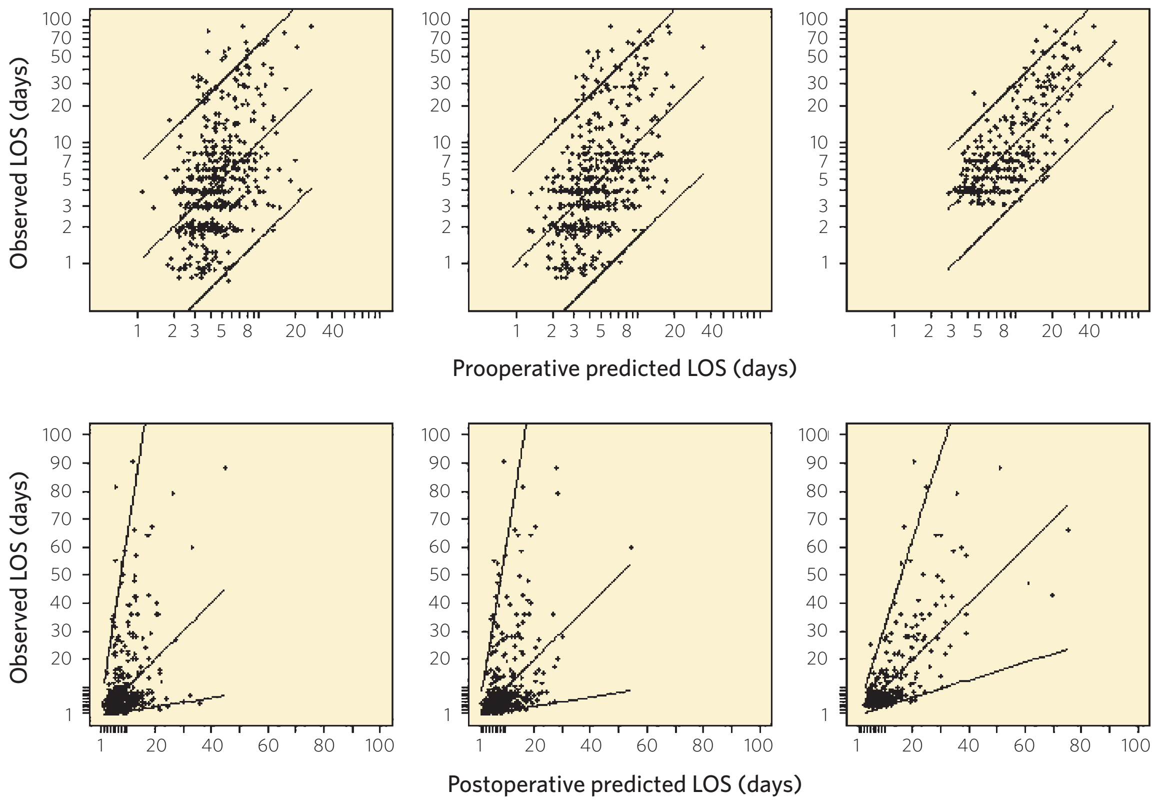


Figure 2 *Calibration plots of the observed LOS against predicted LOS. These plots were constructed using the multivariable preoperative, postoperative and intra-ICU linear LOS models, on logarithmic scales (upper panel) and on untransformed scale with smearing factor (lower panels). ICU, intensive care unit; LOS, length of stay.*

The LOS predictions of the intra-ICU model vary less, however, and are symmetrical. In addition, the prediction bounds of the intra-ICU model are much smaller than those of the preoperative and postoperative models.

Model application Because the intra-ICU model has the highest r^2, this model was assessed for the second, prospective cohort of patients. This model included only patients who stayed in the ICU for at least three days. Of the 65 patients, 46 had a LOS longer than three days. The mean LOS for patients staying longer than three days in the ICU was 14.6 days, and so the 'remaining mean LOS' of 11.6 days (after subtraction of the first 3 ICU days) was used for comparison with the intra-ICU model. In the old situation (remaining mean LOS) and the new situation (prediction model), 10 out of 46 patients had an observed LOS longer than predicted (underestimation of ICU days), and the remaining 36 had an observed LOS shorter than predicted (overestimation of ICU days). In the old situation, these 10 patients together accounted for an underestimation of 220 ICU days; in the new situation they accounted for an underestimation of 155 ICU days. The other 36 patients together occupied the ICU for 213 days longer than predicted in the old situation, but in the new situation they occupied the ICU for 236 days longer than predicted (Table 3).

All in all, the total underestimation of ICU days decreased by 65 in favour of the prediction model; this is equal to 11% of the total ICU capacity of the study group. The total overestimation ICU days increased by 23 with the prediction model (in favour of prediction based on mean LOS). LOS was underestimated by the prediction model in 10 patients; this underestimation was less than in the old situation, however. Ultimately, 10 patient cancellations were prevented, which is equivalent to 15% of included patients.

	Situation	
	Old	New
UPatients with underestimation (*n*)	10	10
Total underestimated days	220	155
Patients with overestimation (*n*)	36	36
Total overestimated days	213	236
Difference in underestimated days	–	65
Difference in overestimated days	–	23

Table 3 *Application of the model: underestimation and overestimation of old and new situation. (In the old situation, estimation of LOS was based on the remeaning LOS after three days. The new situation used the prediction model. LOS, length of stay.)*

Discussion

We showed that a predictive model incorporating characteristics of individual patients who underwent oesophagectomy for cancer enhanced the accuracy of estimated LOS. Key prognostic variables included patient's age, presence of gastroesophageal reflux disease, respiratory minute volume, transthoracic rather than transhiatal approach, and complications within the first 72 postoperative hours. We assessed three models and found that the intra-ICU model, which uses data from the first 72 hours in the ICU, had the best predictive performance. We found that use of this model in our clinical setting would have resulted in a gain of 65 ICU days over a 12-month period. This is equivalent to 11% of the ICU capacity for this patient group. Moreover, 15% of cancellations of future surgical procedures could have been prevented.

Three types of related studies have been reported in the literature. First, we found reports of LOS prediction models that suggest specific therapeutic interventions in patient groups that may influence LOS [22,23]. Second, prediction models to determine risks for prolonged LOS have been developed [24-31]. In these studies the investigators used preoperative, intraoperative and postoperative variables to fit a logistic model, with risk for prolonged LOS as the main outcome. This outcome measure is claimed to improve planning and therefore cost-effectiveness of hospitals. However, the results from these studies do not permit scheduling of individual patients on the ICU. They only calculate the risk for prolonged LOS given a certain cutoff point. The third type of study also uses individual patient characteristics to predict LOS, as in our study. However, they apply less sophisticated mathematical techniques (multiple linear regression), whereas the present study used logistic regression [32]. These models can only be used for medically homogeneous patient groups with a shorter and less variable LOS. In summary, the models proposed in these earlier studies are unsuitable for scheduling of individual patients in the ICU. In contrast, the prediction model proposed here does permit individual patient scheduling in the ICU on a day-to-day basis.

A typical example illustrates the value of the prediction model. A 79-year-old oesophagectomy patient without previous radiotherapy was operated on via a transhiatal approach; the measured maximum respiratory minute volume was 9.6 l during the surgical procedure, and various complications occurred within the first 72 hours in the ICU. The predicted LOS for that patient using our model is 84 days. This well exceeds the mean LOS of 8.76 days that was calculated using data in the hospital information system.

Prediction models, such as that proposed here, can improve quality of care and cost-effectiveness in an ICU, as was demonstrated in the prospective second cohort of patients we analyzed. Data required for the development and application of prediction models are typically available in every hospital. Therefore, prediction models can be used in almost any clinical setting, but they must be developed for specific groups of individual patients if the full benefit in terms of capacity gain is to be realized. The ICU typically occupies an important position in patient flow, and discharge of a patient typically allows new patients to enter the ICU. More accurate prediction of ICU discharge date therefore results in a more reliable and predictable care process, not just in the ICU but throughout the patient care pathway, including the operating room and the ward.

There are some limitations to our study. It was conducted among just one group at a single centre, which may limit the generalizability of our results to other centres. In addition, classification of variables will not be the same in all centres. The development of models like those proposed here requires effort. Also, some variables may change over time, and so the model should be updated periodically to maintain accuracy of prediction. Moreover, data such as pathological stage and how radical the surgical procedure is are typically only available during the second week postoperatively at our hospital, and so this information cannot be used as variables in a model during the first week. The extent of lymph node dissection was standardized in the surgical approach, and so an extra variable was not needed for this type of operation [33]. In the present study, the mean LOS at the ICU after oesophagectomy was long. The majority of patients had a LOS of more than 3 days. Estimates of ICU LOS in the literature vary, but the ICU LOS at our institution appears to be reasonable in comparison with those reports [8,34,35]. Clearly, ICU LOS prediction models are of greater value to hospitals with patient groups that have longer mean LOS. In our study, the particular patient group chosen was selected for that reason, so that we could experiment with the creation of such a model.

A strength of our study is that there is no selection bias; all patients were admitted to the ICU postoperatively according to protocol. Although our application sample differs statistically from the construction sample for some variables, the model appeared robust enough to make accurate predictions. Furthermore, a multiple imputation method was used to impute missing values, and so all patients were indeed included in the analysis.

Conclusions{Level 1 heading}

We constructed, validated and tested three models, with incrementally enhanced precision, to predict LOS for individual patients in the ICU. The intra-ICU model proved able to predict LOS most accurately. For the highly variable LOS of oesophagectomy patients, this model appears to counter the commonly held view that LOS is unpredictable. Moreover, comparing the predictions of the model with historically determined mean LOS yielded significant improvement in terms of ICU capacity.

References

1 Duke GJ: Metropolitan audit of appropriate referrals refused admission to intensive care. *Anaesth Intensive Care* 2004, 32:702-706.

2 Garrouste-Org, Montuclard L, Timsit JF, Reignier J, Desmettre T, Karoubi P: Predictors of intensive care unit refusal in French intensive care units: a multiple-cen-

ter study. *Crit Care Med* 2005, 33:750-755.
3 Levin PD, Worner TM, Sviri S, Goodman SV, Weiss YG, Einav S, Weissman C, Sprung CL: Intensive care outflow limitation-frequency, etiology, and impact. *J Crit Care* 2003, 18:206-211.
4 Bakker J, Damen J, van Zanten AR, Hubben JH: Admission and discharge criteria for intensive care departments. *Ned Tijdschr Geneeskd 2003*, 147:110-115
5 American Society of Anesthesiologists: New classification of physical status. *Anesthesiology* 1963, 24:111
6 Rempe-Sorm V: Management of postoperative complications. In *Barrett's Esophagus.* Edited by Tilanus HW, Attwood SE. Dordrecht: Kluwer Academic Publishers; 2001:357-366.
7 Sobin LH, Wittekind CH (editors). *TNM Classification of Malignant Tumours (UICC)*. New Jersey: John Wiley & Sons; 2002:60-65.
8 Avendano CE, Flume PA, Silvestri GA, King LB, Reed CE: Pulmonary complications after esophagectomy. *Ann Thorac Surg* 2002, 73:922-926.
9 Bartels H, Stein HJ, Siewert JR: Preoperative risk analysis and postoperative mortality of oesophagectomy for resectable oesophageal cancer. *Br J Surg* 1998, 85:840-844.
10 Ferguson MK, Martin TR, Reeder LB, Olak J: Mortality after esophagectomy: risk factor analysis. *World J Surg* 1997, 21:599-603.
11 Law S, Wong KH, Kwok KF, Chu KM, Wong J: Predictive factors for postoperative pulmonary complications and mortality after esophagectomy for cancer. *Ann Surg* 2004, 240:791-800.
12 Tsutsui S, Moriguchi S, Morita M, Kuwano H, Matsuda H, Mori M, *et al.*: Multivariate analysis of postoperative complications after esophageal resection. *Ann Thorac Surg* 1992, 53:1052-1056.
13 Spiegelhalter DJ: Probabilistic prediction in patient management and clinical trials. *Stat Med* 1986, 5:421-433.
14 Steyerberg EW, Eijkemans MJ, Harrell FE Jr, Habbema JD: Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. *Stat Med* 2000, 19:1059-1079.
15 Steyerberg EW, Eijkemans MJ, Harrell FE Jr, Habbema JD: Prognostic modeling with logistic regression analysis: in search of a sensible strategy in small data sets. *Med Decis Making* 2001, 21:45-56.
16 Rubin DB, Schenker N: Multiple imputation in health-care databases: an overview and some applications. *Stat Med* 1991, 10:585-598.
17 Strum DP, May JH, Vargas LG: Modeling the uncertainty of surgical procedure times: comparison of log-normal and normal models. *Anesthesiology* 2000, 92:1160-1167.
18 Duan N: Smearing estimate: a nonparametric retransformation method. *J Am Stat Assoc* 1983, 383:605-610.
19 Efron B, Tibshirani RJ: *An Introduction to the Bootstrap.* New York, NY: Chapman and Hall; 1993.
20 Steyerberg EW, Harrell FE, Borsboom GJJM, Eijkemans MJC, Vergouwe Y, Habbema JDF: Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol* 2001, 54:774-781.
21 Harrell FE, Lee KL, Mark DB: Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996, 15:361-387.
22 van de Pol MA, van Houdenhoven M, Hans EW, Boersma E, Bax JJ, Feringa HH, Schouten O, van Sambeek MR, Poldermans D: Influence of cardiac risk factors and medication on length of hospitalization in patients undergoing major vascular surgery. *Am J Cardiol* 2006, 97:1423-1426.
23 Van den BG, Wilmer A, Hermans G, Meersseman W, Wouters PJ, Milants I, *et al.*: Intensive insulin therapy in the medical ICU. *N Engl J Med* 2006, 354:449-461.
24 Ammori BJ, Larvin M, McMahon MJ: Elective laparoscopic cholecystectomy: preoperative prediction of duration of surgery. *Surg Endosc* 2001, 15:297-300.
25 Collins TC, Daley J, Henderson WH, Khuri SF: Risk factors for prolonged length of stay after major elective surgery. *Ann Surg* 1999, 230:251-259.
26 Janssen DP, Noyez L, Wouters C, Brouwer RM: Preoperative prediction of prolonged stay in the intensive care unit for coronary bypass surgery. *Eur J Cardiothorac Surg* 2004, 25:203-207.
27 McMeekin DS, Gazzaniga C, Berman M, DiSaia P, Manetta A: Retrospective review of gynecologic oncology patients with therapy-induced neutropenic fever. *Gynecol Oncol* 1996, 62:247-253.
28 Rosenfeld R, Smith JM, Woods SE, Engel AM. Predictors and outcomes of extended intensive care unit length of stay in patients undergoing coronary artery bypass graft surgery. *J Card Surg* 2006, 21:146-150.
29 Stoica SC, Sharples LD, Ahmed I, Roques F, Large SR, Nashef SA: Preoperative risk prediction and intraoperative events in cardiac surgery. *Eur J Cardiothorac Surg* 2002, 21:41-46.
30 Tu JV, Mazer CD, Levinton C, Armstrong PW, Naylor CD: A predictive index for length of stay in the intensive care unit following cardiac surgery. *CMAJ* 1994, 151:177-185.
31 Aronow HD, Peyser PA, Eagle KA, Bates ER, Werns SW, Russman PL, Crum MA, Harris K, Moscucci M: Predictors of length of stay after coronary stenting. *Am Heart J* 2001, 142:799-805.
32 Zimmerman JE, Kramer AA, McNair DS, Malila FM, Shaffer VL: intensive care unit length of stay: benchmarking based on Acute Physiology and Chronic Health Evaluation (APACHE) IV. *Crit Care Med* 2006, 34:2517-2529.
33 Lerut T, Coosemans W, Decker G, De Leyn P, Moons J: Surgical techniques. *J Surg Oncol* 2005, 92:218-229.

34 Chandrashekar MV, Irving M, Wayman J, Raimes SA, Linsley A: Immediate extubation and epidural analgesia allow safe management in a high-dependency unit after two-stage oesophagectomy. Results of eight years of experience in a specialized upper gastrointestinal unit in a district general hospital. *Br J Anaesth* 2003, 90:474-479.
35 Kuo EY, Chang Y, Wright CD: Impact of hospital volume on clinical and economic outcomes for esophagectomy. *Ann Thorac Surg* 2001, 72:1118-1124.

Appendix A

The LOS with the preoperative prediction model can be calculated as:

$$LOS_{preoperative}\text{ (days)} = 1.26 \times 1.10\text{\^{}}exp \times 1.16\text{\^{}}age/10 \times 0.91\text{\^{}}fev \times 1.46\text{\^{}}gerd \times 1.29\text{\^{}}vasc \text{ v } 1.74\text{\^{}}neu \times 0.81\text{\^{}}che \text{ x } 2.13\text{\^{}}tte \times 1.63$$

Where exp is the estimate of the surgeon for the session time per minute; age is patient's age per year; fev is FEV1 in liters; gerd is 1, if patient has gastroesophageal reflux disease, 0 if not; vasc is 1, if patient has a vascular comorbidity, 0 if not; neu is 1, if patient has a neurological comorbidity, 0 if not; che is 1, if patient had chemotherapy, 0 if not; tte is 1, if transthoracic approach, 0 if transhiatal approach; and 1.63 is the smearing factor.

Appendix B

The LOS with the postoperative prediction model can be calculated as:

$$LOS_{postoperative}\text{ (days)} = 0.44 \times 1.20\text{\^{}}age/10 \times 0.85\text{\^{}}fev \times 1.53\text{\^{}}gerd \times 1.32\text{\^{}}vasc \times 1.82\text{\^{}}neu \times 1.79\text{\^{}}tte \times 1.52\text{\^{}}rec \times 1.07\text{\^{}}time \times 1.14\text{\^{}}col \times 0.94\text{\^{}}blood \times 1.12\text{\^{}}urin \times 0.83\text{\^{}}eda \times 1.09\text{\^{}}rmv \times .03\text{\^{}}peep \times 1.58$$

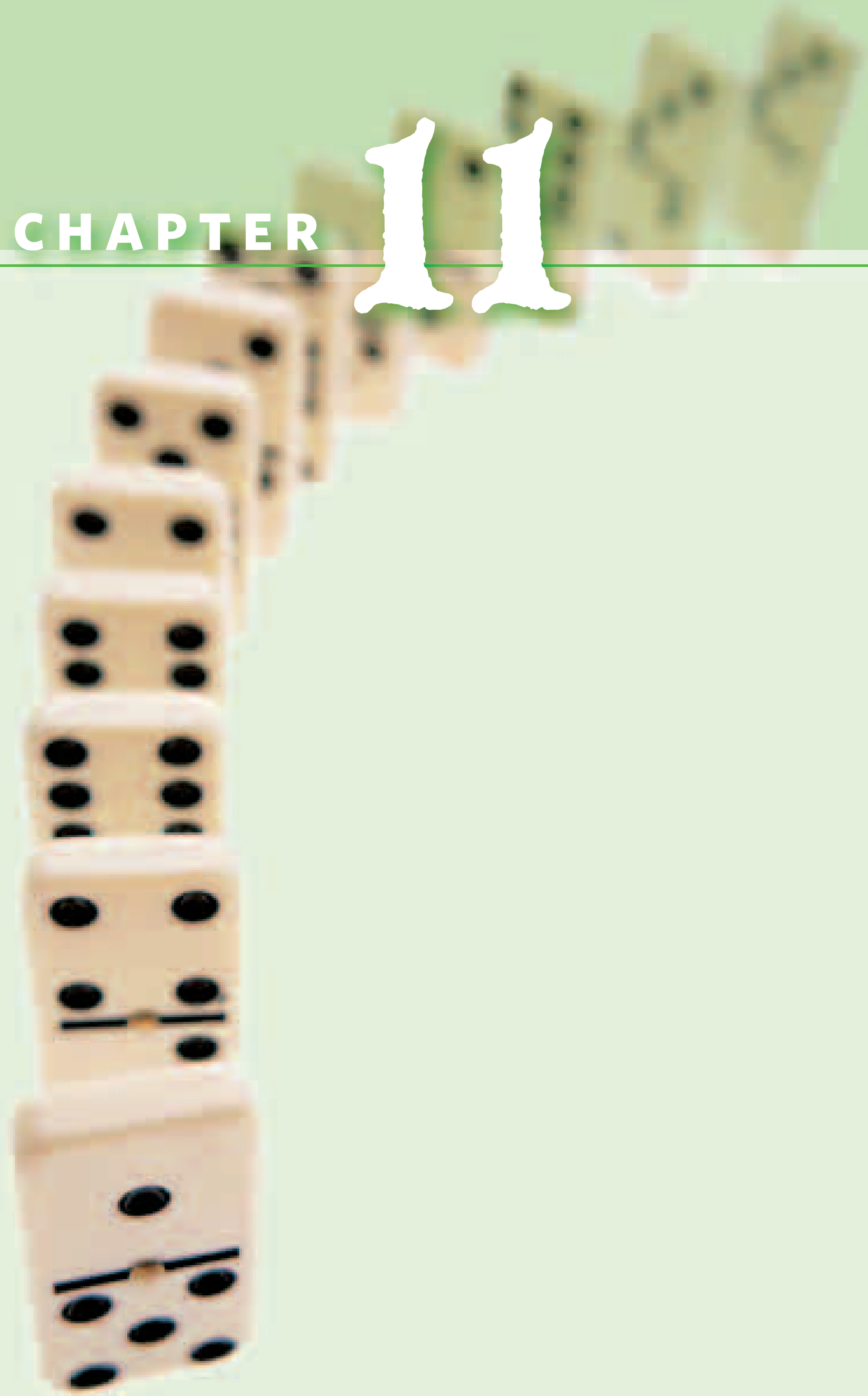
Where age is patient's age per year; fev is FEV1 in liters; gerd is 1, if patient has gastroesophageal reflux disease, 0 if not; vasc is 1, if patient has a vascular comorbidity, 0 if not; neu is 1, if patient has a neurological comorbidity, 0 if not; tte is 1, if transthoracic approach, 0 if transhiatal approach; rec is 1 if the patient had a colon reconstruction, 0 if stomach reconstruction; time is released operating time per minute; col is amount of colloids administration in liters; blood is amount of blood loss during the procedure in liters; urin is amount of urine produced during the procedure in liters; eda is 1, if EDA was administered, 0 if not; rmv was the maximum respiratory minute volume during the procedure in liters; peep is the highest positive end- expiratory pressure in cmH_2O; 1.58 is the smearing factor.

Appendix C

The LOS with the intra-ICU prediction model can be calculated as:

$$LOS_{intra\text{-}ICU}\text{ (days)} = 1.96 \times 1.09\text{\^{}}age/10 \times 1.20\text{\^{}}tte \times 1.05\text{\^{}}rmv \times 0.89\text{\^{}}vas \times 1.33\text{\^{}}chy \times 1.79\text{\^{}}ana \times 1.69\text{\^{}}oths \times 2.02\text{\^{}}pulm \times 1.60\text{\^{}}mi \times 1.58\text{\^{}}inf \times 1.44\text{\^{}}oth \times 1.17$$

Where age is patient's age per year; vasc is 1, if patient has a vascular comorbidity, 0 if not; tte is 1, if transthoracic approach, 0 if transhiatal approach; rmv was the maximum respiratory minute volume during the procedure in liters; vas is 1 if the patient had vasopressor administration, 0 if not; chy is 1, if patient has a chylothorax, 0 if not; ana is 1, if patient has anastomosis leakage, 0 if not; oth is 1, if patient has an other surgical complication, 0 if not; pulm is 1, if patient has pulmonary complication, 0 if not; mi is 1, if patient has a myocardinfarction, 0 if not; inf is 1, if patient has an infection, 0 if not; oth is 1, if patient has an other non- surgical complication, 0 if not ; 1.17 is the smearing factor.



CHAPTER 11

Predicting the Unpredictable: an improved prediction model of operative session times using individual characteristics and the surgeons' estimate

Marinus J. C. Eijkemans, Mark Van Houdenhoven,
Duy-Tien Nguyen, Eric Boersma,
Ewout W. Steyerberg, Geert Kazemier.

Submitted

Introduction

Operating rooms (OR's) are of pivotal importance to a hospital, consuming a considerable part of it's total budget [1]. Typically, over 60 percent of patients admitted to a hospital are treated on the OR (REF). Planning of patients, i.e. the decision to treat a patient and the timing of treatment, is often constrained by limitations in the OR capacity and in the availability of surgeons and qualified OR personnel. For this reason as well as for cost-containment, the planning of care, i.e. planning which patient to operate when, is crucial. Emergency procedures, large variety in processes, dependency on limited capacity in other parts of the care process such as intensive care units (ICU's) and a large number of specialties competing for limited OR facilities make planning complex.

Optimal planning can only be achieved when reliable predictions are available about the time needed for elective operations. When an operation takes longer than predicted, subsequent operations may need to be postponed or even cancelled. When the actual time is shorter than predicted and planned, the OR remains unused for a while. Both are undesirable and could lead to sub optimal use of the OR. Furthermore, in the absence of reliable predictions, the use of advanced planning techniques makes no sense. So far, little progress is made within this area, because of the perceived unpredictability of operative times.

In many hospitals, surgeons make a routine prediction of the time needed, or historical times are taken as a reference [2 3]. If it would be possible to make more accurate predictions of the session time for individual patients, planning will be improved. Potential benefits would be twofold: a) the prediction for an individual patient will be more accurate than the average for the group of patients undergoing the same operation, and b) the uncertainty (or variation) around the prediction will be smaller than the variation for the group as a whole. Previous studies have aimed to develop predictive tools by statistical modeling of operation times [2 4 5]. However, none of them has been able to make predictions for an individual patient undergoing a specific operation. Only broad categories of operations with large variation are taken into account, or only one particular type of operation is considered [4 5]. The role of the prediction made by the surgeon is ignored, or it is compared to predictions made by automated planning software [2].

We aim to predict the total operative session time, using the surgeon's estimate of operative time and procedure, team and patient characteristics of individually coded operations of a general surgical department.

Subjects and methods

Subjects All operative sessions at Erasmus Medical Center (Rotterdam, the Netherlands) are registered electronically since January 1993. For the purpose of this study, data from the 'operation' database (OPERA, operation administration) were matched with global patient data from the general electronic Hospital Information System (HIS) and with more detailed patient data from a previous study on risk factors for complications of surgery [6]. We initially selected 18,838 consecutive elective operations performed by the department of General Surgery until June 2005. Emergency operations were not considered. Operations that had not been performed during the last three years ($n = 1{,}338$), operative sessions for which no matching between the databases could be obtained ($n = 21$) and operations that were performed twenty times or less ($n = 1{,}120$) were excluded. This left 16,359 operations for analysis. Operations were classified in 135 categories, according to the main procedure during the operation. These operations are typical for a surgical department in an academic, tertiary referral center.

The outcome to be predicted was total session time, defined as the time from entry of the patient into the operating room until leaving it. We will systematically use the term operation to characterize a session, and use surgical procedure for the possibly multiple surgical activities that are part of an operation.

Session characteristics were the number of separate procedures within the operation and whether it was a laparoscopic procedure. Team characteristics were the total of the ages of the surgical team, as a measure of combined experience, age of the youngest and oldest surgeon, the number of surgeons and similarly, the ages and number of anesthesiologists.

Patient characteristics were age and sex, the number of admissions to the hospital prior to the operation, and the length of the current admission. For patients who were operated before 2001, additional data were available on the presence or absence of the following cardiovascular risk factors: diabetes, hypercholesterolemia, hypertension, history of heart failure, history of cerebrovascular accident (CVA), history of chronic obstructive pulmonary disease (COPD), history of renal insufficiency and history of coronary artery disease (CAD). Body mass index (BMI, kg/m^2) of the patient was known in 1,419 (8.7%) of the operations, as assessed during a previous study (@@ref).

Prediction by the surgeon Before each session, the surgeons' prediction of the total *surgical* time was routinely registered in the database, and used for planning the session. In an intern evaluation in 2002, it became evident that the time planned in this way systematically underestimated the total session time, because anesthetic time was not taken into account. Starting in 2004, a computerized planning system was used, providing the surgeon with the mean duration of previous operations of the same type. Surgeons made a subjective adjustment when necessary, which was used in planning. We assessed prospectively whether this planning system had improved accuracy by comparing the pre–2004 with the 2004–2005 data.

Recoding of the operations The members of the surgical team and the actual procedures performed during the operation of the 16,359 elective sessions were entered in the database after finishing the operation. Many operations evolve differently than initially intended and planned. Examples are oncological operations with curative intent: the patient may appear to be inoperable only during the operation. Further, operations that are planned laparoscopically may be changed to an open procedure during the operation. Finally, in case of complications during an operation, an extra experienced surgeon may be called in, who is afterwards added to the list of surgeons performing the operation. This makes the data unsuited for prediction modelling, where only factors that are pre-operatively known may be taken into account. Two of the authors (M.J.C. E. and G. K.) have gone through the list of operations, recoding the post-operative code as entered in the database to the preoperative code that was initially planned. Procedure codes that had been changed over time, in particular the codings for laparoscopic procedures, were reassigned to a unique coding.

Statistical analysis We used imputation of missing data, as this is recommended as less biased than dropping cases with missing values when developing multivariable models[7]. The multiple imputation technique, implemented by Harrell's AregImpute function in the Hmisc library in Splus, was used to properly adjust standard errors and confidence intervals for the imputation. Multiple linear regression was used to build the prediction model, with the logarithm of the total session time as dependent variable. The total session time variable was log-transformed, because of its right skewness[8]. First, a base model was fitted, containing the type of operation as a categorical predictor variable with 135 levels. As a screening step before further model building, the non-linearity of the association between the continuous predictor variables and the log (total session time) was assessed by fitting a restricted cubic spline (rcs) function, with knots at the 5th, 35th, 65th and 95th percentiles of the predictor's distribution, as an extension to the base model. In this way, 'learning-curve'-like non-linear patterns, e.g. for the ages of the surgeons, may be detected and incorporated into the prediction model, using only three degrees of freedom[9].

The session- and team-, and patient characteristics were subsequently added to the model and the improvement in predictive ability was assessed, using the non-linear functions when statistically and clinically significant. Selection of variables was applied conservatively, to minimize the risk of over fitting: all predictors with a univariable $P < 0.30$ were included into the model[10]. The predictive ability of the resulting extended models was expressed as percentage of variation in log(session time) that is explained by the model, measured by the model's adjusted R-squared (R^2). To quantify the improvement in comparison with the base model, the gain in R^2 of the extended model was expressed as a percentage relative to the variation left unexplained by the base model: $(R^2_{model} - R^2_{base}) / (1 - R^2_{base})$. The model predictions on the log(session time) were back-transformed to the original time scale, applying a correction for 'back transformation' bias, a 'smearing' factor computed as the mean value of the exponentiated residuals of the model[11].

To assess the potential impact of using the model in planning, we split the data according to the date in 2004 at which the planning was changed from the surgeon's estimate of operative time to the surgeon's estimate based on the mean duration of all previous operations of the same kind. The pre–2004 data were used to estimate the prediction model, and the resulting model was used to predict the durations of operations from 2004 onwards. The difference between predicted and observed session durations was assessed and compared to the difference between the planned and observed duration.

Analyses were performed with S-plus 7.0 (Insightful Corp).

Results

There were 10,712 operations consisting of a single surgical procedure, 3,276 of two, 1,173 of three and 1,202 operations of four or more procedures. Table 1 shows the list of operations, together with their frequency of occurrence and descriptive statistics of their duration. Note that the minimum frequency is 21, because of the selection that was made. The session times shows considerable variation between operations, with the median ranging from 42.5 to 441 minutes. The coefficient of variation (C.V.) illustrates that the variability within the same type of operation may also be considerable. The operation with the highest consistency in duration was *Thoraco abdominal vasculature – suprarenal aneurysm elective* (C.V. = 0.17), whereas *Skin – surgery on large and complicated tumours* had the relatively most unpredictable duration (C.V. = 0.76). After accounting for the operation code (the base model), the predicted session time had a 95% prediction interval with relative bounds between 0.52 and 1.91. For any specific operation, this implies that the session time may be from nearly half as short up to almost twice as long as the median for that operation.

	N	mean 4 SD	C.V.*	median (min-max)
Anus – lateral internal sphincterotomy	24	48425	0.52	42.5 (15 - 130)
Anus – examination under general anesthesia	136	54424	0.44	50 (17 - 137)
Anus – excision condyloma – skintags – fibrous polyps	56	55417	0.31	52.5 (30 - 128)
Anus – incision peri anal abscess	32	57419	0.33	54.5 (30 - 100)
Skin – necrotectomy / debridement	114	59425	0.42	55 (24 - 174)
Anus – fistulectomy – recurrence	181	64429	0.46	58 (30 - 263)
Skin – split skin non meshed	33	63429	0.46	58 (25 - 195)
Foot – amputation or exarticulation toe	139	61420	0.33	60 (25 - 132)
Skin – clinical woundexcision and debridement	60	67434	0.50	60 (20 - 255)
Abdomen – epigastric hernia repair	24	70441	0.59	60 (30 - 220)
Breast – cone excision	36	72449	0.68	60 (35 - 316)
Peritoneum – intraperitoneum cathether – capd dialysis – removal	153	69429	0.42	61 (32 - 270)
Anus - hemorrhoidectomy	100	65418	0.28	62 (15 - 125)
Peritoneum – intraperitoneum cathether – capd dialysis – removal	275	67422	0.33	64 (38 - 216)
Skin – necrotectomy / debridement	30	82439	0.47	65 (41 - 199)
Skin – split skin meshed	62	79428	0.36	70 (40 - 182)
Anus – fistulectomy – recurrence	265	82433	0.40	75 (32 - 310)
Skin – split skin meshed	39	79435	0.44	75 (39 - 225)
Skin – correction scar tissue	57	90444	0.49	75 (35 - 228)
Abdomen – umbilical hernia repair– 12 years and older	96	95448	0.51	77.5 (35 - 297)
Vascualr – introduction port-a-cath	71	83427	0.33	78 (45 - 194)
Lymphatic system – biopsy cervical lymph node	38	89440	0.45	79 (45 - 228)
Breast – cone excision	319	87440	0.46	79 (30 - 400)
Foot – partial amputation	50	83429	0.35	80 (38 - 225)
Abdomen – laparoscopy, (including biopsy)	602	93438	0.40	85 (33 - 407)
Varcosis – subfacial perforating vein surgery	87	92430	0.33	87 (40 - 191)
Abdomen – inguinal hernia recurrent repair	74	96429	0.31	90 (45 - 190)
Varicosis – stripping saphenous vein – unilateral	96	100439	0.39	90 (45 - 229)
Abdomen – incisional hernia repair	166	108452	0.49	90 (35 - 301)
Anus – transanal mucosal advancement plasty	80	95428	0.30	91 (55 - 247)
Small intestines – revision ileostomy	47	99437	0.37	91 (40 - 225)
Thyroid – extirpation unilateral node	44	94422	0.24	92.5 (49 - 150)
Abdomen – inguinal hernia unilateral repair	435	97426	0.27	95 (40 - 242)
Kidney – transplantectomy	54	100424	0.24	97 (63 - 175)
Abdomen – inguinal hernia mesh repair	222	106428	0.27	100 (53 - 234)

	N	mean 4 SD	C.V.*	median (min-max)
Peripheral vascular – correction arterialvenous fistula	43	110444	0.40	100 (25 - 240)
Skin – necrotectomy / debridement	31	1524115	0.76	101 (50 - 500)
Regional lymphanodissection – axillary dissection	42	112428	0.25	105 (45 - 188)
Lower leg – amputation	101	112431	0.28	105 (56 - 215)
Adrenal – laparoscopic extirpation	309	114438	0.34	105 (30 - 319)
Abdomen – mesh repair recurrent inguinal hernia	86	114434	0.30	106 (55 - 258)
Thorax – exploratory thoracotomy	41	120452	0.43	108 (30 - 303)
Abdomen – relaparotomy – compication – lavage – bleeding	39	126469	0.55	108 (20 - 425)
Varicosis – stripping saphenous vein – unilateral	297	111433	0.30	109 (35 - 255)
Thorax – exploratory thoracotomy	148	114429	0.26	110 (65 - 270)
Upper leg – amputation	41	112437	0.33	110 (60 - 250)
Thoracoscopy – diagnostic (including adhesiolysis)	86	124446	0.38	111.5 (55 - 305)
Adrenal – laparoscopic extirpation	78	128454	0.42	112 (50 - 351)
Anus – anterior anal repair	121	115430	0.26	113 (40 - 242)
Varicosis – excision of three or more varicous veins	34	117433	0.28	114.5 (45 - 199)
Breast – cone excision	81	138465	0.47	115 (55 - 415)
Varicosis – stripping saphenous vein and varicous veins	31	120432	0.26	116 (58 - 211)
Abdomen – incisional hernia mesh repair	379	128445	0.35	116 (45 - 378)
Bladder – permanent electrostimulation implant in sacral foramen	38	123441	0.34	116.5 (56 - 253)
Varicosis – stripping saphenous vein – unilateral	22	122426	0.21	119.5 (80 - 191)
Thorax – cervical rib resection	80	130449	0.38	120 (75 - 411)
Vagina – correction recto-vaginal fistula	43	126452	0.41	120 (60 - 382)
Colon – revision colostomy	41	143467	0.47	120 (65 - 395)
Abdomen – incisional hernia repair	40	164495	0.58	120 (70 - 375)
Colon – colostomy	30	139473	0.53	121 (34 - 372)
Thyroid – thyroidectomy: excision residual part	41	120435	0.29	122 (60 - 195)
Varicosis – stripping saphenous vein	125	130437	0.28	123 (60 - 278)
Small intestines – introduction feeding jejunostomy	45	128455	0.43	124 (50 - 355)
Gall bladder – open cholecystectomy	913	131438	0.29	125 (45 - 361)
Peripheral vasculature – exploration	51	139463	0.45	125 (50 - 330)
Thyroid – hemi-thyroidectomy	258	136436	0.26	130 (70 - 288)
Colon – colostomy	109	146450	0.34	135 (60 - 376)
Laparoscopy – diagnostic lymphatic system?	71	152471	0.47	135 (55 - 411)
Gastric – gastro enterostomy – open procedure	40	151463	0.42	138 (60 - 310)
Thorax – exploratory thoracotomy	80	143442	0.30	138.5 (65 - 267)

	N	mean 4 SD	C.V.*	median (min-max)
Thyroid – subtotal thyroidectomy	66	156457	0.37	144 (75 - 366)
Thorax – exploratory thoracotomy	150	153434	0.22	145 (92 - 248)
Rectum – enterocel repair	38	153439	0.26	145.5 (95 - 294)
Rectum – rectopexy	31	158436	0.23	149 (93 - 261)
Small intestines – resection	77	170471	0.41	152 (80 - 430)
Carotic artery desobstruction – endarteriorectomy	635	156431	0.20	153 (60 - 395)
Breast – radical mastectomy	215	164460	0.36	153 (70 - 590)
Abdomen – laparotomy	294	165467	0.41	153 (35 - 484)
Abdomen – peritoneal adhaesiolysis	116	172486	0.50	154.5 (60 - 610)
Small intestines – closing ileostomy	171	187496	0.51	160 (66 - 498)
Peripheral vasculature – aneurysm repair	30	169464	0.38	160.5 (50 - 335)
Colon – ileocoecal resection	137	172452	0.30	166 (80 - 415)
Lymphatic system – modified cervical block	25	179451	0.29	169 (95 - 285)
Splene – open splene surgery including extirpation	132	183460	0.33	170 (60 - 435)
Renal – renal transplant	438	183450	0.27	174 (109 - 640)
Peripheral vasculature – desobstruction endarterectomy	63	190475	0.39	175 (80 - 440)
Colon – right hemicolectomy	192	191462	0.33	177 (80 - 452)
Peripheral vasculature – iliac-fem-fem-cross over prosthetic bypass	110	181452	0.29	178 (89 - 385)
Regional lymphanodissection – axillary dissection	111	186456	0.30	180 (66 - 459)
Adrenal – open adrenalectomy	263	197471	0.36	182 (87 - 488)
Colon – revision colostomy	23	184461	0.33	186 (95 - 385)
Colon – subototal colectomy and ileostomy	110	201455	0.27	187.5 (120 - 364)
Colon – hartmann's procedure	37	194451	0.26	189 (103 - 300)
Rectum – intersphincterial rectal amputation	91	200463	0.31	190 (78 - 415)
Colon – right hemicolectomy	67	214474	0.34	190 (114 - 430)
Lower leg – amputation	25	200477	0.38	190 (105 - 427)
Thorax – exploratory thoracotomy	21	198447	0.24	192 (112 - 314)
Biliary system – redo biliary digestive anastomosis	49	198455	0.28	195 (126 - 345)
Gastric – total gastrectomy	24	201460	0.30	197.5 (89 - 317)
Small intestines – entero-enterostomy	35	215490	0.42	200 (82 - 480)
Gastric – resection	68	223483	0.37	203 (89 - 507)
Colon – sigmoidectomy with primary anastomosis	137	207456	0.27	205 (80 - 390)
Thyroid – total thyroidectomy	86	227484	0.37	206 (100 - 457)
Ureter – reconstruction after renal tranplant	27	2394114	0.48	209 (90 - 485)
Colon – revision colostomy	45	223466	0.30	210 (125 - 468)

	N	mean 4 SD	C.V.*	median (min-max)
Peripheral vasculature – iliac-femoro-popliteal bypass – prosthetic	244	221472	0.33	210 (100 - 635)
Colon – closing colostomy	123	222480	0.36	212 (75 - 537)
Gall bladder – laparoscopic cholecystectomy	25	2314117	0.51	219 (95 - 615)
Biliary system – choledocho jejunostomy	86	248469	0.28	242 (132 - 430)
Rectum – anterior resection	44	250454	0.22	245 (144 - 395)
Biliary system – redo biliary digestive anastomosis	42	2734109	0.40	255 (135 - 669)
Rectum – rectal sigmoid resection	68	262462	0.24	258.5 (131 - 435)
Gastric – total gastrectomy	85	261457	0.22	260 (165 - 431)
Renal – open donor nephrectomy	434	268464	0.24	260.5 (135 - 495)
Thoracoscopy – diagnostic (including adhesiolysis)	27	262482	0.31	265 (105 - 470)
Peripheral vasculature – iliac-femoro-popliteal bypass – venous	149	278467	0.24	271 (149 - 495)
Pancreas – resection body-tail	102	2884108	0.37	275 (85 - 667)
Peripheral vasculature – iliac-femoro-tibial bypass – prosthetic	32	275484	0.30	282.5 (114 - 505)
Thorax – exploratory thoracotomy	44	308497	0.31	289.5 (170 - 494)
Peripheral vasculature – iliac-femoro-tibial bypass – venous	105	313471	0.23	298 (165 - 490)
Oesaphagus – resection and reconstruction with colon	274	310494	0.30	298.5 (145 - 623)
Oesaphagus – resection and reconstruction with stomach	373	307479	0.26	300 (155 - 878)
Rectum – rectal amputation	72	3114115	0.37	300 (93 - 732)
Abdominal vasculature aneurysm repair	437	311474	0.24	305 (117 - 627)
Abdominal vasculature – aortic-iliacal bypass – bifurcation – tube prosthesis	187	319471	0.22	307 (195 - 730)
Liver – partial resection non-traumatic	126	3234118	0.37	310.5 (105 - 780)
Biliary system – hepatico jejunostomy	48	3254120	0.37	312 (140 - 618)
Small intestines – rectal resection and removal ileostomy and implacing a ileo-anal anastomosis	39	323459	0.18	316 (243 - 520)
Rectum – resection and colo-anal anastomosis with j-pouch	90	324467	0.21	319.5 (209 - 645)
Colon – proctocolectomy and ileo-anal anastomosis	89	3434103	0.30	335 (188 - 1095)
Abdominal vasculature – aortic bifurcation prosthesis and renal artery repair	28	369497	0.26	341 (248 - 755)
Liver – lobe resection	150	3974130	0.33	395 (108 - 885)
Thyroid – total thyroidectomy	72	4514131	0.29	435.5 (125 - 840)
Pancreas – whipple's procedure	292	441490	0.20	438 (141 - 804)
Thyroid – total thyroidectomy	21	468481	0.17	440 (370 - 665)

Table 1 *Operation codes with descriptive statistics of the operative session duration (in minutes; *C.V = Coefficient of variation, Standard deviation (SD) divided by the mean).*

The historical pattern in the difference between the surgeon's expectation of operative time and the observed total session time is depicted in Figure 1. A systematic underestimation is evident until 2004, the median difference was 31 minutes. The use of a computerized planning system providing the surgeon with the mean of previous operations, introduced in planning in 2004, clearly resulted in improved correspondence between expectations and observations.

Table 2 shows the session, team and patient characteristics in our study population. On average, patients were 55 years old, with a range from 11 to 95, and the sex distribution was about equal. The predictive effects of the characteristics on the log(total session time) are also shown in Table 2, as well as the significance of the non-linearity in this association, as tested by the spline function. Figure 2 shows the 5 parameters that had a non-linear association with the log(total session time): surgical team: youngest, oldest and summed age, youngest age of the anesthesiologists and the number of previous hospital admissions of the patient. When the youngest member of the surgical team was aged below 30, the total session time was higher with younger age, reflecting both a learning curve and the teaching function of an academic hospital: the younger the resident, the more time is spent with teaching and practice aspects. Between 30 and 35, the total time increased with older age, reflecting the increasing complexity of cases that a young surgeon is allowed to perform with increasing age. For ages of the youngest surgeon over 35, the duration goes down with increasing age, reflecting the high experience of the team in this

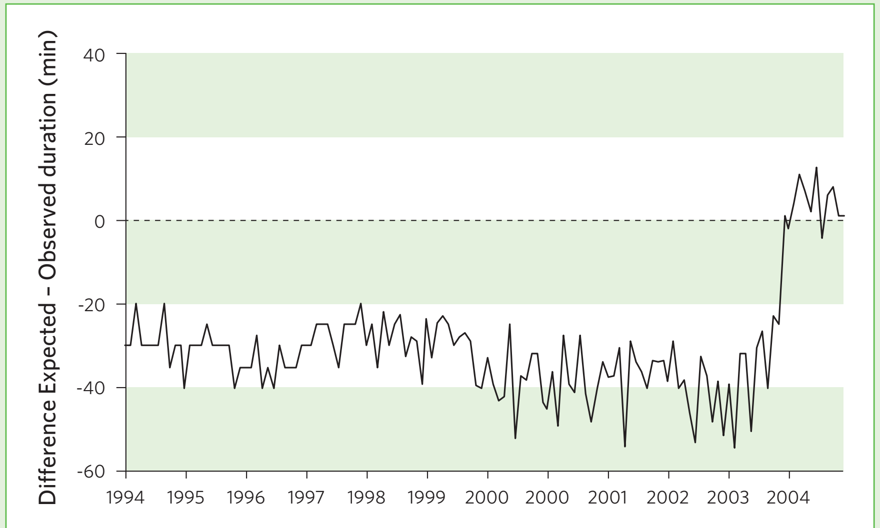


Figure 1 *Difference between the surgeon's pre-operative estimate of operative time ('expected') and the post-operatively observed session duration ('observed'), against calendar time.*

Predictor		Median (min-max) N (%)	Multiplication Factor for session duration
Session characteristics			
Number of separate procedures	1	10,707(65%)	1
	2	3,271(20%)	1.21 (1.20, 1.23)
	3	1,171(7%)	1.36 (1.33, 1.39)
	4	718(4%)	1.49 (1.45, 1.53)
	>=5	492(3%)	1.69 (1.62, 1.75)
Laparascopic procedure?	No	14,320(88%)	1
	Yes	2,039(12%)	1.13 (1.10, 1.15)
Year of surgery (per year)		1998(1993-2005)	1.01 (1.009, 1.011)
Team characteristics			
Number of surgeons	1	701(4%)	1
	2	13,145(80%)	1.16 (1.13, 1.19)
	3	2,513(15%)	1.36 (1.33, 1.40)
Summed ages of the surgical team		75(27-161)	***
Age of the youngest surgeon		33(25-61)	***
Age of the oldest surgeon		40(27-83)	***
Number of anesthesiologists	0	95(0.6%)	1
	1	3,726(22.8%)	1.10 (1.03, 1.17)
	2	12,538(76.6%)	1.11 (1.04, 1.18)
Summed ages of the anesthesiologists		71(27-115)	1.0003 (1.0000, 1.0006)
Age of the youngest anesthesiologist		33(25-64)	0.9990 (0.9981, 0.9998)
Age of the oldest anesthesiologist		41(27-65)	1.0009 (1.0002, 1.0015)
Patient characteristics			
Age		55.7(11.6-95.3)	***
Sex	Female	7,601(48%)	1
	Male	8,370(52%)	1.05 (1.04, 1.06)
Number of previous admissions		2(1-42)	***
Length of the current admission		1(0-322)	1.0004 (1.0000, 1.0008)
First operation*	No	4,291 (39%)	1
	Yes	6,579 (61%)	0.986 (0.973, 0.999)
Body mass index**		24.8(12.8-50.4)	1.005 (1.003, 1.007)

Predictor	Median (min-max) N (%)	Multiplication Factor for session duration
Presence of cardiovascular risk factors*		
Diabetes	438(4%)	0.99 (0.96, 1.02)
Hypercholesterolemia	144(1%)	0.95 (0.91, 0.998)
Hypertension	958(9%)	1.007 (0.98, 1.02)
History of heart failure	346(3%)	0.99 (0.96, 1.02)
History of CVA	127(1%)	0.99 (0.94, 1.04)
History of COPD	174(2%)	1.01 (0.97, 1.05)
History of renal insufficiency	698(6%)	1.00 (0.97, 1.03)
History of CAD	822(8%)	1.00 (0.98, 1.02)

Table 2 *Description of the predictors for session duration, and their significance when added as a single factor to the base model (containing operation code only). Data of 16,359 operative sessions. (* Data available for 10,864 operations, all before 2001. ** Data available for 1,499 operations. *** significant non-linearity, tested by adding 4-knots restricted cubic spline (rcs) to the base model.*

Figure 2 *Graphs of the significant non-linear associations between predictor variables and the total session time. Assessment of non-linearity was performed with a 4-knot restricted cubic spline function, in a regression model for the log(session time) that already contained the operation code as a predictive factor.*

case. The pattern in the effect of the age of the oldest surgeon is reversed. The older the oldest surgeon is, the longer the operation takes: if the oldest surgeon is very young, the operation is apparently of a simple enough type to allow for a relatively inexperienced team. For older ages, the operation is apparently so difficult that a very senior supervisor needs to be present. For anestiologists only the age of the youngest member of the team has a non-linear association with log(session time). A clear learning curve is visible until the age of 35. Finally, the most important patient variable was the number of previous hospital admission, which shows an increase in session time with a greater number of previous admissions. However, more than 5 previous hospital admissions did not make a difference in total session time.

Table 3 summarizes the contribution to the model of the predictive factors. The biggest improvement in adjusted R^2 is due to the session characteristics (the number of separate procedures within the operation, indicating the relative complexity of the operation and the year of surgery), and lesser so the team characteristics. Patient characteristics have only a limited influence, once the session and team characteristics are accounted for. Finally, a substantial improvement came from the surgeon's estimate of the operative time. The model explains close to 80% of the total variation in log(session times), which corresponds to 25% of the variation left unexplained by the base model. When the surgeon's estimate was added as a single factor to the base model, 76.7% of the variation was explained, a relative improvement of 13.9%. For any specific operation, the session time predicted by the final model has a 95% prediction interval with relative bounds from 0.59 to 1.71.

Parameters in the model	Adjusted R2 (%)	Adjusted R^2 gain, relative to the Base model (%)
Base model	72.9	0
+ Session characteristics	76.1	11.7
+ Team characteristics	76.9	14.8
+ Patient characteristics	77.1	15.6
+ BMI	77.2	16.1
+ CVD risk factors	77.3	16.1
+ Expected duration**	79.6	24.8

Table 3 *Predictive models for total operative session time: adjusted model R^2 of the base model containing only the operation code, and subsequent extensions*. (* Model extensions contain the variables that had a univariable P value < 0.30 (Table 2) ** Subjective pre-operative expectation of the surgeon of the operative time)*

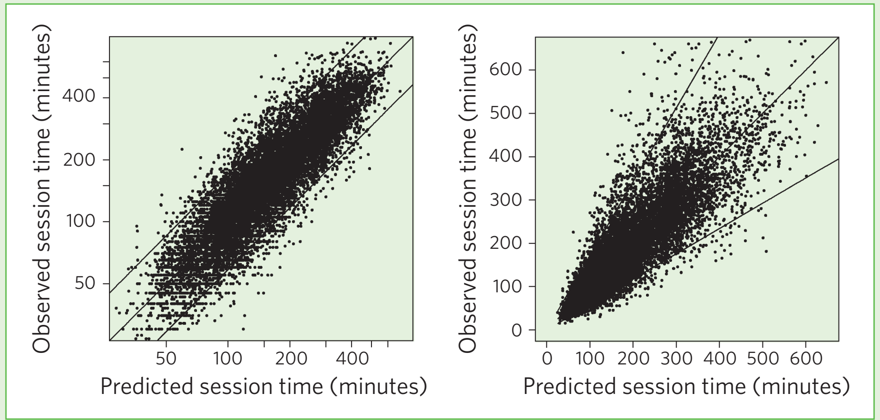


Figure 3 *Graph of predicted operative session time by the model against the observed session time, on logarithmic scales (upper panel) and on the true time scale, after correction for back-transformation bias (lower panel).*

The goodness-of-fit of the model is shown graphically in the upper panel of figure 3 on the log-transformed scale. No substantial deviation from a symmetrical scatter around the regression line is present. The lower panel of the same figure shows the same data on the original scale, where a correction has been used for the 'back transformation' bias [11] (smearing factor: 1.04).

The added value of the model in daily planning is illustrated in Table 4. The total amount of over time as well as idle time is substantially reduced when using the model predictions instead of the surgeon's planning based on historical data. The relative reduction was on average 17% and 19% for idle time and over time respectively

	Sum (minutes)	No. Patients	Mean (minutes / operated patient)
Current method (surgeon's estimates based on historical data)			
Idle time	35,956	789	23
Overtime	40,912	768	26
Total	76,868	1,555	
Prediction model			
Idle time	29,784	802	19
Overtime	32,215	753	21
Total	61,999	1,555	
Difference (current method – prediction model)			
Idle time	6,172		4
Overtime	8,697		5
Total	14,869		

Table 4 *Model performance: Idle and overtime of current method and prediction model.*

Discussion

We have studied the influence of session-, team- and patient characteristics on the duration of operations from a general surgical department in an academic hospital, and we have assessed whether the prediction made by the surgeon had a predictive effect independent of the other factors. Given an individual operation, the session characteristics were the most important predictors of total session time, and lesser so team characteristics. Particular finding in this study were the non-linear patterns in the effects of the ages of the team members on the total session time. Effects of the teaching environment and growing experience were expressed in these patterns. Patient characteristics had hardly any influence, once the session and team characteristics were accounted for. Finally, a substantial improvement came from the surgeon's prediction of the operative time. We assessed prospectively that the model may reduce idle time and over time by 17% and 19 % respectively.

Our results show that the prediction for long operations is less accurate than for short operations. It may seem that the model is therefore not very useful, because the danger of a large deviation from the planned duration is biggest for the long durations. However, very long durations are not common, as can be seen from Figure 2. It is the large bulk of short-duration operations that determines the effectiveness of planning, and these can be predicted quite accurately. Further, operations that are anticipated to take a long time are usually planned as the only operation in the OR on that day.

The database used in the current study was designed for administrative reasons, to have access to the production realized by the department. Therefore, it does not contain the operations that were planned, but instead the operations that were actually performed. For the purpose of scientific research there is a clear registration deficit in this respect: using such a database to predict operative time for scheduled procedures will introduce errors related to deviations from the intended operative strategy during surgery. Oncological procedures are particularly prone to these errors because of unforeseen metastases that can force surgeons to refrain from a curative resection. The results should therefore be used with care for planning oncological operations.

Another issue of usefulness of all statistical modeling is that the factors in the model need to be available online. Particularly, the coding system of the operations needs to be implemented electronically, patient data should be available online and the calculations should preferably be performed electronically.

Year of surgery appeared to be important in our analysis; the median operative time for all procedures increased significantly from 1996 to date (data not shown). This difference cannot be explained by changes in the operative portfolio because this was already corrected for. However from 1996 onwards an active fellowship in upper and lower gastrointestinal surgery and hepatobiliary surgery with three junior surgeons was established at our institution. This had implications for the operative time of these procedures apparently. Also the number of attending surgeons increased in that same period because these fellows were more often supervised during surgery for at least parts of the operation.

The surgeons' estimate of operative time was a strong predictor of total session time and a significant addition to the more objective factors already in the model. Even when very specific cardiovascular risk factors were included that resemble overall comorbidity, the surgeons' estimate remained a very important factor. A potential problem could be the reproducibility of this estimate: it is a subjective assessment by a surgeon, not an objective factor. To a lesser extent this objection could also be raised against our use of the surgical procedures classification: a surgeon enters it into the database, which might lead to some misclassification or non-uniform coding. However, there were many different surgeons involved, and it is hard to imagine that these factors can be such strong predictors, when the data entered are subjective or random to any extent. Moreover, it is unlikely that surgeons from the academic hospital in Rotterdam, the Netherlands would do worse or better than surgeons from elsewhere in planning their operations.

Planning of operations is often considered difficult, because of the unpredictability of operations. Now, the reverse may become true: because we can predict, serious planning becomes feasible. The amount of detail of the current model, using operation codes at the lowest level plus session-, team- and patient characteristics, allows for operational planning of care: The predictions provided by the model are directly applicable in daily planning. Here, it is not the best estimate of the duration that is of interest, but rather the chance that the operation will be finished within a certain amount of time. The prediction interval provided by the model can give just this information. Previous studies on predictive factors for total session time cannot be used for daily planning: Strum et al. [4] developed a prediction model with a couple of patient characteristics applied to very broad categories of operations. The variation between operations within such categories is so large that this model will not be useful for daily planning.

The model can also be used for planning at the tactical level (REF), i.e. how many OR's should we keep open on a given month? There, it is the sum of the best estimates of duration that determines the demand per day. In that case the back-transformation bias is disturbing and a correction should be made, as we have done. Even in strategic planning, e.g. how many and which kind of new OR's to build, the model could be used, by projecting the consequences for OR capacity of changing characteristics of the population.

We conclude that a prediction model could be developed containing detailed procedure codes and session, team and patient characteristics. The surgeons' estimate together with specific aspects of the session and the experience of the surgical team are the best predictors of the session time of a given operation. Use of prediction models can improve the planning of operating rooms. This approach can readily be extended to other specialties and other parts of the care process.

Acknowledgements This study is financially supported by 'ErasmusMC Doelmatigheidsonderzoek'.

References

1 Gordon T, Paul S, Lyles A, Fountain J. Surgical unit time utilization review: resource utilization and management implications. *J Med Syst* 1988;12:169–79.
2 Wright IH, Kooperberg C, Bonar BA, Bashein G. Statistical modeling to predict elective surgery time. Comparison with a computer scheduling system and surgeon-provided estimates. *Anesthesiology* 1996;85:1235–45.
3 Zhou J, Dexter F, Macario A, Lubarsky DA. Relying solely on historical surgical times to estimate accurately future surgical times is unlikely to reduce the average length of time cases finish late. *J Clin Anesth* 1999;11:601–5.
4 Strum DP, Sampson AR, May JH, Vargas LG. Surgeon and type of anesthesia predict variability in surgical procedure times. *Anesthesiology* 2000;92:1454–66.
5 Ammori BJ, Larvin M, McMahon MJ. Elective laparoscopic cholecystectomy: preoperative prediction of duration of surgery. *Surg Endosc* 2001;15:297–300.
6 Boersma E, Kertai MD, Schouten O, Bax JJ, Noordzij P, Steyerberg EW, et al. Perioperative cardiovascular mortality in noncardiac surgery: validation

of the Lee cardiac risk index. *Am J Med* 2005;118:1134–41.
7 Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;15:361–87.
8 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.
9 Harrell FE, Jr., Lee KL, Pollock BG. Regression models in clinical studies: determining relationships between predictors and response. *J Natl Cancer Inst* 1988;80:1198–202.
10 Steyerberg EW, Eijkemans MJ, Harrell FE, Jr., Habbema JD. Prognostic modeling with logistic regression analysis: in search of a sensible strategy in small data sets. *Med Decis Making* 2001;21:45–56.
11 Duan N. Smearing Estimate: A Nonparametric Retransformation Method. *Journal of the American Statistical Association* 1983;78:605–610.

CHAPTER 12

Uniform time registration in surgical suites. Definition system provides for comparison of work processes

Mark Van Houdenhoven, Erwin Hans,
Arjan van Hoorn, Nelleke Pullen, Robert van Barneveld,

Translated from Dutch
Medisch Contact, 13 January 2006; 61 (2), p. 52–54.

Introduction

Surgical suites will have different routines. Comparison of operating room processes enables mutual learning on effectiveness. To this aim we need uniform definitions for performance indicators. The university medical centers in the Netherlands together with the University of Twente developed a model.

The surgical suite is a fundamental feature of any hospital; on average 60% of admissions is for the purpose of surgical intervention.

Medical procedures in the operating room (OR) are usually intricate, and performed under high pressure. Availability of expensive equipment and scarcity of staff result in difficult planning problems that may jeopardize efficacy. The surgical suite therefore is a bottleneck in patient flow, and characterized by great turbulence, as strictures tend to be. In other words, when the surgical suite sneezes, the whole hospital catches a cold.

In the spirit of mutual learning, it would seem wise to compare effectiveness of work processes in various surgical suites. Valid comparison is only possible, however, if all participants speak the same language, that is: use a uniform registration system that defines the relevant moments in the OR process in an unambiguous manner.

A joint project group of the university medical centers in the Netherlands and the University of Twente developed such a time registration system. The definitions used are in part based on international definitions.[1 2] Formulating uniform definitions that meet the requirements for different patient groups and different types of operating and anesthesia techniques appeared to be a difficult task. Nevertheless, the time registration system proposed here is complete and relatively simple.

During the operation

A series of registration moments during the operation form the building stones for effectiveness performance indicators (see figure 1). It is up to the managerial level to decide whether all these moments and periods will be recorded to serve as a basis for analysis and control. The eventual decision is a trade-off between the registration efforts required and the level of completeness that will satisfy the management's information demands.

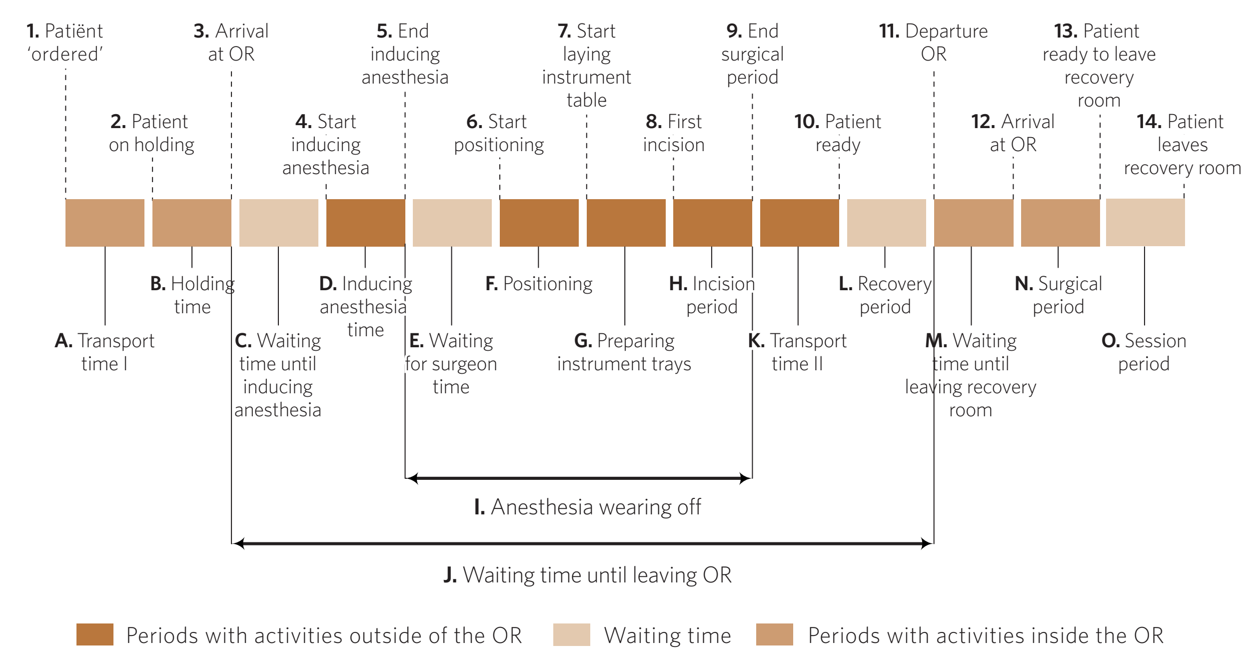


Figure 1 *Time registration OR*

Nevertheless, a minimum number of moments and periods from the complete set is needed to provide the minimally required information. These are: Arrival at OR, Start inducing anesthesia, End inducing anesthesia, End surgical period and Leaving OR. The corresponding definitions have been formulated pragmatically, aiming at achieving uniform registration.

For example, the item 'Start inducing anesthesia' is defined as: "the first time point after 'Arrival at OR' on which the anesthesiology team provides direct care to the patient". A definition like this will not raise any doubt about the method of registration when a locoregional anesthesia technique is used. Likewise it will be clear that Start inducing anesthesia precedes the very injection of the anesthetics in the case of general anesthesia. Using the proposed measuring moments, 'Session period' is defined as: "the period between Arrival at OR and Leaving OR". 'Surgical period' is defined as: "the period between 'End inducing anesthesia' and 'End surgical time'; this is the period in which operations are performed. 'Session period' minus 'Surgical period' is 'Anesthesiology period'.

It appeared necessary to introduce the term 'session' for the period between 'Arrival at OR' and 'Leaving OR', so as to distinguish from the term 'operation', seeing that a single session may comprise multiple operations, which in their turn may comprise multiple procedures. This information enables to determine per type of session the proportion of time needed to induce anesthesia relative to the total time needed for the session. This parameter could serve as input for evaluation among the medical staff of the department of Anesthesiology as well as in talks between this department and the surgical departments.

Planning of the various interventions will also be facilitated by this kind of information. Having recorded the actual 'Anesthesiology period', 'Surgical period' and 'Session period' we can compare these with the planned times. Possible discrepancies may then serve as leads for the analysis and improvement of the planning procedure. The minimum basic registration also provides for making an analysis per specialty and per type of intervention, as well as analyzing developments in these determinants over a certain period.

During the day

For the purpose of performance indicators such as those referring to OR capacity usage, and under- and over-utilization, we also need to define – next to the session-specific ones – registration moments during the day. Take for example the so-called OR time-slot: this is the OR time available for specialties to plan their sessions. An OR time-slot is the OR time delimited by 'Start OR time-slot' and 'End OR time-slot'. With the aid of such broader registration moments we can define performance indicators that relate to OR days, to specialties, to the entire OR suite, to longer periods, etc, on different levels of control.

Stimulus

The OR suite time registration model proposed here, encompassing session-specific and day-specific registration moments, facilitates a process analysis relating to the OR process at all levels: from OR capacity usage on a yearly basis to characteristics of narcosis time on a patient level. The model is useful for all hospitals.

The university medical centers do not propose this time registration system be the method of registration which every hospital intent on comparing OR performances should comply with. After all, not all registration periods are relevant to all OR suites. Recording specific data is only useful when they actually will serve as a basis for comparison, analysis and finally control of the activities and processes in the OR suites. They do propose, however, that these definitions for time registration are used to compare performances of OR suites mutually. This is thought to improve transparency, uniformity and comparability within the Netherlands and possibly on an international level.

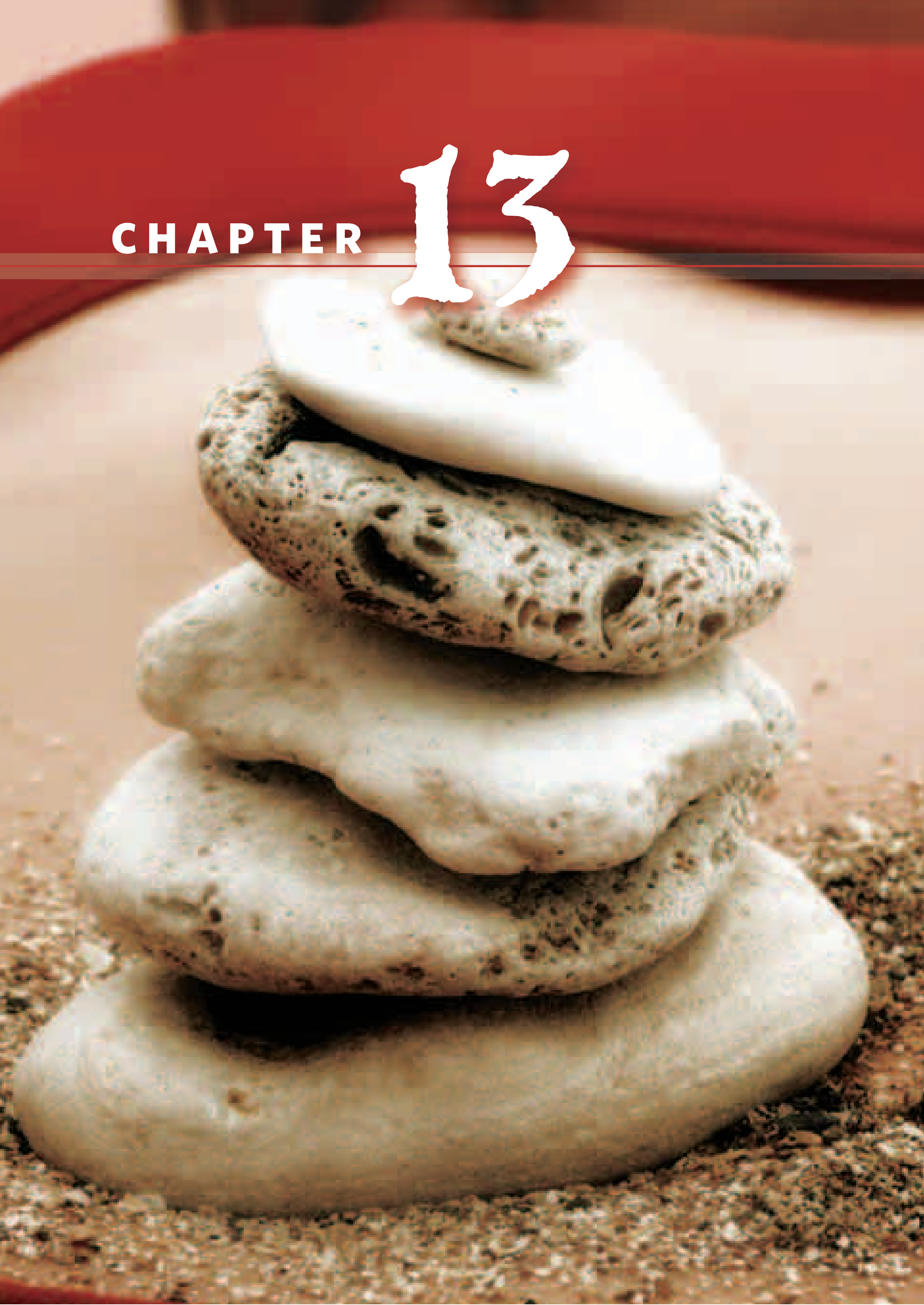
This model provides an essential step to enhance transparency in (differences between) performances of OR suites. Comparison of uniformly defined performance indicators stimulates collaborative learning about good working methods.

References

1 http://aacdhq.org/Glossary.htm

2 Audit commission, Operating Theatres, review of national findings, Audit Commission for local authorities and the Nationals Health Service in England & Wales, London 2003

CHAPTER 13



A progressive scheme for benchmarking. Dutch University hospitals find each other in the operating departments

Arjan van Hoorn, Mark Van Houdenhoven, Gerhard Wullink,
Erwin Hans, Geert Kazemier.

Translated from Dutch,
Tijdschrift Controlling, nr. 3, maart 2007, p. 28 – 31

Benchmarking compares performances of organizations in view of achieving lasting improvement. Still too often, however, comparison is 'between apples and oranges', with less favorably performing organizations unnecessarily falling prey to naming and shaming. In this article we present a new model to overcome these flaws. Developed for the healthcare sector, it is nevertheless applicable to other sectors as well. As the performance indicators chosen and the assessment procedure take into account differences in the organizations to be assessed, this model for benchmarking offers concrete starting points for lasting improvement of one's own business.

Better performance of healthcare providers has been high on the administrative agenda for quite some time. The introduction of market function – contracting out in free market – is thought to be valuable in this respect. Successful market function requires a transparent market offering freedom of choice. An essential prerequisite for having true freedom of choice from the healthcare providers, both for the patient and the healthcare insurers, is good information on the healthcare institutions' products, quality, service and tariffs. This will also benefit the government in its role as 'care' quality controller and co-funder. Benchmarking can provide a contribution to this aim.

At set times the media will publish lists of 'best' and 'worst' healthcare institutions. This practice, however, is a far cry from proper benchmarking. Real benchmarking is a continuous and structured method that uses performances and processes of the best performing organizations as points of reference to set daring goals and to improve processes.

The current benchmarking concepts, though, show two major drawbacks. First of all they do not take into account the influences of internal and external factors on an organization's process management and performance. Thus, it is wrongly suggested that organizations can be compared in all cases and to the fullest extent. The second drawback relates to 'naming and shaming': organizations are ranked according to the measured performances without providing insight into the causes of differences in performance. These two flaws may undermine stability of partner relations, so that benchmarking often is a single event rather than a continuous and collaborative process. In other words: a new benchmarking model is needed to really exploit the potential of benchmarking. This article presents such a benchmarking model, based on three principles:

1. It is no use comparing performances when insight into the differences and similarities between the benchmarking partners is lacking.
2. Benchmarking is not intended to call each other to account for performances, but to sustain a dialogue on improvement potentials in a blame-free environment.
3. Benchmarking is to provide concrete pointers to this aim.

Benchmarking in its present form

The way in which benchmarking is applied will differ for organizations, dependent on the specific situation and goals (Camp, 1993). Nevertheless, irrespective of how different applications may be, all benchmarking instruments share the following steps:

1. Selection of process, based on thorough analysis of one's own process.
2. Selection of partners we can learn from.
3. Selection of performance indicators, dependent on the process elements to be compared.
4. Measuring; the basis for determining rank order of the partners.
5. Publication of rankings.
6. Analysis of performances and processes, taking into account the underlying process and working methods.
7. Drawing up improvement plans; optimal working methods are adapted to one's own process and culture.
8. Implementation of the working methods.

A first drawback of this method is the prominent position of appraisal. As such, this will almost automatically result in a ranking list. A ranking list, however, does not provide insight into the backgrounds of the differences in performance, as the features of the system in which the performances are realized remain unknown. So we are left to guess whether the best rated working method actually is the best, or whether we are presented with a distorted picture produced by special features of the organization. Furthermore, a ranking list does not provide concrete starting points for improvement, leading to less motivation to apply benchmarking. Finally, a ranking list presents the oversimplified picture of one winner. All other partners are therefore seen as losers, with all negative effects on public image building. Consequently, staff will feel less motivated to work on improvements, being more easily inclined to plead incomparability of the obtained benchmarking results.

The hospital sector; an industry like no other

We set out to develop a new benchmarking model for the healthcare sector that would do away with the drawbacks of the current benchmarking concepts.

Taking the hospital sector as an example, we accounted for the features that make it unique.

The hospital does not exist

The hospital sector is not a uniform entity. Hospitals may differ in level of care, they may not offer all the functions of other hospitals, such as a trauma-, transplantation or training/teaching function, and there are clinics that specifically aim at the treatment of just a few clinical pictures.

Delivery of care is always a trade-off between quality of care, costs and quality of labor: a trade-off triangle in which emphasis on one factor has effect on the other factors. Various aspects play a role here: strategic choices to execute certain functions, but also non-influenceable factors such as environmental characteristics or physical and organizational structures. Demographic or cultural characteristics of the region as well as vested functions may be implicated as well. Such differences between hospitals may bear on the complexity, predictability and performances of care processes. We definitely need to view performances in the light of these differences if eventually we are to learn from one another.

The hospital: a complex political system

Glouberman and Mintzberg (2001) describe a hospital's anatomy in terms of four C's: Cure, Care, Control and Community. The four C's represent the physicians, the nurses, the managers and the trustees, respectively, each with their own task, position and impact on the care process. Benchmarking activities need to anticipate to this complex political system, which may be associated with coalition formation between factions, for instance regarding distribution of scarce resources.

Various interested parties are involved in assessing the outcomes of the benchmarking procedure. It is advisable to take this into account when defining performance indicators.

Unpredictability

Unpredictability and variability are inherent to the care process. Occasionally, a patient's treatment course will not become clear until after the first steps have been taken, and surgery may last considerably longer or shorter than expected, in spite of good planning. Furthermore, care processes typically require time-related resources that cannot be stored, such as deployment of specialists and high-quality medical equipment. Unused capacity will go lost. In other words: a hospital is characterized by a complex production environment with a high degree of uncertainty. Each hospital has its own way of dealing with this. So, benchmarking activities should not be aimed at demonstrating performances only, but at mapping out work processes as well. This will then allow explaining differences in performance in the light of differences between these processes.

A new progressive scheme

The essence of the new benchmarking model is early description of both the processes and the relevant internal and external organizational characteristics of the institutions involved. Table 1 shows the steps proposed by the new model alongside the steps in the current benchmarking models.

1. **Selection of comparable process** The healthcare sector often compares institutions as a whole. A valid option is to compare a single operational process, or a process encompassing multiple departments. Comparability is facilitated by first selecting a comparable process.
2. **Selection of comparable benchmarking partners** The hospital sector is characterized by strong pluralism. Realizing improvements will therefore require selecting partners with similar care functions and comparable patient populations. For collaborative learning to be really effective, partners must pay great attention to each other's process set-up, and this will take time. It is recommended therefore to select a limited number of benchmarking partners.
3. **Description and analysis of process- and contingency variables** This description includes physical characteristics of the building, number and location of resources, but also staff availability, (financial) structure of the organization and environmental variables. Mapping out of these 'contingency variables' (Lawrence and Lorsch, 1969) and their relations with pro-

cesses and performances will help to find explanations for differences in performance, and will provide concrete starting points. Striking differences and similarities can then be discussed and placed in the context of regional or historical differences. Recognizing and discussing unique differences is conducive to creating a safe learning environment, mutual trust and – consequently – the durability of the benchmarking process.

4 **Development of comparable indicators** The core of each benchmarking procedure is the quantitative comparison of performances expressed in indicators. Insight into the above-mentioned contingency differences offers the opportunity to account for these in choice of indicators. For example, benchmarking in home care institutions revealed that support functions are either embedded in the primary departments or centralized in a separate support facility. Thanks to step 3 this will be known in step 4, and thus we can prevent an 'apples and oranges' comparison when developing, for example, an indicator for staff productivity. Thus, measuring correct comparables requires testing the comparability of performance indicators against each partner's features as described in step 3.

5 **Selection of performance indicators by stakeholders** The performances to be measured should be influencable and relevant. Particularly when multiple stakeholders are involved it is advisable to have the stakeholders select the performance indicators. This way we ensure not only that stakeholders will remain involved in the process, but also that really important outcomes are compared.

6 **Uniform and integral measurement** Obviously, good comparability of performances requires the use of similar and unambiguous performance indicators. In addition, the underlying index figures are of importance. Utilization of a scarce resource like an operating room, for example, can be considered a performance indicator. Equally interesting, however, is the background of the utilization rate, in terms such as time trends and utilization rates per specialty. The measurement of performances should always be underpinned by an integral comparison. In this respect, a hospital's process set-up has always evolved from a trade-off between quality of care, quality of labor, and effectiveness. In the event that choices result in high effectiveness, but also in a high absenteeism and low quality of labor, we cannot label this as best practice.

7 **Analysis of differences in performance** This step combines the description (step 3) of processes and contingencies with comparable (step 4), relevant (step 5) and uniformly and integrally (step 6) measured performance indicators. Insight into the organizational differences between partners enables refining of differences in performance, and thus true appreciation of performances. In this way areas for improvement as well as improvement targets will manifest themselves in a blame-free safe environment. The established benchmarking models carry the risk that not until the analysis stage it will become clear we are comparing 'apples and oranges'.

8 **Drawing up improvement plans** The integral comparison of results offers partners with very good performance in the one field (e.g. quality of care) pointers for improvement in other fields (e.g. efficacy). Only an integral perspective will provide learning points for all partners. And so we will see growing motivation to keep learning from one another within a lasting collaboration.

9 **Implementation** The ultimate goal of each benchmarking procedure eventually is to implement improvements in one's own organization. This progressive scheme for benchmarking was designed to prepare the partners for this final step as purposefully as possible and without unnecessary repetitions of steps. Continuation of the benchmarking routine is by way of going through the steps again in good cooperation.

Conclusion

In what ways does this new comparative scheme for benchmarking differ from the established models? For one thing, process analysis already starts in step 3 of the new model, where before this was part of step 6. In addition, a new element in our model provides for discounting process and contingency differences in the development of comparable indicators and implicates stakeholders in the selection. Finally, integral assessment of differences in performance prevents oversimplified opinion making. This facilitates the continuous search for and implementation of points for improvements.

References

1 Camp, R.C., A bible for benchmarking, by Xerox, Financial Executive 1993; 4.

2 Glouberman, S. en H. Mintzberg, Managing the care of health and the cure of disease – Part I: Differentiation, Health Care Management Review 2001; 26: 56–69.

3 Lawrence, P.R. en J.W. Lorsch, Organization and environment: managing differentiation and integration, Irwin, Homewood (Ill.), 1969.

CHAPTER 14 General discussion and implications

This chapter reviews the research presented, and places the results in a wider framework. The necessity of careful weighing up considerations at the different levels of the care organizing process is discussed first. Next, the results of the studies serve to illustrate how an optimal balance can be achieved; in other words, how balancing becomes the "art of balance".

Necessity of weighing up considerations

Health care is for all of us. This very statement, hides a number of fundamental discrepancies, however, including the discrepancy between:

- the sake of an individual patient and the sake of all patients;
- the sake of patients with acute health care need and the sake of patients with plannable health care need
- the sake of the individual employee and the sake of the individual patient;
- the sake of the individual hospital and the sake of a group of hospitals in a region;
- the sake of medical technological developments and the control of health care costs;
- the sake of rapid access and the sake of high utilization;
- the sake of efficiency and the sake of the employees.

This thesis offers a methodology and instruments that allow weighing up the mentioned discrepancies. Besides, a conceptual framework was developed and partially worked out for particularly the OR/ICU setting. This involves policy making, on the basis of the balancing of the different interests, by hospital boards, specialists, partnerships, planners, OR managers, financial managers, health insurers and governmental bodies. Nevertheless, a hospital board that would expect to find answers to all questions will be disappointed. That which is available, is a conceptual framework for balancing and broadening logistics processes by means of operations research, with the use of an arsenal of techniques.

Health care is generated in the individual relation between patient and specialist. In order to give shape to this individual relation the specialist will use various scarce and costly hospital resources. Examples include operating rooms, intensive care units, clinics, outpatient departments, radiology, laboratories, financial resources, et cetera.

These resources are also used, however, for other patients treated by the same specialist, treated by other specialists, treated in other hospitals, and treated in other countries.

In addition these resources may also be used by future potential patients. Some future potential patients will announce themselves 'unplanned' and for reason of medical urgency directly demand use of scarce available resources and the individual specialist. Furthermore, patients already being treated may show unplanned complications, extra care need, et cetera.

At the level of the individual relation between patient and specialist, the operational level, we see therefore in fact a constant competition for scarce and costly resources. The purpose of the decisions at the operational level, on the basis of the weighing up of the individual patient's interests, is the best medically required care for every (potential) patient at the right moment.

This individual relation between specialist and (potential) patient does not provide, however, for any fall-back arrangement at the operational level. When a new patient presents and no or insufficient capacity has been reserved, then the only options are saying "no" to a earlier planned patient or saying no to the new patient. The possible consequence of this is, for example, a cancelled operation, sending away the ambulance, or a longer waiting list. In other words, a suboptimal utilization of the available capacity at the operational level. The solution for this problem at the operational level, without any flexibility for the actors, is found at tactical and strategic levels.

All developments in society and health care result in constant competition for the scarce nationally available resources. Other than the operational individual level, the societal level provides many possible answers to the question of how to meet the growing national demand for scarce resources. Here the balancing of the interests does not so much concern individual patients interests at a given moment, but rather the interests of groups, the hospitals, the regions, the patients, the national income, the labor market, the medical specialists, et cetera.

Possible answers are, for example, raising of national insurance contributions, admitting market regulation elements in health care, regulating health care, regulating educational capacity, accreditation policy, admitting private initiatives, regional decentralization, the new financing system, cost containment by benchmark competition, et cetera. National health care policy making will create both room and limitations for the hospitals in balancing and selecting options at a strategic level. In other words, more and more responsibility will be given to hospitals themselves considering the options for setting up their processes and using instruments and methods to weigh up options at the various levels.

The relation between the choices of the attending specialists and the political choices in society should ideally be synchronized in the hospital's strategic and tactical policies. The hospital board translates the political answers from society into strategic policy in which the partnerships and the management together establish the tactical policy. All this involves decisions to be made on utilization of the scarce and costly resources, on treatment plans, on defining emergency categories, on possibilities to call in staff in emergencies, staggering of holidays, required capacity in the weekends, collaboration with other partnerships, availability of ICU beds during the week, reserving OR time for emergency surgery, et cetera. These tactical considerations constitute the boundary conditions within which the individual specialists and other staff make their operational decisions when individual patients present themselves to them. If these boundary conditions should be lacking in adequate definition, the individual specialists and staff will not have at their disposal the right tools when confronted with clashing individual interests. The resulting ethical dilemmas for the individual physicians and planners, such as having to cancel a planned operation or having to send away the ambulance if no ICU bed is available, clearly spring from inconclusive balancing of interests at tactical and strategic levels, as illustrated in chapter 3, chapter 4, chapter 5, chapter 6 and chapter 7. Each of the choices in such a dilemma by the individual specialist is actually wrong, and it is almost unethical to force such a decision in this manner without any leeway at that level. Especially in view of the fact that techniques are available that would prevent having to make these impossible decisions. The cause of this dilemma, but at the same time the solution as well, is not be found with them, but rather at the tactical and strategic levels.

The art of balance

This thesis offers the approaches and techniques that enable to formulate at an operational, as well as tactical and strategic level, the answers within the various areas of planning and control. This can only be successful, however, when specialists and the management work together and are willing to share their knowledge with experts in operations research who can offer methods for the balancing of interests and risks in the developed framework, and taking into account the specific features of the hospital in question.

The framework for health care planning & control, chapter 2, presents four levels on which interests are weighted and choices are made. The problem is the balancing of the interests of (potential) patients, the specialists and other staff, the quality of care, and the costs. There is no such thing as "the best balance". The eventual balance is always the result of the trade-off and choices by the responsible actors at the level concerned, on the basis of their competences and the information, tools and support made available to them. The outcomes of the balancing problem at a higher level are decisive and constitute boundary conditions for the trade-off at the lower levels. Actually the available leeway for weighting the various interests is decreasing for every lower level. Should it have become too small, the actors involved are no longer able to make their own trade-offs of the various interests. The outcome of the trade-off at the higher level then no longer facilitates making a proper trade-off, but rather impedes this, as we saw earlier.

In fact there is a fifth level as well, that of society as a whole. It is actually the level that provides the greatest flexibility in balancing the stakes in health care. The actors involved at this level, such as political parties, ministries, health insurer umbrella organizations, order of medical specialists, pharmaceutical industry lobbies, and care provider umbrella organizations, make decisions on the continuity, accessibility, costs, and level and amount of health care on the long term. Choices are made for macro financing models, care control models, accreditation requirements for health care institutions, national insurance contributions, inspection, educational requirements, available educational volumes, et cetera. The choices made at this level create room for hospital boards to decide on building plans, size of the ICU, private initiatives, profit sharing, regional collaborations, cooperation with the general practitioners, etcetera.

Chapter 3 brings out that hospital boards under pressure to increase their utilization may hence decide to achieve higher targets by refusing complex and emergency care that involves a high variance. From a societal perspective, however, such an eventuality is highly undesirable. To prevent this calculating behavior, hospitals should be judged on their utilization with respect to their own norm utilization. Chapter 3 thus establishes a link between the way in which performances at the societal level are assessed and making strategic choices for a hospital. Chapter 4 then shows that from the perspective of the logistics of a single hospital the allocation of capacity to region-wide capacity for trauma patients would not always be advantageous, particularly with regard to acceptance of planned patients. Governmental stimulation of market incentives in health care carries the risk of less cooperation between hospitals with regard to intake of regional trauma patients, seeing that competitive considerations will become more important. The choices made by the authorities therefore lead regional hospital boards to make decisions pressurizing overall quality and efficiency.

The highest organizational level incorporated in the framework for healthcare planning & control is the strategic hospital level. The actors involved at this level, such as the hospital board, client council, works council, and staff council, will weigh up the interests in the context of the political frames. They deal with strategic problems for which balancing of the stakes involved results in a hospital balance that guarantees continuity of management and the availability, accessibility and funds for the resources for longer periods, i.e. from one year to maximally 20 years. This thesis provides the methods to evaluate the effects of the trade-offs at the strategic level. Should the authorities persist in taking utilization of the scarce hospital resources as a measure of success, then we may well see that hospitals will lead up to this policy by means of selection at the gate. The effect of this strategy is lowering the overall quality of care. The introduction and growth of Independent Treatment Centers, Physicians Owned Hospitals, and so-called treatment lines indeed confirm that this trend is already ongoing. While from the points of view of the individual boards and entrepreneurs this would seem understandable, from a societal perspective it nevertheless results in a reverse effect.

The third level is the tactical organizational level. The actors involved at this level, such as the hospital management together with the partnerships, weigh up the interests in the context of the strategic organizational frames. They are dealing with tactical problems for which balancing of the stakes involved results in departmental balances that guarantee continuity of the management and availability of resources for periods ranging from several days to one year.

Chapter 5 shows that a discrete simulation model is of value in determining the best size and composition of an emergency team, taking into account the patients' safety. Its flexibility provides for varying the input variables, such as safety intervals frequencies, which indicated the sensitivity of the outcome measures to the safety intervals. Moreover, the approach allows evaluating different scenarios as a means to support complex managerial decision-making. Any hospital that reconsiders its staffing during night shifts should carefully consider the balance between the safety intervals of the hospital's patient mix, reduce staffing during the night and the burden for the staff during the night. Strategic considerations may lead a hospital to opt for a more predictable patient mix. In this case the methodology presented in chapter 7, a master surgical scheduling approach, provides the possibility to model the OR planning at a tactical level. Here the management aimed at maximizing the OR utilization in combination with achieving leveled flow of patients to clinical wards. Essential in all this is accurate prediction of length of stay. And then, possible reduction of length of stay gives room for production growth. The second is the operational off-line patient level. The actors involved at this level, such as the individual patients, planners, individual specialists and heads of production-units like the OR department or ICU balance the interests in the context of the tactical organizational frames. Operational questions are at stake here, for which the balancing process results in decisions on time expenditure of the individual specialist and utilization of the available resources for diagnosis and/or treatment of an individual patient.

Provided the specialists and specialties that use the operating room suites are no longer seen as separate, independent entities, but rather as team partners in the planning of patients, there are thus new possibilities to improve utilization. Chapter 9 demonstrates that rearranging planning elements can indeed improve utilization. It is the availability of the specialists that determines the flexibility and consequently the possibility of planning rearrangement. Utilization improvement therefore requires finding a balance between the flexibility in availability of specialists to perform surgery and his or her needed availability in other units in the hospital, such as the outpatient department.

This thesis also makes clear that advanced knowledge of expected duration of admission (chapter 10), and of the operation (chapter 11), including their variabilities, is quite essential for the off-line planning. In addition, chapters 11 and 12 point out the importance of uniform collection of data around these logistics variables, as a basis for this operational planning knowledge. This means that the method of data collection needs to be changed, not only for the ORs and the ICU, but also for other care processes. After all, the present method has resulted from financial and medical considerations, so in fact without casting an eye on requirements dictated by care process control.

The first level is the operational on-line patient level. The actors involved at this level, such as the individual patients, planners, individual specialists and heads of production-units like the OR department or ICU, are balancing the interests in the context of the tactical organizational frames and the operational off-line framers. At stake here are operational on-line questions for which balancing results in decisions on time expenditure of the individual specialist and utilization of the available resources for **unplanned** (additional) diagnosis and/or treatment of an individual (new or earlier) patient.

With the use of the methods developed in this thesis it appears that the possible leeway at this level is an outcome of the balancing at all higher levels.

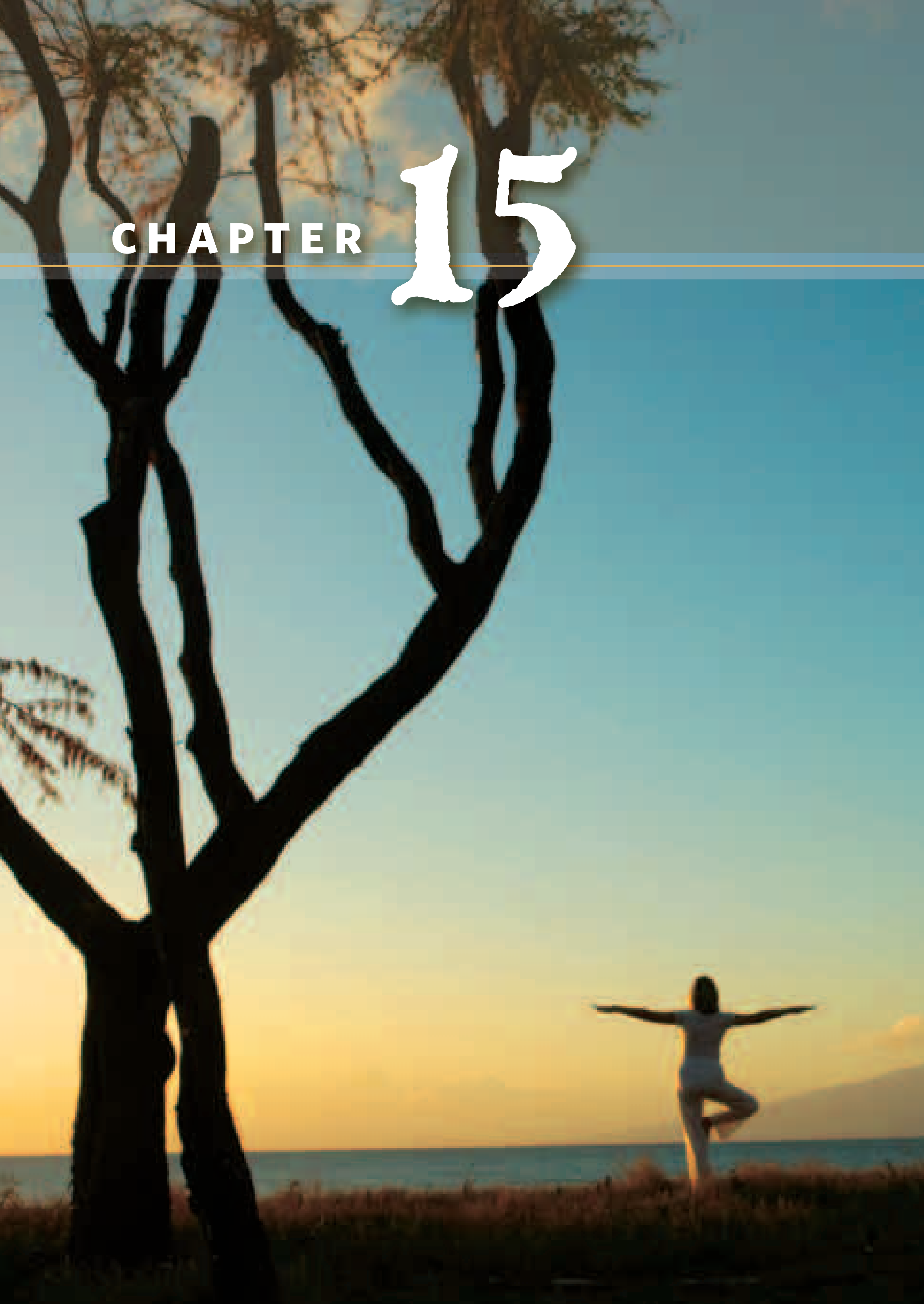
The outcome of the choices made on all levels thus represents in fact the final balance between the patients' interests, the professionals' interests, the quality aimed for, and the costs. It will be clear that outcomes will differ between hospitals. This thesis and the reported studies therefore offer the hospitals an impression of the possible choices and the available techniques. True, while a clear-cut solution for one's own hospital cannot be found in this thesis, it might be formulated with the use of this thesis. It would nevertheless seem to require, first and foremost, the management's dedication to fathom the processes of the medical planning at all levels, so as to establish boundary conditions for the choices in resource capacity planning, material coordination, and financial planning. In addition, the specialists would need to gain insight into the areas of planning and control mentioned. Not until a uniform language and conceptual framework has evolved in this area, it will be possible to apply other knowledge fields, such as operations research, for the facilitation of balancing of interests. The field of operations research has had a place in health care for quite some time, but is relatively unknown in the Netherlands, with application still in its infancy. As early as 1975, the European Working Group on Operational Research Applied to Health Services (ORAHS), www.management.soton.ac.uk/orahs, was founded, a platform providing for European researchers to exchange insights on application of operations research in health care. This application has already a long tradition in the United Kingdom, and other countries have expressed an interest as well: France, Germany, Austria and Canada. Meanwhile, an Operations Research Health Group in the Netherlands is in formation.

Nevertheless, it was not until recent years that actual collaboration and the implementation of operations research in hospitals has expanded. It is particularly the external pressure of all developments in health care that has necessitated a quest for other solutions. Here the field of operations research seems eminently suited to really adapt policies to the threats and developments health care is confronted with.

When doing the research presented here and implementing the outcomes in practice it appeared that patience and cooperation are essential prerequisites for the introduction of these new methods and the proposed framework in hospitals. Cooperation and achieving that experts from different fields will really listen to one another takes time. Notably to familiarize themselves with each other's concepts, and in fact setting up a new collective reference framework. For hospitals to assure successful integration, it would be essential to invest in the time needed.

The absolute basis for this cooperation and the search for balance is knowledge of one's own hospital product in terms of logistics features such as variability, predictability, costs, capacity utilization, duration, and association between capacities and for example successive demand. This requires setting up the data registration and information supply in such a way that it indeed represents the translation of the collectively developed conceptual framework.

From the perspective of this thesis, comparison of organizations would therefore imply more than just comparing the outcomes at an operational level. Benchmarking in this context would then be learning from the way in which comparable organizations provide answers to the different balancing questions, and how, departing from this, policy making took place. This involves learning how the various actors at the various levels in the hospital systematically and methodically arrive at balancing their different interests. Mere comparison of the outcomes of these balancing processes carries the real risk, however, that the underlying methodologies will remain submerged. The reason is the fact that "bad" outcomes may verily lead the participants in a benchmarking process to adopt a defensive stance, on account of which the learning goal will vanish into the background. On the other hand, the learning goal may also vanish when so-called "good" outcomes give rise to a form of euphoria in the organization, or in other words, the impression that "we don't have to learn anything". There are no good and bad outcomes, however. They are the outcomes of a trade-off of the things at stake, and learning will develop from finding the mechanisms behind the outcomes, the methods to balance the interests. The article in question has in it that learning involves comparing and discussing these methods. The novel framework and methodologies thus provide a hold for the new benchmarking.



CHAPTER 15

Summary

Chapter 1

Introduction In chapter 1 a general introduction of this thesis and an introduction to the topics are lined out.

Chapter 2

A Framework for Healthcare Planning and Control As a result of the increasing costs of health care and the introduction of (managed) competitive health care in western countries there is a great need for new and adequate approaches to hospital management. As in traditional manufacturing, OR/MS can fulfill an important role. Many managers and consultants that work in health care have recognized this development. This, however, did not yet result in a structured approach to hospital management. Efforts to adopt hyped concepts from manufacturing frequently resulted in failures and misunderstandings between managers and professionals in health care. The main reasons are that the concepts focus on a part of the areas of interest, and were developed for a system that is entirely different. They do not account for that hospitals can be very different, and have several (generally conflicting) objectives.

In this paper we propose a reference framework for hospital planning and control. It hierarchically structures all planning and control functions of a hospital in all areas of interest. This offers a common language for all stakeholders that are involved in hospital management: clinicians, managers, and experts on planning and control. Any research that focuses on hospital process optimization can use this framework to position problem areas, analyse the control functions that are involved, and analyse the relations between adjacent and related control functions. Also, new techniques from for instance the area of OR/MS or economics can be applied in a structured way. The second contribution of this paper is a typology for hospital types. This typology enables the formulation of different objectives for different types of hospitals, and accordingly the selection of different instruments for planning and control.

The strength of the approach introduced in this paper is that it can not only be applied to hospitals, but also to specific hospital departments, such as the operating room department. With the combination of the reference framework for planning and control and the typology for hospitals we believe that hospital managers and clinicians are better suited for managing the competitive hospitals of the future.

Chapter 3

A Norm Utilization For Scarce Hospital Resources: Evidence from Operating Rooms in a Dutch University Hospital Utilisation of operating rooms is high on the agenda of hospital managers and researchers. Many efforts in the area of maximising the utilisation have been focussed on finding the holy grail of 100% utilisation. The utilisation that can be realised, however, depends on the patient mix and the willingness to accept the risk of working in overtime.

This is a mathematical modelling study that investigates the association between the utilisation and the patient mix that is served and the risk of working in overtime. Prospectively, consecutively, and routinely collected data of an operating room department in a Dutch university hospital are used. Basic statistical principles are used to establish the relation between realistic utilisation rates, patient mixes, and accepted risk of overtime.

Accepting a low risk of overtime combined with a complex patient mix results a low utilisation rate. If the accepted risk of overtime is higher and the patient mix is less complex, the utilisation rate that can be reached is closer to 100%.

Because of the inherent variability of health-care processes, the holy grail of 100% utilisation is unlikely to be found. The method proposed in this paper calculates a realistic benchmark utilisation that incorporates the patient mix characteristics and the willingness to accept risk of overtime.

This paper showed that reserving capacity results in a lower utilisation rate. Hospital boards under pressure to increase their utilisation may hence decide to achieve higher targets by refusing complex and emergency care that involves a high variance. From a societal perspective, such an eventuality is highly undesirable. To prevent this calculating behaviour, hospitals should be judged on their utilisation with respect to their own norm utilisation.

Chapter 4

Regional Synchronization of Intensive Care Capacity Being private institutions, hospitals are themselves responsible for generating revenues. The importance of adequate patient care cuts straight through the hospitals' independent responsibility. Regional care of trauma patients calls for synchronization and transparency between the hospitals so as to achieve they can admit all regional trauma patients as well as their own elective patients without the need to increase capacity. Mathematical models for planning purposes and calculating minimally required capacity at maximal intake will help to secure backing from hospitals participating in regional capacity. We developed a model for the regional intake of patients that adequately calculates proportions of refused patients. Regional synchronization of Intensive Care capacity for trauma patients enables to improve efficiency for all hospitals, without increasing the total number of Intensive Care beds in the region. Also, as fewer beds need to be reserved for elective patients, region-wide capacity is better utilized.

Chapter 5

A Simulation Model for Determining the Optimal Size of an Emergency Team on Call in the Operating Room at Night Hospitals dedicated to perform emergency surgery during the night, from 11:00 P.M. to 7:30 A.M., are facing decisions on optimal operating room (OR) staffing. Emergency patients need to be operated on within predefined safety intervals to avoid risk of life or limbs. This study was designed to find the optimal OR team composition during the night, such that staffing costs are minimized and surgery still will start within safe time limits.

A discrete event simulation in combination with modeling safety intervals was used. Emergency surgery was therefore allowed to be postponed. The model was tested based on data of the main operating room department of Erasmus University Medical Center (Erasmus MC). Two outcome measures were calculated, namely violation of safety intervals and frequency of calling in from home operating and anesthesia nurses. We used available Erasmus MC data on arrival times of emergency patients, durations of surgical procedures, length of stay on the recovery, and transportation times to estimate distributions of all relevant parameters in our model. In addition, safety intervals were determined by surgeons and OR staff of Erasmus MC.

In the scenario in which the number of team members on call was brought down from nine to five, only 2.5 percent point more patients were treated too late as compared to the basic scenario. At the same time the number of operating and anesthesia nurses substantially increased.

The use of safety intervals benefits operating room management during nights. Modelling safety intervals substantially influences the proportions of emergency patients treated on time. Based upon modelling of safety intervals and application of computer simulation in a case study, an operating room department could reduce its team on call with acceptable longer waiting time for a few patients.

A simulation model combined with modeling of patient safety intervals was successfully used to determine the optimal composition and size of emergency teams on the OR from 11:00 P.M. to 7:30 A.M. A case study showed that this approach results in fewer staff required with longer waiting times for only a few cases.

Chapter 6

Closing Emergency Operating Rooms Improves Efficiency Long waiting times for emergency operations increase a patient's risk of postoperative complications and morbidity. Reserving Operating Room (OR) capacity is a common technique to maximize the responsiveness of an OR in case of arrival of an emergency patient. This study determines the best way to reserve OR time for emergency surgery.

In this study two approaches of reserving capacity were compared: (1) concentrating all reserved OR capacity in dedicated emergency ORs, and (2) evenly reserving capacity in all elective ORs. By using a discrete event simulation model the real situation was modelled. Main outcome measures were: (1) waiting time, (2) staff overtime, and (3) OR utilisation were evaluated for the two approaches.

Results indicated that the policy of reserving capacity for emergency surgery in all elective ORs led to an improvement in waiting times for emergency surgery from 74 (± 4.4) minutes to 8 (± 0.5) minutes. Working in overtime was reduced by 20% , and overall OR utilisation can increase by around 3%.

Emergency patients are operated upon more efficiently on elective Operating Rooms instead of a dedicated Emergency OR. The results of this study led to closing of the Emergency OR in the Erasmus MC (Rotterdam, The Netherlands).

Chapter 7

Fewer ICU Refusals and a Higher Capacity Utilization by Cyclic Case Scheduling Mounting health care costs force hospital managers to maximize utilization of scarce resources and simultaneously improve access to hospital services. This paper assesses the benefits of a cyclic case scheduling approach that exploits a Master Surgical Schedule (MSS). An MSS maximizes operating room (OR) capacity and simultaneously levels the outflow of patients towards the Intensive Care Unit (ICU) to reduce surgery cancellation.

Relevant data for Erasmus MC have been electronically collected since 1994. These data are used to construct an MSS that consisted of a set of surgical case types scheduled for a period or cycle. This cycle was executed repetitively. During such a cycle, surgical cases for each surgical department were scheduled on a specific day and OR. The experiments were performed for the Erasmus university medical centre and for a virtual hospital.

Unused OR capacity can be reduced by up-to 6.3% for a cycle length of four weeks, with simultaneous optimal levelling of the ICU workload.

Our findings show that the proposed cyclic OR planning policy may benefit OR utilization and reduce surgical case cancellation and peak demands on the ICU.

Chapter 8

Influence of Cardiac Risk Factors and Medication on Length of Hospitalization in Patients Undergoing Major Vascular Surgery Major vascular surgery is associated with long length of in-hospital stay (LOS). Cardiac risk factors identify patients with an increased risk. Recent studies found that statin, aspirin and beta-blocker therapy were associated with improved postoperative outcome. However, the effect of all these factors on LOS is not yet defined. Our aim is to determine the effect of cardiac risk factors and (preventive) statin, aspirin and beta-blocker therapy on LOS, and to deduce from these factors a model that predicts LOS. A total of 2.374 patients in the period 1990-2004 was enrolled. Mean LOS was 18±9 days. Cardiac risk factors significantly associated with LOS in the multivariable analysis were age, prior heart failure, hypertension, diabetes mellitus, renal failure, and COPD. Statin and aspirin use were associated with a reduced LOS. Beta-blockers reduced LOS only in patients with underlying coronary artery disease. Together, these factors explain 14,1% of variance in LOS. In conclusion, hospital stay in patients undergoing major vascular surgery can be predicted more accurately by clinical cardiac risk factors. A significant reduction of hospital stay was achieved with statin, aspirin and beta-blocker therapy.

Chapter 9

Improved Efficiency by Applying Bin Packing and Portfolio Techniques to Surgical Case Scheduling An operating room department has adopted an efficient business model and subsequently investigated how efficiency could be further improved. The aim of this study is to show the efficiency improvement of lowering organizational barriers and applying advanced mathematical techniques.

We applied advanced mathematical algorithms in combination with scenarios that model relaxation of various organizational barriers using prospectively collected data. The setting is the main inpatient OR department of a university hospital, which sets its surgical case schedules two weeks in advance using a block planning method. Main outcome measures are the number of freed OR blocks and OR utilization.

Lowering organizational barriers and applying mathematical algorithms can yield a 4.5 percent point increase in OR utilization (95% confidence interval 4.0% - 5.0%). This is obtained by reducing the total required OR time.

Efficient OR departments can further improve their efficiency. The paper shows that a radical cultural change that comprises the use of mathematical algorithms and lowering organizational barriers improves OR utilization.

Operating room departments that cannot improve efficiency by current known scheduling practices can benefit from lowering organizational barriers and adopting mathematical algorithms to their surgical case scheduling. This study demonstrates that operating room utilization can be increased by 4.5 percent point.

Chapter 10

Optimizing Intensive Care Capacity Using Individual Length-of-stay Prediction Models Effective planning of elective surgical procedures requiring postoperative intensive care is important in preventing cancellations and empty intensive care unit (ICU) beds. To improve planning, we constructed, validated and tested three models designed to predict length of stay (LOS) in the ICU in individual patients.

Retrospective data were collected from 518 consecutive patients who underwent oesophagectomy with reconstruction for carcinoma between January 1997 and April 2005. Three multivariable linear regression models for LOS, namely preoperative, postoperative and intra-ICU, were constructed using these data. Internal validation was assessed using bootstrap sampling in order to obtain validated estimates of the explained variance (r^2). To determine the potential gain of the best performing model in day-to-day clinical practice, prospective data from a second cohort of 65 consecutive patients undergoing oesophagectomy between May 2005 and April 2006 were used in the model, and the predictive performance of the model was compared with prediction based on mean LOS.

The intra-ICU model had an r^2 of 45% after internal validation. Important prognostic variables for LOS included greater patient age, co morbidity, type of surgical approach, intraoperative respiratory minute volume and complications occurring within 72 hours in the ICU. The potential gain of the best model in day-to-day clinical practice was determined relative to mean LOS. Use of the model reduced the deficit number (underestimation) of ICU days by 65 and increased the excess number (overestimation) of ICU days by 23 for the cohort of 65 patients. A conservative analysis conducted in the second, prospective cohort of patients revealed that 7% more oesophagectomies could have been accommodated, and 15% of cancelled procedures could have been prevented.

Patient characteristics can be used to create models that will help in predicting LOS in the ICU. This will result in more efficient use of ICU beds and fewer cancellations.

Chapter 11

Predicting the unpredictable: an improved prediction model for better planning of Operating Room capacity using session, team and patient characteristics, together with the surgeons' estimate Variability of operation times leads to sub-optimal use of operating room (OR) capacity, causing increased healthcare costs and cancellation of planned procedures. Reliable prediction of operative time is therefore mandatory, but routine predictions of procedure time made by surgeons or historical mean durations have only limited predictive capacity for individual patients. We aimed to devise a prediction model taking into account the surgeons' estimate and characteristics of the surgical team, the operation and the patient.

16,359 consecutive elective operations from the general surgical department in an academic hospital were analysed. A coding of operations in 135 categories was used. The outcome measure was total session time, and potential predictive factors were: operation code, surgeon's estimate of operative time, number of procedures to be performed during the operation, total number of surgeons and anesthesiologists attending, experience of the surgeons and of the anesthesiologists, age and sex of the patient, number of previous hospital admissions, body mass index (BMI) and eight cardiovascular risk factors. Multiple regression on the logarithm of the total session time was performed. The predictive contribution of the factors was expressed as the gain in R-squared relative to a base model that only contained the operation code.

The operation code explained 72.9% of the variation in (log) session times. Characteristics of the session and of the team gave the largest gain in predictive performance, whereas patient characteristics were of lesser predictive value. The surgeons' estimate had an independent and substantial contribution to the prediction and the final model R-square was 79.6%, a gain of 25% relative to the base model. For a given operation of a given patient, the 95% prediction interval for the session duration ranges from 0.59 to 1.71, relative to the median duration for that operation and patients of that particular profile.

Detailed information on the operative session and on the team gives a substantial improvement of the prediction of operative session times, but the surgeons' estimate remains important. The prediction model is fit for practical use in day-to-day planning of OR capacity.

Chapter 12

Uniform Time Registration in Surgical Suites. Definition System Provides for Comparison of Work Processes Learning about the effectiveness of processes in OR suites by comparing these is worthwhile, but can only be effective if participants use the same definitions for relevant measuring moments and performance indicators.

The university medical centres together with the University of Twente have developed a comprehensive and relatively simple definition- and time registration method. It is up to each hospital itself to determine which intervals from this model to record.

Chapter 13

A new Progressive Scheme for Benchmarking Benchmarking compares performances of organizations in view of achieving lasting improvement. Still too often, however, comparison is 'between apples and oranges', with less favourably performing organizations unnecessarily falling prey to naming and shaming. In this article we present a new model to overcome these flaws. Developed for the healthcare sector, it is nevertheless applicable to other sectors as well. As the performance indicators chosen and the assessment procedure take into account differences in the organizations to be assessed, this model for benchmarking offers concrete starting points for lasting improvement of one's own business.

Chapter 14

General Discussion and implications

Chapter 15

Summary

Chapter 16

Dutch Summary

Appendix 1

Managing the Overflow of Intensive Care Patients Many hospitals in the Netherlands are confronted with capacity problems at their intensive care units (ICUs) resulting in cancelling operations, overloading the staff with extra patients, or rejecting emergency patients. In practice, the last option is a common choice because for legal reasons, as well as for hospital logistics, rejecting emergency patients has minimal consequences for the hospital. As a result, emergency patients occasionally have to be transported to hospitals far away. In this work, we propose a cooperative solution for the ICU capacity problem. In our model, several hospitals in a region jointly reserve a small number of beds for regional emergency patients. We present a mathematical method for computing the number of regional beds for any given acceptance rate. The analytic approach is inspired by overflow models in telecommunication systems with multiple streams of telephone calls. Simulation studies show that our model is quite accurate. We conclude that cooperation between hospitals helps to achieve a high acceptance level with a smaller number of beds resulting in improved service for all patients.

Appendix 2

Robust Surgery Loading We consider the robust surgery loading problem for a hospital's operating theatre department, which concerns assigning surgeries and sufficient planned slack to operating room days. The objective is to maximize capacity utilization and minimize the risk of overtime, and thus cancelled patients. This research was performed in collaboration with the Erasmus MC, a large academic hospital in the Netherlands, which has also provided historical data for the experiments. We propose various constructive heuristics and local search methods that use statistical information on surgery durations to exploit the portfolio effect, and thereby to minimize the required slack. We demonstrate that our approach frees a lot of operating room capacity, which may be used to perform additional surgeries. Furthermore, we show that by combining advanced optimization techniques with extensive historical statistical records on surgery durations can significantly improve the operating room department utilization

Appendix 3

A Master Surgical Scheduling Approach for Cyclic Scheduling in Operating Room Departments This paper addresses the problem of operating room(OR) scheduling at the tactical level of hospital planning and control. Hospitals repetitively construct operating room schedules, which is a time-consuming, tedious, and complex task. The stochasticity of the durations of surgical procedures complicates the construction of operating room schedules. In addition, unbalanced scheduling of the operating room department often causes demand fluctuation in other departments such as surgical wards and intensive care units. We propose cyclic operating room schedules, so-called master surgical schedules (MSSs) to deal with this problem. In an MSS, frequently performed elective surgical procedure types are planned in a cyclic manner. To deal with the uncertain duration of procedures we use planned slack. The problem of constructing MSSs is modelled as a mathematical program containing probabilistic constraints. Since the resulting mathematical program is computationally intractable we propose a column generation approach that maximizes the operation room utilization and levels the requirements for subsequent hospital beds such as wards and intensive care units in two subsequent phases. We tested the solution approach with data from the Erasmus Medical Centre. Computational experiments show that the proposed solution approach works well for both the OR utilization and the levelling of requirements of subsequent hospital beds.

CHAPTER 16

Dutch Summary

Hoofdstuk 1

Introductie Hoofdstuk 1 geeft een algemene inleiding op dit proefschrift, en vervolgens worden de te behandelen onderwerpen in het kort uiteengezet.

Hoofdstuk 2

Een denkraam voor planning en beheer in de gezondheidszorg Tengevolge van de toenemende kosten van de gezondheidszorg en de invoering van marktwerking op dit gebied in westerse landen is er een grote behoefte aan nieuwe en effectieve benaderingen van het ziekenhuismanagement. Net als in traditionele bedrijfstakken kunnen *operations research* en *management sciences* (OR/MS) hier een belangrijke rol in spelen. Al vele managers en consultants in de gezondheidszorg hebben deze ontwikkeling onderkend. Maar dit heeft echter nog niet geresulteerd in een gestructureerde benadering van het ziekenhuismanagement. Pogingen om opzienbarende concepten uit de industrie over te nemen hebben veelal geleid tot mislukkingen en tot misverstanden tussen managers en professionals in de gezondheidszorg. Dit komt hoofdzakelijk omdat concepten zich richten op een deel van de aandachtsgebieden, en toegesneden zijn op een totaal ander systeem. Er wordt geen rekening mee gehouden dat ziekenhuizen heel verschillend kunnen zijn en verscheidene doelstellingen hebben (die meestal in strijd zijn met elkaar).

In dit artikel introduceren we een referentie-denkraam voor ziekenhuisplanning en –beheer, met een hiërarchieke structuur en gericht op alle aandachtsgebieden. Het biedt een universele taal voor alle betrokkenen bij het ziekenhuismanagement: clinici, managers, planners en beheerders. Bij alle optimaliseringstudies kan dit denkraam dienst doen om probleemgebieden te positioneren en de betreffende beheersfuncties te analyseren, alsmede de relaties tussen aangrenzende en verwante beheersfuncties. Ook voorziet het denkraam in een gestructureerde toepassing van nieuwe technieken bijvoorbeeld op het gebied van OR/MS of uit de economie. De tweede bijdrage van dit artikel is een typologie voor de verschillende ziekenhuizen. Met deze typologie kunnen verschillende doelstellingen worden geformuleerd voor de verschillende types ziekenhuizen, en parallel daaraan verschillende instrumenten voor planning en beheer worden geselecteerd.

De kracht van de benadering die in dit artikel wordt geïntroduceerd is dat deze niet alleen geldt voor ziekenhuizen als geheel, maar ook voor specifieke afdelingen, zoals het operatiekamercomplex. We denken dat deze combinatie, d.w.z. het referentie-denkraam voor planning en beheer samen met de typologie voor de ziekenhuizen, managers en clinici beter in staat stelt de elkaar concurrerende ziekenhuizen van de toekomst te beheren.

Hoofdstuk 3

Een normering voor aanwending van schaarse ziekenhuismiddelen: bewijs voor het operatiekamercomplex een Nederlands academisch ziekenhuis De bezettingsgraad van operatiekamers staat hoog op de agenda van ziekenhuismanagers en -onderzoekers. Op het gebied van maximale benutting zijn al veel pogingen gedaan de heilige graal van 100% benutting te vinden. De bezettingsgraad echter die realistisch kan worden bereikt hangt af van de patiëntenmix en de bereidheid om kans op overwerk te accepteren.

Dit project betreft wiskundige modellering van het verband tussen de bezettingsgraad en de kans op overwerk. Er wordt gebruikgemaakt van prospectieve, successievelijke, en routinematig verzamelde gegevens van een operatiekamercomplex in een Nederlands academisch ziekenhuis. Elementaire statistische beginselen werden gebruikt om de relatie te leggen tussen realistische bezettingsgraad, patiëntenmix, en acceptabele kans op overwerk.

Het combineren van een lage acceptabele kans op overwerk met een minder voorspelbare patiëntenmix levert een lage benutting op. Bij een hogere kans en meer voorspelbare patiëntenmix komt de haalbare benutting meer in de richting van de 100%.

De variabiliteit die processen in de gezondheidszorg eigen is maakt het onwaarschijnlijk dat de heilige graal van 100% benutting ooit wordt gevonden. De methode in dit artikel berekent een realistische íjkpunt voor de benutting gebaseerd op de kenmerken van de patiëntenmix en de bereidheid om kans op overwerk te accepteren.

Het artikel laat zien dat het reserveren van capaciteit uitmondt in een lagere benutting. Ziekenhuisbestuurders die onder druk staan om hun benuttinggraad te verhogen zouden daarom hoger kunnen mikken door af te zien van complexe en spoedeisende hulp die een hoge variatie met zich meebrengt. Maatschappelijk gezien is deze mogelijkheid echter hoogst ongewenst. Om zulk berekenend gedrag te voorkomen zouden ziekenhuizen moeten worden beoordeeld op de benuttinggraad gerelateerd aan de eigen normbenutting.

Hoofdstuk 4

Regionale afstemming van Intensive Care capaciteit Ziekenhuizen zijn private instellingen – en verantwoordelijk voor het eigen resultaat. Het belang van adequate patiëntenzorg loopt dwars door de zelfstandige verantwoordelijkheid van de ziekenhuizen. Het belang van de traumapatiënten in de regio vraagt om transparante afstemming tussen de ziekenhuizen, om zodoende met dezelfde capaciteit zowel alle regionale traumapatiënten als de eigen geplande patiënten te kunnen opvangen. Wiskundige planningsmodellen die de minimaal benodigde capaciteit bij maximale opvang kunnen berekenen, bieden onderbouwing voor het realiseren van draagvlak bij de ziekenhuizen die participeren in de regionale capaciteitsvoorziening. Voor de regionale opvang van patiënten is een model ontwikkeld dat adequaat fracties geweigerde patiënten kan berekenen. Het blijkt dat een systeem voor regionale Intensive Care capaciteit voor traumapatiënten gekoppeld aan goede samenwerking voor alle ziekenhuizen de efficiëntie kan verhogen, zonder het totale aantal Intensive Care bedden in de regio te vergroten. De capaciteit in de regio wordt beter benut, mede omdat het aantal voor geplande patiënten te reserveren bedden kan worden verminderd.

Hoofdstuk 5

Een simulatiemodel voor het bepalen van de optimale omvang van een 's nachts beschikbaar OK-team voor spoedoperaties Ziekenhuizen die 's nachts spoedoperaties uitvoeren, d.w.z. van elf uur 's avonds tot half acht in de ochtend, zien zich voor de vraag gesteld hoe de OK optimaal bemand kan worden. Spoedpatiënten dienen te worden geopereerd binnen een bepaalde tijd om complicaties en morbiditeit te voorkomen.

Deze studie had als doel de optimale samenstelling te bepalen van een 's nachts beschikbaar OK-team, met als uitgangspunten minimale personeelskosten en start van de operatie binnen de veilige tijdsgrens.

Een *discrete event* simulatie in combinatie met intervalmodellering vond plaats. Dit houdt in dat de spoedoperatie eventueel mocht worden uitgesteld. Het model werd getest met gegevens van het centrale operatiekamercomplex van het Erasmus MC. Twee uitkomstmaten werden berekend, namelijk het niet voldoen aan de veilige tijdsgrens en de frequentie waarmee operatie- en anesthesieverpleegkundigen van huis moesten komen. We gebruikten de beschikbare gegevens van het Erasmus MC voor de aankomsttijden van spoedpatiënten, de duur van de operaties, de duur van het verblijf in de uitslaapkamer, en transporttijden om de distributie van alle relevante parameters in ons model te schatten. Daarnaast gaven de chirurgen en OK-assistenten in het Erasmus MC aan welke veilige tijdsgrenzen moesten worden gehanteerd.

In een bepaald scenario waarin het aantal oproepbare teamleden was teruggebracht van negen tot vijf, werden slechts 2,5 procentpunt meer patiënten te laat behandeld in vergelijking met het basisscenario.

Het gebruik van veilige tijdsgrenzen werkt gunstig op het beheersaspecten van het operatiekamercomplex gedurende de nacht. De tijdsintervalmodellering heeft een aanzienlijk invloed op de percentages patiënten die op tijd worden geopereerd. Op basis hiervan en de toepassing van computersimulatie in een gevalsstudie zou een operatiekamercomplex de omvang van het 's nachts oproepbare team kunnen verminderen, met een acceptabele langere wachttijd voor enkele patiënten.

Een simulatiemodel gecombineerd met intervalmodellering van tijdveiligheidsmarges werd met succes toegepast om de optimale samenstelling en omvang van 's nachts beschikbare OK-teams voor spoedoperaties te berekenen. Uit een gevalsstudie bleek dat men met deze methode met minder mensen toe kan, en dat slechts in enkele gevallen sprake is van langere wachttijden.

Hoofdstuk 6

Het sluiten van operatiekamers voor spoedgevallen verbetert de doelmatigheid Lange wachttijden voor spoedoperaties vergroten de kans op postoperatieve complicaties en morbiditeit. Het reserveren van OK-capaciteit is een gebruikelijke techniek om de slagvaardigheid van het operatiekamercomplex in geval van de komst van een spoedpatiënt te maximaliseren. Dit onderzoek richt zich op de beste manier om OK-tijd voor spoedoperaties te reserveren.

Er werden twee verschillende benaderingen met elkaar vergeleken: (1) alle gereserveerde OK-capaciteit concentreren in speciale OK's voor spoedoperaties, en (2) evenredig capaciteit reserveren in alle electieve OK's. Met behulp van een discrete event simulatiemodel werd de werkelijke situatie gemodelleerd. Als belangrijkste uitkomstmaten voor beide methoden golden: (1) wachttijd, (2) overwerk van het personeel, en (3) OK-benutting.

Uit de analyse kwam naar voren dat de benadering waarbij capaciteit voor spoedoperaties werd gereserveerd in alle electieve OK's leidde tot een verbetering in de wachttijd: die liep terug van 74 (± 4,4) minuten naar 8 (± 0,5) minuten. Overwerk kon worden verminderd met 20%, en de totale OK-benutting kon worden verhoogd met zo'n 3%.

Spoedpatiënten kunnen in electieve OK's ook effectiever worden geopereerd dan in speciale spoed-OK's. Naar aanleiding van de resultaten van dit onderzoek werd de speciale operatiekamer voor spoedgevallen in het Erasmus MC gesloten.

Hoofdstuk 7

Minder patiënten geweigerd op de Intensive Care afdeling en hogere capaciteitsbenutting door toepassing van cyclische operatieplanning De steeds stijgende kosten in de gezondheidszorg dwingen ziekenhuismanagers het gebruik van schaarse middelen te maximaliseren en tegelijkertijd de beschikbaarheid van ziekenhuisfaciliteiten te verbeteren. In dit artikel wordt gekeken naar de voordelen van een methode voor cyclische planning van operaties met gebruikmaking van een zogenaamd *master surgical schedule*. Hierbij wordt de OK-capaciteit vergroot en wordt tegelijkertijd de patiëntenstroom naar de Intensive Care afdeling afgevlakt om zodoende het aantal vervallen operaties te verminderen.

Sinds 1994 worden gegevens die nodig zijn bij deze planning in het Erasmus MC elektronisch vastgelegd. Deze gegevens werden gebruikt om een *master surgical schedule* te construeren dat bestond uit een set van verschillende types van operaties voor een bepaalde periode of cyclus. Deze cyclus werd herhaaldelijk uitgevoerd. In zo'n cyclus werden de operaties van de afzonderlijke afdelingen op bepaalde dagen en in bepaalde OK's gepland. De experimenten werden gedaan voor het Erasmus MC en voor een voor een virtueel ziekenhuis.

Het blijkt dat de niet-benutte OK-capaciteit met maximaal 6,3% kan worden verminderd voor een cyclus van vier weken, waarbij tegelijkertijd de werkdruk op de Intensive Care afdeling optimaal wordt afgevlakt.

Uit onze bevindingen blijkt dat introductie van deze cyclische planningsmethode voor het operatiekamercomplex tot betere benutting kan leiden waarbij er minder geplande operaties hoeven te vervallen en piekbelasting van de Intensive Care afdeling wordt verkleind.

Hoofdstuk 8

De invloed van cardiale risicofactoren en medicatie op de duur van de ziekenhuisopname voor patiënten die ingrijpende vaatoperaties ondergaan Patiënten die ingrijpende vaatoperaties ondergaan worden lang opgenomen in het ziekenhuis. Aan de hand van cardiale risicofactoren kunnen degenen met een hoger risico worden geïdentificeerd. Recentelijk is ontdekt dat het gebruik van statines, aspirine en bètablokkers met een betere postoperatieve uitkomst is geassocieerd. Het effect van al deze factoren op de duur van de ziekenhuisopname is echter nog niet bekend. We stelden ons als doel na te gaan in hoeverre cardiale risicofactoren en het (preventieve) gebruik van statines, aspirine en bètablokkers de opnameduur beïnvloeden, en vanuit deze factoren een model af te leiden waarmee de opnameduur kan worden voorspeld. In de periode 1990-2004 werden totaal 2.374 patiënten geïncludeerd. De mediane opnameduur was 18±9 dagen. Bij multivariable analyse bleken de volgende cardiale risicofactoren significant geassocieerd te zijn met de opnameduur: leeftijd, eerder optreden van hartfalen, verhoogde bloeddruk, diabetes mellitus, nierfalen, en COPD. Het gebruik van statines en aspirine was geassocieerd met kortere opname. Voor bètablokkers was dat laatste alleen maar het geval als er sprake was een onderliggende aandoening aan de kransslagader. Gezamenlijk verklaren deze factoren 14,1% van de variantie in opnameduur. We concluderen dat de opnameduur voor patiënten die ingrijpende vaatoperaties ondergaan accurater kan worden voorspeld aan de hand van klinische cardiale risicofactoren. Het gebruik van statines, aspirine en bètablokkers ging gepaard met een significante verkorting van de opnameduur.

Hoofdstuk 9

Betere doelmatigheid door het gebruik van *bin packing* en *portfolio technieken* bij de operatieplanning Voor een operatiekamercomplex dat al werkte met een efficiënt bedrijfsmatig model werd onderzocht hoe de doelmatigheid nog verder verbeterd zou kunnen worden. Het doel van dit onderzoek was de doelmatigheid aan te tonen van het verlagen van barrières in de organisatie en het toepassen van geavanceerde wiskundige technieken.

Verfijnde mathematische algoritmen werden toegepast in combinatie met scenario's waarin minder strikte barrières in de organisatie worden gemodelleerd aan de hand van prospectief verkregen gegevens. Dit werd gedaan voor het centrale operatiekamercomplex van een universiteitsziekenhuis dat met behulp van een *block planning* methode de operatieplanning twee weken vooruit maakt. Als belangrijkste uitkomstmaten golden het aantal vrijgekomen OK-blokken en de OK-benutting.

Het verlagen van de barrières in de organisatie gecombineerd met de toepassing van wiskundige algoritmen kan een 4.5 procentpunt (95% betrouwbaarheidsinterval 4,0% - 5,0%) betere OK-benutting opleveren. Dit wordt bereikt door verkorting van de totaal benodigde OK-tijd. Zodoende kan een operatiekamercomplex dat al efficiënt werkt de doelmatigheid nog verder opvoeren. Dit artikel toont aan dat een radicale cultuurverandering in de vorm van het gebruik van wiskundige algoritmen en het verlagen van barrières in de organisatie tot betere OK-benutting leidt.

Operatiekamercomplexen waarvoor de tot dusverre toegepaste planningsmethoden de doelmatigheid niet kunnen vergroten zullen zeker baat hebben bij het verlagen van barrières in de organisatie en het toepassen van wiskundige algoritmen op de operatieplanning. Dit onderzoek laat zien dat hiermee de OK-benutting met 4.5 procentpunt kan worden verbeterd.

Hoofdstuk 10

Het optimaliseren van Intensive Care capaciteit met behulp van voorspellingsmodellen voor individuele opnameduur Electieve operaties waarvoor de patiënt aansluitend intensieve zorg behoeft dienen effectief gepland te worden om het niet doorgaan en niet-bezette intensive care unit (ICU) bedden te voorkomen. Met het oog op betere planning hebben we drie modellen ontworpen, gevalideerd en getest voor het voorspellen van de opnameduur in de ICU van individuele patiënten.

Retrospectieve gegevens werden verzameld voor 518 achtereenvolgende patiënten die tussen januari 1997 en april 2005 oesofagectomie ondergingen met reconstructie vanwege een carcinoom. Aan de hand van deze gegevens werden drie multivariable lineaire regressiemodellen voor de opnameduur geconstrueerd, namelijk preoperatief, postoperatief en intra-ICU. Interne validering werd gedaan met bootstrap sampling om gevalideerde ramingen te verkrijgen van de verklaarde variantie (r^2). We wilden ook bepalen wat de mogelijke winst zou zijn voor het best presterende model in de dagelijkse praktijk. Hiervoor hebben we dit model gevoed met prospectieve data van een zelfde cohort tussen mei 2005 en April 2006, en de voorspellende waarde van het model vergeleken met voorspellingen gebaseerd op de gemiddelde opnameduur.

Het intra-ICU model had een r^2 van 45% na interne validering. Belangrijke voorspellende variabelen voor de opnameduur zijn onder andere: hogere leeftijd van de patiënt, co-morbiditeit, type operatiemethode, intra-operatief ademvolume, en het optreden van complicaties binnen 72 uur na opname in de ICU. De mogelijke winst van dit model in de dagelijkse praktijk werd bepaald in vergelijking met de gemiddelde opnameduur. Bij toepassing van dit model werd voor de cohort van 65 patiënten het deficiet aantal (onderschatting) opnamedagen verminderd met 65 en het exces aantal (overschatting) opnamedagen met 23 verhoogd. Uit een conservatieve analyse van deze tweede, prospectieve cohort bleek dat 7% meer operaties hadden kunnen worden uitgevoerd, en dat 15% van de afgelaste operaties had kunnen worden voorkomen.

Patiëntgegevens kunnen inderdaad worden gebruikt bij de bouw van modellen die als steun kunnen dienen bij het voorspellen van de opnameduur in de ICU. Dit heeft efficiëntere inzet van ICU-bedden en minder afgelastingen tot gevolg.

Hoofdstuk 11

Het onvoorspelbare voorspellen: een verbeterd voorspellingsmodel voor betere capaciteitsplanning van een operatiekamercomplex met gebruikmaking van de kenmerken van de sessie, het team, de patiënt, en een inschatting van de chirurg De variatie in duur van een operatieve ingreep leidt er toe dat de capaciteit van het operatiekamercomplex niet optimaal wordt gebruikt. Het gevolg is dat de kosten van de gezondheidszorg stijgen en er meer geplande sessies komen te vervallen. Een betrouwbare voorspelling van de operatieduur is daarom hoogst gewenst. De inschatting die de chirurg uit ervaring maakt of de gegevens uit het verleden betreffende gemiddelde duur hebben slechts beperkte voorspellende waarde voor de individuele patiënt. We wilden daarom tot een voorspellingsmodel komen waarin rekening wordt gehouden niet alleen met de inschatting van de chirurg, maar ook met de typische kenmerken van het operatieteam, de operatie, en de betreffende patiënt.

De gegevens rond 16.359 achtereenvolgende selectieve operaties in een algemene afdeling chirurgie in een universiteitsziekenhuis werden geanalyseerd. De ingrepen zelf werden gecodeerd volgens 135 categorieën. Als uitkomstmaat gold de totale duur van de sessie, en de mogelijk voorspellende factoren waren: operatiecategorie, inschatting van de chirurg over de operatieduur, aantal ingrepen uit te voeren tijdens de operatie, totale aantal aanwezige chirurgen en anesthesiologen, ervaring van deze chirurgen en anesthesiologen, leeftijd en geslacht van de patiënt, aantal eerdere ziekenhuisopnamen, *body mass index* (BMI) en acht cardiovasculaire risicofactoren. Multipele regressie voor de logaritme van de totale sessieduur werd uitgevoerd. De voorspellende bijdrage van elk van de factoren werd uitgedrukt als de toename in R^2 ten opzichte van een basismodel met alleen maar de operatiecategorie

De operatiecategorie verklaarde 72,9% van de variatie in (log) sessieduur. De kenmerken van de sessie en die van het team lieten de grootste toename in voorspellend vermogen zien, terwijl de kenmerken van de patiënt minder voorspellende waarde hadden. De inschatting van de chirurg vertegenwoordigde een onafhankelijke en aanzienlijke bijdrage tot de voorspelling, met een R^2 van 79,6% voor het uiteindelijke model, een toename van 25% ten opzichte van het basismodel. Voor een gegeven operatie van een gegeven patiënt, loopt het 95% voorspellingsinterval voor de sessieduur uiteen van 0,59 tot 1,71; in relatie tot de mediane duur van die operatie en patiënten die in dit specifieke profiel passen.

We concluderen dat gedetailleerde informatie over de operatiesessie en het operatieteam de voorspelling van de sessieduur aanzienlijk verbetert, maar dat de inschatting van de chirurg ook belangrijk is. Dit voorspellingsmodel is in de praktijk goed geschikt voor de dagelijkse capaciteitsplanning binnen een operatiekamercomplex.

Hoofdstuk 12

Eenduidige tijdregistratie operatiekamers. Definitiesysteem maakt vergelijking werkprocessen mogelijk Door OK-processen onderling te vergelijken, kan men van elkaar leren wat doelmatigheid betreft. Maar dan moeten er wel eenduidige definities van relevante meetmomenten en prestaties worden gehanteerd.

De universitair medische centra hebben samen met de Universiteit Twente een compleet en relatief eenvoudig definitie- en tijdregistratiesysteem ontwikkeld. Ieder ziekenhuis kan zelf bepalen welke intervallen uit dit model het registreert.

Hoofdstuk 13

Een nieuw stappenplan voor benchmarking Benchmarking vergelijkt prestaties van organisaties met het oog op duurzame verbetering. Te vaak nog worden daarbij 'appels met peren' vergeleken en minder presterende organisaties onnodig in de hoek gezet ('*naming en shaming*'). In dit artikel presenteren we een nieuw model om deze euvels te overwinnen. Dit model is ontwikkeld voor de zorgsector, maar is ook daarbuiten toepasbaar. Door bij de keuze van prestatie-indicatoren en bij de prestatiebeoordeling recht te doen aan verschillen in de te meten organisaties, biedt dit model voor benchmarking concrete aanknopingspunten voor duurzaam verbeteren van de eigen onderneming.

Hoofdstuk 14

Algemene discussie en implicaties

Hoofdstuk 15

Engelse samenvatting

Hoofdstuk 16

Nederlandse samenvatting

Bijlage 1

Het in goede banen leiden van de toevloed van Intensive Care patiënten Vele ziekenhuizen in Nederland hebben te maken met capaciteitsproblemen in de intensive care afdelingen, zodat operaties moeten worden afgelast, het personeel te maken krijgt met extra patiënten, of dat spoedpatiënten worden geweigerd. In de praktijd wordt deze laatste optie vaak gebruikt, omdat dit om wettelijke redenen, maar ook uit logistiek oogpunt, weinig consequenties heeft voor het ziekenhuis. Daarom moeten spoedpatiënten soms naar ver gelegen andere ziekenhuizen worden vervoerd. We hebben de mogelijkheid van een collectieve aanpak van dit capaciteitsprobleem bestudeerd, met een model waarin verschillende ziekenhuizen in een regio gezamenlijk een klein aantal bedden reserveren voor spoedpatiënten in de regio. We presenteren een wiskundige methode voor het berekenen van het aantal regionale bedden bij elke gegeven acceptatiegraad. Deze analytische benadering is geënt op *overflow*-modellen voor telecommunicatiesystemen met meerdere stromen telefoongesprekken. Uit simulatieonderzoeken blijkt dat ons model zeer accuraat is. We concluderen dan ook dat samenwerking tussen ziekenhuizen bijdraagt aan het behalen van een hoge acceptatiegraad bij een kleiner aantal bedden, en betere service biedt aan alle patiënten.

Bijlage 2

Robust surgery loading We hebben gekeken naar het probleem van *robust surgery loading* bij een operatiekamercomplex, dat wil zeggen het toewijzen van sessies aan bepaalde dagen en het plannen van voldoende uitloopcapaciteit. Het doel hiervan is de capaciteitsbenutting te maximaliseren en de kans op overwerk te minimaliseren, en dus minder patiënten te hoeven afzeggen. Dit hebben we gedaan aan de hand van historische gegevens van het Erasmus MC. We introduceren verschillende constructieve heuristische en lokale zoekmethoden met statistische informatie over de duur van operaties die met gebruikmaking van het *portfolio effect* de benodigde uitloopcapaciteit minimaliseren. Het blijkt dat door deze benadering veel OK-capaciteit vrijkomt, die dan gebruikt kan worden voor extra operaties. Ook laten we zien dat het combineren van geavanceerde optimalisatietechnieken met lange reeksen historische gegevens over de duur van operaties de capaciteitsbenutting van een operatiekamercomplex aanzienlijk kan verbeteren

Bijlage 3

Cyclische planning voor een operatiekamercomplex met gebruikmaking van *master surgical scheduling* Dit artikel gaat in op het probleem van de operatieplanning op het tactische niveau van planning en beheer in een ziekenhuis. Ziekenhuizen zien zich genoodzaakt steeds weer opnieuw een operatieschema te maken, een tijdrovende, eentonige en ingewikkelde klus. Het stochastische element in de duur van operaties betekent een complicatie bij het plannen. En dan kan het zijn dat een niet goed uitgebalanceerde planning binnen het operatiekamercomplex de gang van zaken verstoort op andere afdelingen, zoals de gewone verpleegafdelingen en de intensive care afdelingen. We introduceren een vorm van cyclisch plannen, *master surgical scheduling*, als oplossing voor dit probleem. Hierbij worden frequent uitgevoerde electieve ingrepen cyclisch gepland. Aan het probleem van de onzekere duur van procedures wordt tegemoetgekomen door uitloopcapaciteit te plannen. De modellering is in de vorm van een mathematisch computerprogramma met waarschijnlijkheidsvoorwaarden. Aangezien het programma dat hieruit voortvloeit onwerkbaar is, komen we met een benadering op basis van kolomgeneratie waarbij in twee opeenvolgende fasen de OK-benutting wordt gemaximaliseerd en de doorstroom van patiënten naar andere afdeling wordt afgevlakt. We hebben deze methode getest voor het Erasmus MC, en deze bleek goed uit te werken voor zowel de OK-benutting als de afvlakking van de behoefte aan bedden na de operaties.

APPENDIX 1

Managing the overflow of Intensive Care patients

Nelly Litvak, Marleen van Rijsbergen,
Richard J. Boucherie, Mark Van Houdenhoven

European Journal of Operational Research, Article In Press,
Available online 16 October 2006

Introduction

"Each year, hundreds of patients die unnecessarily." This was announced in the Dutch current affairs program NOVA on November 6th, 2001, during the discussion on the capacity shortage at Intensive Care Units (ICUs) in Dutch hospitals [1]. The Dutch minister of Health, Welfare and Sports recognized the problems and initiated studies into the capacity problems of ICUs. A primary report [7] indicated that almost 10% of the severely ill patients were refused, 4% were admitted even though there was actually no space, and 3% were released earlier to make place for new patients. The most important reason for the refusal of a patient was the lack of operational (staffed) IC beds caused mostly by shortage of nurses. The ICU capacity problem for emergency or trauma patients (victims of accidents) is strengthened by the complicated chain logistics of hospitals. In particular, cancellation of planned operations due to ICU capacity shortage is highly expensive. As a consequence, trauma patients are refused to accommodate these planned operations.

In the Netherlands, care for trauma patients is organized in a regional setting. In principle, each trauma patient has to be admitted to an ICU within the region. Only when all ICU beds in a region are occupied, a trauma patient may be transported to a hospital outside that region, with obvious degradation of the quality of health care due to, e.g., extended transportation times. In the current situation, where each ICU decides independently whether or not a trauma patient is admitted, it may be that a trauma patient is transferred outside the region due to simultaneous reservation of capacity at some of the ICUs, while other ICUs have actually reached their capacity. An initial capacity study [2] indicates that indeed in the Rijnmond Region sufficient ICU capacity seems to be available, and that lack of cooperation is a major cause for trauma patients to be transported outside the region.

This paper focuses on solutions for cooperation among ICUs so as to minimize the number of trauma patients transported outside the region while maintaining a sufficient amount of ICU beds for planned operations. We show that reservation of several ICU beds for regional trauma patients and sharing these beds among the hospitals (so-called *regional beds*) results in a higher acceptance rate for emergency patients with a smaller number of beds in the region, without serious degradation of the fraction of cancelled operations. This is mainly due to the more efficient use of ICU capacity. Cooperation among hospitals thus helps to achieve a high acceptance level with a smaller number of beds resulting in improved service for all patients.

This paper provides a mathematical model for regional capacity allocation at ICUs under constraints on the number of refused patients. The model includes regional ICU capacity for regional emergency patients, and contains a detailed description of patient classes admitted to ICUs, and of solutions to accommodate bed shortages. Typical solutions in case of bed shortage are: transferring a patient to another hospital/region; postponing a planned operation; and releasing another patient earlier. These solutions have serious drawbacks, and the solution also depends on the patient class. Patients arriving at an ICU are of three classes, which mainly differ in the decision for admittance to the ICU. An elective patient may require an ICU bed following a planned operation. A planned operation can start only when an ICU bed is available. When all ICU beds are occupied, the operation is cancelled. An internal trauma patient, due to e.g. an emergency at the ward, must always be admitted to the ICU. When all ICU beds are occupied, a so-called *over-bed* is created. An over-bed is an originally non-staffed bed which is forcefully brought into operation thus loading the staff with an extra patient. This results in a decreased level of care at the ICU. A regional trauma patient, due to e.g. an accident in the region, is accepted only when an ICU bed is available. Otherwise the patient is not admitted and sent to another ICU. From a mathematical perspective, a regional model for ICUs shows major similarities with queueing theoretical models developed for circuit switched telephone systems with overflow capacity. For such systems, the highly accurate Equivalent Random Method (ERM) allows us to approximate the fraction of blocked telephone calls [11]. Unfortunately, internal emergency patients placed in over-beds cannot be included in the ERM. Therefore, in this paper, we develop a generalisation of the ERM that also allows for these patients. A detailed simulation study indicates that our generalisation of the ERM accurately approximates the fraction of refused patients.

A case study focusing on the Rijnmond Region indicates the capacity gain that may be achieved. In this region, the Erasmus Medical Centre (Erasmus MC) is appointed as one of the ten trauma centres in the Netherlands (see the National Atlas of Public Health [3]). According to a strategic analysis of cluster 17 of the Erasmus MC, responsible for Anesthesiology, ICUs, and Operating Theatres, the number of trauma patients sent to the ICU of the Erasmus MC has increased since its recognition as a trauma centre [2]. Some of the capacity problems at the ICU of the Erasmus MC are presumably caused by other hospitals in the region, which are not willing to cancel elective (planned) operations to allow for admission of emergency patients. As indicated in [2], it seems that the operational IC capacity in the Rijnmond Region reasonably approaches the demand for IC beds. At present however, emergency patients are occasionally sent outside the Rijnmond Region because no operational bed can be found in the region. If all hospitals in the region allocate several IC beds as emergency beds, the region can most likely take care of most of the emergency patients in the Rijnmond Region [2]. This does imply that sometimes hospitals might have to cancel elective operations while having an empty operational bed. Our case study indicates that the increased ICU bed occupation due to regional cooperation indeed reduces the fraction of regional trauma patients not accepted at an ICU in the region, but that this is *not* at the cost of an increased fraction of cancelled planned operations.

The paper is organized as follows. In the next section we describe the structure of the ICU, and available data. In Section 3, we present an overflow model of an ICU inspired by closely related models in telecommunications systems. In Section 4 we carry out the analysis of the model and provide the method for computing the fraction of rejected patients. Section 5 provides a simulation study to indicate the accuracy of our approximation, and is devoted to the case study for the Rijnmond Region. Conclusions and recommendations are given in the final Section 6.

Patient flows in the ICU

Intensive Care is specific medical treatment and nursing to severely ill patients who require intensive monitoring, mostly elaborate pharmacological treatment and in many cases support with artificial ventilation. The admission and release of a patient in the ICU is subject to a number of rules [5]. There are, however, no unambiguous agreements on how to deal with an arriving patient when no operational IC bed is available. An IC bed is operational when sufficient staff is available.

In practice, one can roughly distinguish three patient types: elective patients, internal emergency patients and external/regional emergency patients. Elective patients arrive from the operating theatre after undergoing a planned operation. If no operational IC bed is available, the operation is cancelled. An exception is made for operations that involve many people (staff and patients), for example a liver transplantation with a living donor. For such patients, beds are reserved that will not be taken by another patient.

Emergency patients arrive unexpectedly and require immediate care. Internal emergency patients arrive from a nursing ward. Regional emergency patients arrive through the emergency room, mostly brought by ambulance. The ambulance nurse does not have information on the availability of IC beds. If there is no bed available for an emergency patient, an attempt is made to create a place. For instance, another patient may be predischarged from the ICU but only if the discharge of the patient was already imminent. Also, a patient who came from a different hospital for some special procedure may be sent back if the special procedure is finished. If none of these options is available, the solution depends on the type of patient.

An internal emergency patient should be kept in the hospital mostly because it is not desirable to transport a critically ill patient, but also because legally, a patient can only be transferred if it is beneficial for the patient. Therefore, for an internal emergency patient an over-bed is created, which is an IC bed that was not staffed. The drawback of the over-bed is that physicians and nurses have to work harder as they have an extra patient to take care of, which requires flexible staff and negatively affects the quality of care. As soon as a patient is discharged, the over-bed is cancelled. For regional patients an over-bed is generally not an option because the hospitals tend to give priority to already admitted patients, and legally, a patient not yet admitted to the hospital can be sent to another hospital. Thus, for a regional emergency patient, generally an operational bed in another hospital is sought, and sometimes an available bed can be found only outside the region.

Figure 1 schematically depicts the patient flows for two ICUs. Flow 1 reflects the regional emergency patients, who are transferred to another hospital/region if all the beds are occupied. Flow 2 is the flow of elective patients. If no operational bed is available at their arrival, they are sent home to return later. Flow 3 corresponds to the flow of internal emergency patients who are not transferred in case of a full ICU, but are placed in an over-bed. Flow 4 depicts the patients whose discharge is imminent and who can be predis-

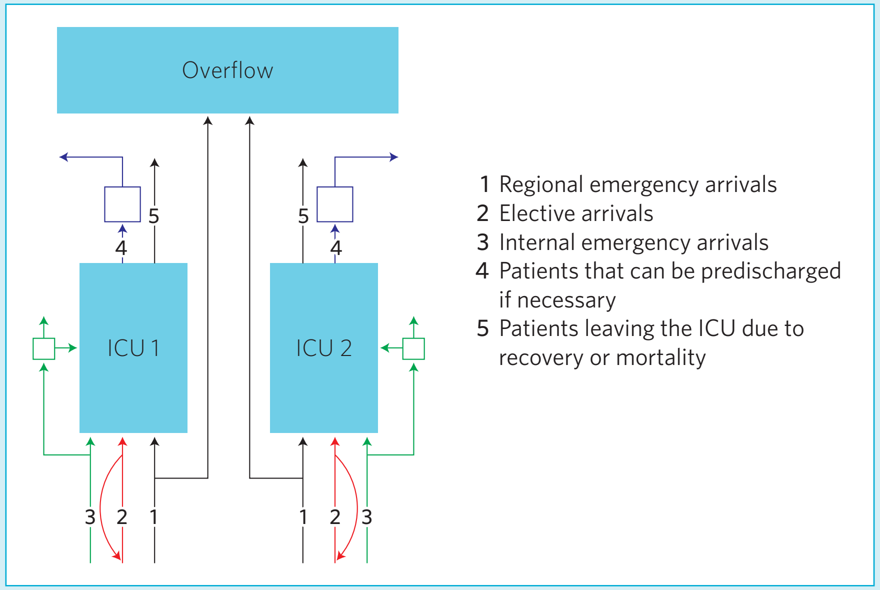


Figure 1 *Overview of the erm including all patient streams.*

Type of arrival	Mean	Standard deviation
Elective	0.58	0.92
Elective excluding weekend days	0.42	0.79
Internal emergency	0.62	0.74
Regional emergency	0.46	0.60
Total	0.18	0.39

Table 1 *Interarrival times in days for Erasmus MC*

Type of arrival	Mean	Standard deviation
Elective	3.88	6.44
Internal emergency	8.15	12.69
Regional emergency	7.95	13.78
Total	6.93	11.90

Table 2 *Mean length of stay in days for Erasmus MC*

charged in case of an incoming emergency. We do not take this flow into account. Flow 5 is the flow of patients who leave the ICU (because of recovery or mortality). The Overflow block denotes the patients who are rejected at the ICU.

The order of magnitude of the number of arriving patients, and the length of stay (LOS) are required for the selection of a proper approximation. According to the data presented in [9] for the Erasmus MC, the average interarrival time is 0.18 days. More detailed data on different patients types are given in Table 1. As the elective patients never arrive at weekends, we also provide the mean and standard deviation of the interarrival times for elective patients leaving out the weekend days.

The total mean LOS given in [9] for the Erasmus MC is 6.93 days. Table 2 contains the mean LOS for the three types of patients. The mean LOS of the elective patients differs significantly from the two types of emergency patients. The LOS is measured in whole days and includes the arrival and the release days not taking into account the time of release/arrival. It is also shown in [9] that the data on the LOS fits a Log Normal distribution.

The number of operational IC beds ranges from 5 for small hospitals to 40 for larger hospitals. Typically, the occupation degree of IC beds is above 80%.

Overflow model for regional ICU capacity

Consider a region containing multiple ICUs that jointly reserve beds (regional beds) for regional emergency patients only. In this case, the overflow block in Fig. 1 depicts this regional emergency capacity, consisting of an extra ICU that is intended for regional emergency patients who are refused at an original ICU. In practice, these beds will be distributed over the ICUs in the region, but will be reserved for regional emergency patients, thus creating a virtual ICU. Our goal is to compute the fraction of rejected patients (rejection probabilities).

We assume that all hospitals have similar patient stream structure. Assume that patients arrive to the hospital according to a Poisson flow. For the emergency arrivals this assumption is reasonable and is supported by statistical data [9]. The elective arrivals, however, are scheduled and therefore most likely do not constitute a Poisson flow. However, a surgeon is not aware of the occupation of the ICU when planning operations. As only a fraction of 5% of operated patients require Intensive Care after the operation, the assumption of Poisson arrivals is reasonable. In our model, ICUs may have a different mean arrival rate, reflecting the size of the area immediately surrounding a hospital from which patients are sent to the ICU. Let λ_i denote the total arrival rate (average number of patients arriving per time unit) at ICU i. The fraction of regional emergency patients, elective patients, and internal emergency patients is denoted as $p_{1,i}$, $p_{2,i}$, and $p_{3,i}$, respectively, with $p_{1,i} + p_{2,i} + p_{3,i} = 1$. The return of elective patients after a cancelled operation is modelled as a new arrival.

For analytical tractability, we assume that the LOS is exponentially distributed. A large class of queueing loss models is insensitive to the distribution of the service time. In Section 5 we present simulation results that support this kind of insensitivity in our model, and justify the assumption of exponential LOS.

To simplify notation, we do not discriminate between the mean LOS of different patient types. The mean LOS for patients at ICU i is denoted as . Data indicate that the LOS is indeed similar for internal and regional emergency patients [9]. For elective patients the LOS is generally smaller and less variable. Nevertheless, the model with equal mean LOS provides a good approximation and can be readily extended to the case of different mean LOS for different patient types.

Analysis

From a mathematical perspective, the behaviour of ICUs with shared regional capacity closely resembles that of a circuit switched telephone system with common overflow. In that system, a telephone call occupies a circuit during its call-length, and a call generated when all circuits are occupied is blocked and clear. To see the resemblance, identify patients with calls, beds with circuits (servers), and LOS with call-length. By capacity of the system we mean the number of servers, and we use these terms interchangeably throughout the paper. The computation of call blocking probabilities in such systems is an important research question that has received considerable attention in the literature. In the simplest case of one telephone switch with one incoming flow and c circuits, the system is referred to as the Erlang loss system, and the blocking probability can be computed using the famous Erlang loss formula [6]:

$$\text{Blocking probability} = B(c,\rho) = \frac{\rho^c/c!}{\sum_{k=0}^{c}\rho^k/k!}$$

where $\rho = \lambda\mu^{-1}$ is the load, with λ the call arrival rate, and μ^{-1} the mean call length.

Real-life systems, however, require analysis that is far beyond this basic model. For instance, the problem becomes much more complex when several multi-server units share a common overflow. To approximate the blocking probabilities in this model, the Equivalent Random Method (ERM) introduced by Wilkinson [11] can be efficiently applied. The idea of the classical ERM and its numerous modifications is to replace several multi-server units by one Equivalent Random unit that generates the same expectation and variance of the overflow as in the original system. Then the Erlang loss formula can be applied as for a classical loss system with equivalent random load ρ and capacity $c + r$, where r is the capacity of the overflow buffer, and c is the capacity of the Equivalent Random unit (see Figure 2).

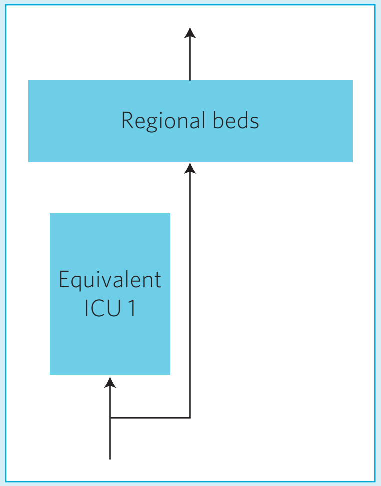


Figure 2 *Equivalent Random IC with regional emergency patients.*

More formally, consider a system of I multi-server units. In order to apply the ERM we need to find an equivalent random load ρ and capacity c. To this end, first consider an overflow with unlimited capacity. The mean and variance of the number of calls in the overflow from unit $i = 1, \dots, I$, with load $\rho_i = \lambda_i/\mu_i$, and capacity c_i, are [6]

$$E_i = \rho_i B(c_i,\rho_i), \quad V_i = E_i\left(1 - E_i + \frac{\rho_i}{c_i + 1 + E_i - \rho_i}\right).$$

The mean and variance of the *total* number of calls in the overflow buffer, assuming that the latter has an unlimited capacity, is

$$E = \sum_{i=1}^{I} E_i, \quad V = \sum_{i=1}^{I} V_i. \tag{1}$$

The Equivalent Random system is the Erlang loss queue with capacity c and load ρ that satisfy [6]

$$E = \rho B(c,\rho), \quad V = E\left(1 - E + \frac{\rho}{c + 1 + E - \rho}\right). \tag{2}$$

System (2) can readily be solved numerically. We can also find a solution using analytic approximations such as the equations given by Rapp [10]:

$$\rho = V + 3\frac{V}{E}\left(\frac{V}{E} - 1\right), \quad c = \frac{\rho\left(E + \frac{V}{E}\right)}{E} + \frac{V}{E-1} - E - 1. \tag{3}$$

Cooper [6] states that these estimates of ρ and c are generally on the high side of the exact values. Rounding c down to an integer c and then finding ρ by

$$\rho = \frac{(\lfloor c \rfloor + E + 1)\left(E + \frac{V}{E} - 1\right)}{E + \frac{V}{E}}, \tag{4}$$

gives a better approximation.

Let r be the capacity of the overflow determined above. Once ρ and c for the Equivalent Random unit are defined, we can compute the approximate average number that is rejected at the overflow

$$\overline{E} = \rho B(c + r,\rho) = \rho\frac{\rho^{c+r}/(c+r)!}{\sum_{k=0}^{c+r}\rho^k/k!}. \tag{5}$$

Now consider a region with regional ICU. Compared to the known versions of ERM, our model is different because (i) internal emergency patients cannot be rejected, and (ii) elective patients are never sent to the overflow. In order to apply the ERM, we have to be able to compute the mean E and variance V of the overflow for our model with three patient flows and the possibility of overbeds.

From (1), it is sufficient to find E_i and V_i for the ith ICU. For $j, k \geq 0$, let $P_i(j, k)$ be the steady-state probability that there are j patients at ICU i and k

patients at the overflow, $j \geq 0, k \geq 0$. Setting $P_i(-1,k)=0, k \geq 0$, we can write the global balance equations that uniquely determine these probabilities as follows:

$$(\lambda_i + j\mu_i + k\mu_i)P_i(j,k) = \lambda_i P_i(j-1,k) \tag{6}$$
$$+ (j+1)\mu_i P_i(j+1,k) + (k+1)\mu_i P_i(j,k+1), \quad j < c_i; \; k \geqslant 0,$$
$$((p_{1,i}+p_{3,i})\lambda_i + c_i\mu_i + k\mu_i)P_i(c_i,k) = \lambda_i P_i(c_i-1,k) + p_{1,i}\lambda_i P_i(c_i,k-1) \tag{7}$$
$$+ (c_i+1)\mu_i P_i(c_i+1,k) + (k+1)\mu_i P_i(c_i,k+1), \quad k \geqslant 0,$$
$$((p_{1,i}+p_{3,i})\lambda_i + j\mu_i + k\mu_i)P_i(j,k) = p_{3,i}\lambda_i P_i(j-1,k) + p_{1,i}\lambda_i P_i(j,k-1) \tag{8}$$
$$+ (j+1)\mu_i P_i(j+1,k) + (k+1)\mu_i P_i(j,k+1), \quad j > c_i; \; k \geqslant 0.$$

The left-hand side of (6) represents the probability flow out of state (j, k) due to arrivals at rate λ_i, departures of patients from the ICU at rate $j\mu_i$, and departures from the overflow at rate $k\mu_i$. The right-hand side represents the probability flow into state (j, k) due to arrivals from state $(j-1, k)$, due to departures from the ICU from state $(j+1, k)$, and due to departures from the overflow from state $(j, k+1)$. The other equations have a similar interpretation.

We are interested in the mean and variance. To obtain expressions for these measures, let $G_{i,j}(z)$ be the marginal generating function

$$G_{i,j}(z) = \sum_{k=0}^{\infty} P_i(j,k)z^k, \quad |z| \leqslant 1.$$

Multiplying the balance Eqs. (6), (7) and (8) by z^k, z 1, and summing both sides of the equations over k, we obtain the following relations:

$$(\lambda_i + j\mu_i)G_{i,j}(z) = \lambda_i G_{i,j-1}(z) + (j+1)\mu_i G_{i,j+1}(z) \tag{9}$$
$$+ \mu_i(1-z)\frac{d}{dz}G_{i,j}(z), \quad j < c_i,$$
$$((p_{1,i}(1-z) + p_{3,i})\lambda_i + c_i\mu_i)G_{i,c_i}(z) = \lambda_i G_{i,c_i-1}(z) \tag{10}$$
$$+ (c_i+1)\mu_i G_{i,c_i+1}(z) + \mu_i(1-z)\frac{d}{dz}G_{i,c_i}(z),$$
$$((p_{1,i}(1-z) + p_{3,i})\lambda_i + j\mu_i)G_{i,j}(z) = p_{3,i}\lambda_i G_{i,j-1}(z) + \tag{11}$$
$$(j+1)\mu_i G_{i,j+1}(z) + \mu_i(1-z)\frac{d}{dz}G_{i,j}(z), \quad j > c_i.$$

The expectation and variance of the overflow can be now calculated by using first and second order derivatives of $G_{i,j}(z)$ with respect to z as follows:

$$E_i = \sum_{j=0}^{\infty} \frac{\partial}{\partial z} G_{i,j}(z)\bigg|_{z=1},$$
$$V_i = \sum_{j=0}^{\infty} \frac{\partial^2}{\partial z^2} G_{i,j}(z)\bigg|_{z=1} + E_i - (E_i)^2.$$

Therefore, differentiating both sides of Eqs. (9), (10) and (11) with respect to z and substituting $z=1$ and denoting $E_i(j) = \frac{\partial}{\partial z}G_{i,j}(z)|_{z=1}$ we obtain

$$(\lambda_i + (j+1)\mu_i)E_i(j) = \lambda_i E_i(j-1) + (j+1)\mu_i E_i(j+1), \quad 0 \leqslant j < c_i, \tag{12}$$
$$(p_{3,i}\lambda_i + (c_i+1)\mu_i)E_i(c_i) = \lambda_i E_i(c_i-1) + (c_i+1)\mu_i E_i(c_i+1) + p_{1,i}\lambda_i P_i(c_i), \tag{13}$$
$$(p_{3,i}\lambda_i + (j+1)\mu_i)E_i(j) = p_{3,i}\lambda_i E_i(j-1) + (j+1)\mu_i E_i(j+1) + p_{1,i}\lambda_i P_i(j), \quad j > c_i, \tag{14}$$

where $P_i(j) = G_{i,j}(1)$ is the probability that there are j patients at ICU i, and $E_i(-1) = 0$, $i = 1, \ldots, I$. To obtain E_i, we sum the above equations over j. It now follows that

$$E_i = p_{1,i}\rho_i \sum_{j=c_i}^{\infty} P_i(j). \tag{15}$$

To find the variance of the number of patients in the overflow, we take the second order derivatives in (9), (10) and (11) with respect to z and put $z=1$. Summing up the resulting equations, we obtain the second factorial moment of the number of patients in the overflow, which leads to the following expression for the variance:

$$V_i = p_{1,i}\rho_i \sum_{j=c_i}^{\infty} E_i(j) + E_i - (E_i)^2. \tag{16}$$

It now remains to find $\sum_{j=c_i}^{\infty} P_i(j)$ and $\sum_{j=c_i}^{\infty} E_i(j)$. The probabilities $P_i(j)$, $j \geq 0$, can be found by iteratively solving (9), (10) and (11) with $z=1$. This gives

$$P_i(j) = \begin{cases} \frac{1}{j!}(\rho_i)^j P_i(0), & 0 \leqslant j \leqslant c_i; \\ \frac{1}{j!}(p_{3,i})^{j-c_i}(\rho_i)^j P_i(0), & j > c_i. \end{cases} \tag{17}$$

From (17) and the normalizing condition $\sum_{j=0}^{\infty} P_i(j) = 1$ we obtain $P_i(0)$:

$$P_i(0) = \left[\sum_{j=0}^{c_i} \frac{(\rho_i)^j}{j!} + \sum_{j=c_i+1}^{\infty} \frac{(\rho_i)^j}{j!}(p_{3,i})^{j-c_i} \right]^{-1}. \tag{18}$$

It now follows from (17) and (18) that

$$\sum_{j=c_i}^{\infty} P_i(j) = 1 - \sum_{j=0}^{c_i-1} \frac{(\rho_i)^j}{j!} \cdot \left[\sum_{j=0}^{c_i} \frac{(\rho_i)^j}{j!}(1-(p_{3,i})^{j-c_i}) + (p_{3,i})^{-c} e^{p_{3,i}\rho_i} \right]^{-1}. \tag{18}$$

The expression for $\sum_{j=c_i}^{\infty} E_i(j)$ can be found similarly. Iterating (12), (13) and (14), we can express $E_i(j)$ via $E_i(0)$ for all $j > 0$. Then $E_i(0)$ can be found from the normalisation $E_i = \sum_{j=0}^{\infty} E_i(j)$. This will give the values of $E_i(j)$ for all $j > 0$ and thus we can compute the sum on the right-hand side of (16). The resulting analytical expression is, however, cumbersome and will not be given here. Instead, we suggest a simpler computational approach. Since $E_i(j)$ decreases quickly with j, we may iterate (12), (13) and (14) only up to some sufficiently large value of $j = M$. Then $E_i(0)$ can be found from $E_i = \sum_{j=0}^{M} E_i(j)$, and the required expression will be $E_i = \sum_{j=c_1}^{M} E_i(j)$. This approach reflects reality, for instance, if M is the number of constructional beds. On the other hand, if we want an accurate solution of the proposed model with the unlimited over-bed capacity, we can choose M large enough so that the resulting values of $E_i(0)$ are sufficiently close for M and $M-1$, and $E_i(M)$ is close to zero.

Having computed the mean E and the variance V of the overflow, we can use (2) or (3) and (4) to define ρ and c of the Equivalent Random ICU. An expression for the number of patients rejected at the overflow of capacity r is then given in (5). The loss probability $B^{(r)}$ for regional emergency patients is

$$B^{(r)} = \frac{\overline{E}}{\sum_{i=1}^{n} \rho_i p_{1,i}}, \qquad (20)$$

where $\rho_i p_{1,i} = \lambda_i p_{1,i}/\mu_i$ is the load of regional emergency patients at ICU i. The blocking probability for an emergency patient arriving at ICU i can be approximated as follows. For $i = 1, \ldots, I$, let

$$B_i = \sum_{j=c_i}^{\infty} P_i(j)$$

be the probability that ICU i is full. According to the PASTA property this is the probability of rejection of a regional patient at the original ICU. Thus, the probability that an emergency patient attempts to access a regional bed equals $\sum_{i=1}^{I} B_i(p_{1,i}\lambda_1/\lambda_{1\bullet})$, where $\lambda_{1\bullet} = p_{1,1}\lambda_1 + \cdots + p_{1,I}\lambda_I$ is the total arrival rate of regional emergency patients, and $p_{1,i}\lambda_i/\lambda_{1\bullet}$ is the probability that an emergency patient claiming a regional bed comes from the ICU i. Assume that the rejection probability at the regional ICU, B_0 say, is the same for patients originating from any ICU. Hence, using the total probability formula for the blocking probability $B^{(r)}$, for any $i = 1, \ldots I$, $r \geq 0$, we write:

$$B^{(r)} \approx \left[\sum_{i=1}^{I} B_i(p_{1,i}\lambda_i/\lambda_{1\bullet})\right] B_0,$$

so that

$$B_i^{(r)} \approx B_i B_0 \approx \frac{\lambda_{1\bullet} B_i B^{(r)}}{\sum_{i=1}^{I} B_i p_{1,i}\lambda_i}. \qquad (21)$$

Note that the equivalent random load ρ and capacity c are defined from the mean and variance of the overflow which consists only of emergency patients.

Thus, the Equivalent Random ICU has load and capacity related only to the regional emergency flow. However, the blocking probability is also related only to the regional emergency patients. Therefore, one may hope that $B^{(r)}$ in (20) provides a good approximation for the real percentage of rejected regional patients. The numerical results in the next section show that this is indeed the case.

Simulation model and numerical results

This section contains both a simulation study of patient flows to investigate the accuracy of the ERM approximation, and a case study for the Rijnmond Region in the Netherlands. Data for the simulation study are obtained from a database of the Erasmus MC containing detailed information on patients, operations, and LOS for the years 1994–2004.

Accuracy of the erm approximation To investigate the accuracy of the ERM approximation, a simulation model has been developed in eM-Plant, version 7.0.2. eM-Plant is software for object-oriented, graphical modelling for simulating and visualizing systems and business processes [4]. Our simulation model is generic in the sense that the number of ICUs in the region, the number of beds per ICU, the arrival times and Length Of Stay (LOS) can all be adjusted. The simulation study includes detailed acceptance rules, and closely mimics the actual patient flows in ICUs including general LOS. The aim of the simulation study is to (i) investigate the influence of the distribution of the LOS, and (ii) investigate the accuracy of the ERM approximation.

The main frame of the simulation model represents the region which contains several ICUs and a unit with a number of regional beds. The three types of patients arrive at an ICU according to a Poisson process, each with its own rate. Elective patients do not arrive on weekends. If a bed is available the patient is treated at this ICU. The length of stay of the patient is modelled through a LogNormal distribution, each patient type having a different mean LOS. In the case when no beds are available and an internal emergency patient arrives, an over-bed is created for this patient. When no bed is available upon arrival of an elective patient, the patient is deleted from the system. When a regional emergency patient arrives and no bed is available, the patient is sent to a regional bed (in the frame of the region). Figure 3 shows the basic patient streams in the simulation model with one ICU.

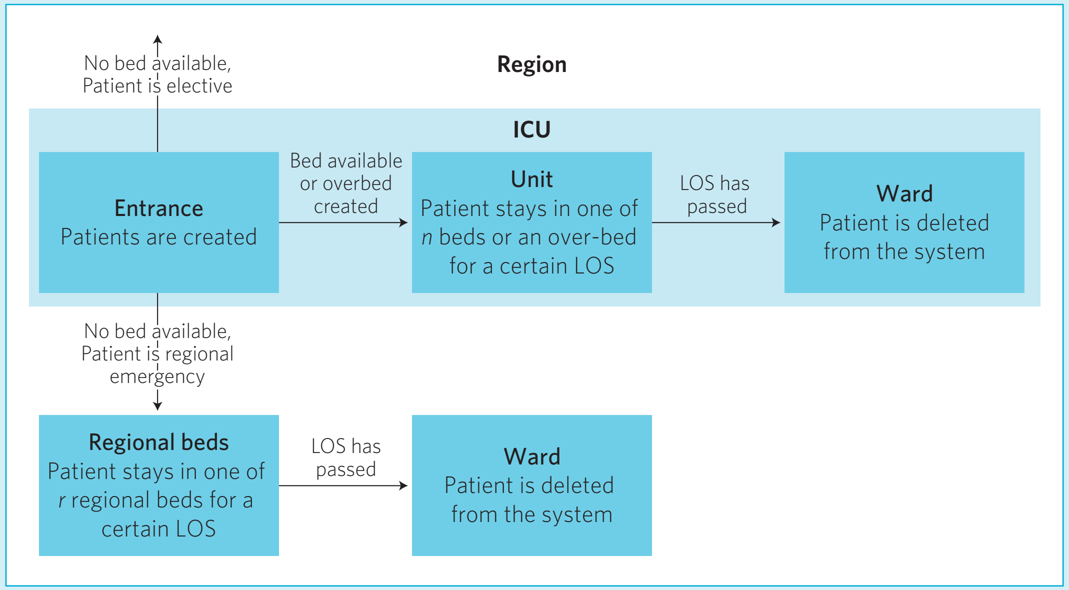


Figure 3 *The basic patient streams in the simulation model.*

In the simulation model, we used the actual data from the IC department of the Erasmus MC and the estimated data from three other hospitals in the Rijnmond Region, the Albert Schweizer hospital, hospital Dirksland, and the Sint-Franciscus hospital. Further hospitals in the Rijnmond Region have not been taken into account. We have made 20 simulation runs for each test. Results of the first 10 runs are used to determine the warm-up period. This is necessary because the system starts empty, which is not the case in reality. Results of the second 10 runs are used to calculate the 95% confidence intervals with at least 10% relative precision by the Replication/Deletion Method [8].

The first aim in the simulation study is to compare the results for LogNormal and exponential length of stay. As was mentioned in Section 2, the LogNormal distribution fits the real-life data, whereas the exponential distribution is used for analytical tractability. Table 3 gives 95% confidence intervals for the mean bed occupation, the proportion of cancelled operations, the proportion of refused emergency patients and the average number of over-beds, for exponentially and LogNormally distributed LOS, as well as the confidence interval for the difference, based on Common Random Numbers (CRN) [8]. For all performance measures except mean bed occupancy, 0 is contained in the confidence interval based on CRN for the difference between performance measures under LogNormal and exponential LOS. This result supports the use of the exponential distribution in our analytical model to approximate the performance measures for a region with cooperating ICUs. Insensitivity to the LOS distribution is a common property of the Erlang loss model, the mathematical model underlying our approximation, which suggests that the accuracy of our results will not change when compared to results obtained using the empirical distribution of the LOS.

The second aim of the simulation study is to verify the quality of the ERM approximation of the blocking probability $B^{(r)}$ for regional emergency patients. Table 4 contains the results for different numbers of regional beds. As can be seen from these results, the ERM provides an engineering approximation (roughly 10% accuracy) of the fraction of rejected regional emergency patients. ERM overestimates the loss probability for a small number of beds and underesti-

Confidence intervals	Exponential	LogNormal	Difference (CRN)
Mean bed occupation	0.90 ± 0.002	0.89 ± 0.001	0.003 ± 0.002
Proportion of cancelled operations	0.26 ± 0.006	0.26 ± 0.005	0.003 ± 0.009
Proportion of refused regional patients	0.18 ± 0.005	0.18 ± 0.004	0.002 ± 0.007
Average number of over-beds	0.08 ± 0.004	0.08 ± 0.004	0.002 ± 0.008

Table 3 *Exponential versus LogNormal LOS, based on the Erasmus MC data*

Number of regional beds	ERM blocking probability	Proportion refused regionals in simulation
0	0.255	0.232 ± 0.006
1	0.215	0.195 ± 0.003
2	0.177	0.162 ± 0.006
3	0.142	0.134 ± 0.005
4	0.112	0.107 ± 0.004
5	0.085	0.083 ± 0.004
6	0.063	0.065 ± 0.002
7	0.045	0.049 ± 0.001
8	0.030	0.036 ± 0.001
9	0.020	0.026 ± 0.002
10	0.013	0.018 ± 0.001
11	0.008	0.011 ± 0.001

Table 4 *Blocking probability regional emergency patients in the region with cooperation*

mates the loss probability for a large number of beds. The reason may be that ERM smooths the discrepancy between distinct ICUs. We conclude that ERM captures the loss probability and the required number of regional beds with good precision.

We have also used simulation to verify analytical formulae for blocking probabilities at each hospital separately, with and without cooperation. For the case with cooperation, we used the formula (21) to determine the probability that an emergency patient arriving at ICU *i* eventually has to be sent outside the region. In case without cooperation, we assumed that a hospital reserves several emergency beds, and we used ERM involving only one unit in order to compute the blocking probability. The error in the analytical approximation in these two cases turns out to be of a similar order as in Table 4. In the case study below we use the analytical approximation.

Case study for Rijnmond region

The objective of this case study is to investigate the advantage of cooperation between the hospitals. For that, we used the ERM to compute the blocking probabilities in each hospital separately assuming that they handle the emergency patients on their own, without the regional beds capacity. We will illustrate the advantage of cooperation within the region by means of the following example.

The goal of the management of the ICUs in the region is that at most 1% of the regional patients are rejected and transferred to an ICU outside the region. Table 4 indicates that 11 regional beds are required to achieve this goal. Table 5 provides the fraction of rejected regionals per hospital, where the approximate blocking probabilities computed using formula (21). The row with 11 beds indicates that this results in a blocking probability of approximately 0.6% for regional patients arriving at the Erasmus MC, approximately 2% for the Albert Schweizer Hospital, 2% for the Sint Franciscus Gasthuis, and approximately no rejected regionals for the Dirksland Hospital. Notice that these numbers at the Albert Schweizer and Sint Franciscus hospitals exceed those of the Erasmus MC. As the total number of regionals arriving at the ICU of the Erasmus MC is considerably larger than that number at the other hospitals, the total rejection probability is 0.8%. Furthermore, note that the Albert Schweizer and Sint Franciscus hospitals seem to benefit more than the Erasmus MC from the introduction of regional beds. This is due to the fact that the initial rejection rate at these hospitals is much higher than at the Erasmus MC.

Now consider the hospitals without regional cooperation. Table 6 presents the fraction of rejected regionals for each hospital. To achieve at most 1% of

Nr of regional beds	Erasmus MC	Albert Schweizer	Dirksland	Sint Franciscus
0	0.207	0.689	0.004	0.715
1	0.174	0.478	0.003	0.602
2	0.144	0.496	0.003	0.478
3	0.116	0.385	0.002	0.399
4	0.091	0.302	0.002	0.313
5	0.069	0.230	0.001	0.239
6	0.051	0.169	0.001	0.176
7	0.036	0.120	0.001	0.125
8	0.025	0.082	0.000	0.085
9	0.016	0.054	-	0.056
10	0.010	0.034	-	0.035
11	0.006	0.020	-	0.02

Table 5 *Blocking probability regional emergency patients for each hospital with cooperation*

Nr of regional beds	Erasmus MC	Albert Schweizer	Dirksland	Sint Franciscus
0	0.207	0.732	0.016	0.742
1	0.168	0.230	0.001	0.357
2	0.133	0.049	0.000	0.135
3	0.102	0.007	-	0.039
4	0.077	0.001	-	0.009
5	0.056	0.001	-	0.002
6	0.039	0.000	-	0.000
7	0.026	-	-	-
8	0.017	-	-	-
9	0.011	-	-	-
10	0.006	-	-	-
11	0.004	-	-	-

Table 6 *Blocking probability of regional emergency patients for each hospital, without cooperation*

rejected regionals without cooperation, the Erasmus MC needs 10 emergency beds, the Albert Schweizer hospital 3 beds, the Sint Franciscus Gasthuis 4 beds, and the Dirksland Hospital one bed, resulting in 18 beds in total. Only a slightly higher fraction of rejected regionals will be guaranteed with 9 beds at the Erasmus MC and 0 beds at the Dirksland Hospital. Further decreasing of the number of reserved beds results in a much higher rejection rate. Thus, at least 16 reserved beds are required in the region without cooperation. Accordingly, cooperation between the hospitals can save at least 5 beds (31%). We note that simulation results (that we do not present here) come down to the same numbers.

In the case study reported above, the reservation of regional beds does not influence rejection of elective patients or the use of over-beds. In practice, an elective operation sometimes will have to be cancelled although there is an empty regional bed available. However, since the LOS of elective patients is more predictable and their arrivals can be controlled, exact knowledge about how many emergency patients can be present in the ICU may help to decrease the number of cancellations by better planning of elective arrivals and thus result in a smaller number of cancellations. Cooperation helps to decrease the total capacity required for regional emergency patients, which eventually is advantageous for elective patients, too.

Conclusions and further research

Hospitals in the Netherlands are responsible for their own budget. In contrast, efficient care for patients within a region covered by multiple hospitals requires coordination among hospitals. A strong basis for coordination is provided by proper insight into the benefits and drawbacks of cooperation. To this end, this paper has investigated the effect of regional Intensive Care capacity on the quality of patient care, in particular focusing on the fraction of regional emergency patients not admitted to an ICU in the region, and the fraction of cancelled operations. Reserving IC beds for regional emergency patients seems to increase the number of cancelled operations. As is demonstrated in a case study for the Rijnmond Region in the Netherlands, cooperation may both lead to a reduction of the fraction of rejected regionals, and a reduction in the fraction of cancelled operations. Both reductions are due to the more efficient use of IC capacity.

Establishing an IC bed is extremely costly. Therefore, making the trade-off between regional and local IC capacity requires an adequate tool to quantify the number of required IC beds for each hospital in various scenarios taking into account aspects including the expected number of patients, the division of beds over hospitals, but also the fraction of cancelled operations and rejected regional patients allowed by the management, by health insurers, or by the government. Based on mathematical methods developed for circuit switched telephone systems, this paper has developed an extension of the Equivalent Random Method that allows us to quantify both the local (for each hospital) and regional fractions of rejected patients. The advantage of the Equivalent Random Method over simulation is that the ERM provides insight into the nature of the regional overflow problem, and that the ERM allows for a fast evaluation of all different combinations of the number of beds at each hospital and the number of regional beds. This allows for optimisation of the distribution of beds over hospitals. The model may also be applicable to other departments such as Radiology, or the wards.

There is room for improvements. Our results for the blocking probabilities seem to be too high. In part, this is due to the data provided by the hospitals. In particular the length of stay is on average one day too long for the Erasmus MC, since both the day of arrival and the day of departure are included. Furthermore, we have assumed that patients at peripheral hospitals have the same LOS. As the Erasmus MC is an academic hospital that also serves as a regional trauma centre, the LOS for other hospitals seems to be overestimated. A detailed data analysis, including data for hospitals in the region is beyond the scope of the current paper, and is among our aims for further research. The aim of the current paper is to show that the developed mathematical model provides an adequate predictions for required capacity in the given setting.

A second improvement may be to include non-Poissonian arrivals of elective patients. Although this seems to be an important improvement, in practice the assumption of Poisson arrivals may be reasonable, since only 5% of patients from the operating theatre require an IC bed. Therefore, the arrival process of elective patients to the ICU is more variable than the scheduled arrival of patients to the operating theatre. From a mathematical perspective, however, the generalisation is very interesting.

References

1 www.novatv.nl (in Dutch).

2 *Schakels in een keten*, Strategische visie 2004–2008 van Cluster 17 (in Dutch).

3 Rijksinstituut voor Volksgezondheid en Milieu, *Nationale Atlas Volksgezondheid*. Available from: <http://www.rivm.nl/vtv/generate/objv2atl_atlas_0.htm> (in Dutch).

4 Technomatrix, Tutorial version 5.5, eM-Plant, 2002.

5 J. Bakker, Criteria voor opname en ontslag van Intensive Care afdelingen in Nederland, Nederlandse Vereniging voor Intensive Care, 2001 (in Dutch).
6 R.B. Cooper, Introduction to Queueing Theory, Elsevier North Holland, New York (1981).
7 J.L.A. Hautvast, J. Bakker, N. Boekema-Bakker, J.A.J. Faber, D.E. Grobbee, A.J.P. Schrijvers, Plaats in de herberg, een studie naar determinanten van opname- en ontslagproblemen in IC-afdelingen in Nederland, Julius Centrum voor Huisartsgeneeskunde en Patiëntgebonden Onderzoek, Utrecht, 2001 (in Dutch).
8 A.M. Law and W.D. Kelton, Simulation Modeling and Analysis, McGraw-Hill Higher Education (2000).
9 R. Lesscher, E. Meutstege, J. Rouhof, Data Analyse van de Intensive Care, Universiteit Twente, 2004 (in Dutch).
10 Y. Rapp, Planning of junction networks in a multiexchange area, *Ericsson Technics* **20** (1964) (1), pp. 77–130.
11 R.I. Wilkinson, Theories for toll traffic engineering in the USA, *The Bell System Technical Journal* **35** (1956), pp. 421–514. MathSciNet

APPENDIX 2

Robust surgery loading

Erwin Hans, Gerhard Wullink,
Mark Van Houdenhoven, Geert Kazemier.

European Journal of Operational Research,
185 (2008), p. 1038–1050,

Introduction

In the Netherlands, long waiting lists and ageing population will result in huge expenses for health care in the near future (TPG, 2004). Application of operations research techniques, which is common practice in manufacturing industry, is expected to lead to an enormous improvement of efficiency in health care organizations. Erasmus MC in the Netherlands is one of the largest academic hospitals in Europe, with over 10,000 employees and an operating theatre with 16 operating rooms for inpatient surgeries. Every year approximately 9500 inpatient surgeries are performed, of which 85% are planned (elective). Since for a patient almost all costs occur on the day of surgery (Dexter, 2001), one of the most critical and expensive resources in a hospital is the operating theatre.

This paper addresses the problem of assigning elective surgeries to operating rooms (ORs), in such a way that not only the utilization of the OR theatre department is optimized, but also the total overtime is minimized. The latter prevents surgery cancellations. To make a surgery schedule more robust against overtime, we assign planned slack in addition to the planned surgeries on each OR-day. The planned slack is based on historical statistical data concerning the (stochastic) durations of the planned surgeries. We refer to this operational planning problem as robust surgery loading.

Erasmus MC has extensively collected data concerning the surgical process for more than 10 years. Various points in the process were consequently registered, which allows statistical analyses with respect to, e.g., surgery durations.

As is the procedure in many hospitals, in Erasmus MC each specialty has been assigned a number of ORs, to which they assign surgeries on a weekly basis. Unless there is a medical indication, these surgeries are assigned "FCFS," using a "First Fit" priority rule (we shall discuss this method in Section 3.1). The surgery assignments are then verified by the OR-management, as to whether they are feasible with respect to capacity constraints, and as to whether the risk of overtime is not too high. To prevent overtime, the specialties must plan slack, which is based on the variability of the surgery durations of the specialty. The OR-management has decided that the amount of slack is such that the probability of overtime is approximately 30%. This amount is determined by assuming that the sum of the planned surgery durations is normally distributed: the amount of planned slack is then 0.5 times the standard deviation of the total planned surgery duration. This standard deviation is calculated from the given standard deviations of the individual planned surgery durations.

Several approaches for OR-planning have been proposed in the literature. Guinet and Chaabane (2003) first assign patients to ORs, and then reschedule the surgeries assigned to an OR to satisfy the material and personnel constraints. They use an extended version of the Hungarian method to solve an integer linear program that minimizes the patient waiting time between the hospitalization date and the actual start of the operation. Nevertheless, they do not consider the surgery duration variability.

Roth and Van Dierdonck (2005) propose an MRP-based approach that uses the concept of diagnostic-related groups (DRG) as a Bill of Materials. DRGs define health care as product classes, which comprise a relative standard set of activities and required materials. In manufacturing planning and control literature it is generally recognized that MRP, which is material and not capacity oriented, has serious drawbacks, especially in environments characterized by much variability (see e.g. Hopp and Spearman, 2000 and Zijm, 2000).

Beliën and Demeulemeester (2007) study the problem of building robust cyclic master surgery schedules, to minimize the shortage of beds. Demand constraints ensure that each surgeon obtains a specific number of ORs, and capacity constraints limit the number of OR blocks on each day. They formulate a stochastic MIP (with stochastic surgery durations), which they linearize and solve heuristically. De Kreuk et al. (2004) study the problem of developing a cyclic schedule for the activities of medical specialists. They use a composite objective function to optimize a schedule from the viewpoint of the specialist, by considering the location, preferred day-part, and the preferred sequence for each specialist. The resulting quadratic integer program (QIP) is solved using simulated annealing.

Kuo et al. (2003) deal with the surgery loading problem from the specialty's viewpoint, and propose an LP model to optimize the revenues of the specialties. They assume that there are always sufficient cases to plan, intensive care capacity is unlimited, and surgeons are eager to operate more if is this is possible. They do not consider surgery duration variability.

Dexter et al. (1999) study the problem of assigning add-on surgeries to ORs. Surgeries are planned based on their expected duration; no slack is planned to decrease the risk overtime. They compare 10 algorithms on the OR utilization for the on-line and off-line case. In the on-line case the surgeries are not known in advance, whereas in the off-line case they are. The algorithms are based on either "Best Fit" or "Worst Fit," optionally with first sorting the surgeries on their duration (increasing or decreasing), and optionally with fuzzy constraints. Best (Worst) Fit assigns the surgery to the OR in which the available capacity is sufficient, but as small (large) as possible. The fuzzy constraints allow surgeries for which there is no OR in which they fit to be assigned to an OR in which at most 15 minutes of overtime is generated. Dexter et al. (1999) demonstrate that Best Fit Descending with fuzzy constraints performs best with respect to OR utilization, and they plead for using an off-line approach whenever possible. Goldman et al.

(1969) demonstrated as early as 1969 that "longest cases first" scheduling priorities yields the highest OR utilization rate, the lowest amount of overtime, and the largest number of delayed cases being transferred to another room.

Summarizing our short literature review on OR-planning, we conclude that there are several interesting approaches to planning problems in hospitals. The robust surgery loading problem has, in our opinion, not been addressed properly in the literature. While many approaches do recognize the presence of uncertainty in health care processes in general and OR-planning in particular, most do not deal with it explicitly. In this paper, we propose several constructive and local search heuristics for the robust surgery loading problem, which we regard as a "stochastic knapsack" problem. The objective is to optimize the assignment of surgeries by the specialties in such a way, that the risk of working in overtime is minimized, no surgeries are cancelled, and at the same time the OR capacity utilization can be improved.

The robust surgery loading problem is a generalization of the so-called "general bin packing problem with unequal bins," which deals with assigning a set of items to a set of bins with different sizes. If necessary, the size of each bin can be extended. The objective is to minimize the sum of the sizes of the used bins. This problem (and hence also the surgery loading problem) is strongly NP-hard, which can be proven by reduction from three-partition. Dell'Olmo and Speranza (1999) prove that the longest processing time (LPT) heuristic has a worst-case bound of $2 \cdot (2-\sqrt{2}) \approx 1.17$ in the off-line variant of this problem, and a worst-case bound of 5/4 for list-scheduling (LS) in the on-line variant. Ye and Zhang (2003) propose an improved on-line algorithm, and discuss the problem in which the maximum item size may be larger than the smallest bin. They present a worst-case bound of 6/5. The analogy with the surgery loading problem is that the ORs are bins with unequal capacities, which may be extended by planning overtime. The additional difficulty is that slack must be planned within the available capacity, which is based on the duration distributions of the concerned surgeries.

The remainder of this paper is organized as follows: Section 2 gives a formal problem description. Section 3 discusses the constructive heuristics, and Section 4 the local search heuristics. Section 5 gives the test results, and we conclude and outline directions for further research in Section 6.

Formal problem description

We consider a discretized planning horizon of T days ($t = 1, \ldots, T$), which in practice is usually a regular week of 5 days. On each day t, K parallel identical operating rooms are available (index $k = 1, \ldots, K$). The capacity of operating room k on day t is c_{kt}. All operating rooms start at the same time. There are S specialties (index $s = 1, \ldots, S$). On each day t, each specialty s has a number of operating rooms at its disposal, indicated by the set K_{st} ($\sum_s |K_{st}| = K, \ \forall t$). We refer to an element k of K_{st} as an "OR-day," also denoted by a triplet (s, k, t). For each OR-day an "OR-team" is available, which is a team of personnel required to perform surgeries. The personnel from whom an OR-team is put together comes from a so-called "unit." A unit is an organizational unit, which contains personnel capable of supporting a number of specialties. We denote the set of specialties that can be supported by a unit u as O_u ($u = 1, \ldots, U$). We assume that each specialty has a sufficient number of surgeons on each OR-day to perform the surgeries. Hence on each day t, each specialty s has at least as many surgeons as it has OR-days at its disposal.

N is the set of elective surgeries that are to be loaded ($|N| = n$). Each surgery i ($i = 1, \ldots, n$) has an expected duration μ_i (includes set-up time), and duration standard deviation σ_i. N_s is the set of surgeries performed by specialty s, and $N = \bigcup_s N_s$. In a given solution, N_{skt} ($k \in K_{st}$) indicates the set of surgeries assigned by specialty s in operating room k, on day t. We do not consider emergency surgeries, or patients that must be planned on certain OR-days for a medical reason. These are planned *on-line*, while robust surgery loading is an *off-line* problem. Therefore, in our experiments, we have decreased the OR-day capacity c_{kt} to correct for these omitted emergency surgeries.

Without loss of generality, we assume that the elective surgeries N originate from an initial "base" solution, which specifies a set of assigned surgeries N_{skt} ($k \in K_{st}$) for each OR-day (s, k, t). The base solution serves as a reference point: it represents the OR program that Erasmus MC would use, if it would not have any of the techniques proposed in this paper at its disposal. The base solution is found using the First Fit dispatching rule, and some basic statistical data. In Section 3.1 we will explain this method, which is in fact the same as the one currently used by Erasmus MC for planning surgeries for regular (non-urgent) patients that are on the waiting list. The reason that we use the surgeries from the base solution is that our loading methods will plan the same surgeries as the medical specialists would do in practice. In other words, no surgeries are cancelled, and no surgeries are added. By reloading and rescheduling surgeries, we are able to free capacity. The question whether this freed capacity should be used to plan additional surgeries, whether operating rooms should be closed earlier, or whether less personnel is required, is left to the board and the OR-manager.

The amount of planned slack on each OR-day to prevent overtime is based on the expected variance of the durations of the surgeries planned on that OR-day. The expected duration of the planned surgeries on OR-day (s, k, t), $k \in K_{st}$ is

$$\mu_{skt} = \sum_{i \in N_{skt}} \mu_i. \tag{1}$$

The variance of the planned surgeries on OR-day (s, k, t), $k \in K_{st}$ is

$$\sigma^2_{skt} = \sum_{i \in N_{skt}} \sigma^2_i. \tag{2}$$

We assume that the surgery durations are mutually independent. The planned slack size δ_{skt} on each OR-day (s, k, t), $k \in K_{st}$, is calculated as follows:

$$\delta_{skt} = \beta \cdot \sqrt{\sum_{i \in N_{skt}} \sigma^2_i}, \tag{3}$$

in which β ($\beta \geq 0$) is a parameter that influences the probability that the surgeries will complete on time, i.e., so overtime will not occur. In accordance with the central limit theorem in statistics (see e.g. Kallenberg, 1997) we assume that the sum of the durations of the planned surgeries on an OR-day is normally distributed with mean μ_{skt} and standard deviation σ_{skt}. As a result, if for example $\beta = 0.5$, the surgeries will finish on time with a probability of 69.15%. Our approach can easily be adapted for other distributions, in which case the percentage will likely be different. In the literature often a lognormal distribution is chosen for the surgery duration (Strum et al., 2000 and Zhou and Dexter, 1998). However, the distribution of the sum of the surgery durations must then be approximated, since there is no known exact result for such a distribution. The value of β is typically chosen by management, since a higher β will lead to a lower OR utilization on the one hand, but may lead to less overtime, less costs, and higher quality of labor and health care on the other hand.

Given a surgery allocation N_{skt}, the OR-day capacity constraint is as follows:

$$\sum_{i \in N_{skt}} \mu_i + \delta_{skt} \leqslant c_{kt} + O_{skt} \quad (\forall s, k \in K_{st}, t), \tag{4}$$

in which O_{skt} ($O_{skt} \geq 0$) is the overtime on OR-day (s, k, t). Observe we consider planned slack after regular time as overtime.

We distinguish between six degrees of freedom for the allocation of surgeries to OR-days, which we shall refer to as scenarios:

1 Surgeries must be planned on the day t that they were assigned to in the base solution, within the OR-days assigned to their specialty s (K_{st}).
2 Surgeries must be planned on the day t that they were assigned to in the base solution, within the OR-days assigned to all the specialties supported by unit u ($\bigcup_{s \in O_u} K_{st}$).
3 Surgeries must be planned on the day t that they were assigned to in the base solution ($\bigcup_s K_{st}$).
4 Surgeries must be planned on the OR-days assigned to their specialty s ($\bigcup_t K_{st}$).
5 Surgeries must be planned on the OR-days assigned to all the specialties supported by unit u ($\bigcup_{t, s \in O_u} K_{st}$).
6 A surgery may be planned on any OR-day ($\bigcup_{t,s} K_{st}$).

Given a scenario, the problem of assigning surgeries to operating rooms decomposes. For example in scenario 4, the problem decomposes into a subproblem for each specialty. Scenarios 1–3 (as compared to 4–6) leave the surgery on the day it was in the base solution. Furthermore, the higher the scenario number, the more allocation freedom there is, and the more flexible the OR-personnel must be. More allocation freedom may give practical personnel and surgeon problems, as follows. In scenarios 1 and 4, OR-team and surgeon restrictions are automatically satisfied. In scenarios 2 and 5, OR-team restrictions are satisfied (since they can be exchanged freely between OR-days of a unit), but solutions may require more surgeons than available. In scenarios 3 and 6, both OR-team and surgeon restrictions may be violated. Since surgeries may be planned freely over all OR-days, this may result in solutions where more than K_{st} surgeries of specialty s are planned in parallel on day t. The sequence in which the surgeries are performed on the OR-day is not of interest in this paper, since it does not affect the required OR-time. However, surgeon or OR-team restriction violations may be solved by changing the sequence in which the surgeries are performed, and/or by exchanging entire OR-day surgery assignments between different days. This is a subject for further research.

To evaluate a given solution, we use the following three ranked optimization criteria (in order of importance):

1 To minimize the total overtime: $\sum_{s,k,t} O_{skt}$.
2 To maximize the total number of free OR-days: $\sum_{\{s,k,t \mid N_{skt} = \emptyset\}} 1$.
3 To maximize the total free capacity: $\sum_{s,k,t} \max\left(0, c_{kt} - \sum_{i \in N_{skt}} \mu_i - \delta_{skt}\right)$.

If we compare two solutions, and both solutions score equally on criterion 1, then we use criterion 2 to compare, etc.

Constructive methods

Base solution determination using First Fit In this section, we describe the algorithm that we use to find a base solution. It is the loading procedure currently used by each specialty at Erasmus MC. There it is performed on Friday in week "x," and applies to 5 working days (Monday–Friday) of week "$x + 2$."

Each specialty s allocates a number of surgeries (N_{skt}) to each OR-day that has been assigned to them by management as follows. Each specialty has a waiting list of patients, which is sufficiently long to fill all available OR-days. The First Fit dispatching rule basically assigns the surgery from the top of the waiting list into the first OR in which it fits. A set of surgeries N_{skt} fits on an OR-day if the total expected surgery time plus the planned slack does not exceed the capacity c_{kt}. For the base solution (and currently in Erasmus MC), the planned slack is calculated differently than in Eq. (3). Instead of using the standard deviation σ_i of a surgery type i, each specialty s uses a standard deviation σ_s that is the same for all the surgeries performed by this specialty. As a result, the planned slack is computed as follows:

$$\beta \cdot \sqrt{\sum_{i \in N_{skt}} \sigma_s^2} = \beta \cdot \sqrt{|N_{skt}|} \cdot \sigma_s \quad (\forall s, k \in K_{st}, t),$$

in which β, just like in (3), is a parameter that determines the probability that the surgeries will complete on time, i.e., so overtime will not occur. The capacity constraint thus becomes

$$\sum_{i \in N_{skt}} \mu_i + \beta \cdot \sqrt{|N_{skt}|} \cdot \sigma_s \leqslant c_{kt} \quad (\forall s, k \in K_{st}, t). \tag{5}$$

The list of surgeries N_{skt} for an OR-day (s, k, t) is complete if no patients can be selected from the waiting list and added to N_{skt} without violating constraint (5), and enough slack can be planned during the OR-day. The algorithm requires $O(n)$ time.

Our goal is to determine whether or not using more statistical information (i.e., the standard deviation σ_i per surgery type) and using more advanced planning techniques lead to an improved OR utilization.

LPT-based dispatching The longest processing time (LPT) dispatching rule is a list-scheduling (LS) variant in which the list of candidate items is first sorted in non-increasing order of size (for a comparison of LS and LPT, see Brehob et al., 2000). LS is commonly used for on-line parallel machine scheduling problems (we refer to the survey paper by Chen et al., 1998). LS is an on-line algorithm, because it permanently assigns the current job to a machine, before it is aware of the next job. Because it knows the job list in advance LPT is an off-line algorithm, which has a worst-case ratio of 4/3–1/3m (Graham, 1969), where m is the number of machines. Without sorting the jobs, the worst-case ratio is 2–1/m (Graham, 1966). "First Fit Decreasing" is an LPT-variant which assigns the item to the machine with the most time assigned so far, and the first one that fits.

We describe the LPT-based algorithm for scenario 4. The procedure is the same for the other scenarios, however with different ORs and surgeries. In scenario 4 the surgeries must be planned on the OR-days assigned to their specialty s, i.e., on all OR-days in: $\bigcup_t K_{st}$. This means we have to perform the algorithm precisely S times, since we have to solve an independent loading problem for each specialty s. As argued in Section 2, without loss of generality, we assume that the surgeries originate from the base solution (see Section 2). The list of surgeries is thus formed by the surgeries on the OR-days in $\bigcup_t K_{st}$ in the base solution. We first sort these surgeries in non-increasing order of their expected duration. Hence, we essentially perform a longest expected processing time (LEPT) rule. We then assign each surgery in this sequence to the first OR in which the surgery fits, i.e., for which constraint (4) still holds, without using overtime. If there is no OR in which the surgery fits without yielding overtime, it is assigned to the OR in which the additional overtime is as small as possible. The algorithm running time is $O(n \log n)$.

Sampling procedures Sampling procedures have been successful as randomized constructive heuristics for resource constrained project scheduling problems Hartmann and Kolisch (2000). Just as list-scheduling, sampling procedures (generally) use a priority rule. The difference with LS is that sampling procedures use multiple passes. Different solutions are obtained by biasing the selection of the priority rule through a random device, and the best solution is kept after a number of passes. So, in addition to the surgery scheduling priority, a selection probability P_i is computed. Each pass is referred to as a sample. We consider three sampling methods: random sampling, biased-random sampling, and regret-based random sampling. The differences between these methods are only in the way the surgery drawing probabilities are calculated.

General procedure in each sample The procedure in each sample is as follows. The scenario determines what surgeries and OR-days are involved. We sort the list of surgeries in non-increasing order on their expected duration. Each iteration considers at most Z surgeries for dispatching from the beginning of the list (Z is a non-zero integer). If $Z = 1$, the algorithm is precisely the same as the LPT-based algorithm of Section 3.2, since only one surgery is considered in each iteration. For higher Z, the algorithm increasingly abandons the LPT-idea, and scours neighborhood solutions. The following procedure is carried out for each of the (at most) Z surgeries. If there are ORs in which surgery i fits without generating overtime, a scheduling priority v_{iskt} is calculated (we demonstrate how this is done later), and a most suitable OR is selected. If there are no ORs available in which surgery i can be planned without generating overtime, we immediately plan surgery i into the OR in which the generated overtime as in constraint (4) is as small as possible. We then replenish the list of Z surgeries. After a priority has been calculated and a most suitable OR has been selected for each of the (at most) Z surgeries, a drawing probability P_i is calculated for each surgery i. Finally, a surgery is drawn and planned into the most suitable OR.

Surgery scheduling priority calculation We calculate a surgery scheduling priority v_{iskt} as follows. The idea is that we try to assign surgeries in such a way that the total planned slack is minimized. When a surgery is planned, it also introduces additional planned slack. If a surgery i is assigned to an *empty* OR, the additional planned slack is $\beta \cdot \sigma_i$, see Eq. (3). If a surgery i is assigned to a *filled* OR k, the additional planned slack is generally smaller. Suppose N_{skt} are the surgeries already assigned to OR k. Then the additional slack Δ_{iskt} generated by adding surgery i is

$$\Delta_{iskt} = \beta \cdot \sqrt{\sum_{j \in N_{skt}} \sigma_j^2 + \sigma_i^2} - \beta \cdot \sqrt{\sum_{j \in N_{skt}} \sigma_j^2}.$$

The profit of not planning surgery i into an empty OR, but into a filled OR k is thus

$$\Omega_{iskt} = \beta \cdot \sigma_i - \Delta_{iskt}.$$

If (s, k, t) is a filled OR-day, then $\Omega_{iskt} > 0$. The priority of surgery i is

$$v_{iskt} = \max_k \Omega_{iskt}. \tag{6}$$

Of course Eq. (6) only involves the candidate OR-days, which depends on the scenario. The most suitable OR-day for surgery i is the k^* that maximizes $\max_k \Omega_{iskt}$ in (6).

We minimize the total planned slack by exploiting the so-called portfolio effect. In the financial literature this term is used to indicate that portfolio risk falls with increasing diversity, as measured by the absence of correlation (covariance) between portfolio components (Markowitz, 1991). Since surgery durations are not correlated, we can minimize the total planned slack by clustering surgeries with similar variability on the same OR-day. To illustrate this, we give an example. Consider two OR-days, both with two assigned surgeries, one surgery with $(\mu, \sigma) = (100, 10)$ and one surgery with $(\mu, \sigma) = (100, 50)$, see Figure 1. We compare this situation (the left-hand side of Fig. 1) with the situation in which the surgeries with the same variability (σ) are clustered. In the first situation, the standard deviation of the total duration of the surgeries is the same for both OR-days: $\sqrt{(50^2 + 10^2)} = 51.0$. The total planned slack is thus $102.0 \cdot \beta$. Similarly, in the second situation the total planned slack is: $(\sqrt{(50^2 + 50^2)} + \sqrt{(10^2 + 10^2)}) \cdot \beta = 84.9 \cdot \beta$. This means a reduction in the total planned slack time of $17.1 \cdot \beta$, and thus an equal increase in the available capacity. This portfolio profit will increase when the variability of the concerned surgeries is higher.

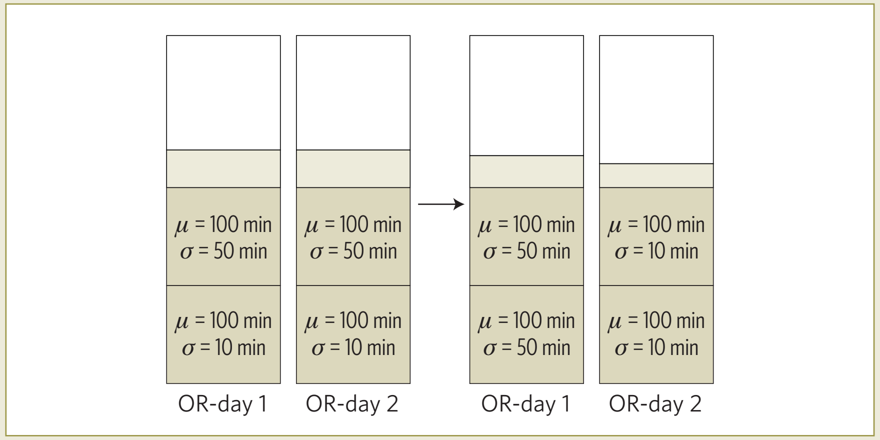


Figure 1 *Example of planned slack reduction as a result of the portfolio effect.*

We shall now describe the differences between the sampling methods.

Random sampling (S1) S1 uses the aforementioned procedure to determine the most suitable OR, but discards the calculated surgery priority v_{iskt}. S1 gives each candidate surgery an equal probability to be loaded. Hence when there are q remaining candidate surgeries, each surgery i has a selection probability: $P_i = 1/q$. The drawn surgery is assigned to the most suitable OR.

Biased-random sampling (S2) Biased-random sampling (Cooper, 1976) sorts the Z surgeries in order of non-increasing priority value. The ith surgery in the list then has a selection probability: $P_i = C \cdot \gamma^i$, in which C is a normalization constant $(C = 1/\sum_j \gamma^j)$, which ensures that the sum of surgery probabilities is 1, and in which γ $(0 < \gamma \leq 1)$ is a "bias factor." Smaller values for γ result in a stronger dominance of the deterministic surgery priority values, and $\gamma = 1$ corresponds to random sampling.

Regret-based random sampling (S3) Biased-random sampling does not take into account the relative difference in priority values of activities in the list. To deal with this disadvantage, Drexl (1991) proposed regret-based random sampling (S3). In S3, the selection probability of a surgery is based on its "regret." The regret w_{iskt} of a surgery i is the difference between its priority v_{iskt}, and the worst of all surgery priorities:

$$w_{iskt} = v_{iskt} - \min_j \{v_{jskt}\}.$$

The probability that a surgery i is selected is now

$$P_i = C \cdot (1 + w_{iskt})^{\alpha}, \tag{7}$$

in which C is a normalization constant that ensures that the sum of surgery probabilities is 1:

$$C = 1 \Big/ \sum_j (1 + w_{jskt})^{\alpha}.$$

Furthermore, α ($\alpha \geq 0$) is the bias factor, i.e., a parameter that measures the bias. $\alpha = 0$ corresponds to random sampling, $\alpha = \infty$ corresponds to deterministic dispatching. Observe that in (7), the regret factor is incremented by 1, to ensure that all surgeries have non-zero selection probability. As a result, any surgery can be selected. Kolisch and Drexl (1996) show that deterministic sampling ($\alpha = \infty$) works best when the sample size is 1. When the number of passes is increased, the sampling should be more biased, so α must be decreased. We determine the best combination of α and the number of samples empirically (see Section 5.3.1). Each sample of S1 and S3 requires O(n) time. Since S2 performs a sorting, it requires O($n \log n$) time.

Local search methods

In addition to the constructive heuristics, we tested some local search heuristics, which focus on solution improvement by swapping two different surgeries between OR-days (a two-exchange), or by moving one surgery to another OR (a one-exchange). In Sections Sections 4.1 and 4.2, we discuss two local search methods, which both use the aforementioned one- and two-exchanges.

Random exchange method The random exchange method (REM) is a greedy local search procedure that uses the following procedure:

1 With probability P^{REM} a one-exchange is evaluated, for which we select a random surgery and a random suitable OR-day.
2 With probability 1-P^{REM} a two-exchange is evaluated, for which we select two random surgeries from two different OR-days.

If an exchange yields an improved solution, it is accepted. The method stops if no improvement has been found during η seconds.

Simulated annealing Simulated annealing (SA) is a generalization of a Monte Carlo method for examining the equations of state and frozen states of n-body systems (Metropolis et al., 1953). Kirkpatrick et al. (1983) propose it as an optimization technique for combinatorial problems. Ever since, SA has been widely and successfully used as a local search method in many applications. For an extensive description of the simulated annealing algorithm, we refer to Aarts and Korst (1989).

Our SA approach uses a proportional cooling scheme: $\tau^{\mathrm{new}} = \tau^{\mathrm{old}} \cdot \theta$, where $\theta < 1$, and τ is the control parameter, which by analogy with the original application is also known as “cooling parameter,” or “temperature.” The initial control parameter is τ^{begin}, and the final control parameter is τ^{end}. The number of one- or two-exchanges evaluated before the control parameter is decreased (also known as the length of the Markov chain) is λ. We do a one-exchange with probability π, and a two-exchange with probability $1-\pi$. Every exchange that yields an improvement of the solution (according to the three optimization criteria in Section 2) is accepted. If an exchange yields a worse solution, let Y be the increase of the solution value of optimization criterion 1, or, if this is 0, the increase of the solution value of optimization criterion 3 (we do not consider neighbor solutions in which the number of free days decreases). We accept the exchange with probability: $e^{\frac{-Y}{\tau}}$. The cooling scheme parameters are such that the probability of accepting a worse solution is almost 1 at the start of the cooling scheme and is almost 0 at the end.

Test results

Test data As argued in Section 2, we test our methods on an instance that is formed by a base solution, which we find using the “First Fit” based algorithm described in Section 3.1. We base the test instance on the operating theatre department of Erasmus MC, which consists of 16 ORs (for non-urgent clinical patients), and of 11 specialties. The instance spans a full year of 52 weeks, each week consisting of 5 working days of 7.5 hours. Hence the number of OR-days per year is $52 \cdot 5 \cdot 16 = 4160$, and the total surgery capacity is $4160 \cdot 7.5 = 31{,}200$ hours. Table 1 gives an overview of the data used for each specialty.

Table 1 shows the OR allocation per surgery per week. For example on Monday, General surgery is assigned to OR 1, OR 2 and OR 3 (3 OR-days), Gynecological surgery to OR 4, etc. Table 1 also gives the standard deviation σ_s for all the surgeries performed by a specialty s, and the unit number to which the specialty belongs. For example Gynecological and Urological surgery are both supported by the OR-personnel of unit 4.

Since more than a decade the Erasmus MC has collected data concerning all processes in the organization, which is stored in a large data warehouse. This

Specialty	# OR-days per day					σs	Unit
	Monday	Tuesday	Wednesday	Thursday	Friday		
General surgery	3	3	3	3	3	105.8	2
Gynecological surgery	1	1	1	1	1	53.7	4
Oral surgery	1	1	1	1	1	71.2	3
ENT surgery	2	2	2	1	2	108.3	1
Pulmonary surgery	0	0	0	1	0	32.7	1
Neurosurgery	2	2	2	2	2	144.5	1
Traumatology	1	1	0	1	1	59.9	3
Eye surgery	1	1	1	1	1	31.8	1
Orthopedic surgery	1	1	2	1	2	73.5	3
Plastic surgery	2	2	2	2	1	110.8	3
Urological surgery	2	2	2	2	2	104.4	4
Total	16	16	16	16	16		

Table 1 *Instance data for the 11 specialties*

data warehouse contains medical information, but also time-related data like the start and completion time of a surgery. Based on this data, we generate our artificial test instances. We generate an artificial instance because we do not consider the non-elective/emergency surgeries.

For each specialty, the surgeries are generated as follows. We made a statistical analysis of 10 years historical data of all elective inpatient surgeries in Erasmus MC. For each specialty, we clustered the surgeries into at least 4, and up to 8 surgery categories. Each surgery category has an expected surgery duration μ_i and a duration standard deviation σ_i. The surgery category data is not given here, but is available from the authors upon request. To generate a surgery, a category is drawn with a probability that is proportional to the share of this category in all surgeries of the specialty. Consequently, the surgery characteristics are the same as the category characteristics. We use the procedure of Section 3.1 to generate a base solution, along with the aforementioned procedure to generate surgeries. If a generated surgery does not fit into the available OR-days, the OR-days are considered "filled."

For the parameter β of constraint (3), which determines the probability that the surgeries will complete on time, we choose a value of 0.5. Assuming that the sum of the surgeries that are assigned to one day is normally distributed, this gives a probability of 69.15% that the surgeries will finish before the end of the planned slack. This is actually the probability that the board of Erasmus MC has chosen. The total number of surgeries in the test instance is 11,380. Of course, when β is decreased, more surgeries can be planned. For example, if $\beta = 0$ (50% overtime probability), the number of surgeries is 13,470. On the other hand, if $\beta = 3.9$ (0% overtime probability), the number of surgeries is 2158. Table 2 summarizes the characteristics of the base solution.

Number of surgeries	11,380
Total surgery time (hours)	23,177.45
Total OR capacity (hours)	31,200.00
Total overtime (hours)	0.0
Total unused/free capacity (hours)	5240.29
Total planned slack (hours)	2782.25

Table 2 *Summary of the base solution characteristics*

The total planned slack is computed using formula (3). Annual planned OR utilization rate (total surgery time plus slack/total capacity) is 83.20%.

Test approach All algorithms are implemented using the Borland Delphi programming language, and tested on a Pentium III–1600 desktop PC. We base all our experiments on the base instance described in Section 5.1. Depending on the scenario, an algorithm is applied to the OR-days and corresponding surgeries that belong to a specialty (scenarios 1 and 4), or a unit (scenarios 2 and 5), or all OR-days (scenarios 3 and 6), of a day (scenarios 1–3) or of a week (scenarios 4–6). So, for example in scenario 1, an algorithm must be applied precisely #weeks · #days · #specialties = 52 · 5 · 11 = 2860 times.

We first perform experiments in which we determine the best parameter settings for all the methods (Section 5.3). Given the best parameter settings for all methods, we then first compare the constructive methods (Section 5.4.1). We use the base solution and the best solution of all constructive methods and each scenario as starting solutions for the local search methods (Section 5.4.2). Finally, Section 5.5 gives the results of a Monte Carlo simulation of a solution found by the best algorithm.

Algorithm parameter determination

5.3.1. Sampling methods S1, S2, S3 For S1 we must set the parameter Z, the number of surgeries considered in each iteration, and the number of samples. We performed an experiment in which we test all values of $Z \in \{1, \dots, 10, 20, 50\}$ in each scenario, and 10 samples. We compute the average objective values over all scenarios. We find the best average solution performance (over all three optimization criteria) is achieved for $Z = 4$. To determine the sample size, we make a trade-off between computation time and solution improvement. In an experiment in which we perform 1500 passes, we evaluate for each sample size in $\{50, 250, 500, \dots, 1500\}$ the number of free days (optimization criterion 2), averaged over all scenarios (see Table 3). The total average overtime (criterion 1) is 0.0 in all samples. Table 3 shows that the solution improvements after 500 samples are marginal, so this is the value we choose for S1.

For S2, we must set the parameter Z and the bias factor γ. We performed an experiment in which we test all combinations $Z \in \{1, \dots, 10, 20, 50\}$ and $\gamma \in \{0.1, 0.2, \dots, 0.9\}$ in each scenario, and 10 samples. We compute the average objective values over all scenarios. We find the best average solution performance for $Z = 6$ and $\gamma = 0.5$. To determine the number of samples, we performed the same experiment as for S1. This gave very similar results, so we choose 500 samples for S2.

The parameters that we must set for S3 are: α (bias factor), the number of samples, and Z. To determine these, we perform experiments with S3 in which we test all combinations of: $\alpha \in \{1, \dots, 5, 10, 50, 100, 1000\}$, 1500 samples, and $Z \in \{1, \dots, 10, 20, 50\}$ for each scenario and the base instance. To determine the best values for α and Z, we compare the average solution criteria values after 1500 samples for each (α, Z) combination and in each scenario. The best solution performance is found for $(\alpha, Z) = (10, 9)$. Table 4 gives the results for S3 and the bias factor. Each cell is the average solution value (over all scenarios) for the optimization criterion of the row, and the bias factor of the column.

To determine the number of samples, we perform an experiment in which we vary the number of samples just as for S1 and S2, and with $(\alpha, Z) = (10, 9)$. We found that after 500 samples, the solution values only marginally improve, so this is the number of samples we choose for S3.

	Number of samples						
	50	250	500	750	1000	1250	1500
#Free days	336.0	341.3	343.2	344.7	345.0	345.5	345.8
Exec. time (seconds)	49.2	246.0	492.0	737.9	983.9	1229.9	1426.7

Table 3 *Computational results for S1*

Optimization criteria (avg.)	Bias factor (a)									
	0	1	2	3	4	5	10	50	100	1000
Overtime (hours)	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
# Free days	341.9	344.2	344.3	344.7	344.4	344.4	344.5	343.6	343.2	343.0
Free cap. (hours)	5370.8	5371.9	5371.9	5372.0	5372.0	5372.1	5372.4	5371.2	5371.7	5372.3

Table 4 *Average scenario objective values for S3 and various α*

Random exchange method For REM, we must choose the parameter P^{REM}, which determines the probability that an exchange is a one-exchange, and the stop-criterion, which is the number of seconds that no improvement was found (η). The best solution performance was found for $P^{\mathrm{REM}} = 0.10$ and $\eta = 2$. To set η, we made a trade-off between solution performance and computation time, both of which increase with η. For $\eta > 2$, the solution performance improvement is only marginal.

Simulated annealing For SA, we determine the best values for the length of the Markov chain (λ), the parameters of the proportional cooling scheme ($\tau^{\mathrm{begin}}, \tau^{\mathrm{end}}, \theta$) and the fraction of one-exchanges (π). For λ, we use the guideline suggested by Aarts and Korst (1989), the number of neighbor solutions. In our case, the number of neighbor solutions is

$$\lambda = \pi \cdot \varepsilon + (1 - \pi) \cdot \binom{\varepsilon}{2},$$

in which the first term is the number of one-exchanges, the second term is the number of two-exchanges, and ε is the number of surgeries considered for an exchange.

It is generally advised that the cooling schedule is such that the initial acceptance ratio (the number of accepted exchanges/the number of suggested exchanges) is close to 1 (see e.g., Van Laarhoven and Aarts, 1987). To determine the initial cooling parameter (τ^{begin}), we use the method suggested by Kirkpatrick et al. (1983). We choose an initial value 1 for τ^{begin} and perform λ exchanges. If the acceptance ratio is below 0.8, we multiply τ^{begin} by 2, and repeat the procedure until the acceptance ratio is above 0.8. The average best τ^{begin} for the six scenarios we found is 256. Empirically, we also determined $\pi = 20\%$, $\tau^{\mathrm{end}} = 0.001$, and $\theta = 0.995$.

Algorithm comparison

Constructive methods In Table 5 we compare the performance of the constructive methods, using the algorithm parameter settings discussed in Section 5.3. The last column gives the computation time for solving the entire base solution, averaged over all six scenarios. The other columns give the average solution performance for the three solution criteria.

Table 5 clearly shows that many OR-days can be freed by the constructive methods. This is especially achieved by the scenarios in which surgeries of different specialties are allowed to be combined into OR-days. Specialties that generally have long surgeries (such as neurosurgery) initially have a poor OR-utilization, because they lack short surgeries to fill their OR-days better. In scenarios in which surgeries of different specialties are allowed to be combined into OR-days, short surgeries from other specialties can complement these long surgeries, thus leading to free OR-days. S3 gives the best solution performance, albeit that its computation time is significantly higher. However, recall that the base solution spans a whole year. In a practical setting, the planning horizon is generally only a week. On average, S3 frees 6.7 OR-days every week, i.e., 8.4% of all OR-days.

From Table 6, in which we compare the objective values of S3 in each scenario, the profit of allowing more allocation freedom can be read. Compare, for example, scenarios 1 and 4. The fact that we allow surgeries to be reassigned to OR-days of the specialty on another day makes that an additional 257 OR-days can be freed.

	Avg. overtime (hours)	Avg. #free days	Avg. free capacity (hours)	Avg. comp. time (seconds)
Base solution	*0.0*	*0.0*	*5240.3*	*0.0*
LPT	0.5	316.5	5359.7	0.0
S1	0.0	343.2	5373.4	492.0
S2	0.0	344.5	5376.5	632.8
S3	0.0	348.3	5372.9	2226.7

Table 5 *Average solution performance for the constructive algorithms over all scenarios*

	Base sol.	Scen. 1	Scen. 2	Scen. 3	Scen. 4	Scen. 5	Scen. 6
Overtime (hours)	0.0	0.0	0.0	0.0	0.0	0.0	0.0
#Free days	0.0	5	135	552	262	508	628
Free capacity (hours)	5240.3	5258.9	5326.4	5445.3	5343.1	5409.9	5453.9

Table 6 *Objective values of S3 in each scenario*

Local search methods Table 7 compares the solution performance (avg. objective value over all scenarios) of the local search methods REM and SA, which start from the base solution.

SA clearly outperforms REM, and slightly outperforms S3, albeit with significantly more computation time. This might not be an issue in practice, if the OR-manager has sufficient time. We performed an additional experiment in which we executed REM after S3, to determine to what extent the S3 solution can be improved further. We found that on average only 3.2 additional OR-days and 14.2 hours of capacity can be freed. The combination S3+REM thus performs similar to SA, with much less computation time.

Simulation In this section, we give the results of a Monte Carlo simulation of a solution found by the best algorithm. For each surgery, we draw the outcome of its duration, by sampling a value from its distribution function. Per OR-day, we then add up the surgery durations of all the planned surgeries, and compute the overtime or unused capacity. Table 8 gives the results of the Monte Carlo simulation of the base solution, and the solution for scenario 3 found by S3.

The third column in Table 8 gives the probability that any OR-day will result in overtime. The last column is the average OR-utilization of the *used* ORs, i.e.

	Avg. overtime (hours)	Avg. #free days	Avg. free capacity (hours)	Avg. comp. time (seconds)
Base solution	*0.0*	*0.0*	*5240.3*	*0.0*
REM (base sol.)	*0.0*	*336.0*	*5380.1*	*1734.0*
SA (base sol.)	*0.0*	*352.5*	*5388.2*	*6315.6*

Table 7 *Performance comparison of the local search methods*

	Total overtime (hours)	Avg. overtime per OR-day (hours)	OR overtime probability	Total free capacity (hours)	#Free OR-days	Avg. OR-utilization (%)
Base solution	330.1	0.08	0.094	8378.2	0	73.8
S3	886.0	0.21	0.255	8934.1	552	74.7

Table 8 *Monte Carlo simulation results for scenario 3*

the average utilization of all OR-days with assigned surgeries. Observe that in the S3 solution, the probability that overtime occurs is higher, but still smaller than $1-0.69=0.31$, i.e. the selected probability that no overtime will occur (see Section 5.1). The reason that the actual probability is lower than $1-\beta$ is that there is still some unused capacity after the surgeries. We conclude that we can free 552 OR-days in scenario 3 (see Table 6 for the other scenarios) without exceeding the probability of overtime chosen by management.

Conclusions and further research

We have proposed constructive and local search heuristics for the robust surgery loading problem. The heuristics aim to minimize the total planned slack. Due to the portfolio effect, this automatically leads to freeing OR-days and -capacity. We have shown that a given plan that was made by specialists (the "base solution") can be improved significantly by these methods.

The best constructive approach is regret-based random sampling (S3). The random exchange method (REM) can be used to further improve its solutions. Simulated annealing has a similar performance as S3, but uses much more computation time. All discussed approaches show that many OR-days and a lot of OR-capacity can be freed, without canceling surgeries, and without introducing more overtime than is allowed by the OR-management.

In the optimized solutions, we observed that as a result of the portfolio effect, surgeries with similar duration variability are often clustered on the same OR-day. Computational results show that smart algorithms and the exploitation of the portfolio effect frees much OR-capacity, within the limits of the accepted risk of overtime. This is only possible if extensive statistical data on surgery durations is available. This stresses the importance for hospitals to register surgery durations (see also Ozkarahan, 2000), and OR-planners to recognize that surgeries have different duration variability, and account for this in OR-loading.

The clustering of surgeries with similar duration variability often leads to OR-days on which several surgeries of the same type are performed. This may have an additional advantage that surgeons can reduce surgery time because of the repetitive nature of their work on such OR-days. The clustering of surgeries can also be realized by using a so-called "master surgical scheduling" (MSS) approach (Blake and Donald, 2002 and Beliën and Demeulemeester, 2007), in which surgeries or surgery types are clustered in a (cyclic) schedule. In a forthcoming paper we shall present an MSS approach that accounts for both OR-capacity restrictions, and capacity restrictions imposed by subsequent departments, like ICU and wards (Van Oostrum et al., 2006).

In scenarios that allow much surgery allocation freedom, we ignored the surgeon and OR-personnel capacity restrictions. By optimizing the sequence in which the surgeries are performed on each OR-day, and by exchanging entire OR-day surgery assignments between days the extent to which these restrictions are violated can be decreased. This is also a subject for further research.

References

1 Aarts and Korst, 1989 E.H.L. Aarts and J.H.M. Korst, Simulated Annealing and Boltzmann Machines, John Wiley & Sons, New York (1989).

2 Beliën and Demeulemeester, 2007 J. Beliën and E. Demeulemeester, Building cyclic master schedules with leveled resulting bed occupancy, *European Journal of Operational Research* 176 (2007) (2), pp. 1185–1204.

3 Blake and Donald, 2002 J. Blake and J. Donald, Mount Sinai hospital uses integer programming to allocate operating room time, *Interfaces* 32 (2002) (2), pp. 63–73. Full Text via CrossRef | View Record in Scopus | Cited By in Scopus

4 Brehob et al., 2000 M. Brehob, E. Torng and P. Uthaisombut, Applying extra-resource analysis to load balancing, *Journal of Scheduling* 3 (2000), pp. 273–288. Full Text via CrossRef | View Record in Scopus | Cited By in Scopus

5 Chen et al., 1998 B. Chen, C.N. Potts and G.J. Woeginger, A review of machine scheduling: Complexity, algorithms and approximability. In: D.Z. Du and P.M. Pardalos, Editors, *Handbook of Combinatorial Optimization* vol. 3, Kluwer Academic Publishers (1998), pp. 21–169.

6 Cooper, 1976 D.F. Cooper, Heuristics for scheduling resource-constrained projects: An experimental investigation, *Management Science* 22 (1976) (11), pp. 1186–1194. View Record in Scopus | Cited By in Scopus

7 De Kreuk et al., 2004 A.C.C. De Kreuk, E.M.M. Winands and J.M.H. Vissers, Master scheduling of medical specialists. In: J.M.H. Vissers and R. Beech, Editors, *Health Operations Management – Patient Flow Logistics in Health Care*, Routledge, London (2004).

8 Dell'Olmo and Speranza, 1999 P. Dell'Olmo and M.G. Speranza, Approximation algorithms for partitioning small items in unequal bins to minimize the total size, *Discrete Applied Mathematics* 94 (1999), pp. 181–191.

9 Dexter, 2001 F. Dexter, Cost implications of various operating room scheduling strategies, *American Society of Anesthesiologist's Clinical Update Program* 52 (2001) (262), pp. 1–6.

10 Dexter et al., 1999 F. Dexter, A. Macario and RD. Traub, Which algorithm for scheduling add-on elective cases maximizes operating room utilization?, *Anesthesiology*

91 (1999), pp. 1491–1500. Full Text via CrossRef | View Record in Scopus | Cited By in Scopus

11 Drexl, 1991 A. Drexl, Scheduling of project networks by job assignment, *Management Science* 37 (1991), pp. 1590–1602. View Record in Scopus | Cited By in Scopus

12 Goldman et al., 1969 J. Goldman, H.A. Knappenberger and EW. Moore Jr., An application of OR scheduling policies, *Hospital Management* 107 (1969), pp. 40–51. View Record in Scopus | Cited By in Scopus

13 Graham, 1966 RL. Graham, Bounds for certain multiprocessing anomalies, *Bell System Technical Journal* 45 (1966), pp. 1563–1581.

14 Graham, 1969 R.L. Graham, Bounds on multiprocessing timing anomalies, *SIAM Journal on Applied Mathematics* 17 (1969), pp. 416–429. MathSciNet | Full Text via CrossRef

15 Guinet and Chaabane, 2003 A. Guinet and S. Chaabane, Operating theatre planning, *International Journal of Production Economics* 85 (2003) (1), pp. 69–81. SummaryPlus | Full Text + Links | PDF (231 K) | View Record in Scopus | Cited By in Scopus

16 Hartmann and Kolisch, 2000 S. Hartmann and R. Kolisch, Experimental evaluation of state-of-the-art heuristics for the resource-constrained scheduling problem, *European Journal of Operational Research* 127 (2000), pp. 394–407. SummaryPlus | Full Text + Links | PDF (137 K) | View Record in Scopus | Cited By in Scopus

17 Hopp and Spearman, 2000 W.J. Hopp and M.L. Spearman, Factory Physics – Foundations of Manufacturing Management (second ed.), McGraw-Hill/Irwin, New York (2000).

18 Kallenberg, 1997 O. Kallenberg, Foundations of Modern Probability, Springer-Verlag, New York (1997).

19 Kirkpatrick et al., 1983 S. Kirkpatrick, C.D. Gerlatt and M.P. Vecchi, Optimization by simulated annealing, *Science* 220 (1983), pp. 671–680

20 Kolisch and Drexl, 1996 R. Kolisch and A. Drexl, Adaptive search for solving hard project scheduling problems, *Naval Research Logistics* 43 (1996), pp. 23–40.

21 Kuo et al., 2003 P.C. Kuo, R.A. Schroeder, M. Mahaffey and R.R. Bollinger, Optimization of operating room allocation using linear programming techniques, *Journal of American College of Surgeons* 197 (2003) (6), pp. 889–895.

22 Markowitz, 1991 H.M. Markowitz, Portfolio Selection: Efficient Diversification of Investments (second ed.), Blackwell, Cambridge, MA (1991).

23 Metropolis et al., 1953 N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller and E. Teller, Equation of state calculations by fast computing machines, *Journal of Chemical Physics* 21 (1953) (6), pp. 1087–1092.

24 Ozkarahan, 2000 I. Ozkarahan, Allocation of surgeries to operating rooms by goal programming, *Journal of Medical Systems* 24 (2000) (6), pp. 339–378.

25 Roth and Van Dierdonck, 2005 A.V. Roth and R. Van Dierdonck, Hospital resource planning, *Production and Operations Management* 4 (2005), pp. 2–29.

26 Strum et al., 2000 D.P. Strum, J.H. May and L.G. Vargas, Modeling the uncertainty of surgical procedure times: comparison of log-normal and normal models, *Anesthesiology* 92 (2000) (4), pp. 1160–1167.

27 TPG, 2004 TPG, 2004. Het kan echt: betere zorg voor minder geld. Eindrapportage TPG voor Sneller Beter – De logistiek in de zorg (in Dutch).

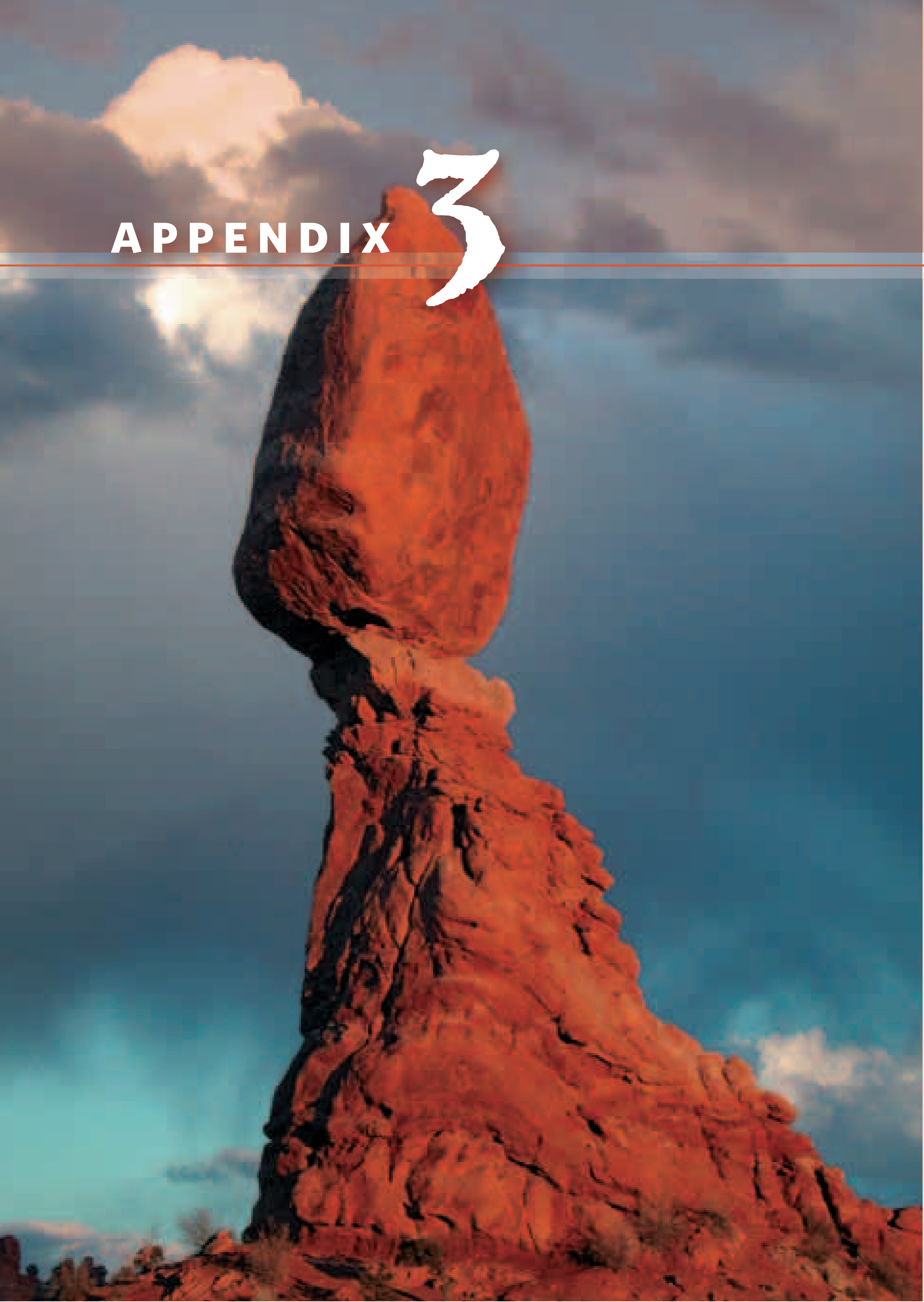
28 Van Laarhoven and Aarts, 1987 P.J.M. Van Laarhoven and E.H.L. Aarts, Simulated Annealing: Theory and Applications, Reidel, Dordrecht (1987).

29 Van Oostrum et al., 2006 Van Oostrum, J.M., Van Houdenhoven, M., Hurink, J.L., Hans, E.W., Wullink, G., Kazemier, G., 2006. A master surgical scheduling approach for cyclic scheduling in operating room departments. Working paper, Erasmus MC, Rotterdam.

30 Ye and Zhang, 2003 Ye, D., Zhang, G., 2003. On-line extensible bin packing with unequal bin sizes. Proceedings of the First Workshop on Approximation and Online Algorithms (WAOA 2003). Springer LNCS 2909, pp. 235–247.

31 Zhou and Dexter, 1998 J. Zhou and F. Dexter, 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* 89 (1998), pp. 1228–1232.

32 Zijm, 2000 W.H.M. Zijm, Towards Intelligent Manufacturing Planning and Control Systems, OR *Spectrum* 22 (2000), pp. 313–345.



A master surgical scheduling approach for cyclic scheduling in operating room departments

J.M. van Oostrum, M. Van Houdenhoven, J.L. Hurink, E.W. Hans, G. Wullink, G. Kazemier.

OR Spectrum, Published online: 21 September 2006

Introduction

Increasing costs of health care imply pressure on hospitals to make their organization more efficient. Recent studies show that operations research provides powerful techniques in this context (Carter 2002). One of the most expensive resources in a hospital is the operating room (OR) department. Since up to 70% of all hospital admissions involve a stay in an OR department (OECD 2005), optimal utilization of OR capacity is of paramount importance.

Operating room utilization is typically jeopardized by numerous factors and various players are active in OR planning, such as individual surgeons, ORmanagers, and anesthesiologists (Weissman 2005). All players have autonomy, and can have conflicting objectives with respect to productivity, quality of care, and quality of labor (Glouberman and Mintzberg 2001). As a result, OR planning is constantly under scrutiny and pressure of potentially competing objectives.

A further complicating factor of the OR planning is the stochastic nature of the process. There are many uncertainties, such as stochastic durations of surgical procedures, no-shows of patients, personnel availability, and emergency surgical procedures. In addition, because surgeons tend to plan their procedures independently from others, this results in peak demands at subsequent hospital resources such as intensive care units (ICU). As a result, unavailability of for example ICU bed capacity can result in cancelation of surgical procedures (McManus et al. 2003).

In this paper we consider the problem of scheduling elective procedures, which is an operational planning problem that concerns the assignment of elective procedures to ORs over the days of the week. Due to the aforementioned difficulties, the planning process is complex, time consuming, and often under a lot of pressure. However, a lot of elective procedures tend to be identical during consecutiveweeks in the year. In a regional hospital it is not uncommonthat this is for more than 80% of the total volume the case (Bakker and Zuurbier 2002). In manufacturing as well as in health care, repetitive production is common practice. In such environments a cyclic planning approach is often used (e.g., Tayur 2000; Schmidt et al. 2001; Millar and Kiragu 1998). This reduces planning efforts considerably, and leads to reduced demand fluctuations within the supply chain, and higher utilization rates.

We propose in this paper a model for a cyclic scheduling approach of elective surgical procedures. We refer to such a cyclic surgical schedule as a *master surgical schedule* (MSS). An MSS specifies for each "OR-day" (i.e. operating room on a day) of the planning cycle a list of recurring surgical procedure types that must be performed. We demonstrate that our approach is generic: it not only allows to level and control the workload of the involved surgical specialties, but also from succeeding departments such as ICUs and surgical wards. It optimizes OR utilization without increasing overtime and cancelations. Furthermore, our approach accounts for the stochastic nature of the surgical process, such as stochastic durations of surgical procedures.

The approach for generation of MSSs was tested with data from the Erasmus Medical Center in Rotterdam, The Netherlands, which is a large university hospital. Approximately 15,000 patients annually undergo surgery in the OR departments of Erasmus MC. Since 1994, Erasmus MC has collected their surgical data in a database of 180,000 surgical procedures. The hospital actively supported the research project and affirms the applicability of this study.

The remainder of the paper is structured as follows. Section 2 presents an overview of studies relatedto the problem of construction MSSs. Section 3 presents a base model that represents the problem of constructing MSSs. Section 4proposes a solution approach to solve the problem. In Sect. 5we evaluate the solution approach. Section 6draws conclusions from this research.

Related literature

There exist a strong interest in OR scheduling problems, resulting in a wide range of papers on this subject. These studies can be separated into short-term operating room scheduling (e.g., Gerhak et al. 1996; Sier et al. 1997; Ozkaraham 2000; Lamiri et al. 2005; Jebali et al. 2006) and mid-term planning and control (e.g., Guinet and Chaabane 2003; Ogulata and Erol 2003; Kim and Horowitz 2002). Studies about MSS are, however, scarce. Moreover, various definitions of a MSS are used. Blake and Donald (2002) construct MSSs that specify the number and type of operating rooms, the hours that ORs are available, and the specialty that has priority at an operating room. They use an integer programming formulation for the assignment of specialties to operating rooms. The objective function minimizes penalties related to the total undersupply of operating rooms to specialties. The authors implement a straightforward enumerative algorithm, which results in considerable improvements. Beliën and Demeulemeester (2005) use a nonlinear integer programming model to construct MSSs. The model assigns blocks of OR time to specialties in such a way, that the total expected bed shortage on the wards is minimized. After linearization of the model the authors examine and compare several heuristics to solve the resulting mixed integer program. They conclude that a simulated annealing approach yields the best results, but since this heuristic requires much computation time they propose a hybrid algorithm that combines simulated annealing with a quadratic programming model. This approach yields the best results concerning solution quality and computation times. Vissers et al. (2005) propose an MSS approach for a cardiothoracic department. At an aggregate level they form surgical procedure types and level resource requirements such as bed requirements. The objective of their approach is to minimize the deviation of target utilization rates for the OR, the ICU, and the wards. The approach focuses on capacity planning and does not account for the stochastic nature of health care processes.

The aforementioned authors propose various approaches for cyclic OR planning, some of them taking into account succeeding or preceding hospital departments. These approaches are designed for a higher level of aggregation than what we focus on. None actually constructs OR schedules in which actual surgical procedures or procedure types and their stochasticity are incorporated.

Problem description

The aim of this paper is to develop methods to generate MSSs, i.e., OR schedules that are cyclically executed in a given planning period. The cyclic nature of an MSS requires that not surgical procedures of concrete patients but surgical procedures of a certain type are scheduled. The concrete assignment of patients to the planned procedure types has to be done in a latter stage. To make such an approach applicable, the types of surgical procedures must represent surgical procedures, which are medically homogeneous in the sense that they share the same diagnosis and are performed by the same surgical department. In most hospitals there are three categories of types of procedures:

- Category A: elective procedures that occur quite frequent,
- Category B:elective procedures that occur rather seldom,
- Category C: emergency procedures.

Following the above discussion, an MSS can concern only Category A procedures. More precisely, we define Category A procedures as elective procedure types, which have a frequency such that they occur at least once during the cycle time of the MSS. The chosen cycle length thus determines the number of surgical procedure types incorporated in an MSS. Category B procedures consist of all other elective procedures and cannot be planned in an MSS, whereas Category Cprocedures cannot be planned due to their nature. However, in the construction of an MSS, capacity for the procedures of types B and C will be reserved. An MSS is part of a cyclic OR planning strategy, which has three stages. First, clinicians and managers determine the MSS cycle length. Correspondingly, they determine how the ORcapacity is divided over the three categories. Second, before each cycle, clinicians assign actual Category A patients to the procedure type "slots" in the MSS, and Category Bprocedures to their reserved capacity. Third, during execution of the elective schedule, Category C (emergency) procedures are scheduled. Widely used approaches are to assign these to reserved capacity (Goldratt 1997), or to capacity obtained by canceling elective procedures (Jebali et al. 2006).

In this paper we propose a model for the construction of MSSs for Category Aprocedures. Scheduling Category Band Cprocedures is beyond the scope of this paper. An MSS can be used repetitively bya hospital until the size and the content of the three categories change. Then, the MSS must be reoptimized.

The goal of our MSS is to generate a cyclic schedule, in which all Category A procedures are scheduled according to their expected frequency, in such a way that the workload of subsequent departments like wards and IC is leveled as much as possible. This leveling results in reduction of peak demands on hospital bed departments caused by elective surgical procedures and, as such positively influences resource shortages and minimizes the number of cancelation of surgical procedures McManus et al. 2003. The number of available ORs restricts constructing the MSS as well as the available operating time and the capacity of succeeding departments (i.e., number of available beds). Personnel restrictions are not taken into account. We assume that sufficient flexibility remains for personnel scheduling at the operational level when the scheduling of Category B procedures is done. To avoid the probability of overtime, planned slack is included in the construction of MSSs. The amount of slack depends on the accepted probability that overtime occurs, which is determined by the management, and the variance of procedure durations. We use the portfolio effect to minimize the total amount of required slack (Hans et al. 2006). The portfolio effect is the tendency for the risk of a well-diversified range of stochastic variables to fall below the risk of most and sometimes, all of its individual components. This principle can be applied with respect to the stochastic surgery durations. Exploiting the portfolio effect can thus reduce the required amount of slack.

Formal problem description The surgical procedures to be incorporated into an MSS (Category A procedures) are categorized into I different types of medical and logistical similar procedures. From type i, $i = 1,\dots, I$ we have s_i procedures to be added in the MSS. The duration of a surgical procedure of type i is a stochastic variable ξ_i, and based on Strum etal. (2000). We assume that ξ_i has a lognormal distribution. Let B be the number of different hospital bed types. The various hospital bed types differ in importance and to indicate the relative importance of hospital bed type b we introduce priority factor c_b. The duration of hospital bed requirements of type b for a procedure of type i is denoted by $l_{ib} \in \mathbb{N}$, $i = 1,\ldots, I; b = 1,\ldots, B$. We assume that only one patient per day can use a bed.

The MSS has a fixed duration, the cycle length T. This cycle length is measured in days and typically is a multiple of 7 days. The given surgical procedures have to be carried out in J identical ORs, where OR j on day t has a capacity of o_{jt}, $j = 1,\ldots, J; t = 1,\ldots, T$. For creating an MSS, procedures have to be assigned to the ORs. The total sum of the duration of procedures assigned on a single OR on a specific day may not exceed the available capacity with probability α, i.e., with probability α that no overtime occurs. We refer to OR j on day t as OR-day (j, t).

The combined objective of the problem is to construct MSSs such that both the required OR capacity is minimized and the hospital bed requirements are leveled over the cycle.

Base model In this subsection we give a base model of the MSS problem. The aim of the model is to create a precise description of the objectives and the constraints.

To distinguish between minimization of OR capacity and hospital bed requirement leveling we define a weighted objective function, in which $\theta 1$ is the weight of minimization of the required OR capacity and $\theta 2$ is the weight of the hospital bed leveling. The weights may for example be related to the costs of the reduction of required OR capacity relative to the costs of peak demand on hospital beds.

We introduce an integer decision variable V_{ijt} to indicate the number of surgical procedures of type i that is assigned to OR-day (j, t), and an auxiliary binary variable W_{jt} to indicate whether an OR j is used on day t. An OR is considered to be used on day t if at least one surgical procedure is assigned to this OR-day. The total amount of OR capacity that is made available on day t is the sum of the available capacity of all used ORs. This is given by

$$\sum_{t=1}^{T}\sum_{j=1}^{J} o_{jt} \cdot W_{jt}.$$

To calculate the number of beds that is required from hospital bed type b,we introduce parameters $\psi_{t\tau ib}$ that denotes the requirements for hospital bed type b on day τ for a surgical procedure of type i, if this procedure is scheduled on day t. More specific, parameter $\psi_{t\tau ib}$ is $\left\lceil \frac{l_{ib}}{T} \right\rceil$ if $\min\{(t-1) \bmod T, (t+l_{ib}-2) \bmod T\} \leq (\tau-1) \leq \max\{(t-1) \bmod T, (t+l_{ib}-2) \bmod T\}$and $\left\lfloor \frac{l_{ib}}{T} \right\rfloor$ otherwise. To illustrate this expression, suppose an MSS has cycle length $T = 7$ days. On day $t = 5$, a procedure of type i is scheduled that subsequently requires an IC bed for 8days ($l_{ib} = 8$). This results in the requirement of two ICU beds on day $\tau = 5$ of the cycle and one IC bed on all other days. On day 5 the requirement is two beds, because the patient of the previous cycle is still occupying an ICU bed.

To level the hospital bed requirements, we minimize the maximum demand for hospital beds during an MSS cycle. This min–max type of resource leveling objective is generally used for problems where resource usage is very expensive (for this and other types, see: Brucker et al. 1999; Neumann and Zimmermann 2000). The presented approach is not specific for beds but can be used similarly for other types of hospital resources.

The maximum demand for hospital bed type b in acycle is: $\max_{\tau\in T} \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{t=1}^{T} \psi_{t\tau ib} \cdot V_{ijt}$. To ensure that the objective function is not influenced by the total requirement of different hospital bed types, but only by their relative importance, we normalize the maximum demand for any hospital bed. The normalization factor is the total demand for an hospital bed type b during one cycle: $\left(\sum_{i=1}^{I} l_{ib} \cdot s_i\right)/T$. This yields the normative sum of the maximum demand of all hospital bed types:

$$\sum_{b=1}^{B}\left[\frac{c_b}{\left[\sum_{i=1}^{I} l_{ib} \cdot s_i\right]/T}\right] \cdot \max_{\tau\in T}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{t=1}^{T} \psi_{t\tau ib} \cdot V_{ijt}$$

The overall objective function consisting of the weighted sum of needed OR capacity and the peak demands of hospital beds is given by formula (1) in the base model presented below.

To ensure that an operatingroom is considered to be used if at least one procedure is assigned to that operating room, constraints (2) are introduced. Constraints (3) ensure that all surgical procedures of all types are assigned. To model the bound on the probability that overtime occurs, we introducea function $f_{jt}(V)$. It denotes the probability distribution of the total duration of all procedures that are scheduled on OR-day (j, t) by V, where V is the vector of all variables V_{ijt} (a possible way to deal with this function, is given in the following section). Using the function $f_{jt}(V)$, the restriction that the total duration of procedures on an OR-day may not exceed the available capacity with probability α, can be expressed bythe probabilistic constraints (5). We refer to Charnes et al. (1964) for detailed information on probabilistic constraints. Summarizing, the base model becomes:

$$\min \theta_1 \cdot \sum_{t=1}^{T}\sum_{j=1}^{J} o_{jt} \cdot W_{jt} + \theta_2 \cdot \sum_{b=1}^{B}\left[\frac{c_b}{\left[\sum_{i=1}^{I} l_{ib} \cdot s_i\right]/T} \cdot \max_{\tau\in T}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{t=1}^{T} \psi_{t\tau ib} \cdot V_{ijt}\right] \quad (1)$$

subject to

$$V_{ijt} \leq s_i \cdot W_{jt}, \quad i = 1, \ldots, I, \;\; j = 1, \ldots, J, \;\; t = 1, \ldots, T \tag{2}$$

$$\sum_{t=1}^{T} \sum_{j=1}^{J} V_{ijt} = s_i, \quad i = 1, \ldots, I \tag{3}$$

$$\Pr[f_{jt}(V) \leq o_{jt}] \geq \alpha, \quad j = 1, \ldots, J, \;\; t = 1, \ldots, T \tag{4}$$

$$V_{ijt} \in \mathbb{N}, \quad i = 1, \ldots, I, \;\; j = 1, \ldots, J, \;\; t = 1, \ldots, T$$

$$W_{jt} \in \{0, 1\}, \quad i = 1, \ldots, I, \;\; t = 1, \ldots, T.$$

The min–max objective can be reformulated (see Williams 1999, p. 23) such that the base model is an integer linear program (ILP) with additional probabilistic constraints. The size of instances from practice gets extremely large (the Erasmus MC instances approximately have 1.9×10^5 decision variables), such that even without the probabilistic constraints this is far too large to solve the model to optimality within reasonable computation time. The MSS problem itself is NP-hard even if the probabilistic effects are neglected. The first part of the objective function together with the packing constraints contains e.g. the bin-packing problem and the second part of the objective function contains e.g. the three-partitioning problem. Based on this, we concentrate on a heuristic approach to solve the MSS problem.

Solution approach

The main decision in the MSS problem is to fill OR-days (j, t) according to the imposed restrictions. Since in practice the given capacities o_{jt} are of ten the same for different ORs and for different days, we introduce the concept of so-called *operating room day schedule* (ORDS). An ORDS for capacity o is a set of surgical procedures of various types, which is feasible with respect to the OR-capacity constraint (5) with $o_{jt} = o$. As a consequence, an ORDS for capacity o can be assigned to all OR-days (j, t) with $o_{jt} = o$. MSS comprises of assigning one ORDS to each OR-day $(j, t,)$ in the cycle, such that the objective function (1) is minimized.

We propose a two-phase decomposition approach. In Phase 1 hospital bed requirement leveling is ignored, and a set of ORDSs that covers all procedures is selected. These ORDSs have capacities fitting to the capacities of the OR-days, and minimize the required OR capacity. We discretize the probabilistic OR capacity constraints, and formulate an ILP that we solve with an implicit column generation approach. In Phase 2 we assign ORDSs to concrete OR-days in such a way, that the hospital bed capacity demand is leveled. For this purpose, the problem is formulated as mixed integer linear program (MILP).

Phase 1 The problem in Phase 1 consists of selectinga set of ORDSs that coversall surgical procedures and all OR-day capacities and minimizes the required OR capacity. In Sect. 4.1.1 we formalize the problem as an ILP problem where the variables correspond to ORDSs of given capacities. Afterwards, in *Phase 1 model* we propose a column generation approach to generate possible ORDSs. In this part we discretize the probabilistic constraints on the ORDSs.

Phase 1 model The available capacity of ORs in the MSS cycle may differ from day to day. Let R be the number of different OR capacity sizes (sorted in non-decreasing order). The actual capacity of an OR of capacity size type r is given by d_r, $r = 1, \ldots, R$. Let U be the set of possible ORDSs, and let U_r be the subset of U that contains all the ORDSs that belong to the rth capacity size. In this context an ORDS u belongs to U_r if the rth capacity size is the smallest available capacity size where the ORDS fits in. Hence, $U = \cup_{r=1}^{R} U_r$. Let m_r be the number of OR-days within one cycle length that have the rth capacity size and let φr be the set of corresponding tuples (j, t). For a given ORDS $u \in U$ we denote the number of surgical procedures of type i that are scheduled in u by $a_{iu} \in \mathbb{N}$.

To formulate the Phase 1 model, we introduce integer decision variables X_u $(u \in U)$ that represent the number of times that ORDS u is selected. The objective function (5) corresponds to the first part of the objective function (1) of the base model: minimization of the required OR capacity. Constraints (6) impose that all procedures are selected. The number of ORDSs generated for every OR capacity size that we can select is restricted by the number of available OR-days m_r of capacity type r. This restriction is imposed by constraints (7). Summarizing, in Phase 1we must solve the following ILP:

$$\min \sum_{r=1}^{R} \sum_{u \in U_r} d_r \cdot X_u \tag{5}$$

subject to

$$\sum_{r=1}^{R} \sum_{u \in U_r} a_{iu} \cdot X_u \geq s_i \quad i = 1, \ldots, I \tag{6}$$

$$\sum_{u \in U_r} X_u \leq m_r \quad r = 1, \ldots, R \tag{7}$$

$$X_u \in \mathbb{N} \quad u \in U.$$

This model has two main drawbacks. The set of possible ORDSs U grows exponentially with the number of procedure types, and due to the probabilistic constraints, the identification of all possible elements of U is difficult. To overcome this, a column generation approach for this problem is presented where furthermore the check on containment of an ORDS in a set U_r is discretized.

Column generation Column generation is an often-used approach to solve complex optimization problems with a large number of variables (e.g. cutting stock, capacity planning, and crew scheduling, e.g., Barnhart et al. 1998; Pinedo 2005). The outline of our approach is as follows. We use column generation to solve the LP relaxation of the Phase 1 model, and round this solution to obtain a feasible solution. In the column generation procedure we iteratively generate subsets of U (i.e., subsets of ORDSs) and solve the Phase 1model for these subsets. The Phase 1 model restricted to such a subset of U is called the restricted master problem. In each iteration, solving the restricted LP-relaxation (i.e. the LP-relaxation of the restricted master problem) yields shadow prices. These are used as input for the sub-problem (the pricing problem), which revolves around generating ORDSs that are not included in the restricted master problem, but that may improve its solution. The reduced costs of the corresponding variables X_u are negative. These ORDSs are added to the restricted master problem, and the LP-relaxation is re-optimized. This procedure stops if no ORDSs exist that may improve the restricted LP-relaxation solution. The restricted LP-relaxation solution is then optimal to the LP-relaxation. We then apply a rounding procedure to obtain a feasible Phase 1 solution.

Initialization We use an initialization heuristic to generate subsets of U_r for all ORcapacity sizes $r = 1,\ldots, R$. More precisely, for each $r = 1,\ldots, R$ we generate subsets $\bar{U}_r \subset U$ of ORDSs that cover all surgical procedures. This initial set of ORDSs serves as a starting point for the column generation procedure.

Let the variable $Z_i^r \in \mathbb{N}$, $(i = 1,\ldots, I,)$ denote the number of procedures of type i that is scheduled in an ORDS for OR capacity size r. Any vector $Z^r = (Z_1^r, \ldots, Z_I^r)$ must satisfy the probabilistic bin-packing constraint (8) to be a feasible ORDS for capacity size r, where $f(Z^r)$denotes the distribution function that represents the stochastic sum of the duration of all surgical procedures in the ORDS.

$$\Pr[f(Z^r) \leq d_r] \geq \alpha \tag{8}$$

The probabilistic constraints (8) impose difficulties on the generation of ORDSs. We discretize constraints (8) using prediction bounds. A prediction bound n_i^α denotes that the duration ξ_i of procedure type i is smaller than or equal to n_i^α with a probability α. These prediction bounds are used to replace the stochastic variables ξ_i, and can be calculated using the primitive of the distribution function of ξ_i. The total required OR capacity for an ORDS given by the vector Z^r is given by $\sum_{i=1}^{I} n_i^\alpha \cdot Z_i^r$. The difference between the value of a prediction bound and the average surgical procedure duration is used to compute the planned slack.

As discussed by Hans et al. (2006) the total amount of planned slack for a multiple of surgical procedures is reduced by the portfolio effect. This portfolio effect may be approximated by a function g, which only depends on the number of procedures that are scheduled in the operating room and on the average standard deviation of all types of surgical procedures. The reduction of required planned slack $g(\sum_{i=1}^{I} Z_i^r)$, as a result of the portfolio effect, is subtracted from the sum of the prediction bounds. This results in the following OR capacity constraints:

$$\left(\sum_{i=1}^{I} n_i^\alpha \cdot Z_i^r\right) - g\left(\sum_{i=1}^{I} Z_i^r\right) \leq d_r \tag{9}$$

All vectors $(Z_1^r, \ldots, Z_I^r)$ that satisfy constraints (9) are possible elements of U_r. Since the generation of ORDS is basically a bin-packing problem, we may apply bin-packing heuristics such as First Fit Decreasing (FFD), Best Fit Decreasing (BFD) and Minimum Bin Slack (MBS) (Gupta and Ho 1999) or a heuristic such as Randomized List Scheduling Heuristic (van den Akker et al. 1999) to generated initial set of ORDSs. Since in a study of off-line bin-packing algorithms by Dell'Olmo and Speranza (1999) Longest Processing Time (LPT) performs well, we use this heuristic for the generation of an initial set of ORDSs for an OR capacity size r. LPT first sorts all procedures of all types in decreasing order of their prediction bound n_i^α and then it creates an ORDS in which itplans the longest procedure that fit, i.e., that satisfy constraints (9). If the heuristic reaches the end of the ordered list it closes the ORDS. This is repeated until no surgical procedures remain in the ordered list. The heuristic is executed for all OR capacity sizes.

Pricing problem An optimal solution of the LP relaxation of the restricted problem is optimal for the LP relaxation of the complete master problem if the corresponding dual solution is feasible for the dual problem of the LP relaxation of the master problem. The pricing problem is thus to determine whether there exist ORDSs that are not in the restricted LP relaxation that violate the dual constraints from the LP relaxation of the master problem. Such ORDSs are added to the restricted LP relaxation and a next iteration starts. If such ORDSs do not exist, column generation terminates, and the current restricted LP relaxation solution is optimal to the LPrelaxation of the master problem.

The dual constraints of the LP relaxation of the Phase 1 model are:

$$\pi_r + \sum_{i=1}^{I} \lambda_i \cdot a_{iu} \leq d_r \qquad r = 1, \ldots, R$$

$$\pi_r \leq 0 \qquad r = 1, \ldots, R \tag{10}$$

$$\lambda_i \geq 0 \qquad i = 1, \ldots, I,$$

where λi are the dual variables corresponding to constraints (6), and πr the dual variables corresponding to constraints (7) of the Phase 1LP.

As input for the pricing problem we obtain two vectors $(\bar{\pi}, \bar{\lambda})$ of shadow prices from the restricted LP relaxation. The pricing algorithm now examines whether for this solution $(\bar{\pi}, \bar{\lambda})$ an ORDS $u \in U_r$, represented by $a_{1u}, \ldots, a_{Iu}$, exists that violates the dual constraint (10), i.e. values $a_{1u}, \ldots, a_{Iu}$, with:

$$d_r - \bar{\pi}_r - \sum_{i=1}^{I} \bar{\lambda}_i \cdot a_{iu} < 0 \tag{10}$$

The left-hand side of constraints (11) are the reduced costs for variable X_u $(u \in U_r)$. We evaluate each OR capacity size r separately to determine whether an ORDS exists, formed by a vector $(Z_i^r, \ldots, Z_I^r)$, that violates the dual constraints (10). In the rth problem we thus need to maximize

$$\sum_{i=1}^{I} \bar{\lambda}_i \cdot Z_i^r$$

over all vectors $(Z_i^r, \ldots, Z_I^r)$ representing a new ORDS, i.e. satisfying constraint (9).

To solve the pricing problem as an ILP we write the term: $g(\sum_{i=1}^{I} Z_i^r)$ as a telescopic sum. For this purpose, we introduce additional notation. The binary variable Ae indicates whether there are at least $e$procedures in an ORDS ($e \leq E$, where E is the maximum number of procedures that can be performed during 1day in one operating room). The function $g(e) := g_1 + \cdots + g_e$ provides the correction for the portfolio effect for e surgical procedures. Using this function and the binary variables A_e,the rth pricing problem ILP becomes:

$$\max \sum_{i=1}^{I} \bar{\lambda}_i \cdot Z_i^r \quad \text{subject to} \quad \left(\sum_{i=1}^{I} n_i^{\alpha} \cdot Z_i^r \right) - \sum_{e=1}^{E} g_e \cdot A_e \leq d_r \qquad r = 1, \ldots, R$$

$$\sum_{i=1}^{I} Z_i^r = \sum_{e=1}^{E} A_e$$

$$A_e \geq A_{e+1} \qquad e = 1, \ldots, E$$

$$A_e \in \{0, 1\} \qquad e = 1, \ldots, E$$

$$Z_i^r \in \mathbb{N} \qquad i = 1, \ldots, I.$$

After this problem is solved for all capacity sizes r, the resulting ORDSs with negative reduced costs are added to the restricted LP relaxation of the Phase 1model. This model is reoptimized to obtain new shadow prices. Column generation stops if no such ORDSs are found any more. If in practice this process takes very long and generates a large number of extra columns, one might incorporate some of the stopping criteria like the amount of improvement in the L Presulting from the newly generated columns. This may have some effect on the quality of the LP-solution, but since afterwards still an integer solution has to be constructed, the effect on the solution after Phase 2 might be only marginal. In our test instances, we always were able to solve the LP-relaxation to optimality.

Rounding heuristic The solution to the restricted LP relaxation does not directly lead to a starting point for the second phase, since ORDSs may have been selected fractionally. To obtain an integer solution we use a rounding heuristic that rounds down the fractional solution. This results in an integer solution with a small number of surgical procedures that are not assigned to selected ORDSs. These procedures are assigned to newly created ORDSs using an LPT heuristic. There may also be some redundant surgical procedures due to the "≥" signin constraints (6). We remove these redundant procedures randomly. In general, this approach does not guarantee to result in a feasible solution. However, for the tested instances a quite large fraction of procedures was planned before rounding, only a fraction had to be planned by the LPT heuristic. We never got stuck with infeasible solutions at this stage. Ifinfeasibility might get an issue, the simple rounding heuristic leave room for algorithmic improvements and may be replaced by more elaborate approaches. Summarizing, the output of Phase 1 consists of a set of ORDSs that cover the set of all surgical procedures to be assigned within the MSS.

Phase 2 In Phase 2the actual MSS cycle is constructed. We propose an ILP in which the set of ORDSs is assigned to OR-days such that the hospital bed requirements are leveled over the days.

Phase 2 model Given is a set $\bar{U}$ of ORDSs to be assigned to the OR-days of the MSS. Let $\bar{U}_r \subset \bar{U}$ denote the ORDSs which are of capacity size r. To model the assignment of an ORDS u to an OR-day (j, t) we introduce binary decision variables Y_{ujt} for all $u \in \bar{U}_r$ and $(j, t) \in \varphi_r$. We ensure that the ORcapacity sizes match and that at most one ORDS is assigned to an OR on a day. The objective function takes into account the requirements for all hospital beds for all days within one MSS cycle, thus also requirements of surgical procedures that have taken place in previous cycles. Corrected by a normalized priority factor (see

Sect. 3.2), we minimize the maximum requirements for hospital beds. The objective function is the second term of the objective function (1) of the base model. This objective function is a minimax objective and can be rewritten to Eq. (12) and constraints (13) in which HB_b is the maximum requirement of hospital bed type b on a given day in the cycle.

All selected ORDSs from Phase 1 must be assigned to an operating room and a day. This is ensured by constraints (14). No more than one ORDS can be assigned to an operating room on a day, which is imposed by constraints (15). Summarizing, the model of Phase 2 is the following ILP:

$$\min \sum_{b=1}^{B} \left[\frac{c_b}{\left\lceil \sum_{i=1}^{I} l_{ib} \cdot s_i \right\rceil / T} \right] \cdot HB_b \tag{12}$$

$$\sum_{r=1}^{R} \sum_{u \in \bar{U}_r} \sum_{(j,t) \in \varphi_r} \sum_{i=1}^{I} \sum_{t=1}^{T} \psi_{t\tau ib} \cdot a_{iu} \cdot Y_{ujt} \leq HB_b \qquad \tau = 1, \ldots, T, \quad b = 1, \ldots, B \tag{13}$$

$$\sum_{(j,t) \in \varphi_r} Y_{ujt} = X_u \qquad r = 1, \ldots, R; \quad u \in \bar{U}_r \tag{14}$$

$$\sum_{u \in \bar{U}_r} Y_{ujt} \leq 1 \qquad r = 1, \ldots, R; \quad (j,t) \in \varphi \tag{15}$$

$$Y_{ujt} \in \{0,1\} \qquad u \in \bar{U}_r; \quad (j,t) \in \varphi_r$$

$$z_b \geq 0 \qquad b = 1, \ldots, B.$$

Solving the Phase 2 model We solve the Phase 2model using the commercial solver ILOG CPLEX 9.0. We use lower bound on the values *HBb* to determine the quality of an intermediate solution and to speed up the computation. These lower bounds are calculated by rounding up the sum of the total requirements of hospital beds during one cycle divided by the cycle length:

$$\left\lceil \frac{\sum_{i=1}^{I} l_{ib} \cdot s_i}{T} \right\rceil$$

This represents a theoretical minimum of the maximum requirements for hospital bed type b on 1day in a cycle. The lower bounds are multiplied by the normative sum used in the objective (1) of the base model:

$$\sum_{b \in B} \left(\frac{c_b}{\left(\sum_{i=1}^{I} q l_{ib} \cdot s_i \right) / T} \right) \cdot \left\lceil \frac{\sum_{i=1}^{I} l_{ib} \cdot s_i}{T} \right\rceil \tag{16}$$

This overall lower bound (16) is given as an initial lower bound to CPLEX to speed up the branch-and-bound process.

Computational experiments

We implemented the two-phase approach in the AIMMS mathematical modeling-language 3.5 (Bisschop 1999), which interfaces with the ILOG CPLEX 9.0 LP/ILP solver. We test our approach with realistic data instances from the Erasmus MC based on the available database of surgical procedures that has been collected from 1994 until 2004. This data consists of the frequency of surgical procedures, procedure durations, and data about the usage of hospital beds after surgical procedures.

Instance generation Since 1994 Erasmus MC has been collecting data on the frequency of surgical procedures, the duration of procedures, and standard deviation of the duration of procedures. In cooperation with surgeons we defined procedure types by grouping medically homogeneous procedures, which results in the Erasmus MC instance. The data consist for each surgical procedure type i of the frequency of a surgical procedure type during one cycle s_i, the prediction bound n_i^{α} , and the length of a request of a hospital bed l_{ib}. We vary the parameter values of the cycle length T, the number of operating rooms J, and the number of hospital bed types B (see Table 1), which results in 36 instances types. For each parameter combination 9 additional instances are generated, this yields a total of 360 instances. The additional instances are generated by randomly drawing data from the intervals in Table 2 and rounding them to the nearest integers (the values with a tilde in the table represent the values of the parameters resulting from the Erasmus MCinstance).

The cycle length influences the number of procedure types and the number of surgical procedures that can be incorporated into the MSS (Category A procedures). Table 3 shows the dependency between the cycle length and the number of surgical procedure types in Category Atogether with their numbers and total duration.

We assume that all ORs are available during weekdays and are closed for elective procedures in weekends. For the computational experiments in this paper we use one OR capacity size ($R = 1$) of 450 min ($d_r := 450$). Furthermore, we assume that procedures are finished before their prediction bound in 69% of the cases, i.e., $\alpha := 69\%$. This value is taken from the current practice of Erasmus MC. The priority factors of hospital beds are given by: $c(1) := 5$ $c(2) := 2$ $c(3) := 1$.

Cycle length in days	$T \in \{7, 14, 28\}$
the instances Number of operating rooms	$J \in \{5, 10, 15, 20\}$
Number of hospital bed types	$B \in \{1, 2, 3\}$

Table 1 *Parameter values for the instances*

$$s_i \in [0.9 \cdot \tilde{s}_i, 1.1 \cdot \tilde{s}_i]$$
$$n_i^\alpha \in [0.9 \cdot \tilde{n}_i^\alpha, 1.1 \cdot \tilde{n}_i^\alpha]$$
$$l_{ib} \in [0.5 \cdot \tilde{l}_{ib}, 1.5 \cdot \tilde{l}_{ib}]$$

Table 2 *Intervals for creating instances*

Cycle length in days	Number of procedure types	Total number of procedures	Total duration of all procedures (in hours)
7	42	56	126
14	109	177	398
28	203	423	952

Table 3 *The relation between the cycle length and procedures in Category A*

e	1	2	3	4	5
g(e)	$0.00 \cdot \bar{\sigma}$	$0.10 \cdot \bar{\sigma}$	$0.22 \cdot \bar{\sigma}$	$0.36 \cdot \bar{\sigma}$	$0.48 \cdot \bar{\sigma}$

Table 4 *Parameter values for function g, to model the portfolio effect*

The function g, which we use to model the portfolio effect, depends on the number of procedures that is scheduled in an ORDS and the average standard deviation $\bar{\sigma}$ of all surgical procedures. We approximate the portfolio effect using the function $g(e)$ that takes the values indicated in Table 4. The value for the average surgical procedure standard deviation $\bar{\sigma}$ is 36, based on the database of the Erasmus MC.

Test results In the tests we focus on three different aspects. Firstly, we study the dependencies of the computation times of both phases on the used parameter combinations. Secondly, we investigate the obtained results of the minimization of the required OR capacity. And finally, we address the hospital bed leveling. For this last issue, we have truncated computations that exceed 600 s and have used the best incumbent solutions as output. These incumbent solutions are, therefore, generally not optimal for the Phase 2model.

Computation times Table 5 presents the computation times in Phase 1 for all parameter combinations. The computation times in Phase 1include the initialization and rounding heuristic.

The computation time increases with T, whereas B and J hardly influence the computation time. Similar results are obtained when computation times of the initialization heuristic are considered solely. Here the computation times vary from 0 to 6 s. We conclude that the initialization heuristic only needs a small fraction of time that is required by the complete Phase 1 computation. Table 6 presents the computation time in Phase 2 for all parameter combinations.

Table 6 shows that all three parameters have considerable impact on the computation time and in all cases the computations time increases with increasing parameter value. Table 7 shows the number of times that the calculation is truncated after 600 s for all parameter combinations. The '–' sign denotes that these test instances are infeasible due to the lack of operating rooms.

The extreme growth of the computation time for some of the test instances in Table 6 results mainly from hard instances, where the calculation is truncated

	T →	7			14			28		
J ↓	B →	1	2	3	1	2	3	1	2	3
5		15.10	17.08	13.76	43.91	47.00	45.90	80.96	78.78	74.56
10		15.29	16.59	13.36	47.12	44.28	45.62	80.24	83.90	87.01
15		16.29	16.12	13.17	47.24	44.70	44.03	80.20	75.96	95.17
20		15.01	16.73	14.35	48.01	45.94	42.39	81.07	75.00	89.70

Table 5 *Computation times of Phase 1 in relation with T, B and J*

	T →	7			14			28		
J ↓	B →	1	2	3	1	2	3	1	2	3
5		0.30	0.49	0.56	1.55	2.79	3.69	6.16	8.94	13.32
10		0.63	0.93	1.11	3.78	5.86	72.04	15.02	30.54	325.08
15		0.96	1.37	121.60	5.39	8.69	72.92	18.87	43.09	517.08
20		1.21	1.81	122.27	7.45	11.26	87.79	24.54	47.25	478.67

Table 6 *Computation times of Phase 2 in relation with T, B and J*

J↓	T→ / B→	7			14			28		
		1	2	3	1	2	3	1	2	3
5		0	0	0	-	-	-	-	-	-
10		0	0	0	0	0	1	0	0	3
15		0	0	2	0	0	1	0	0	7
20		0	0	2	0	0	1	0	0	5

Table 7 *Number of times that computation is truncated*

(see Table 7). Computation times are not high and therefore allow use of the proposed approach in practice.

OR utilization Table 8 shows the average number of required ORs per week in relation to the cycle length T. The number of required ORs increases if the cycle length increases, which may be expected since the total surgical procedure volume increases as well (see Table 3). The rounding gap between the integer solution of Phase 1and the value after rounding up the optimal fractional solution of the LP relaxation denotes the quality of the rounding heuristic. We conclude that the rounding gap is small and decreases if more ORDSs are required. Thus, we may conclude that the achieved OR utilization after Phase 1 is close to the best possible utilization.

Table 8 gives the results of using only the ORDSs generated by the initialization heuristic. These values are found by solving the restricted LP using the initially generated ORDSs and applying the rounding heuristic. They are equal to the values of the complete column generation approach for the construction of MSSs with the cycle length of 7 and 14 days. For larger instances with the cycle length of 28 days, the complete column generation slightly improves the

T↓	Initialization heuristic and column generation		Initialization heuristic only
	Required number of operating rooms during 1 week	Rounding gap(%)	Required number of operating rooms during 1 week
7	16.50	1.25	16.50
14	27.80	0.9	27.80
28	34.18	0.6	34.33

Table 8 *Test results of Phase 1*

initialization heuristic. Thus, in most of the cases, the ORDSs generated by the initial heuristic already contain the ORDSs needed for the optimal fractional solution of the LP-relaxation of the Phase 1 model. But since an MSS is typically constructed once a year, the additional computational effort of the column generation approach should be used to try to improve the initial solution.

Hospital bed leveling In this section we discuss the hospital bed leveling. The relative difference between the objective value of the Phase 2 model and the lower bound [see expression (16)] indicates the quality of the solutions found. Table 9 presents the relative differences.

The results in Table 9 show that the difference between the found solutions and the lower bound is small. Therefore, Phase 2 almost optimally levels the hospital bed requirements. This is the more surprising, since the ORDSs in Phase 1 have been generated with the only goal to optimize resource utilization not taking into account the subsequent problem of hospital bed leveling.

In 22 out of 360 experiments the computation of Phase 2istruncated. Table 10 presents the relative differences between the found solution and the lower bound for the 22 truncated instances.

J↓	T→ / B→	7			14			28		
		1	2	3	1	2	3	1	2	3
5		0.0%	0.0%	0.5%	-	-	-	-	-	-
10		0.0%	0.0%	0.5%	0.0%	0.0%	0.2%	0.0%	0.0%	1.3%
15		0.0%	0.0%	0.5%	0.0%	0.0%	0.2%	0.0%	0.0%	2.4%
20		0.0%	0.0%	0.5%	0.0%	0.0%	0.2%	0.0%	0.0%	1.5%

Table 9 *Average gap between the lower bound and the Phase 2 solution*

J↓	T→ / B→	7			14			28		
		1	2	3	1	2	3	1	2	3
5										
10							1.9%			4.5%
15				2.7%			1.9%			3.4%
20				2.7%			1.9%			3.0%

Table 10 *Average gap between the lower bound and the Phase 2 solution for truncated instances*

Even for these instances the average gap is small; the maximum gap is 10.1%. Based on the presented results we conclude that the constructed MSSs level the hospital bed requirements of the incorporated surgical procedures. This means that the requirements on one day rarely exceed the lower bound.

Conclusions and further research

The computational experiments show that generation of MSSs is well possible within acceptable time bounds by the proposed two-phase decomposition approach. The proposed solution approach generates MSSs that minimize the required OR capacity for a given set of procedures and level the hospital bed requirements well. The chosen solution approach makes it possible to add restrictions imposed by personnel and to consider other types of hospital resources than beds. This flexibility is required to implement an OR planning strategy that includes an MSS. The approach has been successfully tested on real data from Erasmus MC. The hospital management is pleased with the outcomes, and encourages and initiates further research into implementing the MSS-approach in practice.

In further research we will investigate implementation aspects, and scheduling of Category Band Cprocedures as such is required to determine the overall benefits of cyclic scheduling of OR departments. This research should also provide insight into the benefits of a cyclic OR planning approach for hospitals with various patient mixes. Furthermore, we will investigate the leveling of hospital beds when the length of request for beds is assumed to be stochastic.

The repetitive nature of our cyclic surgical planning approach yields that it reduces the overall management effort. In addition, it not only optimizes OR utilization but also levels the output towards wards and ICU. This results in less surgery cancelations, and thus a reduction of the lead-time of the patient's care pathway. Therefore, MSS contributes to an improved integral planning of hospital processes. The intensive cooperation with clinicians and OR managers has lead to a framework for cyclic OR planning and a method for construction of MSSs that can handle constraints imposed by health care processes. This flexibility ensures the applicability of the developed method in OR departments and hospitals.

References

1 Bakker H, Zuurbier JJ (2002) Pareto's law (In Dutch). Zorgvisie 7 Barnhart C, Johnson EL, Neuhauser GL, Savelbergh MWP, Vance PH (1998) Branch-and-price:

2 column generation for solving huge integer programs. Oper Res 46(3):316–329 Beliën J, Demeulemeester E(2005) Building cyclic master surgery schedules with leveled resulting

3 bed occupancy. Eur J Oper Res (in press)

4 J. M. van Oostrum etal.

5 Bisschop J(1999) AIMMS – optimization modeling. Paragon Decision Technology B.V., Harlem

6 Blake JT, Donald J (2002) Mount sinai hospital uses integer programming to allocate operating room time. Interfaces 32(2):63–73

7 Brucker P, Drexl A, Mohring R, Neumann K, Pesch E(1999) Resource-constrained project scheduling: notation, classification, models, and methods. Eur J Oper Res 112:3–41

8 Carter M(2002) Diagnosis: mismanagement of resources. Oper Res Manage Sci Today April 26–32

9 Charnes A, Cooper W, Thompson G (1964) Critical path analyses via chance constrained and stochastic programming. Oper Res 12:460–470

10 Dell'Olmo P, Speranza MG (1999) Approximation algorithms for partitioning small items in unequal bins to minimize the total size. Discret Appl Math 94:181–191

11 Gerhak Y, Gupta D, Henig M (1996) Reservation planning for elective surgery under uncertain demand for emergencysurgery. Manage Sci 42(3):321–334

12 Glouberman S, Mintzberg H (2001) Managing the care of health and the cure of disease Part i: Differentiation. Health Care Manage Rev 26(1):56–69

13 Goldratt EM (1997) Critical chain. The North River Press, Great Barrington

14 Guinet A, Chaabane S(2003) Operating theatre planning. Int J Prod Econ, 85:69–81

15 Gupta JND, Ho JC(1999) Anew heuristic algorithm for the one-dimensional bin-packing problem. Prod Plan Control 10(6):598–603

16 Hans EW, Wullink G, Van Houdenhoven M, Kazemier G (2006) Robust surgery loading. Eur J Oper Res (accepted for publication)

17 Jebali A, Hadj Alouane AB, Ladet P(2006) Operating room scheduling. Int JProd Econ 99:52–62

18 Kim S, Horowitz I (2002) Scheduling hospital services: the efficacy of elective-surgery quotas. Omega 30:335–346

19 Lamiri M, Xie X, Dolgui A, Grimaud F(2005) Astochastic model for operating room planning with elective and emergency surgery demands. In: Conference Proceedings ORAHS, Southampton, UK

20 McManus ML, Long MC,Copper A, Mandell J,Berwick DM, Pagano M, Litvak

E(2003) Variability in surgical caseload and access to intensive care services. Anesthesiology 98:1491–1496

21 Millar HH, Kiragu M (1998) Cyclic and non-cyclic schedulingof 12h shift nursesbynetwork programming. Eur JOper Res 104:582–592

22 Neumann K, Zimmermann J (2000) Procedures for resource leveling and net present value problems in project scheduling with general temporal and resource constraints. Eur J Oper Res 127:425–443

23 OECD (2005) Oecd health data 2005 – statistics and indicators for 30countries

24 Ogulata SN, Erol R (2003) A hierarchical multiple criteria mathematical programming approach for scheduling general surgery operatons in large hospitals. J Med Syst 27(3):259–270

25 Ozkaraham I (2000) Allocation of surgeries to operating rooms by goal programming. J Med Syst 24(6):339–378

26 Pinedo M (2005) Planning and scheduling in manufacturing and services. Springer Series in Operations Research and Financial Engineering. Springer, Berlin Heidelberg New York

27 Schmidt E, Dada M, Adams D(2001) Using cyclic planning to manage capacity at alcoa. Interfaces 31(3):16–27

28 Sier D, Tobin P, McGurk C(1997) Scheduling surgical procedures. J Oper Res Soc 48:884–891

29 Strum DP, May JH, Vargas LG (2000) Modeling the uncertainty of surgical procedure times: comparison of log-normal and normal models. Anesthesiology 92:1160–1167

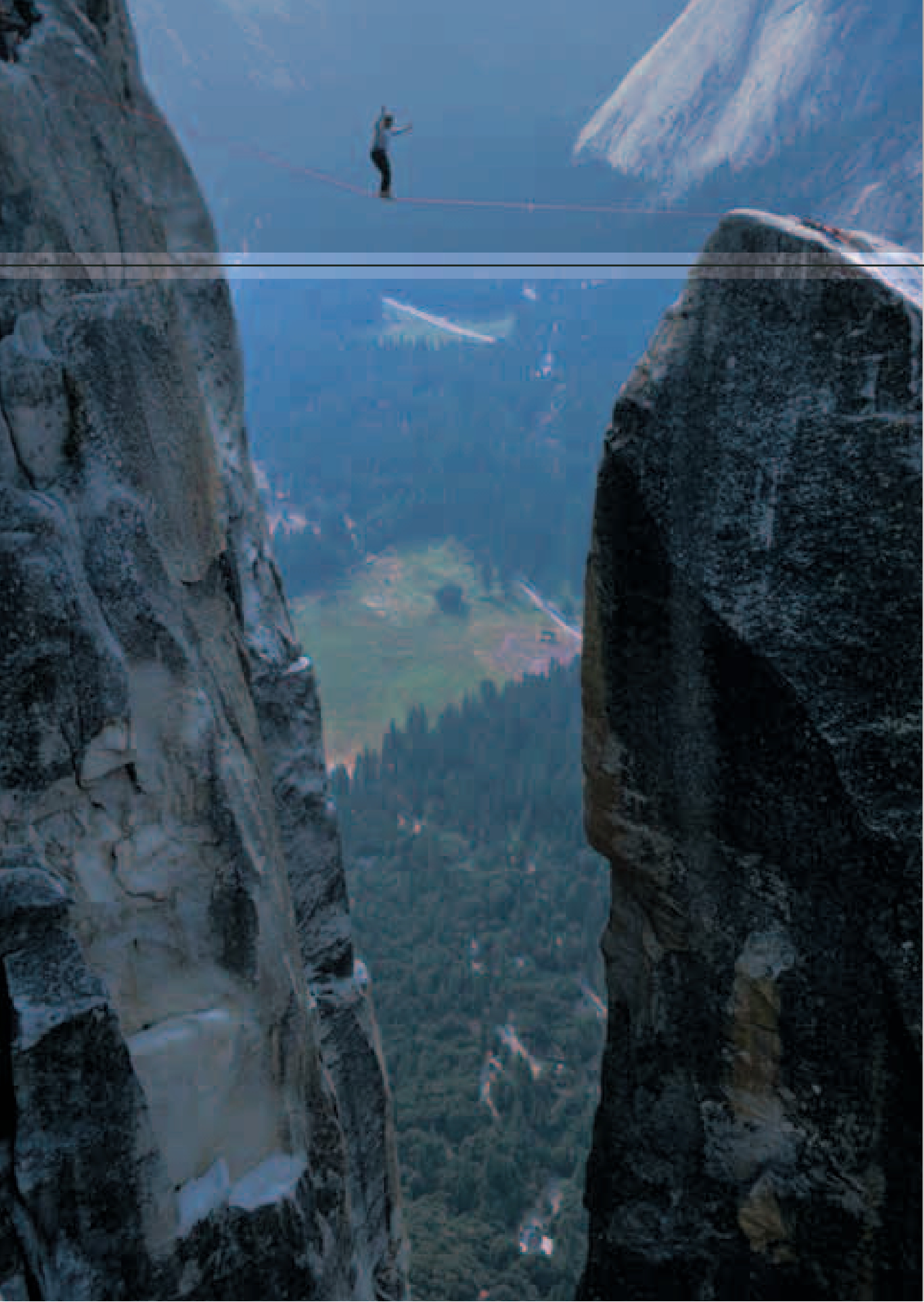
30 Tayur S(2000) Improving operations and quoting accurate lead times in a laminate plant. Interfaces 30(5):1–15

31 van den Akker JM, Hoogeveen JA, van de Velde SL(1999) Parallel machine scheduling bycolumn generation. Oper Res 47(6):862–872

32 Vissers JMH, Adan IJBF, Bekkers JA (2005) Patient mix optimisation in cardiothoracic surgery planning: a case study. IMA JManage Math 16(3):281–304

33 Weissman C(2005) The enhanced postoperative care system. J Clin Anesth 17:314–322

34 Williams HP (1999) Model building in mathematical programming, 4th edn. Wiley, New York .



Acknowledgements

Dit proefschrift is de uitkomst van een jarenlange samenwerking tussen geïnspireerde mensen die verschillende wetenschapsgebieden hebben samengebracht in de gezondheidszorg. Uiteindelijk is dit proefschrift het resultaat van de inbreng, energie, kennis, vaardigheden, inzet, passie voor het eigen vakgebied en passie voor de gezondheidszorg van velen. Zonder het idee te hebben volledig te kunnen zijn hier bij enkele dankwoorden. Velen hebben een bijdrage geleverd zonder het te weten in alle discussies en gesprekken, door jarenlang te registeren, door informatie te genereren en door analyses te maken. Dit proefschrift is er uiteindelijk gekomen omdat wiskundigen, medisch specialisten, managers en bestuurders hun wereld voor elkaar en voor mij hebben geopend. Juist deze opening bleek uiteindelijk een gouden zet en hopelijk vindt deze stap navolging niet alleen in onderzoek maar ook op de verschillende universiteiten en in de gezondheidszorg. Een dankwoord is het afsluiten van een periode die ik eigenlijk niet wil afsluiten. Dit dankwoord is voor mij dan ook geen afsluiting maar de doorstart in een nieuwe periode waarin het onderzoek dat we gezamenlijk zijn begonnen naar mijn stellige overtuiging een gezamenlijk vervolg krijgt.

- Weledelzeergeleerde heer Kazemier, Beste Geert, dank voor je vriendschap en steun tijdens mijn reis door de medisch wereld, het management, de wetenschap en het onderzoek. Zonder jou was het onderzoek en dit proefschrift er niet gekomen en had ik niet zoveel plezier en energie gehad bij het schrijven van de artikelen en in ons werk.
- Hooggeleerde heer Steyerberg, Beste Ewout, je kritische geest, je neus voor nieuwe inzichten en je flexibiliteit om je te verplaatsen van de wereld van de statistiek naar de wereld van de logistiek. Voor mij heb je de wereld van de wetenschap en het schrijven van onderzoeksvoorstellen geopend.

- Hooggeleerde heer Tilanus, dank voor de toegang tot uw wereld. Ik voel me nog steeds vereerd dat u zomaar de jarenlang verzamelde gegevens van uw werk in de chirurgie ongeclausuleerd beschikbaar heeft gesteld voor eigenlijk niet meer dan een wild idee van een onderzoeker buiten uw eigen vakgebied. Hopelijk krijgt uw voorbeeld veel vervolg en gaan anderen ook medische data beschikbaar stellen voor logistieke vraagstukken. Dit is de enige manier om de gezamenlijkheid tussen management en specialisten verder te ontwikkelen en werkelijk inzicht te verkrijgen in elkaars processen.
- Weledelzeergeleerde heer Hans, Beste Erwin, ons eerste gesprek is de basis voor een jarenlange samenwerking gebleken. Vanaf het eerste gesprek hebben we elkaar geïnspireerd en ik ben er van overtuigd dat we dat zullen blijven doen. Wiskunde en management kunnen complementair zijn echter alleen als het echt ergens om gaat.
- Hooggeleerde heer Klein, Beste Jan, vanaf het eerste moment heb je me gestimuleerd om het onderzoek op te starten, er tijd voor vrij te maken (soms teveel) en uiteindelijk af te ronden.
- Hooggeleerde heer Bakker, Beste Jan, vanaf dag een heb ik het gevoel gehad dat we daadwerkelijk samen een klus gingen klaren.
- Hooggeleerde heer Kuipers, Beste Ernst, inspirator pur sang, ooit komt er ons ziekenhuis.
- Hooggeleerde heer Pols, Beste Huib, deze is voor jou: van half vol naar vol. Dank voor je bijna dagelijkse inspiratie om wetenschap, onderwijs en management te blijven combineren.
- Hooggeleerde heer Muffels, Beste Ruud, dank voor mijn start als je onderzoeksassistent in de wetenschap. Je eerste analyse dat ik niet geschikt zou zijn om mijn dagen te slijten als wetenschapper of als manager pur sang bleek een juiste te zijn. De combinatie geeft me het beste van beide werelden.
- Weledelzeergeleerde heer Wullink, Beste Gerhard, techneut met gevoel voor organisaties. Dank voor de samenwerking, de lessen in de logistiek en je drive om er iets van de maken in Rotterdam. Helaas heb je de overstap gemaakt naar de wereld van de adviseurs. Iets in mij zegt me (of hoopt tegen beter weten in) dat dit achteraf een jeugdzonde zal blijken te zijn en we onze succesvolle samenwerking echt kunnen vervolgen.
- Weledelgeleerde heer van Oostrum, Beste Jeroen, je eigenschap te blijven zoeken naar het enige juiste antwoord zorgt ervoor dat je voor mij een super klankbord bent waarbij ik soms met enige schroom mijn eigen analyse ter berde breng. Je bent een echte brug geworden tussen de exacte wetenschap en de soepele praktijk van het management in de zorg.
- Weledelgeleerde heer Meijerink, Beste Rien, dank dat ik altijd met vragen bij je terecht kon en eigenlijk nog steeds kan. Onze gesprekken, met overigens meer nieuwe vragen in plaats van antwoorden van jou kant, boden en bieden me steeds houvast voor een eigen antwoord.
- Weledelgeleerde heer Geerlings, Beste Willem, vanaf ons eerste kennismakingsgesprek, 's morgens om 07:00 samen met Huib rondom de ontwikkelingen binnen de Interne Geneeskunde, tot aan ons laatste gesprek bij een sluis in Voorburg bied je me steeds een nieuw en meestal ook ruimer perspectief op de mogelijkheden in de zorg. Overigens nooit onmogelijkeden.
- Hooggeleerde heer van der Maas, Beste Paul, Hooggeleerde heer van der Meche, Beste Frans, dank voor de kansen die jullie mij hebben geboden vanuit de Raad van Bestuur.
- Hooggeleerde heer Buller, Beste Hans, de openheid waarmee je mij vanaf onze eerste kennismaking en in onze latere gesprekken hebt benaderd heeft me geholpen en geïnspireerd bij de eerste stappen in het Beatrixziekenhuis.
- Hooggeleerde mevrouw Hazes, Beste Mieke, bijna gelijktijdig gestart in Rotterdam krijgt onze samenwerking nu vervolg in Gorinchem.
- Hooggeleerde heer Vissers, Beste Jan, Hooggeleerde heer de Vries, Beste Guus, jullie hebben een begrippenkader opgebouwd in de zorglogistiek, een stevige fundering waarop ik kon voortbouwen.
- Weledelgeleerde heer Peek, Beste Ben, je kennis van de zorg, je liefde voor de zorg en de zorg voor je medewerkers en collega's maakt je tot een bron van inspiratie.
- Hooggeleerde Hordijk, Beste Gertjan, Hooggeleerde Treffers, Beste Frits, bij jullie mocht ik de eerste schreden zetten in het management. Voor het vertrouwen dat daaruit sprak blijf ik altijd dankbaar.
- Weledelgeleerde Klauss Schmitt, na ons laatste gesprek op weg naar "beyond".
- Weledelgeleerde de Kort, Beste Pieter, in korte tijd heb ik veel van je kunnen leren. Dank voor de kans en het vertrouwen bij mijn benoeming in het Beatrixziekenhuis.
- Weledelzeergeleerde Eijkemans, Beste Rene, als het statistisch geweten van mijn onderzoek kwam je op de meest onmogelijke momenten met weer een andere kijk op de werkelijkheid. Toch bleek iedere keer weer dat dit de analyse een stap verder bracht. Echter meestal pas na enkele weken tobben.
- Weledelgeleerde heer Nguyen, Beste Tien, je blijft een mysterie voor mij maar ik kan me geen betere en loyalere collega voorstellen.
- Johan Hurink, Richard Boucherie en Nellie Litvak toegepaste wiskunde vanuit Twente is voor mij een inspiratiebron om steeds verder te blijven zoeken naar antwoorden op nieuwe vragen.
- Arjan van Hoorn, Ingrid Koens, Marleen van der Pol, Marieke van der Lans, Stijn Prinsen, Nelleke Pullen, Manon Vrielink, Marleen van Rijsbergen, Simone Veldhuis, Sjoerd Reterink, dank voor de bijdrage die jullie

hebben geleverd aan de ontwikkelingen binnen het ErasmusMC en het tot stand komen van onze onderzoekslijn.

- Hooggeleerde heer Poldermans, Beste Don, de efficiëntie in je werk is voor mij een voorbeeld.
- Weledelgeleerde heer de Waard, Beste Gerjan, dank voor de zondagen in de regen, dit boek en onze gesprekken op de fiets.
- Geachte leden van het Managementteam van de OK in het ErasmusMC, Beste Joleen, Patty, Theo, Antoine en Jeroen dank voor het bijna oneindige geduld dat jullie moeten hebben gehad met mij.
- Lieve Joleen, dank voor je wijze lessen in inspiratie.
- Lieve Francine, je was mijn geweten in de afgelopen jaren waar ik blind op kon varen.
- Geachte leden van het Stafbestuur, Beste Hielke, Heleen, Rene, Raymond en Ike dank voor jullie ontvangst en de samenwerking eigenlijk vanaf mijn eerste dag in jullie en ons ziekenhuis.
- Geachte Hoofden van het Beatrixziekenhuis, Beste Judith, Eric, Ronald en Arnold, zo ver als nu had ik vooraf niet verwacht maar door jullie wel gerealiseerd.
- Beste Hedzer, Judith, Jurrien, Jan en Yolanda jullie samenwerking gaf mij de inspiratie voor het afronden van dit boekje zonder dat jullie het zelf wisten. Dank!
- Geachte stafleden en medewerkers van het Beatrixziekenhuis, Beste collega's, vanaf de eerste dag heb ik me thuis gevoeld in jullie en ons ziekenhuis.
- Geachte medewerkers van Rivas, Beste collega's, het is een voorrecht om in jullie innovatie organisatie te mogen werken en hopelijk een bijdrage te leveren.
- Lieve Miriam, Debbie, Arjenne en Ine alleen door jullie ondersteuning in het afgelopen jaar is de combinatie van een nieuwe baan, een eerste directiefunctie en het afronden van een promotie mogelijk gebleken.
- Lieve Ellen, zelfs nu ik weg ben voel ik je steun (en die van William) nog iedere dag.
- Beste Winfried, je les op de Dominicaanse eilanden "ga nou eens een keer rustig zitten" ben ik nog steeds niet vergeten.
- Lieve Pa en Ma van Peperstraten vanaf dag één voel ik me bij jullie echt thuis en hebben jullie meegebouwd aan ons liefdevolle gezin.
- Weledelgeleerde heer van Eijk, Beste Martijn, onze jarenlange vriendschap is voor mij een blijvende bron van inspiratie en liefde.
- Lieve Pa en Ma zonder jullie was dit boekje er niet gekomen maar was ik ook niet wie ik nu ben, als zoon, broer, vriend, echtgenoot, vader maar ook als collega en onderzoeker.

- Lieve Robyn, Youri en Gijs iedere dag weer laten jullie mij zien wat echt belangrijk is in het leven. Mijn liefde voor jullie is oneindig maar groeit ook nog steeds.
- Geliefde Mariella, samen bouwen aan onze liefde, ons gezin en ons werk heeft ons gebracht waar we nu staan. Zonder jou heeft het leven voor mij geen waarde.

Curriculum Vitae

Mark Van Houdenhoven werd geboren op 16 november 1968 in Wageningen. Na het behalen van het VWO diploma op het Titus Brandsma Lyceum te Oss in 1987, diende hij in 1987 en 1988 bij de Koninklijke Instituut voor de Marine in Den Helder. In 1989 begon hij met de studie Economie op de Tilburg University. In 1993 werd hij assistent bij Prof. dr. Ruud Muffels, hoogleraar ouderbeleid, en heeft hij in opdracht van de provincie Noord Brabant en verschillende ouderenbonden, meerdere onderzoeken afgerond. In 1995 studeerde hij af op een studie naar de ontwikkelingen in het ouderenbeleid in Nederland.

Na zijn afstuderen werkte hij als onderzoeker en adviseur in de gezondheidszorg en was werkzaam bij verschillende zorgorganisaties. In zijn werk tot heden staat het echte gesprek tussen het primaire proces en de ondersteunende processen, tussen de zorgprofessionals en het management, centraal.

Vanaf 1998 heeft hij in het Universitair Medisch Centrum in Utrecht en in het ErasmusMC in diverse managementfuncties gewerkt. Vanaf 2003 heeft hij zich met name beziggehouden met het optimaliseren van zorglogistieke processen in het ErasmusMC in samenwerking met dr. Geert Kazemier op het operatiekamercomplex, met prof. dr. Jan Bakker op de Intensive Care en met prof. dr. Jan Klein bij de afdeling Anesthesiologie en de gebruikende specialismen binnen het gehele ErasmusMC. In deze periode zijn meerdere onderzoeken gestart rondom zorglogistiek waarbij het combineren van medische kennis, wiskundige kennis en management kennis centraal stond. Dit heeft geresulteerd in diverse wetenschappelijke publicaties.

Vanaf 1993 heeft hij gepubliceerd in verschillende vaktijdschriften op het gebied van bedrijfsvoering en zorg. Vanaf augustus 2006 is hij werkzaam als directeur van het Beatrixziekenhuis onderdeel van Rivas zorggroep te Gorinchem.

Colofon

Uw mild gelaat
blijft over 't kind gebogen;
het wordt voor U geboren en getogen,
vervult zijn wegen naar Uw raad.

(Gez. 335, vers 7)

TEKST Mark Van Houdenhoven (markvanhoudenhoven@wanadoo.nl)
FOTOGRAFIE Dreamstime, Mark Van Houdenhoven, Karel Tomeï (cover)
VORMGEVING Bram Vandenberge (bram@igraph.be)
DRUK www.balmedia.nl

Dit boek is geproduceerd door Uitgeverij Scriptum
in Schiedam (www.scriptum.nl)

ISBN 978 90 5594 578 8