Management-By-Objectives in Healthcare

Traberg, Andreas; Jacobsen, Peter

Publication date:
2011

Document Version
Publisher's PDF, also known as Version of record

Management-By-Objectives in Healthcare

PhD thesis 3.2011

DTU Management Engineering

Andreas Traberg
April 2011

DTU Management Engineering
Department of Management Engineering

Management-By-Objectives in Healthcare

Ph.D. dissertation by Andreas Traberg

Ph.D. dissertation, Technical University of Denmark
DTU Management Engineering

Supervisor: Associate Professor Peter Jacobsen
Department of Management Engineering
Technical University of Denmark
Kgs. Lyngby, Denmark

Opponents: Professor Dariusz Ceglarek
The International Digital Laboratory
University of Warwick
Coventry, United Kingdom

Chief Physician, MD. Kenneth Jensen
Anaesthesiology unit
Bisbebjerg Hospital
Copenhagen, Denmark

Associate Professor Lauge Baungaard Rasmussen (Head of Committee)
Department of Management Engineering
Technical University of Denmark
Kgs. Lyngby, Denmark

The presented dissertation is part of the acquisition of a Ph.D. degree

Title: Management-By-Objectives in Healthcare

Published by:
Department of Management Engineering
Technical University of Denmark
2800 Kgs. Lyngby
Denmark
Phone: (+45) 45 25 25 25

ISBN: 978-87-92706-07-2

Dansk resumé

I takt med den hastige udvikling på de medicinske, teknologiske og organisatoriske områder i sundhedssektoren, implementeres der løbende nye initiativer til at forbedre kvaliteten af de sundhedsfaglige ydelser. Til at evaluere hvorvidt disse tiltag har den ønskede effekt, er brugen af målstyring blevet almindelig praksis i moderne sundhedsinstitutioner. De seneste år er kvalitetsindikatorer, akkrediterings audits, spørgeskemaundersøgelser, arbejdspladsvurderinger, osv. blevet en integreret del af dagligdagen for det sundhedsfaglige personale. Langt de fleste af målemetoderne er velgennemtænkte og veludførte, men alligevel udgør de samlet et problem når de anvendes samtidigt. For lederne sløres overblikket, og målemetoderne pålægger personalet en stigende administrativ byrde. I modsætning til hensigten, så skaber den øgede informationsmængde mindre gennemsigtighed og mindre overblik for lederne i sundhedssektoren. Det resulterer i at flere af evalueringerne ikke finder anvendelse som praktisk beslutningsstøtte.

Forskningsprojektet har derfor haft til hensigt at designe en mere helhedsorienteret målstyringsmodel, der kan medvirke til at det ledere og operationelt personale i højere grad bliver i stand til at overskue performance i relation til de organisatoriske forventninger. Projektet konkluderer at integration af alle betydende indikatorer i et "Performance Regnskab", skaber overblik og gennemsigtighed, uden at detaljerne i de enkelte målinger forsvinder. Performance regnskabets design betyder at de specifikke målinger som lederne finder anvendelse for i deres afdeling kan inddrages, hvilket sikrer fyldestgørende informativ støtte til beslutningsprocesserne på den enkelte afdeling. Regnskabet letter identifikationen af operationelle problemområder, og giver dermed beslutningstageren et pålideligt informationsgrundlag at handle ud fra. Performance regnskaberne har værdi i en travl hverdag, hvor det administrative arbejde tager mere og mere kostbar tid fra det sundhedsfaglige arbejde.

Afhandlingen indeholder fem videnskabelige artikler og en sammenfatning af de vigtigste bidrag og konklusioner fra disse. To af artiklerne er præsenteret på videnskabelige konferencer, og tre er fremsendt til videnskabelige tidsskrifter. Artiklerne beskriver den udvikling som projektet har været igennem, hvor forskellige løsningsforslag løbende har været forsøgt. Sammenfatningen indeholder mere detaljerede beskrivelser af den videnskabelige tilgang som har præget projektet.

Summary

Concurrent to the hasty development within the medical, technological and organizational areas of healthcare, new initiatives are continuously implemented to improve quality of delivered care. To evaluate the effect of these initiatives, the application of performance measurement has become common practice for modern healthcare organizations. During the last decade, vast amounts of quality indicators, accreditation audits, satisfaction surveys etc. have become an integrated part of healthcare professionals' daily work. Most of these measurement structures are well documented and well executed; collectively, however, they pose a significant drawback. The vast selection of self-contained initiatives limits the overview for decision makers and imposes an escalating administrative burden on operational staff members. Contrary to the initial objective, the expanding informational burden limits the overview and transparency for healthcare decision makers; as a result, well-documented initiatives fail to become integrated support in operational decision-making processes.

This research work has thus striven to design a holistic Management-By-Objectives framework that can enable managers and operational personnel to assess performance in relation to the organizational expectations. The work concludes that by integrating all meaningful indicators into a "Performance Account", an overview is established without losing the strength of detailed measures. The design of the Performance Account signifies that managers are able to incorporate those indicators they find useful in their department, and thus secure sufficient informational support for the department's decision-making processes. The Performance Account thereby eases the identification of areas suited for corrective actions, and provides the decision maker with a reliable informational foundation. The account has merits in a hectic environment, where the administrative burden consumes important time from the clinical work.

The dissertation is composed of five scientific articles, together with a synopsis describing the most vital contributions and conclusions. Two articles have been presented at international scientific conferences, and three articles have been submitted to scientific journals. The papers present the development of the research study and successively describe the proposals. The synopsis describes in detail the scientific approach that has guided the study.

Preface

This dissertation is submitted to DTU Management Engineering, Technical University of Denmark, in fulfillment of the requirements for acquiring the PhD degree. The work has been supervised by Associate Professor Peter Jacobsen. The dissertation consists of a recapitulation of the research study and a collection of five research papers prepared during the period from May 2008 to April 2011. Generally, American spelling rules are used in this thesis. All the thesis publications have been submitted under the name `Andreas Traberg`.

Andreas Traberg, Kgs. Lyngby, Denmark, April 2011

Acknowledgements

Hospital of Southern Jutland deserves much gratitude, because of its funding commitment and willingness to allow me be involved at the hospital. Many thanks go to the employees at Southern Jutland Hospital, who helped and supported me during this project. Especially, I would like to thank Jytte Nielsen, Dorte Juhl and Inger Fog, who not only showed patience with my lack of knowledge and difficulty in grasping the concepts, but who constantly supported my research.

Thanks to my fellow PhD researchers. Gudmundur Oddsson, Pelle Jørgensen and Inger Siemsen (IMS) have continuously supported me throughout the study. Thank you all, it has been a pleasure to work with you.

My thanks go to President Keiko Suzuki from Koenji Suzuki Clinic, and President Toshiaki Suzuki and Atsushi Suzuki from the Asagaya Suzuki Clinic, and also to Mr. Satoshi Suzuki at Tokyo Women´s Medical University, Mr. Yukimitsu Satoh from Japan Association for Development of Community Medicine, Professor Hideyuki Sakurai from University of Tsukuba, and Mr. Kiyoshi Yasuoka from the Proton Medical Research Center at Tsukuba University Hospital. Your help is greatly appreciated.

I would additionally like to express my sincere gratitude to Professor Kenji Itoh from Tokyo Institute of Technology. Thank you for your amazing ability to lead me through my stay in Japan, both scientifically and, just as important, also socially and culturally. It has been a tremendous honor to be working with you, and a pleasure to meet your loving family. I genuinely hope that we can maintain our friendship for many years to come.

Also a special thanks to Christina Schnell Christiansen and Else Nalholm, who have aided the process of the study, and been enjoyable and helpful colleagues during the three years at DTU Management.

Nadia Duthiers part in this study have been of remarkable importance. Your winning character and commitment have been an inspiration to me. Besides contributing to my papers, you has been involved in discussing proposals and evaluating the conclusions. Thank you so much.

Finally, I thank Peter Jacobsen for being a constant source of inspiration - the entire way through my academic education. You have been my primary mentor for almost 6 years now, and the most important reason why my education has been successful. I owe you the greatest thanks of all.

Table of Contents

Table of Figures

Table of Tables

Chapter 1 - Introduction

This introductory chapter provides an empirical and theoretical framing of performance measurement systems in the context in which they are applied in this thesis. First, the chapter briefly presents the motivational basis for this research study, and supports this scientifically by reviewing state-of-the-art literature. Emphasizing the focus in the current body of knowledge, this review highlights the historical development in the literary focus. Subsequently, a theoretical elaboration of the fundamental reasons for applying performance measurement systems is provided. The purpose is to provide the reader with a clear understanding of the scientific core in this thesis.

1.1 A quick look at modern healthcare performance measurement

Modern healthcare is characterized by increasing demands for individualized high-quality services, an intensified patient inflow and technological innovations, all resulting in pressure on health expenditures (Strandberg-Larsen et al. 2007;World Health Organization 2008). This trend has led to a growing need for reliable performance evaluation tools to guide the increasingly more complex decision-making processes (Swaminathan, Chernew, & Scanlon 2008). But quality and performance of healthcare services are often difficult to quantify; hence several measurement techniques are applied throughout the healthcare sector (Mohammadi, Mohammadi, & Hedges 2007). Consequently, performance evaluation has developed into a multi-faceted concept, focusing on a variety of aspects such as safety, effectiveness, appropriateness, timeliness and responsiveness of services, along with measures of efficiency and equity (Wait & Nolte 2005). In literature, clinical practice (Hartswood et al. 2002), work environment (Jones et al. 2009), patient satisfaction (Kutney-Lee, McHugh, & Sloane 2009), mortality (Barros 2003), and surgical performance (Treasure et al. 2002) are just a few of the topics within the extensive scientific work conducted in healthcare performance evaluation. Because of a wide acceptance of the strength of both broad high-level strategic measures and detailed clinical measures, most healthcare facilities are adopting both types within their evaluation procedures. This results in inhomogeneous performance measurement systems, which often produce vast amounts of unstructured performance information.

The current challenge is therefore to present performance information in a way that provides the necessary foundation in complex decision-making, without overburdening the decision-makers confronted with this information. This challenge has continued to trouble scientists and decision-makers for decades (Neely 2005), and agreement on how, where and when to evaluate is seldom achieved. This chapter aims to provide an understanding of why Management-By-Objectives is still scientifically interesting, emphasizing the theoretical gaps and practical complications.

1.2 Motivational basis

Management-By-Objectives is a domain filled with simplified assumptions concerning the usage of performance measures (Neely & Al Najjar 2006). Such claims as "what you measure, is what you get" are common and implicitly indicate that performance measurement is a method to control organizations. The growing scientific and practical interest suggests however that there might be more to the matter than this, since the term Management-By-Objectives implies that the intention is to use organizational objectives as guidelines for the management of operations. Approximately 20 years ago Eagle and Davis stated that

> …you cannot manage it until you have a way to measure it, and you cannot measure it until you can monitor it.
>
> (Eagle & Davies 1993)

This statement implies that construction of performance metrics is a necessity for any organization's ability to achieve its objectives, a seemingly simple endeavour that starts with decisions on what to measure, then identifies the proper measures along with their respective data sources, and ends with the analysis and understanding of the results (Loeb 2004). However, as with most assumedly straightforward activities, the difficulty lies in the details. In healthcare, a significant part of the challenge derives from the contrasting viewpoints and variable perspectives represented among the key stakeholders, e.g. patients, employees, relatives, authorities etc., which is why deciding on organizational objectives becomes a difficult task. As a result, numerous methods are used, which are acknowledged as tools for evaluating all the different aspects of healthcare services in relation to different stakeholders (Mohammadi, Mohammadi, & Hedges 2007). It is evident that healthcare organizations need to move beyond a medical view and embrace a more holistic approach to healthcare performance. Accurate diagnosis and proper treatment are not enough; stakeholders need performance in all services (Elleuch 2008). In an attempt to cover all aspects of performance in healthcare organizations, performance measurement systems have become broader in scope and their use more widespread (Curtright, Stolp-Smith, & Edell 2000;Lim, Tang, & Jackson 1999), which leaves each organization with the task of defining their own measurement system, customized to its own setting and with little chance of covering all aspects satisfactorily. Consequently, healthcare is still an sector where no framework is accepted unanimously as the tool for defining and measuring, the quality and performance of healthcare services (Ondategui-Parra et al. 2004).

Thus, the intense employment of various evaluation tools has created another concern for healthcare practitioners. Practitioners experience the cost of this concentrated focus on evaluation in heavy administration, as well as confusing and overwhelming feedback. Performance indicators, quality audits and accreditation standards are gradually becoming fundamentals in the vocabulary of most healthcare professionals. Contradictory to the initial objective, the expanding load of registrations, reports, standards, budgets etc. has limited the individual's ability to comprehend all the information provided. Decision makers are constantly faced with a vast selection of indicators, which in some cases leads to administrative fatigue and information overload (Bovier & Perneger 2003). In some cases, the expansion of the administrative burden does not provide the desired operational value but only more administrative work. As a consequence, performance information is not used as proactive decision support, but as retrospective information. With good reason, operational decision makers are not able to transform all the performance information into effective actions of improvement. Decisions are thereby not always based on objective data but instead on more subjective assessments and risk being out of line with overall organizational objectives. As a result, proactive strategic decision making is moving away from the operational levels to the strategic levels of healthcare organization, thus prolonging the organization's ability to make corrective adjustments and thereby delaying necessary changes.

Assuming that reliable, valid, comprehensible performance information is a necessity in order for healthcare organizations to reach their objectives, the construction of holistic Management-By-Objectives systems is a prerequisite. Therefore, holistic construction of performance measurement systems is a step towards improving the healthcare sector's capabilities. The motivation for this work has been to contribute to the advancement of this process, by focusing on holistic performance measurement with point of departure in strategic objectives, and transforming these into information for operational decision making.

1.3 Setting the empirical scene

Healthcare performance measurement is at least 250 years old (Loeb 2004). While the vocabulary and application of measures and structure have changed, the intent – i.e. obtaining data and providing decision support – has changed little over the years. With respect to terminology, performance measurement is applied in several different contexts, with different meanings in the literature. In its most narrow sense, performance measurement refers solely to the process of measurement. Performance measurement is by this definition limited to applying various techniques for generating performance data, with the measurement process leading to a set of qualitative and/or quantitative data. Additionally, performance measurement is referred to in the sense of performance reporting, e.g. indicator-based league tables, annual reports, internal communications etc. (Greiling 2006). In a wider perspective, some authors even refer to performance measurement with the term "performance management", which is somewhat misleading. Since there is some semantic confusion in the literature concerning performance measurement, explanations of definitions and wording are provided in section 1.5 to provide the reader with insight into the use of terms in this work. Regardless of slight semantic confusion in the field of performance measurement, it is historically considered to be an integral part of the strategic control cycle and a steering instrument for management (Neely et al. 1994). Performance measurement is used to design/modify or even to control an existing system. At all organizational levels, priority setting, system planning, financing and resource allocation, professional recognition and overall quality management often become important aims of modern performance evaluation. As such, the specific terms are a priori or a posterior evaluation respectively, either to assist decision making or to evaluate the quality of recent decisions (Lauras et al. 2010).

The continuing scientific interest in healthcare performance measurement results from the domain's dynamic, unpredictable, ambiguous and uncertain environment. As healthcare systems become more complex, so does the task of developing methodologies that can align organizational objectives with performance measurement (Kocakülâh & Austill 2007). Healthcare decision makers live in a world of conflicting goals with many consequential dilemmas. To choose one side of a dilemma (e.g. enhance production to achieve quota targets) can create a hidden condition in the system on the other side of the dilemma (e.g. quality of care might decrease). It is advocated that by adopting the concepts of performance measurement, healthcare organizations can orient themselves to

systematic evaluation with stakeholder focus, key-process monitoring, data-driven techniques, and team empowerment to address these obvious dilemmas (Klein, Motwani, & Cole 1998). The following sections describe how history has shaped the challenges, and how these challenges have been met by balanced measures and benchmarking approaches.

1.3.1 The evolution of performance measurement challenges

Performance measurement has continued to attract considerable attention among both academics and practitioners, and various ways of framing performance measurement into wider management concepts have been developed (Kollberg, Elg, & Lindmark 2005). While healthcare organizations bear many similarities to industrial organizations and can be subjected to the same forms of analysis, evaluation and improvement, they have some unique factors which challenge the industrial way of perceiving performance measurement. Healthcare involves multiple professional- and stakeholder groups, low reliability processes (Resar 2006), macro and micro system interactions (Mohr, Batalden, & Barach 2004), fragmented leadership, diffuse power and multiple goals (Lozeau, Langley, & Denis 2002). Managing and measuring performance become exceedingly complex as healthcare institutions evolve into integrated health systems comprising hospitals, outpatient clinics and surgery centres, nursing homes, and home health services (Curtright, Stolp-Smith, & Edell 2000), which requires the concept of performance to be interpreted as a multidimensional concept. These preconditions have naturally shaped the challenges that scientists and practitioners have dealt with for decades. Although performance measurement has been a topic of interest for many years, the challenges connected with the measurement of healthcare services have not been altered significantly for some time (Loeb 2004). There are two primary challenges that have been focal points in the literature: 1) the selection of proper measures; and 2) how to support multi-criterion decisions. These represent elements in the meta-hypothesis, which states that successful hospital performance depends on both the clinical and strategic aspects of care in order to provide a satisfactory basis for healthcare decision making.

What to measure

The very first edition of *Administrative Science Quarterly* contained a paper entitled "Dysfunctional Consequences of Measurement" (Ridgway 1956), which explored the strengths and weaknesses of single, multiple and aggregated performance measures, regretting the "strong tendency to state numerically as many as possible of the variables with which management must deal". Even before this, Chris Argyris (1952) stated that managers claimed to "feed machines all the easy orders at the end of the month to meet [their] quota". In continuation hereof, Dweiri and Kaplan argue that diffuse measurements may result in redundant and incompatible performance measurement systems (Dweiri & Kaplan 2006). Here, the essential factor is to consider the impact of each component of a performance measurement system, rather than just compiling random measures (Lauras, Marques, Gourc, & Lauras 2010).

The claim in these papers is similar – that often the selection of performance indicators does not reflect overall business goals, which are either too complex to collect or too

unwieldy to analyze. Furthermore, the performance data that is collected is often considered inconsistent, incomplete and inaccurate, due to the difficulties in selecting the proper metrics. Consequently, failure to make quick corrective action when performance levels begin to drop, and failure to modify management systems as the organization's requirements change, are due to poor selection of proper measures (Norcross 2006). Measures can be seen by staff to be irrelevant, unrealistic, inappropriate or unfair, and in some cases even counterproductive and in some cases, making employees respond to measures in a very different way than intended, leading to poorer service all round (Moullin 2004). Although a basic requirement, performance measurement system almost universally work poorly and are viewed negatively by both managers and managed (Furnham 2004). One-dimensional measures, wrong focus and blurred representations have consistently been central criticisms of the scientific topic of performance measurement. The challenge is that poor selection of metrics constitutes a potential threat to the decision standard, because the information basis does not fulfil the needs of the users; therefore, the selection and construction of measures are argued to be of extreme importance.

Priorities in multiple-criteria measurement

The second key challenge is to know how to interpret or understand the information produced. As described above, the dilemma the decision maker faces constitutes setting difficult priorities in relation to the differing interests of the various stakeholders. If measurement structures are not compiled to align strategic, team and individual goals, the decision maker alone must make these priorities. Thus, operational decision makers must make decisions based in their own perception of importance. Mintzberg (1994) identifies in his book, *The Rise and Fall of Strategic Planning,* three premises behind strategic planning: (1) that strategy making should be a controlled, conscious and formalized process, decomposed into distinct steps and supported by analytical techniques; (2) that responsibility for the overall process rests in principle with the Chief Executive Officer, although in practice, execution is delegated to staff planners; (3) that strategies emerge from this process fully developed, often as a position, to be implemented through detailed attention to objectives, budgets, programmes and operating plans of various kinds (Mintzberg 1991). This idea of a top-down analytical process that produces a fully integrated strategy is theoretically standard for management promoters. For healthcare practitioners, however, clear-cut structures of interlinked decisions about what activities to pursue and how to pursue them constitute idealism. While strategic planning deals with macro-level decisions, operational management is concerned with micro-level decisions, where dilemmas and priorities are not always are clear as portrayed by Mintzberg. But the aim is obvious, and therefore, structures for multiple-criteria decisions are becoming a more and more intensified scientific topic.

1.3.2 The acknowledged solution: balanced sets of measures

The two themes – i.e. the difficulty in selecting proper measures, and deciding what is important – continue to be the foremost debated topic in recent publications. Indeed, the most recognized scientific response appears to be 'rediscovering' or 'adapting' Ducker's

1956 suggestion of balanced measurement systems (Drucker 1956). Through the 'eighties and early 'nineties, authors suggested different performance measurement models suitable for balancing objectives, e.g. Performance Pyramid (Lynch & Cross 1991), Results-Determinants Framework (Fitzgerald et al. 1991), Performance Measurement Matrix (Keegan, Eiler, & Jones 1989), and the most cited, Balanced Scorecard (Kaplan & Norton 1992). Twenty years ago, there would have been little mention of non-financial performance in an organization's strategic reports (Neely 1999). Recently, however, there have been far more explicit reports about the link between financial and non-financial dimensions of performance, especially in healthcare. The introduction of the Balanced Scorecard started reformation in the area of performance measurement. Today, it is broadly accepted that performance must be defined in relation to explicit goals that reflect the values of various stakeholders (such as patients, professions, regulators etc.). Most healthcare organizations are adopting multi-dimensional performance assessment systems to guide operations toward fulfilment of their organizational objectives. The scientific literature suggests that a good Balanced Scorecard contains both leading and lagging measures and indicators. Lagging measures (outcomes) tell what has happened; leading measures (performance drivers) predict what will happen – e.g. employee satisfaction surveys might be a leading indicator for employee turnover, while employee turnover is itself a lagging indicator. These indicators should also establish either correlation relationships across perspectives, or more strongly, cause-and-effect relationships among leading and lagging measures (Evans 2004).

Balanced scorecard 'look-alikes' are increasingly being implemented as a guiding structure, especially as an expanded set of performance indicators, where organizations customize their scorecard to their particular settings (Chen et al. 2006). Although the original Balanced Scorecard focuses on four dimensions (internal business processes, innovation and learning, customers, and finance), the practical implementation of balanced scorecard framework is still criticized for a too intense focus on profit and process outcomes and too little focus on people and the organizational cultures in which they work. In public health, this issue roots itself in the environment, where political agenda defines the public services' desirable outcome or output. The definition of the desirable output or outcome is often under strong political influence, so focus on economic sustainability is a vital priority (Greiling 2006). This is a serious limitation, particularly in the healthcare industry, where employee knowledge, skills and commitment are critical, not only for organizational performance but also for saving lives (Wicks, St Clair, & Kinney 2007).

1.3.3 External influence on state-of-the-art proposals

Historically external influences have had a strong effect on healthcare performance measurement systems. Thus, healthcare performance measurement systems are usually developed at a rather high organizational level, with limited objectives at organizations' tactical and operational levels (Brumback 2003). This is somehow understandable, because healthcare providers are obligated to comply with national guidelines. But this

approach conflicts with the empirical proposals presented, which promote a holistic approach fitted to all organizational levels. The practical tendency to focus on strategic measures has put a significant mark on recent publications, where strategic performance is in focus (see Table 1).

	Strategic level	Tactical level	Operational level
(Pitt 1999)	✸		
(Griffith, Alexander, & Warden 2002)	✸		✸
(Smith 2002)	✸	✸	
(Hartswood, Procter, Rouncefield, & Slack 2002)			✸
(Radnor & Lovell 2003)	✸	✸	✸
(Veillard et al. 2005)	✸		✸
(Yang, Cheng, & Yang 2005)	✸		
(Kollberg, Elg, & Lindmark 2005)	✸	✸	✸
(Schmidt et al. 2006)	✸		
(Cheng & Thompson 2006)	✸		
(Byrne 2006)	✸		
(Dieleman et al. 2006)			✸
(Arah et al. 2006)	✸		
(Dummer 2007)	✸		
(Baker, Beitsch, & Landrum 2007)	✸		
(Lega & Vendramini 2008)	✸	✸	
(Buetow 2008)	✸		
(Rochette & Féniies 2008)	✸		
(Crump 2008)	✸	✸	
(Moullin & Soady 2008)	✸		

Table 1. Focus in performance management literature

Besides the strategic focus, external influence has likewise influenced the techniques discussed in literature. In general, five generic types of measurement techniques cover all known approaches towards measuring organizational performance (Shaw 2003): 1) Surveys of customer experience are used to describe the organization's performance in the eyes of the customer. Surveys differ in size, from local 'paper-and-pencil' surveys to multi-national opinion polls and can be performed by governmental institutions or independent institutions. 2) Third-party assessments are often linked to certification or accreditation by international standards. These assessments are requested by the organization itself and performed by external auditors. 3) Statistical indicators are used as a guideline for performance according to preset criteria. Indicators can be either defined exclusively by an organization or by governmental institutions as a part of a national/international performance report system. 4) Internal assessments are performed by dedicated staff, trained in evaluating the organization according to specified standards. They are often conducted on the initiative of the individual organization; they are most common in large organizations, and seldom appear in small organizations. 5) National inspections are performed by a legal authority according to a set of predefined standards. National authorities use report systems and inspections as a way to verify whether healthcare

providers meet national standards. National inspection is not implemented by organizations but is a governmental outline for performance evaluation. Since these five generic building blocks all constitute strengths and weaknesses, it could be assumed that they are all addressed evenly in literature, but this is not the case. There are clear tendencies that authors treat the issue of statistical indicators more extensively than the other four (see Table 2).

	Surveys of customer expience	Third-party assessment	Statistical indicators	Internal assessment	National inspections
(Pitt 1999)					
(Griffith, Alexander, & Warden 2002)					
(Smith 2002)					
(Hartswood, Procter, Rouncefield, & Slack 2002)					
(Radnor & Lovell 2003)					
(Veillard, Champagne, Klazinga, Kazandjian, Arah, & Guisset 2005)					
(Yang, Cheng, & Yang 2005)					
(Kollberg, Elg, & Lindmark 2005)					
(Schmidt, Bateman, Breinlinger-O'Reilly, & Smith 2006)					
(Cheng & Thompson 2006)					
(Byrne 2006)					
(Dieleman, Toonen, Toure, & Martineau 2006)					
(Arah, Westert, Hurst, & Klazinga 2006)					
(Dummer 2007)					
(Baker, Beitsch, & Landrum 2007)					
(Lega & Vendramini 2008)					
(Buetow 2008)					
(Crump 2008)					
(Rochette & Féniies 2008)					
(Moullin & Soady 2008)					

Description	Indicator
Not mentioned	
Mentioned lightly	
Described	
Described in detail	

Table 2. Measurement techniques in healthcare literature

This is founded in the intense external pressure for quantifiable metrics that can determine the level of performance (Arah, Westert, Hurst, & Klazinga 2006), which makes sense in a sector where national standards govern how hospitals should be performing. And because hospital management is evaluated according to these standards, indicators are used as the backbone of the performance management system. It is therefore not surprising that a majority of publications treat the subject of strategic objectives and statistical measures.

1.3.4 Benchmarking healthcare performance

The intense focus on statistical indicators in the national context also incites the use of benchmarking initiatives as a way of evaluating differences between hospitals. Benchmarking was translated for use in the public sector from the management field and is broadly defined as:

> ...the comparison of similar systems or organizations based on a recognized set of standard indicators
>
> (Wait & Nolte 2005)

Benchmarking within healthcare systems is due to the increasingly intense performance focus, particularly in the form of setting targets for improvement initiatives. This results in hospitals being evaluated not only on their own actual performance but also on their performance in comparison to other hospitals. In addition, healthcare sectors as a whole are evaluated, where such institutions as Organization for Economic Co-operation and Development (OECD) and World Health Organization (WHO) assess hospital performance across national borders. This engagement stems from the "New Public Management" culture, which has transferred expectations for public accountability of healthcare services and encouraged the development of benchmarking initiatives (Nutley & Smith 1998).

Indeed, several authors suggest that it is a complex practice to implement benchmarking initiatives fairly, particularly within a healthcare environment, since it is based on the measurement of diverse performance conceptions that aim to identify best practices. There is a growing literature highlighting the measurement limitations of existing indicator systems in terms of the validity and reliability of measures collected (Hurst & Jee-Hughes 2001;Musgrove 2003). Limited data availability and lack of uniformity of data across different settings plague most benchmarking initiatives. Furthermore, since national legislation limits the usability of these initiatives, they traditionally have been used solely to portray tendencies in modern healthcare.

1.4 Framing the thesis theoretically

Moving from the empirical aspect of Management-By-Objectives, this section tries to elevate the presentation to a more theoretical level. The theoretical aim of performance measurement is to:

> …translate raw data material into decision support information, enabling managers to make necessary corrective actions in the pursuit of organizational excellence.
>
> (Folan & Browne 2005)

Decion makers need ways to monitor changes over time, and the method is the construction of numerical representations of the organizational performance. Lord Kelvin expresses the objective for quantifying matters to gain knowledge;

> When you can measure what you are speaking about, and express it in numbers, you know something about it ...[otherwise] your knowledge is of a meager and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely in thought advanced to the stage of science.
>
> (Lord Kelvin 1824-1907)

In the following, a deeper theoretical description of some underlying aspects of performance measurement systems is presented. Apart from a theoretical framing of performance measurement systems, this section focuses on two central concepts, Reduction & Amplification and Remoteness & Displacement, which are key notions in understanding performance measurement systems.

1.4.1 Framing performance measurement systems

From a theoretical point of view, performance measurement systems can be seen as a multi-criteria instrument based on expressions of performance (Lauras, Marques, Gourc, & Lauras 2010). Performance measurement provides the basis on which an organization can assess how well it is progressing towards its predetermined objectives, helping to identify areas of strengths and weaknesses, and decide on future initiatives that aim to improve organizational performance. Max Moullin formulates the design of performance measurement systems as follows:

> Performance measurement systems are evaluating how well organizations are managed and the value they deliver for customers and other stakeholders.
>
> (Moullin 2005)

This quote implies solely the quantification process of measurement related to a goal or target (effectiveness, efficiency). The collection, computation, and use of performance measurement are however excluded from Max Moullin's definition; therefore, a broader definition of the subject, which considers the whole measurement process from collection to the final usage in managerial work, is more appropriate in relation to the initial motivation of this particular study. In this thesis, performance measurement is therefore theoretically defined as:

> The process of collecting, computing, and presenting quantified indicators for the managerial purposes of following up, monitoring, and improving organizational performance.

The search for ways to improve performance measurement systems has resulted in a number of step-by-step proposals, most of which have been designed to respond to particular aspects of performance measurement systems. (Keegan, Eiler, & Jones 1989) outline three distinct steps for developing performance measurement system: (1) defining strategic objectives of the firm and deciding how they can be translated into divisional goals and individual management actions; (2) deciding what to measure; and (3) integrating the performance measurement system into management thinking. (Bititci, Turner, & Begemann 2000) identified that a performance measurement system requires the following characteristics: 1) sensitivity to changes in the external and internal environment of an organization; 2) reviewing and reprioritizing internal objectives when the changes in the external and internal environment are significant enough; 3) deploying changes in internal objectives and priorities to critical parts of the organization, thus ensuring alignment at all times; and 4) ensuring that gains achieved through improvement programmes are maintained. To roughly condense these proposals – they touch upon the most critical issues in the development and implementation of performance measurement systems, namely determining "what to measure, and how to react to changes". The aim is to present a precise picture of the organization, as a foundation for the decision maker in initiating corrective actions derived from strategic objectives.

1.4.2 Reduction and amplification

Indeed, this concept in itself constitutes a theoretical challenge. Since a performance measurement system is a method to see the 'world' and thereby understand or interpret the 'world' in a given context, this context must be decided and agreed upon by all members of the organization. Portraying the world in all its complexity is a practical absurdity, but just as important, it would probably be meaningless even if possible. Performance measures instead represent a condensed view of the 'world', and use this fraction of reality as a way of indicating tendencies in a given context.

> Representation reproduces the events and objects of the world in a curtailed and miniaturized form so that they can be more easily engaged by the mind and body.
>
> (Cooper 1992)

Thereby, the 'world' is reduced to a representation (e.g. numbers, graphs, league tables etc.), which informs the decision maker about a given reality. This is done by amplifying particular aspects of the 'world' and leaving out other aspects – i.e. deciding what to measure. When reducing the complexity of the world, decision makers are provided with focused information that accordingly should guide their decisions. Consequently, the

representation is what is important, not the 'world' itself, because the informational basis in front of the decision maker is a representation and not the world itself.

The theoretical challenge is of course to establish what is important. In other words, what are the organizational objectives? If priorities are not specified, the risk of representing the 'world' in an incorrect manner is evident. The designer should therefore select structure and components for a performance measurement system that represents the organization as it is to be understood. Furthermore, this should be tied to the context (the nature of decisions) that should guide the decisions. For example, if waiting lists are determined to be an important indicator for a hospital, then the decision maker needs to know that longer waiting time is bad and shorter waiting time good in order for the decision maker to make the right decisions. This simplified example presents no problem for most healthcare practitioners, but more complex indicators may cause difficulties if the context is not specifically addressed. For example, in relation to equipment utilization, where high utilization might indicate low flexibility and low utilization may be a result of poor planning, the decision maker needs to know the organization's aims. Indeed, if the right aspects are not represented in the right context, decisions are bound to be taken on an invalid basis. This does not necessarily mean that decisions are bound to be wrong or inappropriate, but the supportive basis is incomplete in the given context in which it is applied.

The noble challenge is to pinpoint the organizational objectives and translate them into measures of performance. This identification also defines the boundaries of the organizational domain, in other words 'what is out of bounds' in terms of decision support. Not everything is important for the individual decision maker. This would often depend on the organizational level at which the measurement system is applied. The governing boundary for what is important and what is not therefore defines the decision maker's area of responsibility. Theoretically, everything that is not within the area of responsibility would be considered unimportant. This is why separated, specified measurement structures are considered to be the most suitable approach, as they provide specific information about the given context in which they are to be applied (Bourne et al. 2008;Evans 2004;Norcross 2006).

1.4.3 Remoteness and displacement

But why try to portray the 'world' in a reduced version? The explanation lies in the ability to displace decisions from action. By representing performance in a reduced view, the decision is no longer bound to the place of action. Disconnecting decisions from actions makes it possible for managers to manage beyond a physical premise. Robert Cooper elaborates on the managerial use of condensation in his 1992 paper:

> Administrators and managers, for example, do not work directly on the environment but on models, maps, numbers and formulae which represent the environment; in this way, they can control complex and heterogeneous activities at a distance and in relative convenience of a centralized work station.
>
> (Cooper 1992)

Though disconnecting decisions from actions, the completeness and validity of the information becomes even more important. Information would increase with unpredictability in order to compensate for the insecurity. Unless the decision-maker fully trusts the information, the performance information becomes theoretically obsolete.

Even though the disconnection of decisions and actions is a practical challenge, it has become a necessity in modern management. Since areas of responsibility are expanding at present, so does the need for decisions disconnected from actions. Adrian Furnham elaborates on the modern manager's most vital role in large organizations:

> The ability to, and necessity for, evaluating the performance of others and using this information to shape individual and organizational outcomes is one of the central functions of management.
>
> (Furnham 2004)

This signifies that it is a premise for managers to make decisions on the basis of performance information. Therefore, without any physical relation to the action itself, the manager needs to be able to take corrective action; therefore, careful construction of performance metrics is an absolute necessity.

1.4.4 The important nature of the decision maker

> Science measures objects objectively, but interprets the significance of the measurements subjectively.
>
> (Boyd & Gupta 2004)

Because of the diversity of influences involved in decision making, there are no set laws to characterize in fine detail the structures that apply in every decision (Saaty 2008). This constitutes a theoretical challenge; since decision makers are by nature different, their subjective interpretations would theoretically also be different. Indeed, it must be noted that by accepting that the world is reduced (and in some way 'misrepresented'), the decision maker needs to transform the 'incomplete' performance information into desired corrective actions. This constitutes one of the most significant challenges for the theoretical framing of performance measurement systems, since they involve such factors as psychology, educational background, experience, sociological contexts and even visual perception. All these are explicit factors that would partially provide meaning for the

individual, but how does a performance measurement structure portray decision information 'correctly', so good decisions are always made.

The theoretical answer is that this is impossible, given the circumstances described above. But in real life, there is hope. Consensus among employees and managers about how to tackle specific issues is the pragmatic answer to the theoretical challenge. When a given organizational problem arises, decision makers usually comprehend the given root-cause, and often also a possible solution. Of course, there are obviously distinctions between decision makers, but managerial consensus is characteristic (Ormrod 1993;Walley, Silvester, & Mountford 2006). This leads towards performance information systems that play a valuable part in aligning organizational objectives, even though subjective assessments are an inherent part of decision making.

1.5 Definitions and wording

To avoid semantic confusion regarding definitions and use of terms, this section describes in detail the context in which the key terms are used in this thesis. Various ways of framing terms into wider management concepts have been developed over the past decades, and although there is extensive research on performance measurement, there are very few commonly accepted definitions within the field. Therefore, this section aims to clarify how the terms and definitions are applied throughout the thesis.

Performance, or more specifically in this context, *organizational performance,* compares the actual output or results of an organization, as measured, against its intended outputs, goals or objectives.

The essential function of a *Performance Measure* or *Performance Indicator* is to express a given value for a given activity, e.g. input, output or process-related values. Usually, this value is related to a predetermined goal and an assessment of the extent of any deviation from that goal. A target level of performance is usually expressed as a quantitative standard, value, or rate (Ahmad et al. 2005).

Performance Measurement System consists of a number of individual performance measures (see Figure 1). There are numerous ways in which these performance measures can be categorized and structured, e.g. the Balanced Scorecard (BSC) (Kaplan & Norton 1992) and the Competitive Values Framework (CVF) (Quinn & Rohrbaugh 1983).

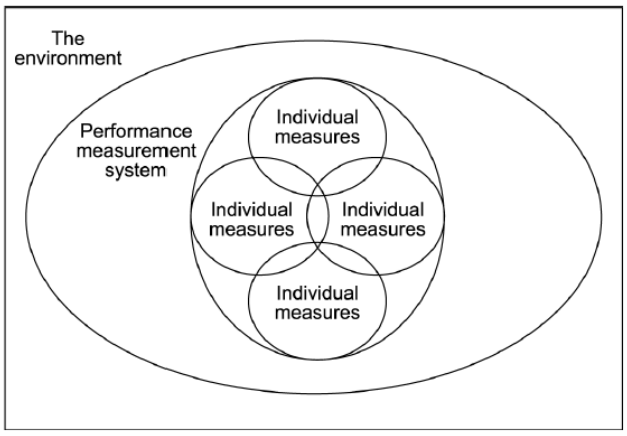

Figure 1. A framework for performance measurement system design (Neely, Gregory, & Platts 2005)

Kaplan and Norton relate the managing task to the navigating of airplanes:

> For the complex task of navigating and flying an airplane, pilots need detailed information about many aspects of the flight. They need information on fuel, air speed, altitude, bearing, destination, and other indicators that summarize the current and predicted environment. Reliance on one instrument can be fatal. Similarly, the complexity of managing an organization today requires that managers be able to view performance in several areas simultaneously.
>
> (Kaplan & Norton 1992)

Performance Management is the process of managing an organization, based on the information provided by the performance measurement system. Wheelen and Hunger (1992) state:

> Control follows planning. It ensures that the organization is achieving what it set out to accomplish. . . . the control process compares performance with desired results and provides the feedback necessary for management to evaluate results and take corrective action.
>
> (Wheelen & Hunger 1992)

Performance management is the alignment of decisions and activities within an organization to ensure that all units are working together to achieve the organizational objectives. This often includes the process of setting expectations, monitoring progress, measuring results, and rewarding or correcting employee performance.

Strategic decisions set the direction for the entire organization and are usually broad in scope and cover the long term. *Tactical decisions* are bound by strategic decisions but have a shorter time frame and are specific in nature. *Operational decisions* are technical

decisions that help execution of strategic decisions. The distinctions are described in Table 3.

Strategic Decisions	Tactical Decisions	Operational Decisions
Strategic decisions are long-term decisions.	Tactical decisions are medium-term decisions.	Operational decisions are taken on day-to-day basis.
Strategic decisions are taken in accordance with organizational aims and visions.	They are taken in accordance with strategic and administrative decisions.	They are taken according to strategic and operational decisions.
They are related to overall planning for the whole organization.	They are related to the work of employees in the organization.	They are related to production issues.
They deal with organizational growth.	They relate to welfare of employees in the organization.	They are related to production and operational growth.

Table 3. Nature of decisions

1.6 Summary

In an attempt to guide decisions within healthcare organizations, the promotion and implementation of performance measurement initiatives is on the rise. Recent literature addresses two major issues that trouble scientists as well as practitioners: 1) what to measure, and 2) the question of setting priorities for multiple-criteria decision making. Selecting and designing proper measures in relation to a given context present both practical and theoretical challenges that have shaped the domain for decades. Balanced sets of measures have been the dominant published answer to the challenge of prioritizing objectives, where the problem of non-financial measures continues to arise. Indeed, with little consensus on the construction of healthcare performance measurement systems, the literature suggests applying statistical indicators derived from national guidelines. This promotes practical application in a highly political environment, in the search for quantifiable justification of performance levels in all aspects.

The motivation for casting an organization into this difficult exercise is to be understood in a theoretical perspective. Performance measurement systems are a way to condense the organization into a representation that portrays progress and regress. This condensed view makes it possible to assess the organization in relation to the measures that are important for the individual decision maker. The representation also makes it possible to separate decisions from actions, which allows the decision maker to not be present when making decisions about performance. Theoretically speaking, performance measurement allows the decision-maker to assess the organization's separate parts without being present at the scene.

Chapter 2 - Research Design

This chapter illustrates the scientific approach that has been applied throughout the research study. It includes a description of the core research problem, stakeholders, research questions, methodology, limitations, and finally, a specification of the expected outcome of the study. This should provide the reader with an understanding of how the problem is analyzed, and what initial steps are taken to provide insight into the research problem at hand.

The present chapter introduces the reader to the specific construction of the thesis' scientific approach. Initially, a short condensed view of the research problem is provided. Secondly, descriptions of key stakeholders are presented, since their identification pin-points the scientific and practical aim. Next, the meta-challenge is crystallized into three specific research questions. The research methodology section presents the approach applied to answer the three questions, and limitations and potential are discussed. The chapter's closing comments outline the expected outcome with regard to all matters discussed.

2.1 Research problem

On the basis of the empirical and theoretical review, it is assumed that efficient settings require some comprehensible balanced measures, if operational managers are to be able to make appropriate decisions. It is also absolutely necessary that the information be presented in such a way that it reflects the strategic objectives of the organization. With this starting point, the meta-challenge for this research is defined as follows:

> How can a holistic healthcare performance measurement system be designed so that it reflects the strategic progress and regress useful in operational decision making?

The thesis focuses on the question of constructing performance information capable of portraying strategic change in an operational context. The aim is to develop a decision support framework that is able to justify that operational performance is aligned with organizational strategies. It is important to note that the research has dual goals; it specifically aims to propose a solution to the given problem, and also to conduct theoretical advancement within performance measurement literature.

2.2 Stakeholders

Two primary stakeholders and two secondary stakeholders are assumed to be the key beneficiaries of the work presented in this study. The scientific community and hospitals are the primary stakeholders, and patients and industry are the secondary ones. These four constitute the dominant stakeholders in modern healthcare, both as active players within the organizations, and as external influencing actors. All have different interests, which also influence the benefits they can derive from the study.

2.2.1 Scientific community

The primary purpose of this research project is to adapt state-of-the-art scientific measurement techniques to the practical context of performance measurement in healthcare. The work intends to provide new knowledge that can contribute to the scientific field of performance measurement. Both theoretical development and practical adaptation are of interest to the scientific community. Through the acceptance of scientific papers for publishing in journals and presentation at international conferences, the work aims to contribute to scientific discussions. The focus when submitting papers is

on contribution to the ongoing scientific discussion rather than beginning a new line of discussion. By gaining in-depth knowledge of recent key publications and using this as the launch pad, the publications presenting the results of this study aim to continue the current trend in the literature and hopefully provoke responses from other scientists. These presentations follow the same line of construction or discuss the drawbacks in recently presented models. Both types of discussions are valuable in the context of science, because they continually contribute to the development of performance measurement systems.

The present work contains five scientific papers that have been submitted, accepted and published in internationally recognized journals and presented at international conferences. Thus, the contributions are scientifically validated by the journals and conference agencies issuing them. The publication strategy has been to cover several journals in order to demonstrate that the scientific appeal goes beyond one single scientific community. The papers are written as successive steps towards the final recommendations, continuously illustrating the development of the study. The submission of the papers likewise follows the development process of the recommendations.

2.2.2 Hospitals, and hospital managers

As the work is conducted in close relationships with clinicians and managers at hospitals, the project's recommendations have aimed throughout the study to benefit them. Consequently, hospitals and hospital managers are obvious stakeholders in the project. In particular, Southern Jutland Hospital, as a sponsoring partner, is presumed to benefit from this research project. The radiology department has been a close collaborative partner throughout the process, which has had a primary influence on the final recommendations. The structure of the framework was adjusted and tested in the environment, which makes the framework particularly applicable in these settings. Although the framework has had Southern Jutland Hospital as the primary collaborative partner, other institutions have also been involved in the development. Since both Danish and Japanese hospitals have participated in the process, it can be assumed that other healthcare institutions would benefit from the recommendations. The generalizing potential of the recommendations is further discussed in section 2.5.3 and includes hospitals and their decision makers.

2.2.3 Patients

As the primary objective of the work is to improve decision support information for hospitals and thus improve decision making, the patients should be an end beneficiary. If better decisions were made within any kind of organization, the clients (in this context, patients) would be a secondary stakeholder. Since patients play a more and more active role in modern healthcare, their gains also increase as decisions improve in quality. Thus, there is reason to suppose that if the research presented in this thesis were able to construct performance information that would enhance the possibility for better decisions, patients would be positively affected.

2.2.4 Industry

The functions that care entails are increasingly dependent on technological equipment to facilitate the progressively more complex processes within healthcare organizations. As a result, healthcare organizations are becoming more reliant on suppliers to provide technological equipment that is customized to their needs. At the same time, suppliers of technological equipment are constantly developing equipment that is well adapted the particular area of application. If this study's recommendations are integrated into an equipment context, suppliers can rely to some extent on scientific evidence in the development of technological equipment.

2.3 Research questions

Since the framing of the research study has so far been somewhat broadly described, it is necessary to crystallize the central theme into more tangible research questions. The formulation of the research questions has been an iterative process throughout the research period, and clarifications of concepts have continuously shaped the research focus. The specification of the theme into concrete questions is based on the discoveries in the literature, together with the theoretical foundation described in the previous chapter. The scientific challenges, practical shortcomings and methodological weaknesses in the current body of literature, constitute the empirical basis for the construction of the research problem. Practical difficulties and frustrations add to the practical depth of the research study. Based on these reflections, three research questions have been formulated. The questions are developed successively: RQ1 deals with which industrial concepts can be adopted; RQ2 treats the issue of construction of performance measurement systems; and RQ3 goes into the area of benchmarking of healthcare outcomes. The underlying basis and the specifics of each research question are described in the following.

2.3.1 Research Question 1 – Using industrial concepts in healthcare

Industrial organizations have been leading the development of performance evaluation models (Baker, Beitsch, & Landrum 2007), but naturally scientists have tried to transform and adapt some of the successful concepts of industrial performance management for use in healthcare. A few of the concepts that have been established are Balanced Scorecard (Yang, Cheng, & Yang 2005), Competing Values Framework (Wicks, St Clair, & Kinney 2007) and Six Sigma methodology (Woodward 2006). Although they are accepted in healthcare, few of these concepts have been equally successful in healthcare as in industry, since the differences between the sectors presumably play a significant role. Since the present work deals specifically with the construction of decision support information, it is necessary to focus attention on the elements of industrial performance measurement that can benefit this particular issue and identify which elements of industrial performance measurement work optimally within healthcare. The first research question is consequently as follows:

> RQ1: Which elements of industrial performance measurement can with benefit be integrated as guiding components for the development of a Management-By-Objectives framework in public healthcare settings?

To answer this question satisfactorily requires deconstruction into several sub-inquiries. First, how do healthcare decision-making processes differentiate from industrial decision-making processes? And which effects do these differences have on the construction of a performance measurement system. Second, are there structural differences between public and private organizations in the context of performance management? And do these differences constitute a change in focus of the measures? During the initial phase of the study, these sub-questions were the primary focus in the identification of suitable concepts and elements that are applicable within healthcare settings, and were considered during the investigation of RQ1.

2.3.2 Research Question 2 – Securing strategic alignment in measurement

Derived from the investigation of suitable elements for a performance measurement system, it is necessary to deal with the particular combination of the elements. In light of the theme of the thesis, the focus is on constructing the measurement system in such a way that it allows decision makers to make strategically aligned decisions. In order to achieve these strategic objectives, all decisions must be aligned. Aligning operational and tactical decision making with the strategic objectives requires performance information suited to the particular context in which it is applied. Therefore, based on the assumption that 'good' decision support information is highly contextual, the next research question can be formulated as follows:

> RQ2: Which construction of decision support information is appropriate on the tactical and operational levels respectively in a healthcare organization in order to secure alignment with strategic objectives?

To elaborate – there are focal points in the question which have to be addressed individually. If there are differences in the construction of the information used for decisions at different levels of the organization, then what constitutes the characteristics on each organizational level? And how are strategic objectives transformed into tactical and operational indicators? And how is a decision that is out of alignment with a strategic objective identified?

2.3.3 Research Question 3 – Benchmarking operational performance

The third research question is formulated on the basis of the previous research question. Given that a performance measurement framework is able to provide valid performance information suited to the strategic alignment in a given organization, can this then be

applied in an external context? In the quest to elevate the investigation of healthcare performance measurement, the research focuses on the globalization process that healthcare is undergoing. At present, it is not enough to be managing closed isolated units; modern healthcare is engaged in globalized competition with other healthcare-providing units. Therefore, an in-house vertical performance information system should be capable of conducting external benchmarking. This changes the context of the information but not necessarily the technique itself; therefore, the third research question is:

> RQ3: To which extent can an internal vertical Management-By-Objectives framework be applied within an external horizontal benchmarking context?

Changing focus from internal vertical performance representation to external benchmarking requires a change in prerequisites. The nature of the information is expected to change, because the context of application has changed. But to dig deeper into the use of the same techniques, internally and externally, several issues have to be addressed. Is all internal performance information suited to external benchmarking, and if it is inappropriate, how is this identified? How do we make sure that benchmarking is conducted fairly with the intent of making the benchmark as valid as possible? Can benchmarks be performed across modalities and departments, or how does the demarcation process function?

2.4 Research methodology

The answers to the three research questions are highly dependent on the choice of the scientific methodology used for gathering the empirical material, and the interpretation of this material. This section tries to provide insight into the choice of methodology that forms the basis of the research conducted throughout the study. The justifications in this section should assist the reader in understanding the rationale behind the results. In addition to contributing structure, consistency and scientific validity, encircling research within a methodological frame. The section elaborates upon the scientific potential and limitations that have evolved throughout the research study. Discussions related to the philosophy of science evolve: Which methods are applied? How is validity realized? What is the generalizing potential, and which limitations and boundaries are evident? All these questions culminate in discussions that are expected to result from this study.

2.4.1 Philosophy of science

The scientific methodology is extremely important with regard to what to expect of the outcome from a research study. The nature of the subject and the paradigm the researcher relates to, are the most critical prerequisites for any research study. The perception of reality or the ontology is an important pre-discussion in any research study. The way in which the research approaches science determines how the data should be obtained, and just as important, how the data should be understood. The paradigm determines which type data it is possible to obtain and consequently how these data could be acquired. The selection thus establishes a frame within which the researcher can decide which methods

to apply in order to gather the data suitable for the particular study. The choice of method imposes discipline and consistency on the analysis- This implicitly defines a 'laboratory' for the analysis. The theoretical assumptions that form the underlying basis for interpretation of data need to be well defined a priori. In this study, the critical realism philosophy has been applied, and the justification for the selection is described in the following, both as a discussion of theoretical applicability and practical suitability.

Critical Realism

Critical Realism is a realist theory that has been applied to explain the fundamental claims obtained through research in both natural and social sciences regarding knowledge, true progress and reality (Connelly 2001). In critical realism, the social variables are a precondition, and it is assumed that it is impossible to quantify these completely. To specifically quantify a certain phenomenon demands total control of all physical and social variables, as in a closed laboratory. This is impossible when dealing with organizations in practice and would therefore be a methodologically misleading approach. Critical realism breaks away from both the perception that "all knowledge is relative" (realism) and the perception that "all knowledge is limited to what can be quantified" (positivism). Critical realism therefore claims that there are independent realities that can be understood, as opposed to social constructivism, which claims that all knowledge is relative (Danerark et al. 2002). But critical realism also emphasizes that the described reality is imperfect. This is opposed to the positivist point of view, which maintains that reality is limited and can be described by objective facts.

Critical realism emphasizes that there is an independent reality that can be understood and described, while recognizing the imperfection of all knowledge.

> Critical realism first of all makes the ontological assumption that there is a reality but that it is usually difficult to apprehend. It distinguishes between the real world, the actual events that are created by the real world and the empirical events which we can actually capture and record.
>
> (Easton 2010).

In critical realism, as in any study, drawing conclusions on the basis of well framed and defined approximations, within the scope of a certain experiment, is a necessity if we are to be able to propose valid recommendations (Wikgren 2005). But awareness that the conclusions and recommendations are based on the researcher's experiences of a given phenomenon under given circumstances is crucial. The more of the same observations, the less the uncertainty concerning the next observation conducted under the same circumstances (Walters & Young 2005). However, it is important to recognize that the mutual interplay between social structure of the object of study and the surrounding society is a factor in any critical realism study. Danerark explains these assumptions as follows:

> The nature of society as an open system makes it impossible to make predictions as can be done in natural science. But, based on analysis of causal mechanisms, it is possible to conduct a well-informed discussion about the potential consequences of mechanisms working in different settings.
>
> (Danerark, Ekström, Jakobsen, & Karlsson 2002)

Context-dependent knowledge, where social factors are significant, gives this type of science a basic limitation when generalizing the conclusions made on the basis of a single experiment (Smith 2006). The question, *"Is there reason to believe that it could be otherwise?"* is therefore extremely important for the researcher using critical realism. If the answer is yes, then additional investigation has to be conducted; if it is no, the proposal can be assumed valid (usable). The arch-critical realist argument would to some extent always have a speculative "as-if"-element as a possible explanatory contribution, due to the underlying assumption that the experiment would always be theoretical 'imperfect'.
In practice, this means that the researcher continuously has to re-conduct the experiments in a well- framed 'laboratory', specifically defining the boundaries of study. These iterative experiments continually justify whether the conclusions are valid in practice, and indicate which elements have to be adjusted. The 'laboratories' that frame the study are limited to a given context, and define the boundaries within which the researcher can conclude from collected data. It is important to note that the researcher needs to respect the fact that data does not in any perfect sense describes reality; therefore, generalizations have to be performed with respect for these conditions. Implicitly, the application of critical realism means that an optimal solution to a problem is unattainable. Assuming that the 'world' is never fully quantifiable, the solution can never cover all possible variables (HARRÉ 2009). Therefore, researchers need to collect data until there are no doubts that they reflect the reality within the boundaries defined for the study.

Application of critical realism in this study
The construction of the study, where Southern Jutland Hospital acts as a collaborative partner throughout, provides the possibility to continuously test the proposals under more or less the same circumstances. The critical realist point of view is thus particularly well suited to case studies,. As Easton maintains:

> Critical realism as a coherent, rigorous and novel philosophical position that not only substantiates case research as a research method but also provides helpful implications for both theoretical and development and research process.
>
> (Easton 2010)

Organizational changes that take place within the same departments can be ignored, compared to conducting repetitive tests across different organizations. Repetitive testing in the same department is therefore ideal for a critical realist study (Walters & Young

2005). Naturally, changes do happen, and such changes signify that the circumstances under which the test is being conducted are not completely flawless. Thus, the conclusions from the tests must be generalized with caution. To be able to add further reliability to the generalized conclusions, multiple cases might be a suitable approach. Widening the scope of the laboratory helps the conclusions' reliability. By first testing a proposal on a single case, and then testing the same proposal on multiple cases, the researcher has an excellent foundation upon which to conclude regarding the proposal's strengths and weaknesses in terms of generalizing potential. But it is important to acknowledge that the proposal should not be modified to any great extent between the two tests, since this would add to the number of altered variables. Small changes from the single case test to the multiple case tests are acceptable, as long as the conclusions take this into account.

In a critical realist study, the application of several methodological tools are often a way to enhance the reliability of the conclusions put forth (Walters & Young 2005). It is often necessary to mix methods in a way that reduces the probability for the proposal to fail. If experiments are continuously performed with the aim to test the generalizing potential, the researcher's recommendations will stand stronger. To achieve this, action research and single-case methodology are applied.

2.4.2 Action research

Action research is regarded as a well suited method to treat the study at hand with the scientific standpoint described. The basic idea of action research stems from Kurt Lewin's thoughts about experimenting in the field rather than in a laboratory. This is particularly well suited to work with practitioners and managers at hospitals, which in addition to being a theoretical challenge also constitutes a practical problem. And as John Dewey explains, a practical problem demands practical solutions (Dewey 1938). Even if a problem is approached scientifically, its solution can only be regarded as viable when it has been demonstrated that it produces the desired outcomes in practice (Reason & Bradbury 2001). Using the notion of Czarniawska, the action research approach signifies the use of the logic of practice (see Table 4). This is an important specification, because applied logic frames the conclusions drawn from the collected material.

Logic of theory	Logic of practice	Logic of representation
– Abstract – Hides its rhetorical accomplishments – Claims to use formal logic – Has methodological criteria of truth	– Concrete (situated in time and space) – Discursively incomplete (tacit knowledge) – Often uses narrative knowledge: narratives are incomplete and not stylized: chronology rather than causality – Pragmatic/aesthetic evaluation criteria: post factum evaluation as a main organizing device ("now it works")	– Abstract (but gladly uses hypothetical examples) – Rhetorically accomplished – Uses stylized narrative knowledge (distinct genres, legitimate repertoire of plots, hero-like characters) – Formal rationality as a main organizing device (purpose-means-effects)

Table 4. The three kinds of logic (Czarniawska 2001)

Applying logic of practice, the researcher limits the answers to research questions to the concrete and pragmatic. Action research is therefore interpreted as being oriented towards inquiry that seeks answers through gathering evidence and testing in practice (Reason & Bradbury 2001). One of the action researcher's key assumptions is that the world is constantly changing and that the researcher and the research itself are part of this change (Collis & Hussey 2003). This indicates that the researcher accepts that tacit knowledge is part of perceived reality. Although there are several definitions of action research, Winter & Munn-Giddings definition clarifies why action research is still particularly well suited for this study:

> Action research is the study of a social situation carried out by those involved in that situation in order to improve their practice and the quality of their understanding
>
> (Winter & Munn-Giddings 2001)

Action research covers a spectrum of research activities; most definitions characterize it as: 1) focusing on change and improvement; 2) involving practitioners in the research process; 3) being educational for those involved; 4) examining questions that arise from practice; 5) being a cyclical process of collecting, feeding back and reflecting on data; and 6) being a process that generates knowledge (Hampshire et al. 1999). Action research is interpreted as an interactive inquiry process that balances problem-solving actions implemented in a collaborative context with data-driven collaborative analysis or research to understand underlying causes, thus enabling future predictions about personal and organizational change (Reason & Bradbury 2001). For this reason, action research is increasingly being used in health-related settings (Meyer, Pope, & Mays 2000). Due to the high rate of involvement of researchers in testing and evaluating ideas with professionals in the healthcare environment, action research seems to be the most obvious methodology for this study. Furthermore, since hospital managers are involved in the development phases, a methodology is called for that takes this into account.

Workshops

A key method within the action research methodology is the use of workshops, which facilitates learning and development in groups. The justification for using workshops is that they often enhance the possibility for reflection at a higher level than using only interviews. As preparation for these workshops, interviews are used to promote discussion topics, since managers may share worries or ideas that are not suited for a discussion context.

Participatory testing

As this research is concerned with the development of performance measurement frameworks, and the process is assumed to be an iterative process, testing is of great importance. And because feedback from managers and employees are the basis for adjustment, participatory testing is needed. Each development/adjustment loop is tested on real live data, so that practitioners evaluate the outcome. In discussing the outcome, another adjustment loop is conducted. Participatory testing is thereby an integrated and highly valuable method for this work. The specifics of both methods are described in more detail in Chapter 3.

2.4.3 Single-case research

Given the construction of this study, with one hospital as the main source of field data, it is important to note the specifics of doing research with a single case. Empirical evidence is used from multiple sources, but the testing and development have particular focus on one hospital. In using single-case study methodology, the concept is to develop new theory and then generalize it on the basis of the findings (Voss, Tsikriktsis, & Frohlich 2002). In action research, the case study is in the phenomenological area of qualitative research. Some authors regard single-case research to be a powerful and effective alternative to some of the more traditionally used methods in social science. In a healthcare perspective, the flexibility and sensitivity of 'local' factors offer substantial benefits to those charged with conducting research in clinical environments (Morgan & Morgan 2009). In addition, the use of a single case often puts the researcher in a position of trust with the subjects of the study. The possibility of developing a relationship is assumed to benefit the collection of data by increasing the volume or specifics.

Interviews

A cornerstone of qualitative research, interviews is recognized as a very powerful method for gathering data. This method is particularly well suited for the use in single cases, where the researcher can conduct in-depth follow-up interviews to help trace changes in the perception of the subject discussed. By using informal interviews continuously throughout the research project, valuable input to the research's development is likely to be uncovered.

2.5 Scientific limitations & methodological constraints

Decisions about scientific approach, methodology and methods constrain the perception of the empirical material upon which the conclusions are drawn. These limitations are

extremely important to acknowledge, because they clarify in which context the conclusions are valid and the generalizing potential of the final recommendations. The discussion in this section emphasizes the limitations of applying critical realism and elaborates on the selection of methods.

2.5.1 Scientific boundaries

It is essential to any field of research to define and be aware of scientific boundaries, the aim being to specify the scientific scope of the research being conducted (Isenmann 2008). The boundaries presented in this section should enable the reader to recognize the limitations and perspectives of the study, implicitly framing the extent of the conclusions. The boundaries are specifically determined in relation to the study and are constituted by the scientific choices made in the research's initial stages (see Figure 2).

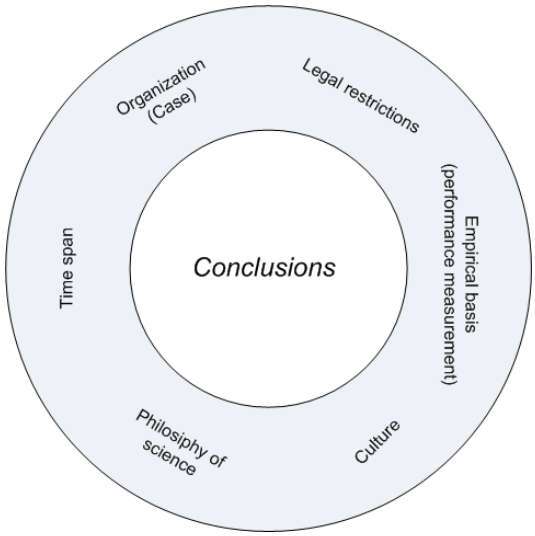

Figure 2. Research boundaries.

As previously discussed, the interpretation and applicability of scientific results are limited by the philosophy applied. In this case, where critical realism is applied, it is necessary to be aware of the restrictions of this approach, which have been discussed in detail in the previous section. The existing body of literature partially constitutes the empirical and theoretical framing of the thesis. When focusing a particular area of literature – in this case, performance measurement – the literary base constitutes a boundary. By remaining within the scope of the body of literature, the research is more likely to be acknowledged among scientists with the same area of interest. By expanding the scope, changing terminology, twisting concepts, the work would constitute a scientific discussion of its own, instead of contributing to ongoing discussions. Since in this study, the aim has been to engage in the ongoing discussion, continuing to work within the frame of performance measurement literature is a prerequisite. Another limiting factor in this work has been national legislation and instructions. Throughout the study, the work has aimed at constructing proposals that would in no way compromise national guidelines

or legislation, thus enhancing the probability that the proposals would be accepted by the healthcare sector.

The development phase has been conducted in close collaboration with the Southern Jutland Hospital; therefore, this particular organization naturally plays a significant part in the framing. When constructing decision support systems, the decision processes of the organization play a key role in framing the work. Since the study primarily deals with operational decision makers, their decision-making scope constitutes the extent of the work. The mutual relationships between organizational stakeholders are also a limiting factor, because they determine the scope of each decision. These relations are not specifically classified, but throughout the study, specific attention is paid to these relations. Where the mutual relations specifically influence the work, this is emphasized in the thesis. Such relations implicitly signify that healthcare workers' culture plays a significant role. As the framework is developed as an action research study on a Danish case, the organizational culture is of course a limiting factor, which means that generalizations beyond the case must be conducted with the utmost respect for any changes in organizational culture.

Finally, attention has been paid to the fact that the work has a time limit of three years, which poses a boundary of its own. This boundary does not limit the scientific work itself, but restricts the amount of investigation that can be achieved. Some obvious enquiries would have further clarified concepts and further validated or improved the final recommendations, but these have been omitted solely because of the time factor. They are discussed in Chapter 7, where possibilities for further research are elaborated upon.

2.5.2 Validity

An important issue for the researcher is to construct a method to prove the validity of the results of the study. In chemistry, performing the same experiment over and over again may in many cases constitute a validity test. In organizational research, where the system (unit of analysis) is never a closed and stable 'laboratory', validity is a more intangible issue. Heraclitus portrays this dilemma this way:

> You could not step twice into the same river; for other waters are ever flowing on to you.
>
> Heraclitus of Ephesus (c. 535 BC – 475 BC)

Therefore, to be able to claim that results are valid and constitute a scientific contribution, the scientific methodology needs to compensate for the fact that it is not a closed system. The critical realist recognizes that there is an objective reality, but this reality can never be described in all its complexity. Consequently, in critical realism, validity is not a question of constructing a truth criterion, because this would be an unachievable goal. Instead, the researcher aims to prove that the results out-perform the state-off-the-art achieved until now. Karl E. Weick discusses this particular topic in his

paper about "Clarity of concepts", where he claims that there is no such thing as validity in critical realism, but only further clarification of concepts:

> … the implication of this is that science should be understood as an ongoing process in which scientists improve the concepts they use to understand the mechanisms that they study. It should not, in contrast to the claim of empiricists, be about the identification of a correlation between a postulated independent variable and a dependent variable. Furthermore, the methodological core perception of critical realism is 'clarity of concepts'.
>
> (Weick 1995)

As the concepts are more and more legitimized, the stronger the claim becomes, which then constitutes the scientific contribution. With respect to terminology, this section discusses the term validity, thus implicitly acknowledging that reliability of recommendations and clarity of concepts may depend on more precise terms.

In organizational studies, this signifies that there are constant changes that cannot be ascribed to the process of the project itself. Therefore, the methodology is guiding, in terms of constructing, this validating procedure. In longitudinal studies, one difficult factor is to separate the effects of the study from the changes that would have taken place even without conducting the study. In this study, the focus is on how to construct performance information that can be used to guide decision making in healthcare. Concurrent with this study, the issue of conducting proper performance evaluation is increasingly becoming a hot topic within Danish healthcare. Because authorities are continuously updating and improving their current systems, the basis for hospital management is undergoing change during the project. Therefore, it is important to separate the changes that would have occurred even if this project had not been started.

This study has therefore been constructed as a sequence of iterative investigations, which should limit the number of unknown factors. By continuously testing the proposals with as few unknown factors as possible, the recommendations' validity is continuously amplified (see Table 5).

Steps	Case		Model		Aim
	Known	Unknown	Known	Unknown	
1		x		x	Understand case, clarify wishes
2	x			x	Adapt model to case
3		x	x		Test generalizability
4	x			x	Adapt model
5	x			x	Test validity and generalizability

Table 5. Sequence of research steps

By selecting either the model to be the unknown factor or the case to be unknown, uncertainty is limited. The term 'known' is of course a misleading term (with reference to Heraclitus and Weick). But if the experiments are conducted for the same case, even though we acknowledge it has changed, it is still a more familiar case than if the experiment were conducted for a new case. This thesis seeks to secure validity by re-performing the testing with either model or case uncertainty, never with both unknown.

2.5.3 Generalizing potential

In applying a critical realist's viewpoint on a single-case study, the generalizing potential of the results would naturally be limited. This implies that results that are validated for a case of contextual nature are limited by the case's contextual borders. In cases with similar contextual foundations, generalizations of the results are appropriate. Thus, any attempt to generalize the results beyond the tested cases would be methodologically misguiding, unless the contextual frame in which the results are applied resembles the environment where the results are achieved.

For this study, this means that generalizing the results to other radiological departments would to some extent be possible. Within Danish healthcare, there would be valid reason to claim that the results could be applied to other radiological departments, since they all share a similar legal foundation and similar structural guidelines, and the employees have had similar training. This would make it probable that the results achieved through this study would to a large extent have been the same for another radiological department in Denmark. Changing the case from a radiological department to another department within Danish healthcare would also be possible to a wide extent.

If applied to another country, the proposals would be challenged further, since legal foundation and cultural issues may influence the strength of the proposals. But testing on alien territory is ultimately necessary to prove applicability; testing for different cases is the only way to find out about the generalizing potential and what works and what does not work, which would then be the basic rationale for any claims concerning generalizing potential.

2.5.4 Paradigm and methodological constraints

The interpretation of data also calls for rigorous evaluation, because faults in interpretation can lead to false conclusions. As stated in Winter and Munn-Giddings' handbook for action research, the collected data should be seen as a possibility for new actions (Winter & Munn-Giddings 2001). But if the foundation for new actions is wrongly interpreted, the next step would be a step in the wrong direction. It is therefore extremely important to be thorough and methodology strict in the process.

The selection of scientific methods also has an essential impact on the nature of the conclusions. In action research, the distinction between research and subject may become blurred in the course of what is usually a lengthy and collaborative relationship (Reason & Bradbury 2001). Bruno Latour describes this difficulty when conducting Science in

Action, where the certification of results often creates a scientific dilemma, which he portrays as a Janus head see Figure 3.

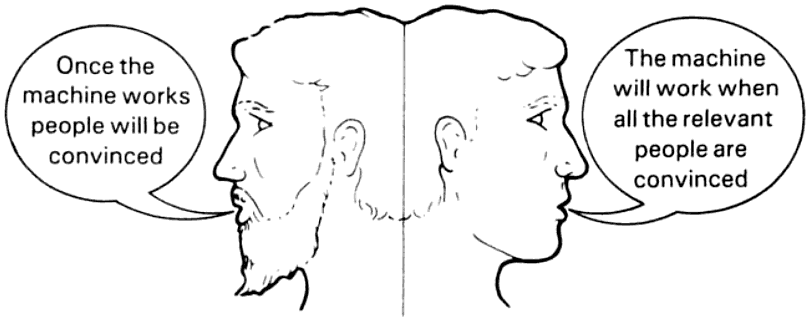

Figure 3. Janus head (adopted from Bruno Latour, *Science in Action* 1987).

In this work, collaboration is a prerequisite for the framework to be constructed, which refers to the right side of the Janus head. But validating and generalizing the results refers to the left side, where a complete framework is applied to other settings. Subsequently, it is necessary to return to a known case to adjust the framework in order to improve the claims, which again refers to the left side. This iterative process signifies the difficulties involved when validation and generalization of results is performed for a case where the unit of analysis is theoretically unstable. This is further enhanced by the difficulties described the literature connected with conducting single-case studies where all the non-discrete events sometimes dramatically change the conditions. Changes in management, retirement of key staff members, or changes in clinical guidelines are all factors that affect organizational behaviour. And when conducting single-case studies, these changes can affect the outcome of the study. Therefore, the single-case researcher has to observe and measure as often as possible, within the practical constraints inherent in this pursuit. This provides the possibility for adjusting for these changes in the basis of the study (Morgan & Morgan 2009).

2.6 Expected outcome

The outcomes that can be expected from a research study are rooted in the scientific approach and the methodological constraints of the study. Here, the expected outcome is divided according to a primary and secondary objective. The contribution to the scientific community and practical application within the case departments are regarded as primary outcomes. A secondary outcome is the attempt to influence the construction of performance evaluation tools for use beyond the cases studied, thus affecting the way commercial hospital information systems are developed.

2.6.1 Primary objective

The primary objective in this study has been to develop a performance measurement framework that would enable operational decision makers to evaluate whether the organization is strategically aligned. This would enhance the decision makers' ability to take strategically appropriate corrective action, which ultimately would improve the

overall performance of the departments. This would, as discussed earlier, have an impact on the scientific community, where the ideas and concepts would be acknowledged as valid input to ongoing discussions. The effect would even broader, if international journals and conferences publish the ideas.

As the primary source of empirical data, Southern Jutland Hospital is also expected to benefit from the study. By involving clinicians and managers within the organization, a rub-off effect is anticipated. The framework that is being developed is customized to the organization, which should ease implementation after the study. This would presumably have an effect with respect to professionalizing the management team.

2.6.2 Side objectives

Since the study deals with the development of systems for application within healthcare, suppliers of technological healthcare equipment may also have an interest in the study. As the recommendations of the study are improved, and its practical applicability demonstrated, construction of new technological equipment may even also result. As data handling systems become more and more advanced, suppliers continuously need new input in the pursuit of market shares. This study might prove to provide valuable input, since it is based in a practical context, where the suppliers systems are also applied.

2.7 Summary

The primary aim of the study has been to propose a framework that enables holistic performance evaluation without compromising strategic consistency. The chapter explains in detail how the scientific approach in this research study has been designed. Three detailed research questions provide the guidelines for the study: RQ1 deals with which industrial concepts can be adopted; RQ2 treats the issue of construction of performance measurement systems; and RQ3 goes into the area of benchmarking of healthcare outcomes. Critical realism constitutes the philosophical backbone of the study. This theory emphasizes that reality can never be described in all its complexity, and that it is indeed important that if there is no reason to suspect that claims are false, then they are considered to be valid. This justification demands strict validity testing, which is not a question of constructing a truth criterion but aims to prove that the obtained results out-perform the present state-of-the-art. Action research and single-case methodology is applied as a method for collecting data. Both have merits in relation to healthcare, where the use of two scientific methods is regarded as a means of compensating for the shortcomings of each of them. To secure validity, the proposals are continuously re-tested with as few unknown factors as possible.

Chapter 3 - Empirical foundation

Chapter 3 aims to clarify the sources of the empirical foundation and how it has been analyzed in the study. The chapter contains an introduction to the cases that constitute the source of data. The specifics regarding the qualitative data collection, i.e. interviews and workshops, are described in detail. The quantitative data used is subsequently described, with emphasis on the methods by which the data was obtained. With regard to both data categories, discussions regarding collection, analysis and application are presented.

3.1 The cases

Throughout the study, Southern Jutland Hospital has constituted the primary case. Hospital management was an integrated part in the initial construction of the study, when goals and sequence of analysis were established. During the study, focus has been on the radiology department, where data collection, operational testing and validation have taken place. In addition, three Japanese hospitals have acted as secondary cases during an external stay in Tokyo, Japan. As secondary cases, they provided a way to test the generalizing potential of the proposals.

3.2 Southern Jutland Hospital

The hospital is a public Danish non-profit hospital situated in the Region of Southern Denmark, consisting of four individual sites (see Figure 4). As a result of the national hospital reform 1 of January 2007, hospitals throughout the country were merged into larger units. The four hospitals were merged at the management level, but the four sites still function as separate operational divisions in the new hospital. Collectively, the hospital currently employs approximately 2,600 staff members and 479 beds, distributed among the four sites. The hospital receives patients from an area with 253,000 inhabitants.

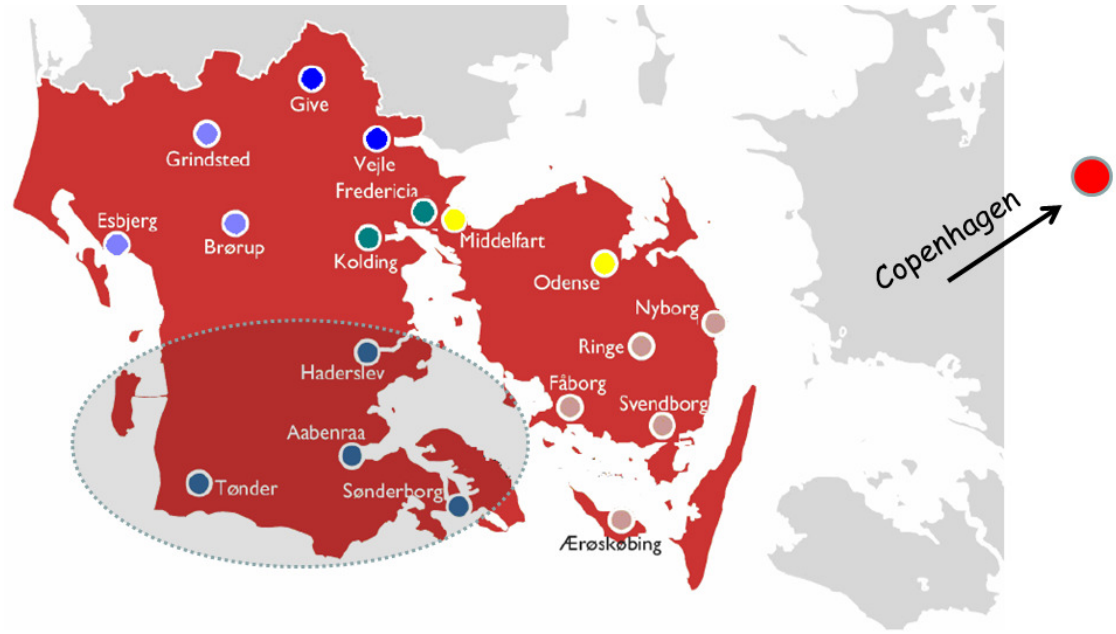

Figure 4. Map of Region of Southern Denmark

The research was carried out at the radiology department, which employs 128 staff members in total, distributed over the four sites. The radiology department treats roughly 145,000 patients per year, where about 40 percent are acute patients. The department performs almost all forms of radiological examination, CT, MRI, ultrasonic, mammography etc. The distribution of patients is dependent on type and availability of equipment, plus patients' geographic location.

Haderslev is the third largest hospital and performs all basic radiological examinations. Due to the re-distribution of patients, Haderslev is scheduled to be shut down in a not so distant future, and employees and patients will be moved to Aabenraa. The hospital in Aabenraa was formerly a private hospital converted to a public hospital. Aabenraa is classified as the Southern Jutland Hospital's acute centre and has a relatively high acute burden compared to the other sites. Aabenraa performs all modalities except MRI examinations. A private clinic located at the Aabenraa site performs MRI examinations, but the clinic is separate from the hospital as such. Aabenraa is planned to be expanded in coming years in order to cope with an increasing acute burden which is a result of the re-distribution of regional patients that affected all hospitals in Region of Southern Denmark. Tønder is the smallest unit within Southern Jutland Hospital and is located in the rural area. The hospital has the organizational role of a local hospital for the inhabitants of the western part of the area the hospital serves. Tønder has equipment for x-ray and MRI examinations. Sønderborg is the largest of the sites, and the primary educational responsibilities are placed here. Since Sønderborg has a long tradition for treating special and difficult cases, these cases are usually moved to Sønderborg from the other three sites. Throughout the study, the management group at the radiology department has been the link between the researchers and the collection of empirical material, both qualitative and quantitative data.

Strategic plan, Quality 24/7

Subsequent to the merging of the four hospitals, the hospital's strategic plan was formulated by the board of directors, and has been official policy for the period 2007 to the end of 2010. The strategic plan, named Quality 24/7, is based on the vision for the hospital:

> Southern Jutland Hospital will under all circumstances deliver quality 24/7.
>
> Own translation from (Sygehus Sønderjylland 2007)

The vision is formulated into four overall strategic perspectives, subdivided into 14 strategic goals (see Table 6). As can be seen, the strategic plan resembles the structure of a Balanced Scorecard, with four perspectives and subjacent goals.

Strategic perspective	Strategic goals
1. Satisfied patients	1.1 Be leading in the implementation of the Danish Quality Model. 1.2 Involve users and collaborators in development of quality. 1.3.1 Patients should have increased admittance to self-service. 1.3.2 Patients should have increased possibility for electronic information. 1.4 Development and use of evidence-based diagnostics and treatment
2. Creative development	2.1 Research and innovative development activities in all departments 2.2 An attractive training environment 2.3 Maintain and develop strong professional environments.
3. Healthy economy	3.1 As much as possible: transfer elective and acute patients to ambulant treatment. 3.2 Working procedures guaranteeing treatment within 4 weeks 3.3.1 Create capacity for new and better offers to patients. 3.3.2 Creation of a bone outpatient department
4. Good colleagues	4.1 Development for staff members with due respect to individual needs and working conditions 4.2 Professional management

Table 6. Strategic goals - Southern Jutland Hospital (own translation (Sygehus Sønderjylland 2007))

Practical implementation at the radiology department

With the 14 goals as general guidelines, each department was to pursue each of these within their own area of responsibility. Since there was no clear procedure for how to realize the goals, responsibility was placed on department management. The radiology department approached this assignment by involving all operational personnel in the task of translating the goals into a set of operational measures. This was performed by altering the strategic goals into a more radiology-specific context, thus making them more applicable in daily management. The plan of involving staff in the development process was to ensure that the measures were understandable and useful to the operational employees, and also gives every employee the possibility to be involved in the process if they wished. The process of translating the strategic goals consisted of a series of workshops, where different aspects of the transformation process were addressed. Figure 5 displays the course of events in the process of translating the strategic plan into operational measures.

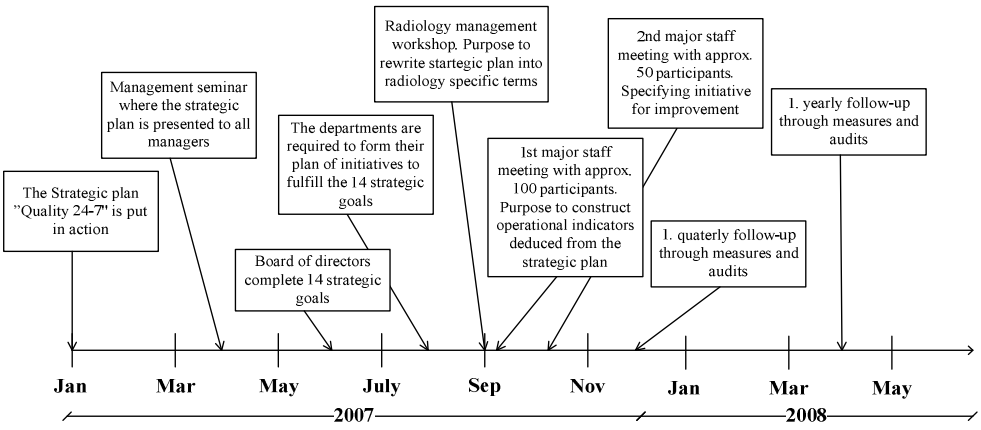

Figure 5. Timeline for strategic plan roll-out

In the beginning of September 2007, radiology management rewrote the content of the strategic plan so that it became more oriented towards radiology. This task was carried out exclusively by management and facilitated by external consultants. During September/October 2007, two major workshops were set up in which the employees participated. The employees were to develop the department-specific indicators related to each of the 14 strategic goals, and these would be the guiding indicators for radiology for the period of the strategic plan.

The transformation of the strategic plan into a guiding operational measurement structure formed the initial motivation for the study. More suitable approaches for transforming strategic goals into operational measures were called for. Therefore, the strategic plan has been a key document throughout the study, and the research has used it as a basis for its proposals.

3.2.1 Three Japanese cases

As described in section 2.4, this study has used three supplementary cases in order to validate and generalize the results obtained from Southern Jutland Hospital. During an external stay in Japan, three hospitals supplied qualitative as well as quantitative data. The three Japanese cases each represent a healthcare sector different from the primary case in the thesis. This enhances the differences between the two healthcare sectors, enlightening which adaptations have to be performed in order to apply the proposals in both sectors. The results of the benchmarking study are presented in section 4.4. These results were presented at the EurOMA conference in Porto, 2010. A short introduction to the three secondary cases follows to provide the reader with an understanding of the specifics of the cases beyond the descriptions in the papers.

Tokyo Women's Medical University (TWMU)

TWMU is a medical university that includes educational, clinical and research environments. Traditionally, all undergraduate schools are devoted to developing women's

professionalism, although they are open to both genders, and is one of the largest hospital complexes in Japan.

The data material collected at the hospital stem from a dialysis department, in contrast to the material from Southern Jutland Hospital. The material used is however primary in relation to employees and patients, so the data is somewhat comparable with other types of departments. The discussion on comparability is elaborated in Paper 3.

Tsukuba University Hospital
Tsukuba Science City is located at the centre of Tsukuba City, 60 km northeast of Tokyo. The university was established in October 1973, due to the relocation of its antecedent, the Tokyo University of Education, to the Tsukuba area. In addition to the normal function as medical facility, Tsukuba University Hospital has both education and research facilities on site.

The case used at Tsukuba University Hospital is the Proton Medical Research Centre. The centre is a radiology department, but a very advanced one. The centre is placed at the high end with regard to technological equipment and as the only facility in the world has two proton scanners for cancer treatment. The use of a radiology department as foreign case provides insight into the differences between similar departments in two very different cultures.

Tagawa Municipal Hospital
Tagawa City Hospital is a regional hospital in Tagawa, with approximately 95,000 discharges per year. The hospital has 334 general beds and a new dialysis department with 50 beds, specifically for dialysis patients. Dialysis is a core specialty in this area, as the effects of the nuclear bombing in 1945 are still evident. Thus, the dialysis department performs comprehensive patient treatment along with children surveys, education of dialysis doctors, and research on dialysis treatment.

The dialysis department is the case at Tagawa. Since the aim was to collect comparable performance data, dialysis-specific data were excluded and data regarding employees were in focus. No interviews were made at Tagawa City Hospital, which means that all data are quantitative.

3.3 Qualitative data
Since the aim of the study is to propose a framework adapted to operational decision makers in the healthcare environment, qualitative enquiries are needed to extract information on which to base the proposals. Two qualitative data collection methods were applied: 1) interviews, as a way to gain insight into the healthcare environment plus provide understanding of differences in perception among key employees; and 2) workshops, as a more discussion-based development method. In the following, the two methods are discussed, along with their potential and their limitations.

3.3.1 Interviews

As a way to gather information, interviews were repeatedly used during the initial and final phases of the study. Interviews function in this study as a method for collecting empirical material in a closed environment with as little interference as possible (Pope & Mays 2006). In the final stages of the study, interviews function as validation of results, since understanding of applicability and validity can be individually assessed. This allows the inexperienced researcher and the subject to discuss matters related to the topic. The interviewing technique applied in the Danish cases was the semi-structured interview (Ulin, Robinson, & Tolley 2005). As opposed to controlled interviews, semi-structured interviews allow discussions to extend beyond the researcher's knowledge. The interviews were intended to allow free-flowing conversation between subject and researcher about the subject's roles and responsibilities in the organization and the activities in which the subject was engaged. By discussing topics of interest, proposals, qualitative data etc., the subject can elaborate his/her viewpoints. This was regarded in the initial states of the study as a suitable approach. In the later stages of the study, discrepancies between the researcher's perception of the organization and the statements gathered from the interviews were discussed and resolved (Kreiner & Mouritsen 2006). Since the researcher has more in-depth knowledge in this phase of the study, more discussion-based interviews could take place and those interviewed could be more engaged in dialogue about proposals than in telling about their daily work.

In the Japanese cases, it was a bit different; more structured approaches were necessary. Because those interviewed in Japan were seldom fluent in English, translation was necessary to some extent. This meant that more structured questions were needed to aid the translation process. This also gave interviewees an impression of professionalism, which was important in a more formal culture like the Japanese. This approach of course changed the nature of the data collected, since conversations never went beyond the scope of the questions.

3.3.2 Workshops

Using one single method for collecting qualitative is regarded as too limited for gaining a holistic view of the organization. Mackay explains this limitation thus:

> Whilst individual detailed case studies based largely on management interviews could potentially provide rich sources of data they lack the capacity as a sole method to inspire more structured sense making debate and generalizeable management process theorizing.
>
> (Mackay et al. 2008)

As response to this, workshop methodology was adopted to extend the discussion about the proposals into a larger context. Workshops were arranged following new realizations, with validation of results as the primary target. Involving operational personnel in the progress of the study increased their interest in the development. During workshops,

experts discussed the background material and proposed new angles for the subsequent investigations. This helped design the next stages of the study, which could thus include input from workshop participants.

3.4 Quantitative data

As a valuable part of the validation process, testing was carried out using the quantitative data is a prerequisite for assessing the reliability of the proposals. For these trials, internal as well as external data were applied. The following sections briefly describe the sources of the empirical data.

3.4.1 Internal data

Internal data stem from various sources. At the hospital level, HR databases and the Hospital Information System (HIS) are the primary sources. HIS is a comprehensive, integrated information system designed to manage the administrative, financial and clinical aspects of a hospital. In the specific case of the radiology department, the Radiology Information System/Picture Archiving and Communication System (RIS/PACS) have supplied modality-specific data. The PACS component is a computer system that interfaces with the medical imaging device (i.e. x-ray, CT scan, MRI, ultrasound etc.) to capture the image in digital format. Once captured, the image can be stored, manipulated and transmitted over a computer network. The RIS component interfaces with the existing hospital information systems to capture patient demographics, scheduling and examination orders. For the cases in Japan and at Bispebjerg Hospital, the data were collected by clinicians at the sites and provided to the researcher.

3.4.2 External data

External data were collected from four federal units and governmental agencies:
1) Unit of Patient-Perceived Quality's survey of patients' experiences in Danish hospitals, a patient satisfaction survey conducted every two years (The Unit of Patient-Perceived Quality's website 2009). The objective of the survey is to benchmark patient experiences by comparing responses across hospitals over time. The survey includes 30 questions which are answered by about 30,000 patients. In addition, Danish Quality Model, a Danish accreditation institution, assesses how well information is distributed to patients (The Danish Institute for Quality and Accreditation in healthcare website 2009). This information is regarded as fundamental for determining the level of patient satisfaction.
2) The Danish Quality Model is an accreditation framework developed by the Danish Institute of Quality and Accreditation in healthcare. The model itself consists of 35 standards related to organizational issues, 54 standards focusing on the continuity of care, and 15 specific disease-related standards. All of these standards contain indicators related to different organizational levels.
3) The National Indicator Project aims to evaluate various forms of treatment: acute surgery, chronic obstructive pulmonary disease, diabetes, heart failure, hip fracture, lung cancer, schizophrenia, and stroke (the National Indicator Project's website 2009).
4) Patient safety records created by the National Board of Health (The National Board of Health's website 2009) and the Danish Patient-Safety Database (The Danish Patient

Safety Database's website 2009). It is important to note that all external data are public and validated by the federal units and governmental agencies issuing them.

3.5 Summary

The research study deals with one main case, a radiology department at a public Danish non-profit hospital. The hospital consists of four sites, which are merged at management level, but each site functions as an independent operational unit with almost all specialties. The case is the primary data source throughout the study, with focus on the specification of the strategic plan. The managerial group that initiated the research project has been continuously involved in validating and developing the proposals throughout the study. In addition, three Japanese have acted as secondary cases, aiding both in the development and the validation of proposals. The three Japanese cases were used to evaluate the generalizing potential, by testing the proposals in relation to cultural and organizational issues. The range of cases utilized in this study aims to secure reliability in generalizing results, whereas a single case would limit this considerably. Regarding cases that bear similarities but have differences in either legal foundation or clinical focus, the reliability of the proposals are tested with respect to generalizing potential.

Interviews and workshops are used for collecting qualitative data, and both internal and external databases have provided the quantitative data. Each data set contains valuable information for different aspects of the study. Qualitative data forms the basis of the formulation of the proposals, while quantitative data is used to test their practical applicability and validity. The data collection in Japan differed from the collection in Denmark. Primarily due to the language barrier, more structured approaches were chosen. In the Danish cases, more open dialogue was possible, which to a great extent allowed the conversation to extend beyond the initial idea and shaped further activities, as interesting new topics were introduced.

Chapter 4 - Model construction

This chapter presents the scientific path that lead to the thesis' final recommendations. Based on five papers published and submitted throughout the research project, the sections summarize the individual contributions of the papers. The final section discusses the fundamentals of the logic and assesses the expected usability and the challenges of implementing the model.

4.1 Introduction

Five papers are included in this thesis. Collectively, they describe the scientific route the research project has followed. The papers stretch over almost two and one-half years, from first to last, and each represent steps towards the final recommendations. They have been submitted and published at international conferences and in international journals to verify their scientific integrity. The papers all share the same main theme – development of a framework for measuring healthcare performance – but cover different aspects of performance measurement. The chapter presents the papers chronologically, which provides the reader with an in-depth clarification of the development of the final recommendations in this research project.

The first paper deals with the concern of aligning and visualizing performance measurement structures. The second paper opens the discussion of using aggregated indicators in a internal benchmarking context. The third paper is based on external research conducted in Japan, where the aggregation approach is tested as the guiding structure in international benchmarking. The fourth paper is an in-depth study of a MRI section at Haderslev Hospital, which describes how the weighed indicator hierarchies can assist decision-makers in obtaining strategic alignment throughout all organizational levels. Finally, paper five combines all the experience into a measurement framework, where Performance Accounts provide the guiding structure for evaluating healthcare performance. The chapter highlights the contributions of the individual papers, and gives details as to which sub-conclusions contributed to the final recommendations. Further elaboration, stretching beyond the conclusions in the papers, is provided, in relation to both empirical and theoretical aspects.

4.2 P1 - The importance of structured visualization

Full paper title: A new approach for translating strategic healthcare objectives into operational indicators

The objective in this work was to analyze the difficulties encountered by operational decision makers when strategic plans are to be translated into operational performance measurement in healthcare organizations. The scientific point of departure was the hypothesis that as complexity in modern healthcare increases, the development of measurement structures likewise becomes increasingly complicated. The task of evaluating operational performance and initiating corrective actions is becoming more and more demanding as the sheer number of high-level indicators increases. The motivation behind this paper was to propose a new approach for structuring and visualizing performance indicator structures for healthcare organizations. The empirical foundation for this work is the deployment of the strategic plan at the Southern Jutland Hospital.

4.2.1 Visualization of measurement structure

The conceptual model consists of a three-dimensional relation matrix, based on the CIMOSA representation. The first axis describes the strategic objectives of the organization; the second axis describes the organizational levels; and the third axis is an evaluation axis (see Figure 6). As a visual platform, the framework is able to portray a strategic plan within the three axes (see Figure 7). The rational for this construction was to accentuate how performance indicators are related to each other, with regard to both internal and external reporting.

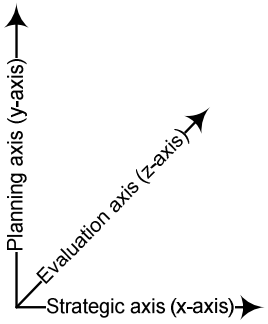

Figure 6. Structural outline (from P1)

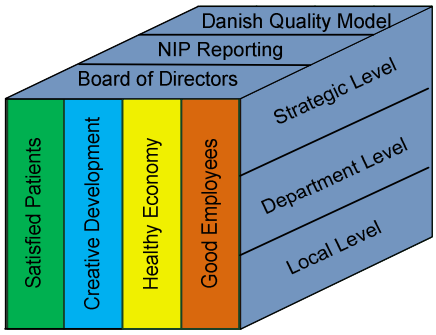

Figure 7. Strategic plan "Quality 24/7" (from P1)

The structure should provide decision makers with a tool to assess whether the current system of indicators are adequate in terms of covering the objective of a given strategic plan. Performance indicators are placed within the "cube" to specify the area of responsibility of the individual indicator. The transparent structure provides a visual representation of which indicators are obtainable in the different aspects of the organization. It also provides insight into where in the organization the reporting responsibility is placed.

Figure 8 presents how the structure of the indicators is positioned, where hierarchies of indicators are deduced from a strategic level to the operational level of the organization. This approach visually illustrates the completeness of the measurement system, as indicators are only put in place if they have a dedicated purpose.

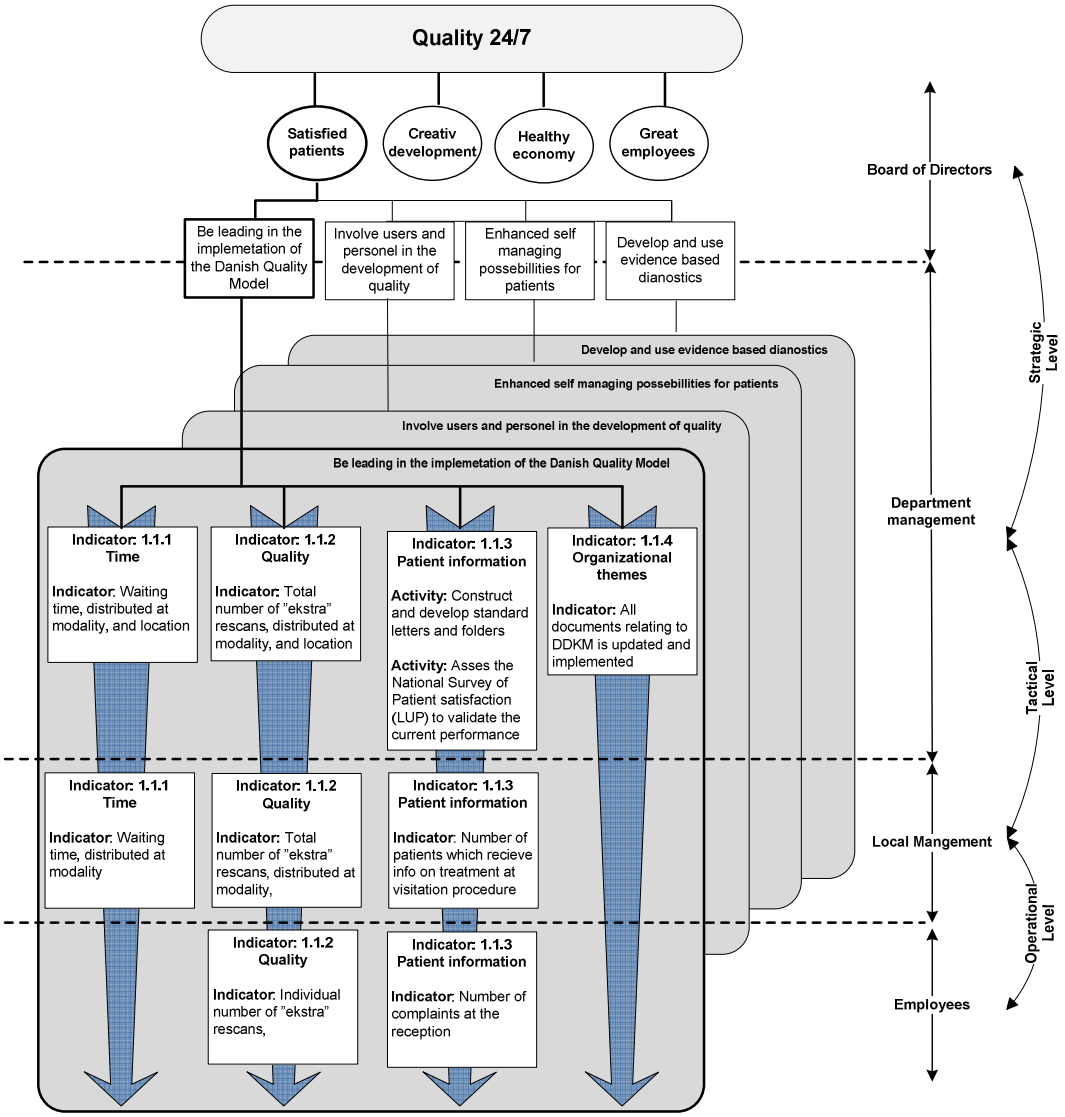

Figure 8. Hierarchical indicator structure

The logic is that indicators are not to be deduced to a level of the organization where they do not have a specific purpose. The justification for creating "stop-rules" is to avoid redundant indicators that do not provide valuable feedback to management. The process of deducing objectives through the organization drives decision makers to explain why a particular indicator is valuable in a given context. The three-dimensional construction visualizes indicators in dedicated "slices" of the cube, presenting the indicator as an integrated part of the measurement framework. Figure 9 illustrates how an indicator for waiting time would appear in strategic context at the hospital and is used in several external reports.

	Description
Indicator name	Waiting List
Purpose	Continuously monitor the maximal waiting time for a non-acute patient, distributed on modalities
Responsible	Head of department
Field of application	Each of four radiology sections of the hospital
Indicator description	Waiting time to the next open examination slot in the booking system for each modality
Displaying guidance	Y-axis: Waiting time in days;: Calendar days 6 month back
Data foundation	Data is collected from RIS (Radiology Information System)
Indicator goal	Waiting time below 20 days, complying with National Treatment assurance (4 weeks)
Timeframe	At all times
Guiding documents	The Danish Quality model (www.ikas.dk) The National Indicator Project (www.nip.dk)
Benchmark	Monthly benchmarked internally between all four locations Bi-annual, waiting time is benchmarked externally between Danish hospitals
References	The Danish Quality Model, Standard 3.1.1- Standard 3.2.1- Standard 3.6.1 - Standard 3.8.1- Standard 3.11.1

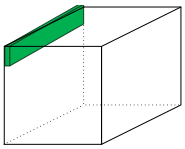

Figure 9. Indicator example (from P1)

The template in Figure 9 also guides the deduction of the strategic objective to indicators. For each of the indicators, a standard template of indicator information has to be provided, all of which contribute to provide transparency in the measurement system. The construction of the conceptual framework accentuates performance indicators in the organizational context in which they are applied.

4.2.2 Contribution

The paper proposes that measurement structures be built in a hierarchical construction, where the indicators are designed in relation to a specific purpose in a specific context. The visual representation put forth in this paper has two primary purposes. First, the proposed three-dimensional structure provides logic and transparent representation of the performance measurement system. Second, inadequate measurement structures become apparent. The aim is to improve the completeness of the performance measurement systems configuration so that redundant indicators are eliminated. The hierarchical construction calls for strategic alignment, since indicator structures are deduced through the organization from the strategic objectives. This consolidates the alignment of indicators operating at the operational levels of the organization with strategic objectives.

4.3 P2 - Aggregated indicators in internal benchmarking

Full paper title: Benchmarking in healthcare using aggregated indicators

Based upon the previous paper, the work focused on the question on how to bring the conceptual ideas into a consistent measurement framework, applicable within the radiology department. The starting point was a desire to design a hierarchical construction of performance indicators, derived from a strategic plan, through to the operational level of the organization. The hierarchical construction aims to describe the levels of measures that apply within a healthcare organization, and at the same time secure transparent representation of performance throughout the hierarchy. In order to meet these conditions, the concept of aggregated indicators has been used as the guiding principle in the model construction. Aggregated indicators use indices of performance as common denominators for all included indicators. Aggregated indicators rely on mathematical summarization of the outcome of individual measures, combined into superior united indicators. Nakajima introduced the use of aggregated indicators in an Overall-Equipment-Efficiency indicator (OEE). In his work, *Availability*, *Performance*, and *Quality* were combined into one single measure of performance. But the challenge in this work is more versatile than the single-stakeholder view presented in Nakajima's paper. This study is constructed as a benchmarking study of the four individual radiology sites that make up the radiology department at Southern Jutland Hospital. The aim was to test whether aggregated measures were a valuable guiding principle in assessing performance differences between the four sites.

4.3.1 Aggregation of healthcare performance

Figure 10 presents the conceptual outline, which aims to provide one aggregated measure that justifies the performance level of the individual site. Performance outcomes are continuously aggregated from lowest level to highest level and present a higher collective expression of performance.

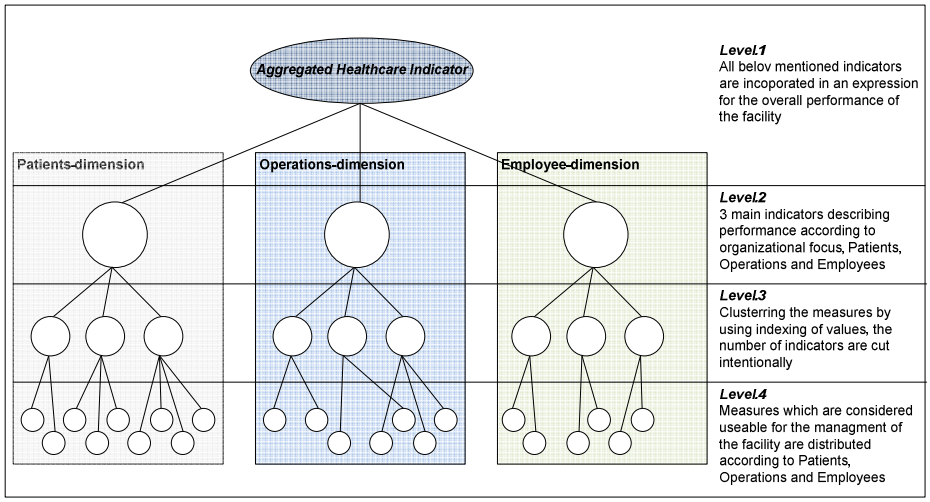

Figure 10. Structural outline

The structure assists the identification of poor performance in a complex system of performance measures. The aggregation of indicators in clusters, and further into a higher level indicator provides the possibility to trace performance 'downwards'. It is important to note that aggregated indicators only make sense in comparisons. The aggregated indicator is a fictional number that represents an estimate of a subsidiary level's outcome. The superior aggregated measure provides a level of performance that only makes sense when compared to other aggregated measures in a similar hierarchy. That is why this approach is suitable for benchmarking between identical sites, because when the benchmarking hierarchy is identical, the aggregated measures become comparative.

Based upon the structural outline, workshops and interviews were used to construct a hierarchy of indicators that would represent three strategic dimensions: Patients, Operations and Employees. As presented in Figure 11, the dimensions are deduced into clusters of performance, which again is deduced into operational indicators.

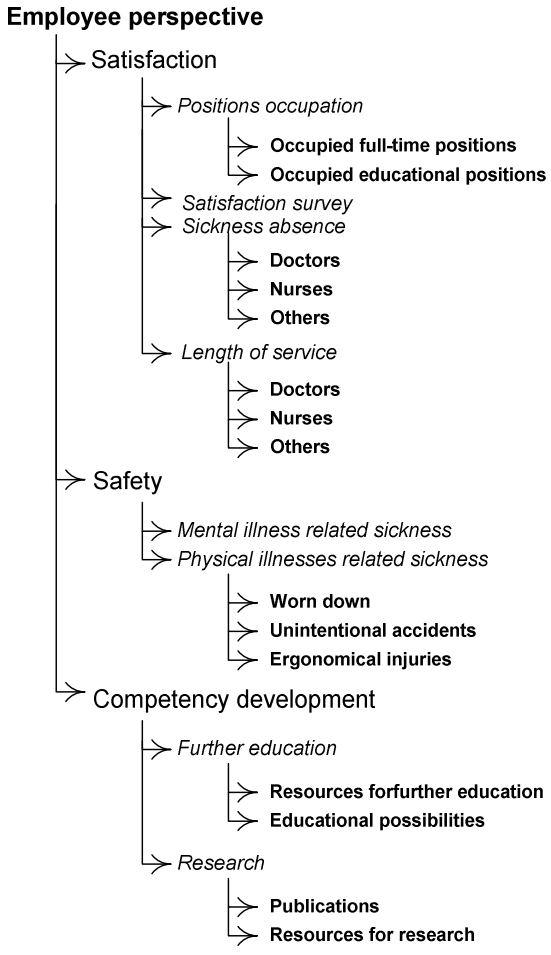

Figure 11. Employee dimension (from P2)

The workshops focused on identifying indicators that would constitute the strategic plan. No specifics were given to participants to decide which indicators were to be used, so the selection was solely based on the perception of the participants. The indicators that were implemented were primarily repetitions of the indicators already used, just clustered according to the hierarchical construction.

4.3.2 Comparing apples and oranges

The next challenge is the construction of an aggregation procedure to mathematically summarize the indicator outcomes into higher-level indices. Because the study was conducted as a benchmarking study, and all sites were evaluated with the same set of indicators, the procedure was constructed on the basis of averages of performance. The calculations (see example in Table 7) were performed in three consecutive steps;

1. For each indicator averages for all involved locations' specific results are calculated. (e.g. patient satisfaction = 81%)
2. Based on this average, a location specific index is calculated (e.g. Location 1 = 1.1)
3. To present the aggregated result for each location, an average of the indexes is calculated (e.g. aggregated result Location 1 = 0.95).

	Loc. 1	Loc.2	Loc. 3	Average	Formula	Index Loc. 1
Patient satisfaction	**90%**	**84%**	**69%**	***81%***	$Index = \frac{90}{81}$	**1.1**
Capacity	**0.3**	**0.5**	**0.7**	***0.5***		**0.6**
Length of service	**5 years**	**2 years**	**6 years**	***4.3 years***		**1.16**
Aggregated Result					$\frac{1,1 + 0,6 + 1,16}{3}$	**0.95**

Table 7. Benchmarking procedure (from P2)

The use of a parametric framework, as opposed to simply reporting a single measure over time, has several advantages. First, because there are multiple measures, estimates of the performance differences of each individual measure are apparent. However, aggregation of this sort only makes sense when the measures themselves are highly correlated, both within and across periods. In cases where there is low correlation among measures, there is a risk of losing information that might be specific to a particular measure (Swaminathan, Chernew, & Scanlon 2008).

To follow up the previous paper's recommendations on transparent representation of performance structure, the benchmarking results were provided as spider charts (see

Figure 12). The representation provides a clear identification of strengths and weaknesses between the sites.

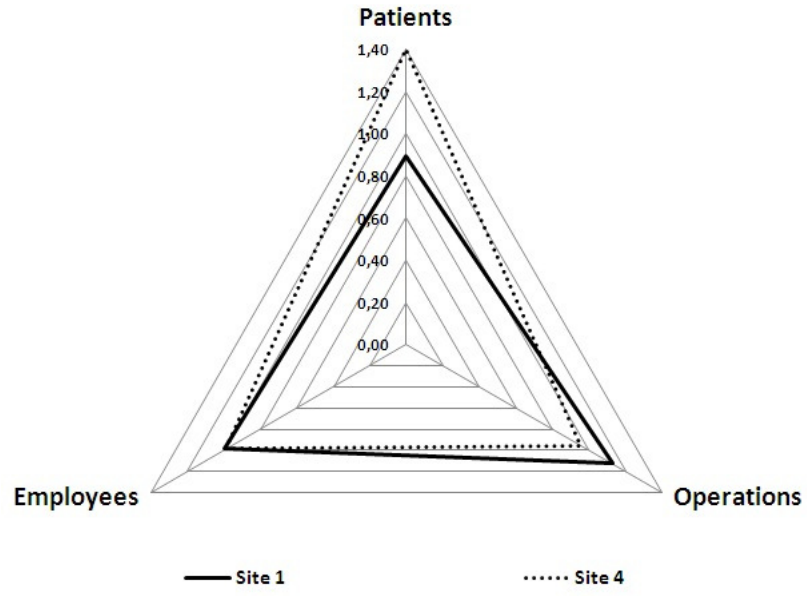

Figure 12. Benchmark result (from P2)

Digging deeper into one dimension, the array of indicators constituting the dimension becomes apparent. In this way, it is possible to trace performance through the levels. This further assists the decision maker to identify suitable areas for improvement. The visual representation also has the strength that it is rather simple to interpret. Employees with no training in management would be able to interpret and presumably identify in which respects a given radiological site is weaker than the other sites.

4.3.3 Contribution

This work has proved that aggregated indicators are a valuable method for creating a guiding structure for internal benchmarking. The presented framework combines measurements from different stakeholders into one unified representation of performance. By benchmarking a department against averages for other departments, the model shows strengths and weaknesses in relation to other departments. The approach tries to represent the holistic nature of healthcare performance by clustering operational indicators in highly correlated groups. The framework is tested as a benchmarking tool, but has obvious potential as an in-house decision support system. The use of the model for performance management in healthcare is thus further legitimized by its not being dependent on the number of indicators used. Mathematical aggregation provides the possibility of including as many indicators as desired, because averages will even things out at the higher levels. However, it must be noted that indicators in 'large' clusters will mathematically have less weight than indicators in 'small' clusters. Therefore, the qualitative construction of the hierarchies is of extreme importance in terms of achieving a superior result.

The transparency is also enhanced in comparison to the previous paper. Tracing of poor performance is made simple, since the representation in hierarchies is intuitive to most people. The result is shown in a spider chart, which further enhances the identification of areas suited for corrective actions without the compromising strategic alignment. The clustering and selection of indicators are of course guides in relation to the superior outcome, but this particular construction allows departments to customize the measurement system to their settings. This gives the framework certain generalizing advantages in terms of further development.

4.4 P3 - International Benchmarking

Full paper title: Operational benchmarking of Japanese and Danish hospitals

To test the consistency of the framework, the paper addresses international benchmarking of operational performance. By applying the hierarchical structure at hospitals in different countries, the framework's ability to identify performance differences is tested. Moving from internal horizontal benchmarking (P2) into competitive benchmarking enhances the value of the performance information considerably (see Figure 13) and intensifies the capability requirements. This further proves the legitimacy of aggregated indicators as a performance evaluation tool.

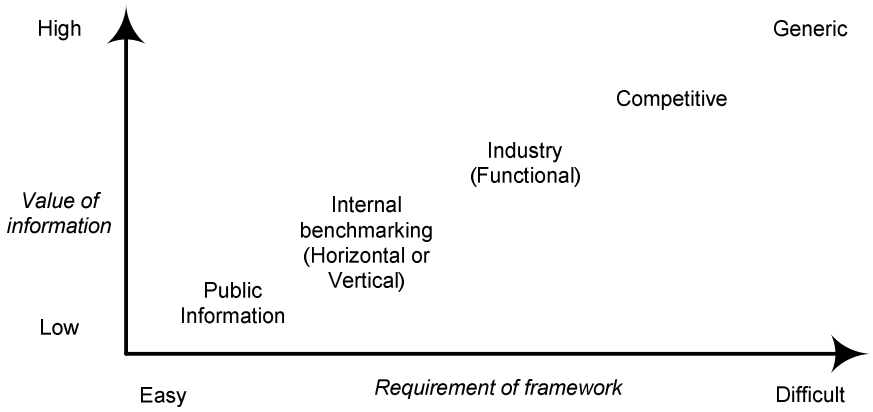

Figure 13. Types of benchmarking

The purpose of this paper is therefore to assess the generalizing potential of the framework by increasing the challenge through testing on international benchmarking. The analysis is conducted to test whether the framework is capable of tackling some of the challenges in international benchmarking, such as cultural differences, jurisdiction, organizational structure etc.

The benchmark was developed in a comparative study, where researchers and clinicians from Denmark and Japan were involved. The development of the benchmarking model was performed as a multiple case study, consisting of seven case departments, four Danish

and three Japanese. The first Japanese hospital is a public hospital of a local municipality; the second and third are university hospitals, belonging respectively to a national and private university. The Danish cases are the four sites at Southern Jutland Hospital, which also constituted the data foundation for the previous paper.

4.4.1 Aggregated indicators as international benchmarking structure

Building upon the same hierarchical structure as the previous paper (see Figure 10), the measurement system intends to highlight the differences in operational performance among the case departments within the study. By evaluating Danish and Japanese hospitals on indicators that are applied in both sectors for decision support, country-specific differences become apparent. The focus is not on high-level indicators but on describing operational performance for the departments. By aggregating performance in a hierarchical structure, the paper tries to compensate for some of the empirically described challenges in international benchmarking.

As explained in the previous benchmarking study, normalization of the data is necessary. But the normalization method is changed from using simple averages of performance to the using the standard score, more commonly referred to as the z-score (see Equation 1). The z-score corresponds to a data point in a normal distribution. The objective is to convert all indicators into a common scale and thereby make them comparable regardless of the initial data.

$$z - score = \frac{(Data\ point_n - Mean)}{Standard\ Deviation}$$

Equation 1. z-score

The justification for changing the normalization method is that the z-score encourages mean scores over high variations, which fulfills one of the primary objectives for healthcare organizations in complying with standards for acceptable performance. It is regarded as more desirable for hospitals to perform acceptably on all indicators than to perform perfectly in some and poorly in others. This constitutes a cornerstone in the reduction of performance inconsistency in delivered care. The benchmarking procedure therefore changed slightly as the standard deviation of the hospitals' mutual performance was calculated as part of the benchmark (see Figure 14).

Indicator cluster

	Data-point					Z-score		
	Hosp. 1	Hosp. 2	Hosp. 3	Mean	St. Dev	Hosp. 1	Hosp. 2	Hosp. 3
Indicator 1	v1	..	..	m1	St.d 1	z1	..	..
Indicator 2	v2	..	..	m2	St.d 2	z2	..	..
Indicator ...	v..	..	..	m..	St.d ..	z..	..	..
Indicator n	vn	..	..	mn	St.d n	zn	..	..
Benchmark result for Indicator cluster No. X						$\sum_{i=1}^{n} zi$	..	..

Step 1 Step 2 Step 3

Figure 14. Benchmarking procedure (from P3)

The procedure is still performed in three successive steps, which continuously form the aggregated indices. The statistical consistency will be further legitimized as more data points are implemented in a benchmark. Uniformity among performance outcomes also contributes to the consistency of the superior aggregated index. However, it is regarded to be a valid approach, even with a relatively low number of participants in the benchmark.

4.4.2 Interpreting aggregated performance data

As the normalization of data is altered, the interpretation is also changed. Since performance is presented as an index related to the standard deviation, the index corresponds to a point in a normal distribution (see Figure 15).

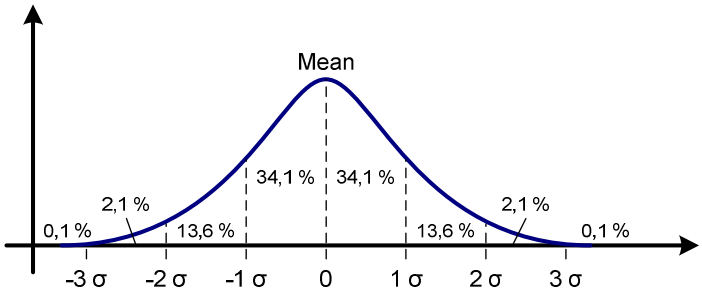

Figure 15. Normal distribution

This implies that the performance outcomes above the mean are indicated by a positive index, and performance outcomes below the mean are represented by a negative index. The magnitude of the index signifies the divergence from the mean, positive or negative respectively.

Table 8 presents the results of the benchmark; red negative numbers indicate poor performance and black positive numbers indicate good performance, compared to the other facilities within the benchmark.

	Japanese hospitals (z-values)			*Danish hospitals* (z-values)			
	Hospital 1	Hospital 2	Hospital 3	Hospital 4	Hospital 5	Hospital 6	Hospital 7
Complaints	0.47	-0.60	-2.05	0.46	0.47	0.63	0.63
Waiting times	0.79	0.79	0.79	-0.37	-0.06	-2.00	0.06
Adverse advents	1.37	-1.15	0.36	-0.14	0.36	-1.66	0.87
Patients dimension	***0.88***	***-0.32***	***-0.30***	***-0.02***	***0.26***	***-1.01***	***0.52***
Sickness absence	0.74	0.74	0.68	0.00	-0.37	-1.40	-0.40
Position occupation	0.34	0.01	0.44	-0.39	0.30	-0.45	-0.25
Staff turnover	-0.13	0.06	-0.12	0.24	-0.13	-0.25	0.33
Length of service	1.34	0.27	0.70	-0.58	-0.58	-0.58	-0.58
Employees dimension	***0.57***	***0.27***	***0.43***	***-0.18***	***-0.19***	***-0.67***	***-0.22***
Equipment utilization	0.62	1.08	0.42	1.22	-1.11	-0.91	-1.31
Clinical errors	1.02	1.11	1.01	-0.96	-0.72	-1.06	-0.39
Overwork	-0.47	-2.00	-0.52	0.77	0.64	0.83	0.76
Operations dimension	***0.39***	***0.06***	***0.30***	***0.34***	***-0.40***	***-0.38***	***-0.31***
Benchmark result	***1.67***	***0.12***	***0.59***	***0.23***	***-0.50***	***-1.86***	***-0.24***

Table 8. Detailed benchmark results (from P3)

From the results, it is apparent that Japanese hospitals perform better than the Danish hospitals. High equipment utilization and few clinical errors are achieved to some extent by a great deal of overtime among Japanese healthcare staff. Danish hospitals pay the price of productivity by focusing on satisfying the caring needs of patients and limiting working hours for employees. These results resemble what could be expected from the comparison on the basis of the conclusions in productivity studies of Danish and Japanese industrial production.

The discrepancy among the indices symbolizes large structural differences between the benchmarked parties. Japanese hospitals manage in-house logistics and patient care differently than Danish hospitals. These differences are highlighted by the proportionally large performance indices, which in several cases exceed 2 σ, signifying a performance discrepancy of at least 95 percent from the mean in a normal distribution. The results also point towards the difficulties in conducting benchmarking when the differences are as obvious as in Danish and Japanese healthcare. The results from the Japanese and Danish hospitals are close to resembling a mirror image that reflects the structural differences. But even though the differences are large, the framework succeeds in presenting the differences between the two sectors.

4.4.3 Contribution

The focus of the paper was not to compare Japanese and Danish healthcare in order to find the 'best' healthcare system. The aim was to test whether the framework was able to

reveal operational performance differences between healthcare sectors. The revealed differences between Japanese and Danish healthcare reliably resemble the structural differences between the two healthcare sectors. This identification of performance differences supports the conclusion that the structure of the framework is suitable for evaluating operational healthcare performance for both internal and external benchmarking. Indicating performance as indices proportional to normal distribution further contributes to transparency in the pursuit of identifying poor performance. Positive and negative numbers provide a logical representation, which most employees can relate to. This contributes to fulfilling the general aim of constructing performance measurement that enables healthcare managers to identify easily areas in need of corrective actions.

Even though the framework is presumed suitable, there are some difficulties with international benchmarking that are not handled by the framework. These challenges are primarily caused by cultural and structural differences and availability of data. By aggregating averages of z-scores, the mutual importance of individual indicators is not accentuated. Therefore, the results are not adjusted for organizational focus, which in the case of Japanese and Danish hospitals is very different. Allocation of weight profiles for within the indicator hierarchies would therefore be a way of enhancing the consistency of the model. Likewise, the indices are not particularly useful for comparisons between diverse organizations, where uniformity in organizational structure would add reliability to the result.

4.5 P4 - Securing strategic alignment

Full paper title: Rethinking performance evaluation in healthcare

The previous two papers both discuss horizontal benchmarking of operational performance, internally and externally respectively. This subsequent paper tries to analyze vertical performance evaluation to further test the generalizing potential of previous conclusions. Vertical performance evaluation aims at securing strategic alignment throughout the organization. The focus in this study is to analyze the capability of the framework to describe operational performance as a function of strategic objectives. Specific information of this kind should place operational decision makers in a position where identification of poor performance becomes simpler. By easing the identification of performance problems in relation to strategic objectives, the probability is enhanced for the right decisions to be made throughout the organization.

The empirical basis is again the radiology department, though the quantitative material is collected exclusively in the MRI at Haderslev hospital. The justification for choosing this particular case is that it represents a borderline between tactical and operational management at the hospital. Since there are two management levels above the MRI unit (Board, Department management), it is obvious to test strategic alignment at this level.

The MRI unit is small, which further limits the relation to the strategic plan, causing difficulties with determining strategic progress or retreat. Accordingly, the aim was to use previous experiences from the study to construct a representation of strategic change at the MRI unit. This could prove whether the MRI unit is in strategic alignment, and indicate where corrective actions could be initiated.

4.5.1 Strategic alignment

Strategic alignment is the adjustment of decision making throughout the organization, in order to optimize performance in relation to overall organizational objectives. This signifies that actions taken on the operational levels should be in line with the desired direction of the organization's strategic plan. This is common sense to most managers. But the challenge lies in the practical accomplishment, which demands a very high degree of transparency of the organizational objectives. To achieve organizational alignment, operational decision makers need to be able to identify performance inconsistency and make their decisions on the basis of this knowledge. But besides a coherent structure of indicators, the strategic importance of the indicators has to be evident if decision makers are to be able to make this identification. Different organizational areas are inevitably of different strategic importance, and the relative priority among these areas has to be clear to the decision maker.

These prerequisites demand several strong points in a performance evaluation framework in order for it to be able to illustrate the extent of alignment within the organization. First, the selection and placement of indicators ought to be performed in order to reflect organizational interests throughout the organization. The selection of suitable indicators is regarded to be of critical importance, because it establishes the organization's goals and priorities. Second, the framework has to incorporate a structure for mutually prioritazation the indicators, assigning weights in accordance with strategic significance. Third, the indicators that are implemented need to be normalized in order to present a unified expression of strategic change. This enables aggregation of performance indices, which enables quick identification of performance problems. Combined, these form the basic requirements for a performance measurement system that is capable of portraying operational performance in relation to strategic objectives and customizing the framework to the specifics of the individual facility. The framework would thus be applicable within most settings, and thereby strive towards a generic structure.

4.5.2 Weighted and aggregated indicators

Normalizing and aggregating performance outcomes have previously been discussed, as well as the construction of suitable indicators into hierarchies of indicators. The lacking element is the weight assignment procedure, which would enable representation of performance indices adjusted for organizational importance. Without individually assigned weights, indicators in 'large' clusters will mathematically have less weight than indicators in 'small' clusters, as long as the comparison is made with simple averages. This arrangement is representatively misleading, because some indicators simply support a decision, while others are governing in terms of which decision is made. To compensate

for this, the concept of Analytical Hierarchy Process (AHP) is adopted. The AHP method provides the ability to make a quantitative distinction between the elements within the framework.

> The analytical Hierarchy Process (AHP) provides a comprehensive framework to cope with the intuitive, the rational and the irrational in us all at the same time when we make decisions. It is a method which we can use to integrate our perceptions and purposes into an overall synthesis. The AHP does not require that judgments be consistent or even transitive.
>
> (Saaty 1982)

The justification for applying the AHP method is that it allows for subjective assessment as well as objective assessment of mutual importance among elements. In healthcare, the subjective assessment is of particular importance, because there is not always a rational or quantitative reason why some areas are prioritized more than others. Issues like political influence, media pressure, patient complaints etc. may change priorities. That is why subjective assessment of mutual importance is most central to the weight assignment procedure within healthcare.

For the framework, this means that after the hierarchy of suitable elements is constructed, a systematic comparison of the incorporated elements is conducted. The elements are compared in pairs, and the decision makers assign values of relative intensity to the individual elements. Subsequent to the assessment of mutual importance, the AHP method is used to perform a mathematical calculation assigning interdependency values. This provides a weight profile throughout the hierarchy, which enables for aggregation as weighted averages of z-scores (see step 4 in Figure 16).

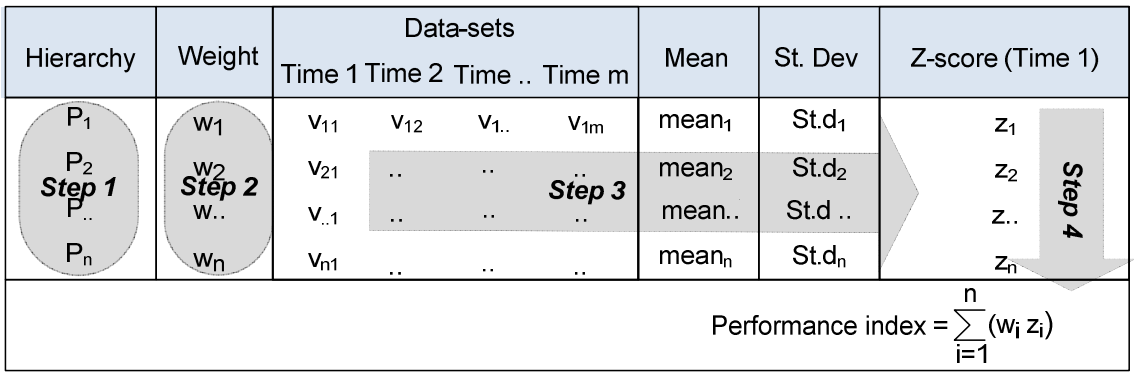

Figure 16. Schematic outline of evaluation framework (from P4)

The performance index that is finally determined as the weighted average now represents performance outcomes as a representation of organizational importance. Low-priority indicators will not have as much impact as high-priority indicators. This paper's results include 27 performance indicators in the hierarchy, distributed in 9 clusters (see Table 9).

The normalization of the performance outcomes is calculated on the basis of past performance data, since the aim was to represent the MRI unit´s strategic progress or regression. The allocation of weights was conducted in workshops where clinicians discussed the individual importance of indicators. Due to the evaluation method, positive values indicate that an organization is performing above average in the retrospective data.

	Dimension *(weight)*	z-score	Cluster *(weight)*	z-score	Indicator *(weight)*	z-score
					Adverse advents *(0.630)*	0.00
			Safety *(0.503)*	0.09	Incorrect treatment *(0.250)*	0.31
					Re-called patients *(0.120)*	0.12
					Received written info *(0.463)*	0.45
	Patient *(0.573)*	0.28	Information *(0.348)*	0.18	Satisfaction (written info) *(0.329)*	0.76
					Satisfaction (oral info) *(0.208)*	-1.33
					Satisfaction survey *(0.586)*	1.88
			Satisfaction *(0.148)*	1.13	Waiting time for treatment *(0.224)*	0.69
					Complaints *(0.190)*	0.00
					Part-time employees *(0.595)*	0.45
			Occupation profile *(0.570)*	0.30	Available posts *(0.277)*	0.13
					Educational positions *(0.129)*	0.00
					Overtime *(0.438)*	-0.95
Result 0.13	Employee *(0.320)*	0.07	Work environment *(0.259)*	-0.36	Sick leave *(0.240)*	0.45
					Turnover rate *(0.202)*	0.20
					Satisfaction survey *(0.120)*	-0.77
			Risk *(0.171)*	-0.08	Reported work hazards *(0.833)*	0.11
					Long-term sickness absence *(0.167)*	-1.00
					Acute load *(0.387)*	-1.06
			Planning *(0.684)*	-0.62	Non-Attending patients *(0.443)*	-0.48
					Cancelled examinations *(0.170)*	0.00
					Operational time *(0.657)*	-0.53
	Operation *(0.107)*	-0.51	Efficiency *(0.244)*	0.06	% procedures (7-15) *(0.207)*	2.14
					Throughput *(0.136)*	-0.27
			Utilization *(0.072)*	-1.38	Employee utilization rate *(0.875)*	-1.37
					Equipment utilization rate *(0.125)*	-1,46

Table 9. Aggregated performance result for the MRI unit (from P4)

To a majority of managers the identification of performance problems is obvious with this particular construction. As transparency within a performance measurement framework is of outmost importance, the construction seems to have the desired advantage. The framework provides a picture of current performance compared to past performance. In addition, it explicitly describes organizational importance, which ultimately indicates whether operations progress according to organizational strategies. In the case with the particular result of 0.13, the MRI unit at Haderslev can be assumed to be making positive strategic progress. The problems are specifically concerned about the operations dimension where corrective actions should be initiated to enhance the result, although to the observer, it is apparent that operations have a lower priority than patients and employees. This might explain why progress has been made in the other areas. Regardless of the reason for differences in performance between the dimensions, transparency is obvious. This also provides a basis for organizational discussions about priorities and selection of indicators.

4.5.3 Contribution

The distinctiveness of this framework lies in the combination of normalization according to past performance and the use of the AHP concept as a method for setting priorities. This allows for monitoring the progress and regression of performance as a function of strategic importance. The framework has potential to include large amounts of information while targeting this information for use in decision support for making strategic decisions. What otherwise would have been a subjective assessment of strategic importance now becomes quantified by representing performance as weighed, aggregated measures. The strength of specific measures is still apparent, because poor performance can be easily identified and corrective actions can be initiated.

The notion of a "perceived reality" is important to emphasize, because there are no absolute values for good or bad performance when aggregating weighted z-scores. The weight profiles are somewhat subjective, since the assignment is conducted on the basis of the interviewees' perception of mutual importance. Thus, the interpretation of performance is biased to represent the "perceived reality" of those who have constructed the hierarchy and designed the weight profile. The advantage though is that the priorities are explicitly formulated, whereas in the present strategic plan, they are implicit. As long as the weights are organizationally accepted, the strategic direction is apparent to decision makers. This constitutes the primary basis for securing strategic alignment throughout the organization, from strategic plans to the daily management of operations.

4.6 P5 - The Performance Account

Full paper title: Performance Account for evaluation of strategic plans

Since the framework in P4 showed strength within detailed performance evaluation at a MRI unit, an attempt is made to expand it in order to comprehend the entire radiology

department. The aim of P5 is accordingly to test whether the methodology has merits in evaluating strategic plans for a whole department. As described in Chapter 3, the radiology department is obligated to pursue the hospital's strategic plan, which is why careful evaluation of strategic progress and regression become a key matter. Hence, this work set out to propose a framework for the structured evaluation of strategic plans by comparing all strategic areas in relation to organizational priorities. The framework accordingly provides decision makers with a map of context, which serves to point out areas suited for corrective actions.

4.6.1 The design of a "Performance Account"

The evaluation of the strategic plan takes its point of departure in the work conducted at the MRI unit, but with intensified focus on the representation of performance. This draws upon the experiences from P1, where visual representation was at the centre of attention. The mathematical construction is assumed to be suitable for evaluating strategic performance, since it showed merit at the detailed operational level. Hence, the same approach is applied to the strategic plan, where the aim was to design a "Performance Account" representing organizational progress and regression. The design phase portrayed in P5 constituted three successive steps:

1. To simplify the expression of performance, clustering techniques are proposed. Indicators are distributed in an indicator hierarchy, determined by their affiliation.
2. To secure strategic alignment, indicators are mutually weighed in order to differentiate according to organizational importance.
3. A superior performance expression is calculated by aggregating normalized performance data. The construction of the framework provides the possibility to present the results in Performance Accounts, which are suitable for identification of performance progress and regression.

Steps 1 and 2 resemble the construction in P4, where a hierarchy is designed in workshops in which healthcare decision makers participate. After the hierarchy of suitable elements is constructed, the decision makers conduct a systematic comparison of the incorporated elements. The elements are compared in pairs, and the decision makers assign values of relative intensity to the individual elements (see Table 10).

Intensity	*Definition*	*Explanation*
1	Equal importance	Two elements contribute equally
3	Moderate importance	Slight favor of one indicator over another
5	Strong importance	Strong favor for one indicator over another
7	Very strong importance	Very strong favor for one indicator over another, demonstrated in practice
9	Extreme importance	Favor for one indicator with highest possible importance
Intensities of 2,3,6 and 8 can be used to express intermediate values		
Reciprocals of above	If activity *i* has one of the above non-zero numbers assigned to it when compared with activity *j*, then *j* has the reciprocal value when compared with *i*	

Table 10. Scale for comparison in pairs

The absolute numbers for each pairwise comparison are shown in the matrix, where inverses are entered in the transposed position. It is possible to approximate the priorities from this matrix by normalizing each column and thus recover the eigenvector from the system of homogeneous linear equations (Saaty 2008) (see Equation 2).

$$\begin{array}{c} \\ E_1 \\ \vdots \\ E_n \end{array} \begin{array}{c} \begin{array}{cccc} E_1 & \cdots & & E_n \end{array} \\ \begin{bmatrix} \frac{w_1}{w_1} & \cdots & & \frac{w_1}{w_n} \\ \vdots & & & \vdots \\ \frac{w_n}{w_1} & \cdots & & \frac{w_n}{w_n} \end{bmatrix} \end{array} \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix} = n \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix}$$

$$\underline{\underline{A}} \cdot \underline{\omega} = n \cdot \underline{\omega}$$

Equation 2. Eigenvalue problem

The eigenvector (ω) thereby constitutes a numerical representation of the relative priority between the elements, similar to the mathematical construction in P4. Because the assessment of relative importance is based on the subjective judgment of the decision makers, the weights would correspond to the decision makers' interpretation of importance. The calculations are performed throughout the hierarchy, constructing a weight profile, in numerals, representing how important each element is to the organization. The aggregation process itself is thereby conducted as a weighted average of the normalized performance outcomes.

The hierarchical design applied in P4 was difficult for clinicians to interpret. Hence, a normal financial account design has been adopted, as it is considered a more intuitive representation (see Figure 17).

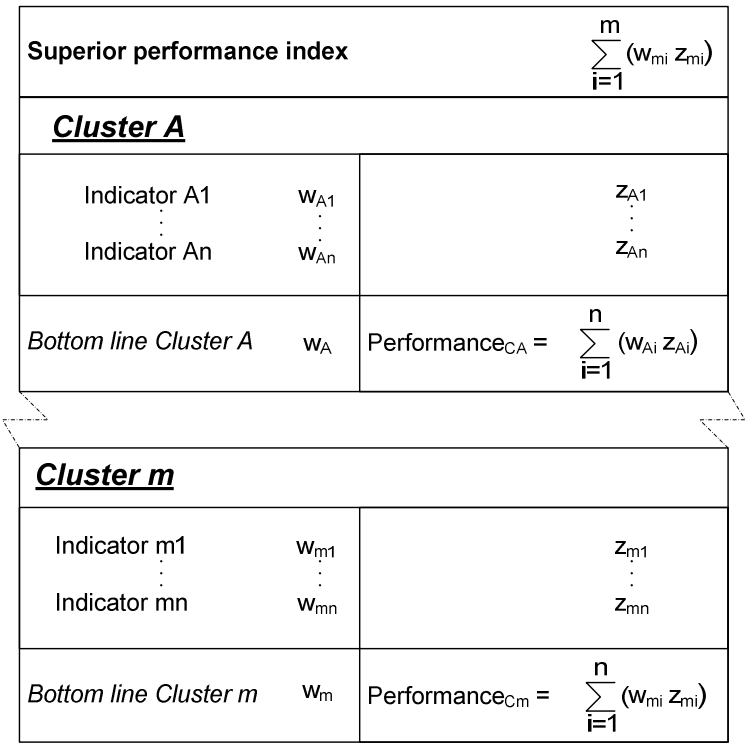

Figure 17. Structural outline of the "Performance Account" (from P5)

The mathematical construction is similar to P4, but the appearance is changed. The design is regarded more suitable, since most clinicians can recognize the layout of a financial account (see Figure 18). The normalized values of performance incorporated here resemble the economic posts in a financial account.

4.6.2 Evaluating "Quality 24/7" with the Performance Account

To prove applicability within strategic plan evaluation, Quality 24/7 is fitted into a hierarchy of indicators, and the performance outcomes are aggregated (see Figure 18). The hierarchy is designed using the indicators currently used at the radiology department along with indicators that were specifically requested by clinicians. Consequently, the Performance Account contains an all-round selection of indicators from different areas of the strategic plan.

Performance account RTG, 2007-2010		
Indicator	***Weight***	***Score***
Pro-/regress Quality 24/7		**0,11**
Patient Bottom-line	**0,68**	**0,12**
Patient satisfaction	***0,41***	***0,15***
Complaints (Satisfaction)	0,25	-0,01
Patients (tilgang)	0,75	0,20
Patient safety	***0,33***	***0,05***
Clinical Quality	0,41	0,01
- Complaints (Safety)	*0,54*	*-0,01*
- Image optimization	*0,2*	*0,00*
- Unintended occurrences	*0,25*	*0,05*
Equipment hygiene	0,33	0,13
Patient process (forløb)	***0,26***	***0,16***
Waiting list	0,47	0,45
Competences	0,43	0,00
Co-operation	0,1	-0,46
Employee bottom-line	**0,09**	**0,24**
Work environment	***0,4***	***0,00***
Psychological work environment	0,5	0,11
- Employee turnover	*0,34*	*-0,08*
- Sickness absence	*0,21*	*0,18*
- Work satisfaction	*0,45*	*0,22*
Physiological work environment	0,5	-0,11
Recruitment	***0,4***	***0,48***
Non-Danish speaking/ Danish speaking	0,33	1,03
Special employments/regular employments	0,33	-0,07
Students/Full-time employees	0,33	0,49
Radiation hygiene	***0,2***	***0,26***
Economy bottom-line	**0,23**	**0,04**
Effective work processes	***0,41***	***0,11***
Capacity utilization	0,67	0,03
- X-Ray	*0,2*	*-0,12*
- Ultrasonic	*0,2*	*-0,26*
- MR	*0,2*	*0,06*
- Biopsies	*0,2*	*0,05*
- CT	*0,2*	*0,40*
Non- attending patients	0,33	0,27
Production plans	***0,33***	***0,00***
Compliance with budget	***0,26***	***0,00***

Figure 18. Performance Account for Quality 24/7 (from P5)

The result of the evaluation of the strategic plan shows slight progress, primarily constituted by a strong patient dimension. The operations dimension has been through a rough time. This is constituted by two major events: 1) a general strike in the summer of 2008, 2) the troublesome implementation of a new RIS/PACS system during the summer of 2010. Both of events hit the work processes hard, resulting in poor average performance.

4.6.3 Contribution

The representation is assumed to provide easy and clear identification of strategic strengths and weaknesses. Hence, the Performance Account offers a substantial contribution to holistic interpretation of healthcare performance. Since the normalized performance outcomes portray progress and regression for all strategic areas, the identification of performance problems in relation to organizational importance is considerably easier. Thus, it is possible, in a holistic way, to assess the full extent of a strategic plan, consequently enabling structured evaluation of all aspects of performance in one single representation. Furthermore, the Performance Account constitutes a detailed foundation for constructing the succeeding strategic plan. Since organizational strengths and weaknesses are easily identified, future objectives can be decided according to the last account. By using the strategic evaluation of the past strategic plan to develop the future plan, more suitable strategic plans are assumed to be developed.

Collectively, the use of Performance Accounts is assumed to facilitate management according to organizational objectives. When considering the framework as a strategic evaluation tool, the paper concludes that there are reason to trust the framework in terms of the scientific advancement within the area of healthcare performance measurement and the progress in practical implementation. However an important discussion upon the implementation of the Performance Account is whether the normalized performance indices comprise a consistent informational basis. Since the validity of raw performance is not affected by the normalizing and aggregating procedure, reliability thus becomes a pivotal point in this discussion. Reliability lies in the performance account being a reflection of reality; that a negative result is an actual indication that something needs to be corrected. The challenge is that weight profiles are subjectively quantified. As a result the weighted aggregation becomes a reflection of the interviewees' priorities. The interpretation of performance is therefore influenced by the "perceived reality" of those who have constructed the hierarchy and designed the weight profile. Indeed the thoroughness of the prioritization according to strategic objectives becomes a key issue regarding reliability, as it determines the end result.

4.7 Summary

This chapter has recapitulated the five papers, aiming to answer the three research questions. RQ1 deals with the identification of measurement methods suitable for public healthcare settings. RQ2 focuses on designing an appropriate measurement structure for

tactical and operational levels respectively. And RQ3 tries to analyze to which extent the structure is applicable in benchmarking settings.

Paper 1 outlines a visual hierarchical structure where indicators are designed to the specific organizational context in which they are applied. The objective was to portray the completeness of the current performance measurement system, enabling decision makers to add or remove indicators to optimize the system. The hierarchical construction calls for strategic alignment, since indicator structures are deduced throughout the organization, from the strategic objectives to operational indicators. Building upon this, Paper 2 converts the hierarchical design into a formalized benchmarking framework, where performance differences among the four radiology sites are analyzed. The work proved that aggregated indicators are a valuable method to use as a guiding structure for internal benchmarking. The presented framework combined indicators from different organizational areas into one unified representation of performance. By benchmarking departmental performance from each site, the results show strengths and weaknesses in relation to different organizational areas. Paper 3 expands the use of the framework to an international benchmarking study between three Japanese and the four Danish sites. The aim was to test whether the framework could accentuate operational performance differences between healthcare organizations in different countries. The differences between Japanese and Danish healthcare resemble reliably the structural differences among the two healthcare sectors. This adds to the general aim of constructing performance measurement that enables healthcare managers to easily identify areas calling for corrective actions. The weaknesses revealed concerning the fairness of the benchmarking result were attempted redeemed in Paper 4 by adapting the framework for internal performance evaluation at the MRI unit at Haderslev Hospital. The distinctiveness of the proposed framework lies in the combination of normalization according to past performance and use of the AHP concept as a method for setting priorities. This allows for monitoring of performance progress and regression as a function of strategic importance. The framework has potential to include large amounts of information while targeting this information for use in operational decision support. What otherwise would have been a subjective assessment of strategic importance can now be quantified by representing performance as weighed, aggregated measures. In Paper 5, the aim was to propose a visually enhanced way of evaluating strategic plans, and the "Performance Account" was developed. The Performance Account constitutes a detailed and holistic foundation for constructing the next strategic plan. As the organization's strength and weaknesses are easily identified, future objectives can be determined on the basis of the account. By using the strategic evaluation of the past strategic plan to develop the future plan, it is assumed that more appropriate plans can be developed. Furthermore, the Performance Account provides a structured way of evaluating implemented initiatives. This final proposition incorporated all the experiences and contributions discovered during the research project. The Performance Account constitutes the final answer to the original motivation, which was to design a Management-By-Objectives model, suitable for operational performance evaluation in healthcare.

Chapter 5 - Discussion

The discussion picks up the most vital issues raised during the course of this research project. The first topic discussed is the research design's appropriateness, implicitly elaborating upon the suitability of the scientific approach. Then the recommendations are discussed in terms of overall advancement to the domain of healthcare Performance Measurement. Both scientific and operational benefits and validity are elaborated upon. Finally, the discussion broadens beyond the scope of the motivation to discuss the issue of good and bad decisions in order to shed light on how the proposals improves the use of Management-By-Objectives in healthcare.

5.1 Elaborations on the research design

As discussed in Chapter 2, research design shapes how investigations are conducted and how conclusions are drawn. Hence, the final recommendations of any research study need to be discussed in relation to the scientific methodology. This study was designed to develop a framework for evaluating strategic healthcare performance in an operational context in order to secure strategic alignment throughout the organization. This challenge was formulated in three research questions, which have guided the investigations. In the following, each of these research questions is discussed individually, and the obtained results are discussed in relation to the expected outcome. Potentials and limitations are discussed in regard to methodology and methods.

5.1.1 RQ1 – Elements adopted from industrial concepts

The first investigation was the identification of industrial performance measurement concepts that have potential in healthcare settings. Essentially, to answer the research question, an analysis of differentiating factors between healthcare and industrial organizations had to be conducted. As described in all the papers, one key factor is that hospitals operate in highly political environments where priorities shift rapidly (Furnham 2004;Griffith et al. 2006). The political agenda is highly influenced by medical, technological, and organizational developments, and this results in an influx of urgent initiatives (Hauck & Street 2007). With every new urgency, the demand for evidence is concurrent, in order to document the effect (Drummond et al. 2006;Moullin & Soady 2008;Stronks & Mackenbach 2006). This causes pressure on performance measurement systems, since they must be able to adapt to a high degree of flexibility, which exceeds the need in industrial organizations. Flexibility in measurement system design is therefore a key to successful implementation in healthcare. Furthermore, this complexity extends even further when moving into benchmarking, because differences are enhanced dramatically when going from internal to international benchmarking.

This recognition has continued to shape the final conclusions throughout the study. Early on, the CIMOSA approach inspired the idea of illustrating performance measurement systems with a visual representation. CIMOSA led to the acknowledgement that a hierarchical construction was needed to portray performance specified at each organizational level. This meant that the hierarchical construction would provide the fundamental basis for the rest of the study. As hierarchies became central, workshops proved to be a suitable forum for discussing indicator structure. The hierarchical construction was intensively discussed. Practitioners came to realize how this construction could enable them to focus performance measurement on their own department. As researcher, the recognition of visual representation became evident, since the discussions of measurement systems became more elaborate through visualization. The early notion of representation was immature, however, but was continuously modified to comply with practitioners' visual perceptions.

The most challenging issue concerning the hierarchical construction was the vast selection of performance measures that needed to be implemented. In several workshops, the integration of indicators into a hierarchical construction in itself could not be beneficial, because this would not reduce the informational burden. Consequently, the OEE methodology became an obvious choice for the mathematical aggregation of performance outcomes. By continuously aggregating performance through the hierarchies, fewer aggregated key measures were obtained. Theoretically speaking, this approach means that an endless number of measures can be aggregated and thereby implemented within a measurement framework. This adds to the need for very flexible and customized measurement systems. The aggregation provides estimates of performance as a common denominator of all lower-level inputs. Indeed, to aggregate performance outcomes, a common unit of all incorporated inputs is needed. The normalization method applied in P2 proved to be weak; normalizing with averages as the sole basis did not provide the necessary robustness. Derived from this, the z-score was proven valuable in the benchmarking study between Japan and Denmark. The z-score has frequently been applied within Six Sigma methodology (Woodward 2006), where stability is in focus. This normalization method helped to distinguish performance in a more elaborate way, since it calls for consistent performance over high variation. This particular strength was accepted well by the clinical personnel, who throughout their clinical education had been taught to avoid variation in delivered quality. This is in alignment with healthcare quality publications, which promote the preservation of evenly distributed quality in health services (Basu, Howell, & Gopinath 2010;Woodward 2006). In the attempt to answer RQ1, the results point to the OEE, the z-score and aggregation techniques as having potential for evaluating healthcare performance. Each of these has distinct strengths that aid the difficult evaluation in such a dynamic and inhomogeneous environment as healthcare. Indeed, it is important to recognize that some of the concepts that have been applied are modified to fit healthcare settings. The OEE, which in its original form includes "Availability, Performance, and Quality", is customized to fit the dimensions in the strategic plan for the case hospital. The z-score is commonly used as a way of evaluating production stability, but here it is applied only as a normalization method. Therefore, several industrial concepts prove to be suitable for healthcare settings, but they often have to be modified to fit the specifics of healthcare organizations.

As the study progressed, the use of interviews and workshops proved appropriate, as continuous discussions with clinicians formed the recommendations. As researchers and clinicians became more and more familiar with one another, the discussions became more and more valuable, because the scientific design required that researchers receive input from clinicians. The growing familiarity between them therefore proved priceless. When multiple cases are used, it is assumed that the chance for the input to be of similar value is reduced. This reasoning is supported by the genuine anxiety about measurement that exists among healthcare practitioners (Loeb 2004); their willingness to participate is assumed to be limited if they feel alien to the researchers or their methods. As the trust between clinicians and researchers increased, their motivation for influencing the

proposals increased. As it became clearer that the aim of the study was to facilitate changing the current way of measuring, the input increased in value. Thus, the use of a single case also proved to be a very valuable methodology in the early stages of the study. There is indeed no scientific justification for stating whether or not the radiology department at Southern Jutland Hospital is the best possible case, but dealing with a case that resembles production (as opposed e.g. to admission wards, geriatric departments etc.) increased the probability that employees would recognize and acknowledge the use of measurements. Since the radiology department constantly has to measure production quantum, throughput times, equipment hygiene etc., the step into performance measurement is not significant. Likewise, orthopedic, dialysis and cardiology departments would also be useful, as they share similar experience with measurement. The choice of radiology is therefore regarded to provide a appropriate case for developing measurement models and afterwards generalizing them to other departments.

5.1.2 RQ2 – Construction suited for healthcare

In accordance with the overall motivation and RQ2, the study intended to clarify how the different industrial concepts should be combined in order to be applied in healthcare settings. This was to be done subsequently to RQ1, after suitable elements were identified. P5 argues that weighted aggregation in hierarchies of normalized performance outcomes provides a detailed and valid performance picture. As presented in P4, the weight profile focuses on the representation of performance, specified according to area of application. This signifies that the framework can be applied at any organizational level, as priorities can be adapted to the exact settings. In this way, it is possible to maintain rigid priorities throughout an organization, and make sure that organizational objectives are prioritized according to the strategic plans. This design is argued to elevate the usage of performance evaluation, since the construction of the measurement system can be configured to the particular settings in which it is applied.

The instantly recognizable challenge is to conduct this prioritization, because it establishes the direction of the organization. As discussed, management theory assumes the prioritization process to be an absolute necessity, if an organization is to achieve decisional alignment.

> Each functional area should develop and utilize a set of performance criteria consistent with its particular operating characteristics and strategic objectives.
>
> (Chen 2008)

Although a difficult task, the process of prioritizing objectives is simplified by the method of pairwise comparison. Instead of balancing many incomparable objectives, comparison of two alternatives makes this process much easier. If the clustering of indicators is performed thoroughly, the comparison can be performed by indicators within the same area of reference, thus aiding the prioritization process. By using input from several key persons (nurses, doctors, managers etc.), an average estimate of mutual

importance was achieved. Moreover, the process clarified the individual differences among management members, naturally revealing an inhomogeneous employee group with different perceptions of organizational importance.

The notion of the Performance Account that was introduced seems to make the visual representation more intuitive to practitioners. For most people, the financial account is an intuitive tool, not necessarily because it is particularly logical, but it has been used for so many years that it has become common sense. Accounts provide practitioners with an easier understanding of how the aggregation was performed. The answer to RQ2 is therefore that by adding weight profiles, which establish the organizational priorities, and aggregating them in indicator hierarchies, performance outcomes are portrayed as a function of strategic objectives. Furthermore, by representing the outcome in Performance Accounts, an intuitive representation is achieved. The AHP method is applied in this work, but other multi-criterion analysis methods might be as suitable. However, the pairwise comparison included in the AHP method has merits that exceed whatever else was investigated during the study.

5.1.3 RQ3 – Design of benchmarking initiatives

The motivation behind RQ3 was to identify to which extent it is possible to employ an internally adapted performance measurement system in an external benchmarking context. This implies evaluating the limitations of applying a vertical measurement structure to a horizontal benchmarking setting. The investigations supporting this were performed in connection with two benchmarking studies – first, the internal benchmarking presented in P2, and second, the international benchmark in P3. Both of these provided valuable insight into the benchmarking potential. It should be noted however that the frameworks applied in the two studies were slightly different, but the insight gained from each of the cases are used as the basis for discussion.

For internal benchmarking (P2), the framework proved to be quite useful, as the result resembled the perceptions of the department's managerial team. Success was achieved to a large extent because the work content, employee combination, and management are roughly identical at the four sites. The result is therefore not surprising; most organizations are able to compare performance results between departments or production sites of similar character. However, the specifics of the cases make the study interesting, even though the results were not that unexpected. Since the four sites were just recently merged, their organizational cultures are still divergent. Because they differ, it could be anticipated that the benchmarking construction would encounter difficulties, and it did. The validation of the study was performed as a blind test, which was biased in that managers knew about the different organizational roles of the sites. It is therefore questionable whether the identification of the sites was made by interpreting performance levels, or just by simple recognizing the differences in organizational roles. At that time, the framework did not incorporate weight allocation; hence, performance indexes were a bit misrepresented. The same applied in the Japan vs. Denmark benchmark (P3), where

identification of performance differences was mainly made by 'unfair' evaluation of performance levels. Distinguishing the performance between the Danish and Japanese hospitals was influenced by the restricted data collection at the Japanese hospitals. Contrary to Danish healthcare, Japanese hospitals operate as closed units, which means that performance data are not public or shared among hospitals unless the hospital chooses to do so. This limits the study, since data collection was constrained to the data available at all seven sites; thus, the Japanese hospitals had a slight advantage, since they controlled the data collection. This does not necessarily mean that the result is useless, but precautions must be taken when elaborating upon the results.

As this became obvious, the weight profiles were introduced as means of compensating for these 'unfair' benchmarks. Due to time restrictions, benchmarking with weight profiles was never completed. In P5, the test was performed solely on the radiology department, and unfortunately not also on an additional case. The paper shows how the weight distribution can distinguish between indicators in relation to strategic importance. The question is then, would the introduction of weight profiles aid in designing more 'fair' benchmarks? As described in section 1.3.4, fairness in healthcare benchmarking is seldom attained, and the reason is often disagreement about importance. Even two similar radiology sites are unlikely to treasure the same values, which makes benchmarking difficult. Agreement about what is most important may never be reached, but a hierarchy of indicators can be designed in consensus among department heads. Therefore, to answer RQ3, the study reveals that adaptation of internal measurement systems to an external benchmarking context can only be achieved to a limited extent. The study reveals that it is possible to decide upon a hierarchy, but fairness would still be difficult to achieve due to disagreement about the indicators' importance. The investigations in this research study have not shown any indications that the frameworks could resolve the healthcare benchmarking meta-problem, although the structure of the framework is presumed suited for benchmarking.

5.2 Has the work generated scientific progress?

To make the claim that the outcome of a research study has been a scientific success in terms of progress, the validity and reliability of the proposals must be trustworthy. Therefore, this section discusses issues of validity and reliability, along with discussions about practical applicability. Elaborations upon these fundamentals in science are used to evaluate the scientific success of the research study, which ultimately determines whether the work has generated a justifiable contribution to the domain. As the study has been highly influenced by real-life scenarios, the practical benefit is discussed in connection with the scientific gains of the project.

5.2.1 Validity and reliability of proposals

Scientific assurance of validity rests on continued testing of the proposals, and its practical applicability rests on how the framework functions in a real-life setting. (Cook & Campbell 1979) define validity as the "*best available approximation to the truth or*

falsity of a given inference, proposition or conclusion". Validity is closely associated with the notion of reliability, which relates to the consistency of the investigations. Reliability is tied to the scientific method, and validity to the interpretation of the collected data. Theoretically, it is possible to present a valid claim on the basis of a method that lacks methodological rigor, but in order to make a sound contribution to the body of knowledge, the reliability of the proposed claims must be irrefutable.

This study has conducted case study research with a critical realist viewpoint. A number of initiatives were carried out to improve the validity of the claims. As described in section 2.5.2, the strength of claims in critical realism is tied to Weick's notion of "clarity of concepts". According to this notion, validity is not a truth criterion; instead, the research performed iterative tests, until "*there are no reason not to trust the proposal*". This is in concordance with Yin (1994), who states that validity in case study methodology is achieved through pattern matching and replication logic, thereby excluding the possibility of false interpretation (see Table 11).

Test	Case study tactic	Phase of research
Construct validity	Use multiple sources of evidence	Data collection
	Establish chain of evidence	Data collection
	Have key informants review draft case study report	Composition
Internal validity	Do pattern matching or explanation building or time-series analysis	Data analysis
External validity	Use replication logic in multiple case studies	Research design
Reliability	Use case study protocol	Data collection
	Develop case study database	Data collection

Table 11. Reliability and validity in case study research (adopted from (Yin 1994))

Hence, this study has been conducted using a series of different tests, thereby adopting multiple sources of evidence, to enable pattern matching of data. Re-testing proposals is one of the most valuable methods to increase the reliability of the claims and lead to sounder scientific statements (Voss, Tsikriktsis, & Frohlich 2002). Table 12 describes which initiatives have been performed in the five papers in order to enhance the scientific validity of the proposals.

Paper	No. of cases applied		Focus	Validity
	Internal	*External*		
P1	1		Initial construct, applied on unknown case. Low detail level.	Multiple sources of evidence
P2	4		Adjusting model on known cases. Low detail level.	Internal
P3	4	3	Testing limitation of known model on known and unknown cases. Low detail level.	External, Replication logic
P4	1		Adjusting model on known case. High detail level.	Internal
P5	1		Testing limitation of known model on known case. High detail level.	Internal Time-series analysis

Table 12. Validity testing during the research study

As the proposals are continuously developed, the testing is modified to adopt multiple sources of evidence. By having as few unknown factors as possible, the testing becomes a solid way to analyze the validity of the proposal. As the model is continuously adjusted, the applicability of the incorporated elements is tested. Furthermore, the detail level in the papers differs – two papers deal with specific performance, and three with overall performance. This allows analysis of the proposals' stability in relation to organizational applicability. As the proposals proved to be stronger and stronger, the scientific reliability is assumed to increase accordingly. Indeed, the question. *'are there reasons to believe that it could be otherwise*', is very contextual. Scientists can interpret the focus on increasing validity differently. This diversity is rooted in divergence in interpretation of the validity threats – that is, how severely a given event is assumed to affect the validity of the investigation.

In general, there are some persistent validity threats that haunt organizational studies. Organizational changes are obvious threats. In longitudinal studies, changes in organizations are expected, which the literature refers to as the "historical validity thread" (Cook & Campbell 1979). This signifies that the subject (the hospital) changes from test to test, which theoretically detracts from the validity. In this study, change in the case is recognized, although it is not considered to have significantly detracted from the validity of the proposals. During all years, the key employees have been the same and also the organization of work among the four radiological sites. Even though the framework itself has changed in form, the incorporated indicators have not changed distinctively from test to test. Hence, the historical thread has not been assessed to be a significant damaging factor, since the radiology department is considered to have been rather stabile during the three-year study. Another common threatening factor is the "instrumentation validity thread"(Gardner & Wright 2009). If the subject (in this case, organizational performance) is measured differently from test to test, then there is no way of comparing the results between the two tests, and thus the two tests cannot validate each other. Instrumentation changes are regarded by some authors as a serious limitation, especially in multiple case

studies. In this study, the framework has changed structure consecutively, which is regarded as an impairing factor to the validity. Since the framework is not entirely the same over the course of investigations, this naturally harms the reliability of the contribution. Changes in instrumentation, along with minor organizational change, slightly weaken the soundness of the claim. But the effects of the validity threats are still regarded as minor, and as a result, the proposals are regarded valid with respect to these weaknesses.

This bold statement is supported by the feedback from clinicians and the review from the scientific community. Clinicians accepted the proposals and recognize their use in practice. This signifies that the proposals out-perform current performance evaluation at the hospital; therefore, local real-life advancement is achieved. The scientific community, represented by journals and scientific societies, acknowledges the ideas by accepting scientific papers and conference proceedings. Since the two primary stakeholders in this study recognize the final proposals, they are regarded valid in the context in which they have been applied. Indeed, testing different cases can reveal validity issues that have not emerged in the study. The claim that the proposals are valid is therefore made with regard to the methodological constraints outlined in section 2.5. Obviously, there is no justification for claiming that the framework would be successful outside the contextual premise stated in the research design.

5.2.2 Applicability of proposals in real life

The obvious question rising in a validity discussion is the issue of real-life applicability. As argued, the Performance Account is presumed scientifically valid, but this does not necessarily imply that it would be a practical success. In general, new technology, new managerial techniques and new clinical methods find their way to implementation in two ways: 1) external requirements of national or regional legislation; or 2) internal requests by employees.

Concurrent with the advancement of measurement systems in healthcare, there is reason to believe that more holistic approaches will be formed in coming years. It is evident that little coherence exists among national monitoring initiatives, which has given rise to the discussion of a higher degree of cooperation. This discussion is primarily based on the desire to benchmark healthcare services and thereby identify state-of-the-art and implement it throughout the sector. The findings in this study do not resolve the primary challenge in benchmarking initiatives. Therefore, political demand to introduce the proposals as they stand today is unlikely, but the proposals may inspire further investigations or future design of national measurement initiatives.

Perhaps the proposals can gain acceptance from within the organizations. Since the managerial team at the radiology department became an integrated part of this study, their wishes and desires is the core of the final proposals. Therefore, it is likely that the proposals reflect to some extent the way the decision makers want their measurement

system to be designed. If this presumption is viable, it is likely that some of the ideas will inspire an internal request for better performance measurement, and adaptations or elements of the framework may thus find their way into management at the radiology department. And if success is achieved in radiology, the concepts could possibly spread throughout the rest of Southern Jutland Hospital. This is indeed a very time-consuming process, as implementation would have to be founded on a single department. As history has shown, bottom-up processes are lengthy and are likely to change in structure. However, conversations with the managerial team suggested that the proposals would be used to assess the strategic plan 2007-2010, with the aim of showing the hospital board the progress that had been made.

5.3 Good vs. bad decisions

This research study has tried to provide the best possible informational basis for healthcare decision makers, but no investigations have been made into the quality of the decisions. The premise for this limitation has been that valid and reliable information is a precondition for high quality decisions, and this limits the initial investigation to securing reliable decision-support information.

In theory, performance measurement systems help decision makers to identify organizational areas of strengths and weaknesses, which serves to support decisions about future initiatives (Rundall et al. 2007). However, it is important to note that the outcome of any performance measurement system indicates what happened, not why it happened or what to do about it. This implicitly signifies that an incorrect decision can be made on "perfect informational ground", in which case the decision error would be assigned to the individual.

> "No management philosophy can resolve management incompetence."
> (Gupta & Snyder 2009)

So wrong decisions can be bound to the system (incorrect information), but also to the individual (management incompetence). As there is no scientific justification to evaluate individuals using the results of this study, this discussion is solely concerned with the measurement system.

But how do we verify that the Performance Accounts provide the right information? First, we evaluate the data that is incorporated. The raw performance data were modified by the normalizing procedure (z-score). This does not alter the information as such, but the interpretation is changed. When aggregating the normalized performance outcomes, the interpretation is changed again, although the empiric data are still equivalent to the initial data. Consequently, to discuss whether the information is appropriate, the notion of "perceived reality" is important. Since there are no absolute values for good or bad performance (Kollberg, Dahlgaard, & Brehmer 2007), there is no way that we can check the correctness of the aggregated performance outcome itself. Therefore, the reasoning has

to rely on the perceived reality of the decision maker. If the decision makers are able to interpret and use the information as desired, then the outcome must be regarded as appropriate information. The justification for arguing that the framework presents appropriate information is that it is based on the conversations and workshops with decision makers at the hospital. As the weight profiles are subjectively quantified, the weighted aggregation is performed as a reflection of the interviewees' perception of mutual importance among indicators. The interpretation of performance is therefore influenced by the "perceived reality" of those who have constructed the hierarchy and designed the weighted profile. As the output portrays performance in relation to the organizational objectives, poor performance in critical areas is highlighted and easily identified. This constitutes the primary basis for securing strategic alignment throughout the organization, from strategic plans to daily management of operations. Furthermore, the value (z-score) in itself is not as important as the identification. If a performance problem is identified, and the measurement system detects it, deeper investigations would be required to arrive at a corrective action. Hence, the most valuable task of the Performance Account is to identify the problematic areas. The attention of the decision maker is focused on the area of weak performance that indicates where action is needed. This supports the statement that the framework aims to assist decision making in a positive manner. Indications suggest that better decision support information is provided; although there is no scientific validation to back the claim that the proposals would provide better organizational decisions, there are reasons to believe that this is so.

5.4 Summary

When evaluating the success of a research study, the obtained results have to be compared to the objectives of the study, implicitly limiting the discussion to the chosen methodology and the methods applied to achieve the objectives. In this study, three research questions guided the course of the investigations. Initially, the study focused on identifying and adapting elements from industrial performance measurement to healthcare performance evaluation. The results showed that aggregation of normalized performance outcomes, enabled a more holistic evaluation of performance. Next, the assembly of these elements was investigated. Weighted aggregation within a hierarchical indicator structure was considered a solid approach that made organizational objectives apparent to decision makers, enabling identification of areas suited for corrective actions to bring them in alignment with organizational objectives. Finally, the study applied the vertical performance measurement system to a horizontal benchmarking situation. The investigations did not solve the benchmarking meta-challenge of unfair benchmarking, since the prioritization of importance continues to be a key problem. As internal benchmarking, however, the design of the Performance Account was found suitable for evaluating performance differences among departments of similar character.

Throughout the study, the scientific approach applied was aimed at enhancing the validity and reliability of the proposals. To heighten reliability of the proposals, the work aimed at collecting evidence from multiple sources to test the potential of the proposals as

many times as possible. The validity of the proposals was tested by continuously testing the proposals in different settings. Two benchmarking studies, along with two vertical measurement evaluations were performed. The papers demonstrate that the framework proved valuable in both contexts, but it was strongest in regard to internal measurement. It is argued that the framework enables better decision making. This is a speculative statement and not a scientifically investigated claim. The statement is based on the construction of the hierarchy, which enables holistic identification of performance problems in relation to organizational objectives. The claim is that since identification is made easier, corrective actions are more appropriate than previously.

Chapter 6 - Conclusion

This concluding chapter recapitulates the recommendations presented in the dissertation. The intention is to summarize and emphasize the overall advancement – both scientific and practical – achieved during the research by relating the scientific outcome to the initial motivation.

The use of performance evaluation is on the rise in modern healthcare, where the political call for accountability is a driving force behind this development. It is nevertheless paradoxical that although the medical community has acknowledged the relevance and importance of performance evaluation, numerous measurement initiatives fail to be an influencing factor for operational decision making. Hence, this study has shown that holistic and comprehendible performance measurement can be achieved by designing indicator hierarchies and subsequently aggregating normalized performance outcomes according to organizational priorities. By aggregating performance from different stakeholder perspectives into one single "Performance Account", the application of performance measures becomes more operable than in current practice. Representing departmental performance as a function of stakeholder perspectives in relation to organizational importance, significantly improves traceability of poor performance as well as the transparency of the measurement process. The design of the Performance Account constitutes a novel approach to the domain of healthcare performance measurement. Combining several scientifically validated methods into one single representation alters the traditional way of interpreting performance. The distinctiveness of the Performance Account lies in the combination of normalization according to past performance and the use of mutual weighting as a method for prioritization. What otherwise would have been a subjective assessment of strategic importance is now quantified by representing performance as weighted, aggregated measures. Consequently, the Performance Account helps decision makers in evaluating the need for corrective actions, ideally ensuring that only organizationally aligned initiatives are commenced. The framework has the potential to include vast amounts of performance information, while targeting this information as decision support in operational decision making. This enables healthcare decision makers to manage their area of responsibility on the basis of the organizational objectives of the organization.

As healthcare moves into an era of competition similar to that experienced by industrial organizations, benchmarking becomes an obvious technique to drive continuous improvements. Since the Performance Account is designed as an internal vertical measurement system, tests on its potential as a horizontal benchmarking model have been investigated. External benchmarking must meet several challenges, such as cultural differences, legislation, organizational structure etc., which the Performance Account does not solve directly. The challenge lies in the alignment of priorities, which in a benchmarking situation is difficult to achieve. However, the design of the Performance Account is suitable for internal benchmarking when evaluating performance differences among departments of similar character. Therefore, the design is considered valuable for internal, vertical, in-house decision support, but with limited potential for external, horizontal benchmarking settings.

The Performance Account is presumed to be able to aid decision making at all levels of public healthcare organizations, since the design is not tied to a specific organizational level or specialty, but its applicability stretches beyond the case that has constituted the

primary empirical foundation. As the work has been conducted as a case study, the generalizing potential of the conclusions is scientifically limited to the cases in relation to which the framework was developed. However, several methodological initiatives have been used to expand the scope of the conclusions, including external benchmarking and changing the data foundation in the testing phases. Furthermore, the intense use of workshops with participation of the practitioners constitutes a reliable empirical basis. Thus, since the participants' input to the development process comprises more than just experiences tied to the single case, the validity and reliability of the proposals are regarded to be high with respect to the scientific boundaries of the research. Since the scientific robustness of the proposals is ensured, there are sound reasons to claim that the proposals constitute a scientific advancement within the domain of healthcare performance measurement.

The research study shows how scientifically developed models can be of great benefit in solving several of the practical problems evident in healthcare. The need for tools to aid the decision-making processes in healthcare can be met, if the increasing quantity of measurement initiatives can become an integrated informational basis for operational decision making. This research study has addressed the issue of utilizing vast amounts of performance information, but several closely related aspects will have to be scientifically dealt with in order to facilitate healthcare decision making. On the basis of the research study, it must be concluded that the work contributes valuable input to the continuous advancement of healthcare performance measurement, but it also accentuates the evident need for more research within this domain.

Chapter 7 - Future research

The study has revealed areas that are alleged to be of great scientific value and would be appropriate research topics to follow up on this work. The identification of indicator correlation could provide valuable scientific insight into the usage of individual indicators. Furthermore, the topic of identifying an 'optimal' quantity and composition of performance indicators is an obvious domain requiring deeper study. In addition, investigation of indicator representation could play a key role in enhancing the quality of decisions.

7.1 Identifying quantitative indicator correlations

The hierarchical structure presented in this thesis implicitly presumes that indicators are snapshots of a given aspect of the organization without correlation to other aspects. Because indicators are distributed in dimensions and clusters, the aggregation procedure imply no mathematical correlation with other parts of the framework. Indeed, this is an acknowledged misrepresentation; consensus exists among practitioners and scientists concerning the mutual relationships among different organizational aspects. Kaplan and Norton elaborate on this matter in their famous Balance Scorecard paper, where the distinction in some cases can be difficult:

> Ideally, companies should specify how improvements in quality, cycle time, quoted lead times, delivery, and new product introduction will lead to higher market share, operating margins, and asset turnover or to reduced operating expenses. The challenge is to learn how to make such explicit linkage between operations and finance.
>
> (Kaplan & Norton 1992)

For decades, scientists have been struggling with the issue of identifying and quantifying the correlations between different areas of organizations and provided insight into both organizational, cultural and structural linkages between the indicators. Identifying the root cause of performance problems involves the ability to distinguish between input and result, as well as an assessment of the strength of the correlation. In the literature, the terms of leading and lagging indicators precisely portray this issue – lagging indicators are results of changes, and leading indicators predict future changes. The theoretical challenge to overcome this is to quantify indicator correlations. By determining the mutual relationship, the practical usage of performance measurement systems is dramatically enhanced, as root causes are more easily detected.

In the quest of mapping some of the most obvious links, a literature study was conducted. The motivation was to construct a spider web of correlations that serve to accentuate the causal relationship between employee and patient satisfaction. Both indicators are recognized to be very context-dependent, which is why they have both attracted considerable attention from academics in recent years. By specifying leading indicators as well as lagging indicators, the understanding of satisfaction surveys is assumed to be enhanced considerably. The study, which is still in the working process, resulted in a correlation map portraying the mutual dependence of indicators (see Figure 19).

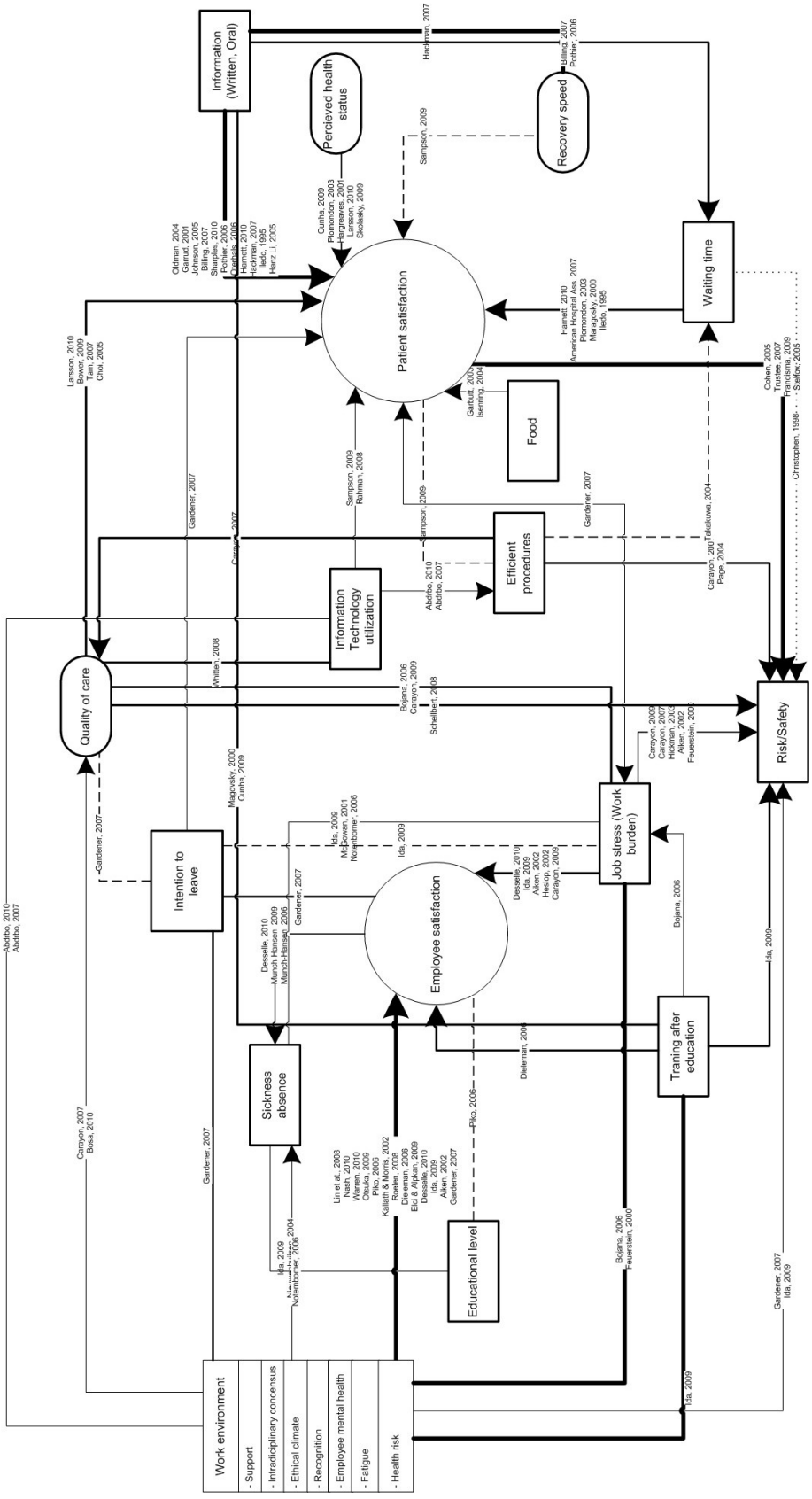

Figure 19. Correlation map

Subsequently, the quantification of these linkages is required to determine the strength of the links. This can be used to test whether the correlations actually apply in practice at hospitals or are only described in the literature. Relating employee satisfaction with the possibility for further education suggests that the link actually applies at Southern Jutland Hospital (see Figure 20).

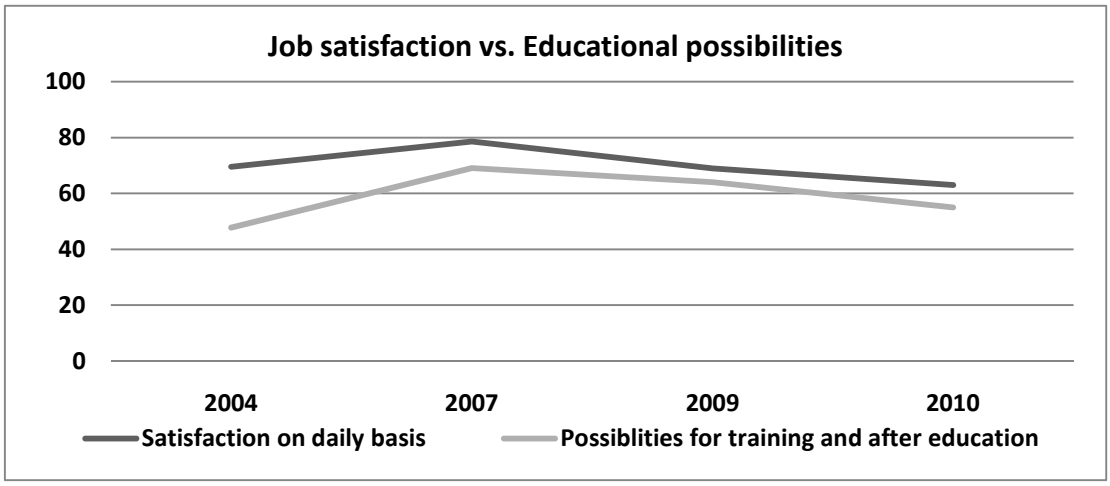

Figure 20. Employee satisfaction vs. Educational possibilities

The determination of these relations is indeed a very difficult task that needs to be investigated much more thoroughly than has been done in this study so far. The identification and then quantification of these causal relations are assumed to provide insight into which context indicators could be applied. Therefore, is it recommended that this subject be pursued more intensively, as it is likely to reveal some undiscovered treasures in the domain of healthcare performance measurement. Such a study should be a high-intensive quantitative study in which a vast amount of performance data would have to be analyzed for correlations. A qualitative assessment of the practical logic is also necessary, since these correlations tend to be intensely context-dependent.

7.2 Identifying an optimal set of measures

After identifying correlations among indicators, it is obvious to focus attention on designing an optimal set of indicators. If there is a strong inter-dependence between two indicators, it may be possible to exclude one of them in the internal measurement system. The quantity and the context in which performance indicators are applied in this thesis were determined on the basis of the perceptions of the workshop participants and researchers. This is because this work has concentrated on proposing a structure for measurement, not determining the specific incorporated indicators. Hence, the selection has taken its point of departure in already applied indicators. This is considered to be a limitation in terms of designing an 'optimal' healthcare performance measurement system. Indeed, if we acknowledge that the present selection is not optimal, then further investigations must be conducted to elevate the scientific and practical value. But identifying an 'optimal' set of measures by interviewing healthcare personnel would be

scientifically unreasonable. Indeed, further investigation into this topic would be of great scientific interest. If a scientific investigation were able to determine which indicators, in which context, would provide most value to decision makers at different organizational levels, this would be the key to unraveling how departments, sections or even whole hospitals should be evaluated. Furthermore, this work assumes that the framework can contain unlimited numbers of indicators and still provide an overview. But is this a legitimate claim? And why include many indicators if a few are enough? To answer these questions, it is necessary to conduct in-depth analysis of the factors decision makers draw upon as their informative basis when making decisions. Implicitly, the context of the decisions plays a key role, as this also may contain information that would clarify whether the quantity of indicators is important. Additionally, there may even be differences depending on the organizational level the decision maker refers to.

7.3 Identifying the most appropriate representation

It would also be interesting to analyze which representation of performance would be most valuable. In this thesis, the 'Standard Score' has been incorporated as normalization method. In the literature, numerous references describe the benefit of this particular method. It is considered appropriate to some extent because practitioners have been integrated into the development process and have thus been gradually introduced and trained in the application and interpretation of the output. Indeed, this may not be the most optimal way to present healthcare performance to healthcare practitioners; other practitioners might interpret the charts differently. Therefore, studies of visual perception could be valuable in terms of future presentation of performance measurement. For example, Figure 21 presents the z-scores of X-ray examinations, which provide insight into the stability of production, and the mean score tells about average progress/regression.

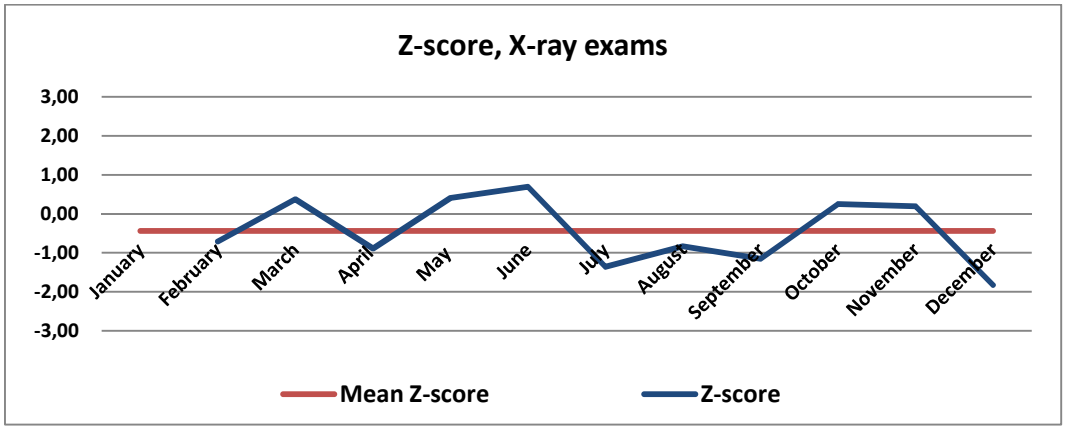

Figure 21. Z-score example

As an alternative, a more traditional representation is presented in Figure 22, where the sheer number of examinations and the average are shown.

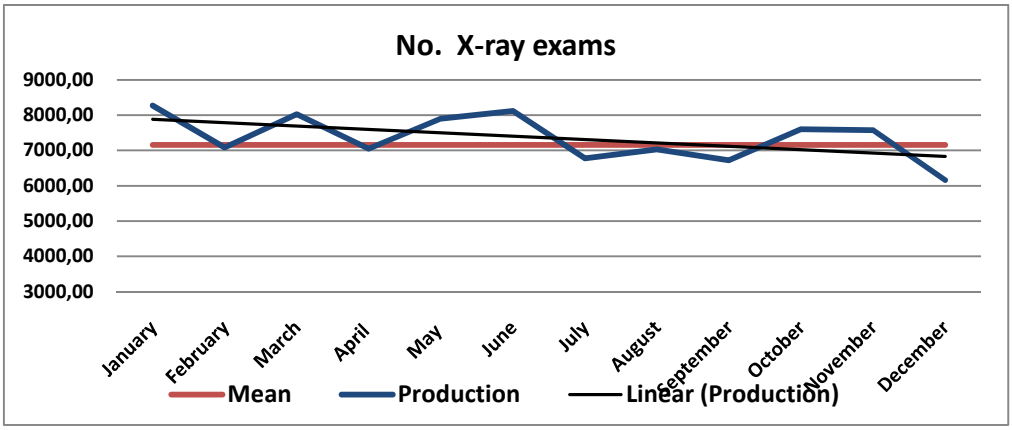

Figure 22. Number of X-ray examinations example

Which of these provide most insight for a decision maker? This may depend very much on the eyes of the beholder. Contributing to the complexity are the countless ways possible to portray performance. But it can be argued that there are some generic features concerning the representation that can affect the interpretation of performance. Indeed, a hypothesis for examining the most appropriate representation could be something like: It is possible to enhance the quality of decision making in healthcare by altering the representation of performance?

7.4 Summary

In the course of this study, three issues have been raised that appear to have great scientific interest, but which have not been investigated. The quantification of the causal relationship among indicators is presumed to be of huge scientific interest. Vast numbers of publications addressing mutual dependence between indicators have been found. Some of these have been mapped in a spider web of correlations to portray this extremely complex topic. Indeed, a deeper analysis into the specifics of the correlations would elevate the use of performance measurement systems. As this chapter outlines briefly, the process of quantifying correlations is very complicated, since the strength of the correlations can shift from one hospital to another, and even from one department to another in the same hospital. Although scientifically and methodically challenging, the quantification, or method for quantification, of causal relationships among performance indicators are seen to be an essential step towards better performance measurement. A second topic of interest is the selection of the 'right' indicators; thus, an investigation of this topic would further enhance the use of performance measurement systems. As a third issue for further analysis into the domain of healthcare performance measurement, this chapter proposes analyzing individual indicators in terms of applicability in decision making. Indicators provide specific information, which can be formulated into a structure that shows where and when a given indicator would provide most value for decision makers. Along with an analysis of appropriate representation, this would enable deeper understanding of why the construction of performance measurement systems constitutes a principal factor affecting the quality of the decisions made and ultimately the success of organizations.

Chapter 8 - Literature

All the references used in this thesis are accentuated in this chapter. The references are divided in three groups, Books, Articles and WebPages. The references are in alphabetic order, using the family name of the first author of the publication. Harvard reference style are applied, where the detail sequence is Author, Year, Publication name, Journal and finally details on volume number.

Ahmad, M., Dhafr, N., Benson, R., & Burgess, B. 2005, "Model for establishing theoretical targets at the shop floor level in specialty chemicals manufacturing organizations", *Robotics and Computer Integrated Manufacturing*, vol. 21, no. 4-5, pp. 391-400.

Arah, O. A., Westert, G. P., Hurst, J., & Klazinga, N. S. 2006, "A conceptual framework for the OECD Health Care Quality Indicators Project", *International Journal for Quality in Health Care*, vol. 18, no. 1-supplement, pp. 5-13.

Baker, S. L., Beitsch, L., & Landrum, L. B. 2007, "The Role of Performance Management and Quality Improvement in a National Voluntary Public Health Accreditation System", *Journal of Public Health Management & Practice*, vol. 13, no. 4, p. 427.

Barros, P. P. 2003, "Random Output and Hospital Performance", *Health Care Management Science*, vol. 6, no. 4, pp. 219-227.

Basu, A., Howell, R., & Gopinath, D. 2010, "Clinical performance indicators: intolerance for variety?", *International Journal of Health Care Quality Assurance*, vol. 23, no. 4, pp. 436-449.

Bititci, U., Turner, T., & Begemann, C. 2000, "The role of Performance measurement in continuous improvement", *International Journal of Operations & Production Management*, vol. 20, no. 6, pp. 692-704.

Bourne, M., Franco-Santos, M., Pavlov, A., Martinez, V., & Lucianetti, L. "Performance management practices and their impact on organisational performance: implications for HR and performance measurement research", in *Euroma 2008*, Springer.

Bovier, P. A. & Perneger, T. V. 2003, "Predictors of work satisfaction among physicians", *The European Journal of Public Health*, vol. 13, no. 4, pp. 299-305.

Boyd, L. & Gupta, M. 2004, "Constraints management: What is the theory?", *International Journal of Operations & Production Management*, vol. 24, no. 4, pp. 350-371.

Brumback, G. B. 2003, "Blending "we/me" in performance management", *Team Performance Management*, vol. 9, no. 7-8, pp. 167-173.

Buetow, S. 2008, "Pay-for-performance in New Zealand primary health care", *Journal of Health Organisation and Management*, vol. 22, no. 1, pp. 36-47.

Byrne, M. 2006, "Implementing Performance Management in the Irish Health Sector", *The Health Care Manager*, vol. 25, no. 2, pp. 114-121.

Chen, C. C. 2008, "An objective-oriented and product-line-based manufacturing performance measurement", *International Journal of Production Economics*, vol. 112, no. 1, pp. 380-390.

Chen, X. y., Yamauchi, K., Kato, K., Nishimura, A., & Ito, K. 2006, "Using the balanced scorecard to measure Chinese and Japanese hospital performance", *International Journal of Health Care Quality Assurance*, vol. 19, no. 4, pp. 339-350.

Cheng, S. M. & Thompson, L. J. 2006, "Cancer Care Ontario and integrated cancer programs: Portrait of a performance management system and lessons learned", *Journal of Health Organisation and Management*, vol. 20, no. 4, pp. 335-343.

Collis, J. & Hussey, R. 2003, *Business Research . A practical guide for undergratuate and post gratuate students*, Second edition edn, Palgrave Macmillan, Hampshire.

Connelly, J. 2001, "Critical realism and health promotion: effective practice needs an effective theory", *Health Education Research*, vol. 16, no. 2, p. 115.

Cook, T. & Campbell, D. 1979, *Quasi-experimentation: Design and analysis for field settings* Rand McNally, Chicago.

Cooper, R. 1992, "Formal organisation as representation: Remote control, displacement and abbreviation," in *Rethinking Organizations*, SAGE Publications.

Crump, B. 2008, "How can we make improvement happen?", *Clinical Governance: An International Journal*, vol. 13, no. 1, pp. 43-50.

Curtright, J. W., Stolp-Smith, S. C., & Edell, E. S. 2000, "Strategic performance management: Development of a performance measurement system at the Mayo Clinic", *Journal of Healthcare Management*, vol. 45, no. 1, pp. 58-68.

Czarniawska, B. 2001, "Is it Possible to be a Constructionist Consultant?", *Management Learning*, vol. 32, no. 2, pp. 253-266.

Danerark, B., Ekström, M., Jakobsen, L., & Karlsson, J. Ch. 2002, *Explaining society - Critical realism in the scocial sciences* Routledge, London.

Dewey, J. 1938, *Logic: The theory of inquiry* Henry Holt, New York.

Dieleman, M., Toonen, J., Toure, H., & Martineau, T. 2006, "The match between motivation and performance management of health sector workers in Mali", *Human Resources for Health*, vol. 4, no. 1, p. 2.

Drucker, P. M. 1956, *The Practice of Management* New York.

Drummond, M. F., Botten, G., HÃ¤kkinen, U., & Pedersen, K. M. l. 2006, "Assessing the quality of Swedish health economics research", *Scandinavian Journal of Public Health*, vol. 34, no. 6, pp. 566-567.

Dummer, J. 2007, "Health care performance and accountability", *International Journal of Health Care Quality Assurance*, vol. 20, no. 1, pp. 34-39.

Dweiri, F. T. & Kaplan, M. M. 2006, "Using fuzzy decision making for the evaluation of the project management internal efficiency", *Decision Support Systems*, vol. 42, no. 2, pp. 712-726.

Eagle, C. J. & Davies, J. M. 1993, "Current models of "quality": An introduction for anaesthetists", *Canadian Journal of Anaesthesia*, vol. 40, no. 9, pp. 851-862.

Easton, G. 2010, "Critical realism in case study research", *Industrial Marketing Management*, vol. 39, no. 1, pp. 118-128.

Elleuch, A. 2008, "Patient satisfaction in Japan", *International Journal of Health Care Quality Assurance*, vol. 21, no. 7, pp. 692-705.

Evans, J. R. 2004, "An exploratory study of performance measurement systems and relationships with performance results", *Journal of Operations Management*, vol. 22, no. 3, pp. 219-232.

Fitzgerald, L., Johnston, R., Brignall, S., Silvestro, R., & Voss, C. 1991, *Performance Measurement in Service Business* CIMA, London.

Folan, P. & Browne, J. 2005, "A review of performance measurement: Towards performance management", *Computers in Industry*, vol. 56, no. 7, pp. 663-680.

Furnham, A. 2004, "Performance management systems", *European Business Journal*, vol. 16, no. 2, pp. 83-94.

Gardner, T. & Wright, P. 2009, "Implicit human resource management theory: a potential threat to the internal validity of human resource practice measures", *International Journal of Human Resource Management*, vol. 20, no. 1, p. 57.

Greiling, D. 2006, "Performance measurement: a remedy for increasing the efficiency of public services?", *International Journal of Productivity and Performance Management*, vol. 55, no. 6, pp. 448-465.

Griffith, J. R., Alexander, J. A., Jelinek, R. C., Foster, D. A., & Mecklenburg, G. A. 2006, "Is Anybody Managing the Store? National Trends in Hospital Performance", *Journal of Healthcare Management*, vol. 51, no. 6, pp. 392-406.

Griffith, J. R., Alexander, J. A., & Warden, G. L. 2002, "Measuring comparative hospital performance / Practitioner response", *Journal of Healthcare Management*, vol. 47, no. 1, pp. 41-57.

Gupta, M. & Snyder, D. 2009, "Comparing TOC with MRP and JIT: a literature review", *International Journal of Production Research*, vol. 47, no. 13, p. 3705.

Hampshire, A., Blair, M., Crown, N., Avery, A., & Williams, I. 1999, "Original Articles - Action research: A useful method of promoting change in primary care?", *Family Practice - Oxford*, vol. 16, no. 3, p. 305.

HARRÉ, R. O. M. 2009, "Saving Critical Realism", *Journal for the Theory of Social Behaviour*, vol. 39, no. 2, pp. 129-143.

Hartswood, M., Procter, R., Rouncefield, M., & Slack, R. 2002, "Performance Management in Breast Screening: A Case Study of Professional Vision", *Cognition, Technology & Work*, vol. 4, no. 2, pp. 91-100.

Hauck, K. & Street, A. 2007, "Do targets matter? A comparison of English and Welsh National Health priorities", *Health Economics*, vol. 16, no. 3, pp. 275-290.

Hurst, J. & Jee-Hughes 2001, *Performance Measurement and Performance Management in OECD Health Systems* OECD Publishing, No. 47.

Isenmann, R. 2008, "Setting the boundaries and highlighting the scientific profile of Industrial Ecology", *Information Technologies in Environmental Engin e e r in g*, vol. 1, pp. 32-39.

Jones, M. K., Jones, R. J., Latreille, P. L., & Sloane, P. J. 2009, "Training, Job Satisfaction, and Workplace Performance in Britain: Evidence from WERS 2004", *LABOUR*, vol. 23, no. s1, pp. 139-175.

Kaplan, R. S. & Norton, D. 1992, "The Balanced Scorecard - Measures That Drive Performance", *Harvard Business Review*, vol. 70, no. 1, p. 71.

Keegan, D., Eiler, R., & Jones, C. 1989, "Are your performance measures obsolete?", *Management Accounting* no. June, pp. 45-55.

Klein, D., Motwani, J., & Cole, B. 1998, "Quality improvement efforts at St Mary's Hospital: a case study", *Managing Service Quality*, vol. 8, no. 4, pp. 235-240.

Kocakülâh, M. C. & Austill, A. D. 2007, "Balanced Scorecard Application in the Health Care Industry: A Case Study", *Journal of Health Care Finance*, vol. 34, no. 1, pp. 72-99.

Kollberg, B., Dahlgaard, J. J., & Brehmer, P. O. 2007, "Measuring lean initiatives in health care services: issues and findings", *International Journal of Productivity and Performance Management*, vol. 56, no. 1, pp. 7-24.

Kollberg, B., Elg, M., & Lindmark, J. 2005, "Design and Implementation of a Performance Measurement System in Swedish Health Care Services: A Multiple Case Study of 6 Development Teams", *Quality Management in Health Care*, vol. 14, no. 2, pp. 95-111.

Kreiner, K. & Mouritsen, J. 2006, "The analytical interview," in *The Art of Science*, Insight@CBS, pp. 153-176.

Kutney-Lee, A., McHugh, M. D., & Sloane, D. M. 2009, "Nursing: A Key To Patient Satisfaction", *Health Affairs*, vol. 28, no. 4, p. w669.

Lauras, M., Marques, G., Gourc, D., & Lauras, M. 2010, "Towards a multi-dimensional project Performance Measurement System", *Decision Support Systems*, vol. 48, no. 2, pp. 342-353.

Lega, F. & Vendramini, E. 2008, "Budgeting and performance management in the Italian National Health System (INHS): Assessment and constructive criticism", *Journal of Health Organisation and Management*, vol. 22, no. 1, pp. 11-22.

Lim, P. C., Tang, N. K. H., & Jackson, P. M. 1999, "An innovative framework for health care performance measurement", *Managing Service Quality*, vol. 9, no. 6, pp. 423-433.

Loeb, J. M. 2004, "The current state of performance measurement in health care", *International Journal for Quality in Health Care*, vol. 16, no. 1-supplement, p. i5-i9.

Lozeau, D., Langley, A., & Denis, J. L. 2002, "The Corruption of Managerial Techniques by Organizations", *Human Relations*, vol. 55, no. 5, pp. 537-564.

Lynch, R. L. & Cross, K. F. 1991, *Measure Up!* Blackwell Publishers, Cambridge, MA.

Mackay, D., Bititci, U., Ackermann, F., Ates, A., Bourne, M., Davies, J., Gibb, S., MacBryde, J., Shafti, F., & Vand der Meer, R. "The Process of Managing Performance: An Inductive Model", in *Euroma 2008*, Springer.

Meyer, J., Pope, C., & Mays, N. 2000, "Using qualitative methods in health related action research", *British Medical Journal*, vol. 320, no. 7228, p. 178.

Mintzberg, H. 1991, *The Rise and Fall of Strategic Planning and Strategic Planning in Education* Technomic, Lancaster, PA.

Mohammadi, S. M., Mohammadi, S. F., & Hedges, J. R. 2007, "Conceptualizing a Quality Plan for Healthcare", *Health Care Analysis*, vol. 15, no. 4, pp. 337-361.

Mohr, J., Batalden, P., & Barach, P. 2004, "Integrating patient safety into the clinical microsystem", *Quality and Safety in Health Care*, vol. 13, no. 2 Supplement, p. ii34.

Morgan, D. & Morgan, R. 2009, *Single-Case Research Methods for the behavioral and health sciences*, 1 edn, SAGE, Thousand Oaks.

Moullin, M. Defining PM - Should the Definition Include Stakeholders. Perspectives on Performance 4[3]. 2005.
Ref Type: Magazine Article

Moullin, M. 2004, "Eight essentials of performance measurement", *International Journal of Health Care Quality Assurance*, vol. 17, no. 3, pp. 110-112.

Moullin, M. & Soady, J. "Outcomes, processes and capability: using the public sector scorecard in public", in *Euroma 2008*, Springer.

Musgrove, P. 2003, "judging health systems: reflections on WHO's methods", *Lancet*, vol. 361, pp. 1817-1820.

Neely, A. 2005, "The evolution of performance measurement research: Developments in the last decade and a research agenda for the next", *International Journal of Operations & Production Management*, vol. 25, no. 12, pp. 1264-1277.

Neely, A. 1999, "The performance measurement revolution: why now and what next?", *International Journal of Operations & Production Management*, vol. 19, no. 2, pp. 205-228.

Neely, A. & Al Najjar, M. 2006, "Management Learning Not Management Control: The True Role of Performance Measurement", *California Management Review*, vol. 48, no. 3, p. 99.

Neely, A., Gregory, M., & Platts, K. 2005, "Performance measurement system design: A literature review and research agenda", *International Journal of Operations & Production Management*, vol. 25, no. 12, pp. 1228-1263.

Neely, A., Mills, J., Platts, K., Gregory, M., & Richards, H. 1994, "Realizing Strategy through Measurement", *International Journal of Operations & Production Management*, vol. 14, no. 3, pp. 140-152.

Norcross, L. 2006, "Building on success [performance management system]", *Manufacturing Engineer*, vol. 85, no. 3, p. 42.

Nutley, S. & Smith, P. 1998, "League tables for performance improvement in healthcare", *Journal of Health Services Research and Policy*, vol. 3, no. 1, pp. 50-57.

Ondategui-Parra, S., Bhagwat, J. G., Gill, I. E., Nathanson, E., Seltzer, S., & Ros, P. R. 2004, "Essential practice performance measurement", *Journal of the American College of Radiology*, vol. 1, no. 8, pp. 559-566.

Ormrod, J. 1993, "Decision making in health service managers", *Management Decision*, vol. 31, no. 7, p. 8.

Pitt, D. J. 1999, "Improving performance through self-assessment", *International Journal of Health Care Quality Assurance*, vol. 12, no. 2, pp. 45-54.

Pope, C. & Mays, N. 2006, *Qualitative research in health care*, Third Edition edn, Blackwell publishing.

Quinn, R. E. & Rohrbaugh, J. 1983, "A Spatial Model of Effectiveness Criteria: Towards a Competing Values Approach to Organizational Analysis", *Management Science*, vol. 29, no. 3, pp. 363-377.

Radnor, Z. & Lovell, B. 2003, "Defining, justifying and implementing the Balanced Scorecard in the National Health Service", *International Journal of Medical Marketing*, vol. 3, no. 3, p. 174.

Reason & Bradbury 2001, *Handbook of Action Research*, First edition edn, SAGE.

Resar, R. K. 2006, "Making Noncatastrophic Health Care Processes Reliable: Learning to Walk before Running in Creating High-Reliability Organizations", *Health Services Research*, vol. 41, no. 4p2, pp. 1677-1689.

Ridgway, V. F. 1956, "Dysfunctional Consequences of Performance Measurements", *Administrative Science Quarterly*, vol. 1, no. 2, pp. 240-247.

Rochette, C. & Féniíes, P. "A Framework to link patient satisfaction with customer satisfaction", in *Euroma 2008*, Springer, pp. 1-9.

Rundall, T., Martelli, P., Arroyo, L., McCurdy, R., Graetz, I., Newwirth, E., Curtis, P., Schmittdiel, J., Gibson, M., & hsu, J. 2007, "The Informed Decisions Toolbox: Tools for Knowledge Transfer and Performance Improvement Practitioner Application -", *Journal of Healthcare Management*, vol. 52, no. 5, p. 325.

Saaty, T. L. 1982, "Priority setting in complex problems", *IEEE Transactions on Engineering Management*, vol. EM-30, no. 3, pp. 140-155.

Saaty, T. L. 2008, "Relative measurement and its generalization in decision making why pairwise comparisons are central in mathematics for the measurement of intangible factors", *Estadística e Investigación Operativa / Statistics and Operations Research*, vol. 102 (2), pp. 251-318.

Schmidt, S., Bateman, I., Breinlinger-O'Reilly, J., & Smith, P. 2006, "A management approach that drives actions strategically: Balanced scorecard in a mental health trust case study", *International Journal of Health Care Quality Assurance*, vol. 19, no. 2, pp. 119-135.

Shaw, C. 2003, *How can hospital performance be measured and monitored*, WHO Regional Office for Europe, Copenhagen, Denmark.

Smith, M. L. 2006, "Overcoming theory-practice inconsistencies: Critical realism and information systems research", *Information and Organization*, vol. 16, no. 3, pp. 191-211.

Smith, P. C. 2002, "Performance management in British health care: Will it deliver?", *Health Affairs*, vol. 21, no. 3, pp. 103-115.

Strandberg-Larsen, M., Nielsen, M. B., Vallgårda, S., Krasnik, A., Vrangbæk, K., & Mossialos, E. 2007, *Health Systems in Transition*, European Observatory on Health Systems and Policies, 9(6).

Stronks, K. & Mackenbach, J. P. 2006, "Evaluating the effect of policies and interventions to address inequalities in health: lessons from a Dutch programme", *The European Journal of Public Health*, vol. 16, no. 4, pp. 346-353.

Swaminathan, S., Chernew, M., & Scanlon, D. P. 2008, "Persistence of HMO Performance Measures", *Health Services Research*, vol. 43, no. 6, pp. 2033-2049.

Sygehus Sønderjylland 2007, *Strategiplan - Kvalitet Døgnet Rundt*.

The Danish Institute for Quality and Accreditation in healthcare website. 2009. 10-12-2009.
Ref Type: Internet Communication

The Danish Patient Safety Database´s website. 2009. 10-12-2009.
Ref Type: Internet Communication

The National Board of Health´s website. 2009. 10-12-2009.
Ref Type: Internet Communication

the National Indicator Project´s website. 2009. 10-12-2009.
Ref Type: Internet Communication

The Unit of Patient-Perceived Quality's website. 2009. 10-12-2009.
Ref Type: Internet Communication

Treasure, T., Valencia, O., Sherlaw-Johnson, C., & Gallivan, S. 2002, "Surgical Performance Measurement", *Health Care Management Science*, vol. 5, no. 4, pp. 243-248.

Ulin, P., Robinson, E., & Tolley, E. 2005, *Qualitative methods in public health*, 1 edn, jossey-Bass, San Francisco.

Veillard, J., Champagne, F., Klazinga, N., Kazandjian, V., Arah, O. A., & Guisset, A. L. 2005, "A performance assessment framework for hospitals: the WHO regional office for Europe PATH project", *International Journal for Quality in Health Care*, vol. 17, no. 6, pp. 487-496.

Voss, C., Tsikriktsis, N., & Frohlich, M. 2002, "Case research in operations management", *International Journal of Operations & Production Management*, vol. 22, no. 2, pp. 195-219.

Wait, S. & Nolte, E. 2005, "Benchmarking health systems: trends, conceptual issues and future perspectives", *Benchmarking: An International Journal*, vol. 12, no. 5, pp. 436-448.

Walley, P., Silvester, K., & Mountford, S. 2006, "Health-care process improvement decisions: a systems perspective", *International Journal of Health Care Quality Assurance*, vol. 19, no. 1, pp. 93-104.

Walters, B. & Young, D. 2005, "Further Reflections on Critical Realism", *Review of Political Economy*, vol. 17, no. 4, p. 601.

Weick, K. E. 1995, "What Theory is Not, Theorizing Is", *Administrative Science Quarterly*, vol. 40, no. 3, pp. 385-390.

Wheelen, T. & Hunger, J. 1992, *Strategic Management and Business Policy* Addision-Wesley.

Wicks, A. M., St Clair, L., & Kinney, C. S. 2007, "Competing Values in Healthcare: Balancing the (Un)Balanced Scorecard/PRACTITIONER APPLICATION", *Journal of Healthcare Management*, vol. 52, no. 5, pp. 309-324.

Wikgren, M. 2005, "Critical realism as a philosophy and social theory in information science?", *Journal of Documentation*, vol. 61, no. 1, pp. 11-22.

Winter, R. & Munn-Giddings, C. 2001, *A handbook for action research in Health and social care*, 1 edn, Routledge, New York.

Woodward, T. 2006, "Addressing variation in hospital quality: is six sigma the answer?", *IEEE Engineering Management Review*, vol. 34, no. 1, p. 25.

World Health Organization 2008, *The world health report 2008 : primary health care now more than ever.*, WHO Press.

Yang, C. C., Cheng, L. Y., & Yang, C. W. 2005, "A study of implementing Balanced Scorecard (BSC) in non-profit organizations: A case study of private hospital", *Human Systems Management*, vol. 24, no. 4, pp. 285-300.

Yin, R. 1994, *Case study research* Sage Publications, Beverly Hills, CA.

Chapter 9 - Appended papers

The five full papers which this thesis is built upon are appended in this section of the thesis. Two papers are conference proceedings which both have been presented and published as part of international conferences. Three papers have been submitted to international recognized journals, where two have been accepted for publication, and one is still undergoing 1st review phase after submission. Full publication details are provided for each paper.

P1: A new approach for translating strategic healthcare objectives into operational indicators

Submitted to: 16th International Annual EurOMA Conference, Implementation - Realizing Operations Management knowledge, Göteborg, Sweden
Paper ID: F2, no. 242
ISBN: (*no ISBN number is assigned the proceedings)*
Submission date: 16th January, 2009
Acceptance date: 13th February, 2009
Publication date: 14th June, 2009
Type: Full conference paper published in proceedings
Status: Published

A new approach for translating strategic healthcare objectives into operational indicators

Andreas Traberg*, Peter Jacobsen

DTU Management Engineering
Technical University of Denmark
2800 Kgs. Lyngby
Denmark

**Corresponding Author:* Mail: atr@man.dtu.dk, Phone +45 4525 4405

Abstract
This paper proposes a new performance measurement approach enabling healthcare managers to design a performance management system tailored for their individual settings. The approach has been developed over the last two years in cooperation with the radiology department at a Danish hospital. The approach is aiming at compensating for some of the shortcomings in the current strategic process. By incorporating indicators from all organizational levels into an interactive platform, a visual and detailed performance measurement landscape is connected to the strategic plan.

Keywords: Performance Management, Healthcare organizations, Strategic development

Introduction
Raising internal complexity combined with increasing external expectations has put pressure on the healthcare sector. Consequently the need for consistent and transparent performance management is growing (Digital Sundhed 2008). Consequently the development of performance management systems, suited for the healthcare sector has been rapidly evolving in the last decades. (Landrum & Baker 2004). But it is a difficult task to develop structured, impartial, reliable, timely and valid performance management systems. Especially the process of translating strategic objectives into a useful set of operational performance indicators is traditionally a difficult and complicated task. In the healthcare area this is further complicated by the diverse interest of the three main stakeholders, i.e. the grant giving authorities, the patient and finally the employee (Berler, Pavlopoulos, & Koutsouris 2005). In the development of a strategic plan, hospital management is obligated to incorporate strategic objectives, which shows consideration to all stakeholder groups. But to be able to coordinate and manage these different requirements, a performance management system, encompassing performance indicators from all the three stakeholder groups is needed. This regards to both the strategic, tactical and operational level of the organization.

The success of any manager, regardless of organizational level, is his or hers ability to carry out the objectives expressed in the strategic plan. This means carrying out the vision for the hospital management, within his/hers area of responsibility. To be able to realize any strategic plan, it is necessary to know where to take corrective actions, and where operations are on track. In modern healthcare clinical educated staff often is placed in a managerial position. Highly skilled clinical personal without managerial education is

responsible for managing highly complex "production systems". A level of complexity which would put even trained managers to the test.

Therefore the motivation for this new performance model is to provide clinical managers with a tool, which enable them to assess performance of their area of responsibility according to a strategic plan. Thereby managers have enhanced possibilities for taken the necessary corrective actions, on a reliable basis. The approach secures that managers doesn't have to be trained operations managers, to command a series of complex operations within healthcare setting.

Methodology

Our results was derived using the action research methodology (Coughlan & Coghlan 2002). The work is based on a two year study, where information are collected from various data sources, including literary material, interviews, workshops and informal conversations with hospital staff. The approach has been continuously validated by hospital mangers, which should ultimately be the end user. The development cycle has been, authors proposing and presented a framework, testing the framework in healthcare settings, and afterwards redesigned inappropriate elements of the model (Winter & Munn-Giddings 2001). This has resulted in that radiology department at hospital of Southern Jutland are likely to be implementing the approach in the upcoming construction of a new performance structure complementing the new strategic plan 2010-2014.

Proposed performance management approach

Any organizations success depends on its ability to accomplish its objectives, in other words reaching a satisfying level of organizational performance. But managing organizational performance is a complicated task, where it is all about translating results of performance into actions for improvements (Veillard et al. 2005). The basic of this approach is to describe the performance of the organization, according to the context of which the indicator should be evaluated. Performance indicators always have some sort of origin, a reason to be measured. But the output of a specific indicator can be affected by several factors in the organization which needs to be considered in order to make the proper corrective actions. As example can a decrease in X-ray exams be due to lack of personal, which is could be caused by high sickness absence. This high sickness absence could be caused by a not so healthy work environment. So the relation between decreases in production could be caused by bad work environment. It is general knowledge that bad work environment and decrease in production in some cases are connected. But to the untrained eye, the relation between more complex parameters often is blurry. If an "unskilled" manager is focusing on increasing the work speed of the remaining personal to compensate for lack in production, this properly would worsen the problem. Therefore these relations are extremely important to be aware of when assessing indicators and consequently take necessary corrective actions.

By using a visual platform, some of these relations can become apparent for the manager. A visual representation would help managers to be aware of these relations when assessing indicators. As example, Waiting lists. This indicator is properly the most used indicator in modern healthcare (Lega & Vendramini 2008) (Griffith et al. 2006) (Radnor & Lovell 2003). It is often distributed on both location/department and modality. But why is this important? First of all, board of directors often has as a strategic goal to lower the waiting list to a given acceptable level. Secondly the planning levels of the healthcare facility needs the information, to allocate resources for the critical areas. Last but not least, waiting list is incorporated in almost every mandatory report on hospital

performance. For a department manager this means that waiting list is used in three different contexts. First the evaluation of strategic compliance, secondly in capacity planning of personal/equipment and finally in the evaluation according national benchmarks. This simple example shows that the manager carefully needs to considerate how to solve the problem.

To be able to coordinate these three dimensions, the model is based on the idea from the CIMOSA representation (Kosanke 1991). The model consists of a three dimensional relation matrix. The first axis describing the strategic objective of the organization, the second axis describing the organizational levels, and the third axis are an evaluation axis, see Figure 1.

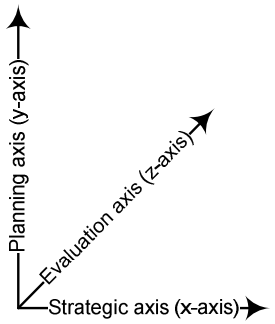

Figure 1: Structural description of performance approach

In a healthcare environment, these three dimensions would always in some way be interrelated, or at least should be. This is because those indicators which have no strategic motivation should not be measured. If the indicator is strategic justified, then one of the planning levels must be responsible for the accomplishment of the goal. Finally the indicator needs to be evaluated and assessed to be useful.

First step of the process is to determine the value of each of the three axes in the matrix. The strategic axis (x-axis) would often be related to Balanced Scorecard or Business Excellence. Each individual healthcare facility would construct a personalized matrix due to the structure of their strategic objectives. The strategic objectives should be listed along the axis, in the order they appear in the strategic plan. The planning levels would be dependent on the management structure. Hospitals are often divided in three levels of management, with board of directors, department management and team management. It should be kept in mind that the planning axis only should contain organizational levels with managerial responsibility. The z-axis or evaluation axis is referring to the internal and external agencies which evaluate the specific department. This can be a range of different organizations either national or regional. These organizations devise guidelines, and monitor indicators inside clinical and patient related quality. These standards/ indicators are to be placed in accordance with the z-axis. Because of the amount of organizations measuring hospital performance, it is important to carefully select which to implement in the matrix. The strategic plan of the individual healthcare facility would often reveal which organizations, board of directors consider most important. If there is a formalized internal evaluation procedure, this should also be implemented as an element on the z-axis. This would help the department management, in evaluation both internal and external performance.

Next step in the process is to load the matrix with indicators. The concept is to develop the indicators in a cascading structure, where the underlying indicators constitute the overlying. This approach suggests that the indicators are developed top-to-bottom, with the strategic objectives and the evaluation axis as baseline, i.e. the x-z level. All indicators which are defined in mandatory reports are distributed according to the strategic plan of

the organization. This will in all cases be possible, because a strategic plan of a hospital is designed to encompass the requirements from national or regional authorities. When the indicators are placed in the x-z level, the indicators should be developed according to the planning levels. As well as the interrelation between strategy and authorities is important, the planning structure of the indicators is just as important. Healthcare facilities are characterized by a high number of planning levels, which demand contiguous multi level indicators (Lemieux-Charles et al. 2003). Each level of the organization would have to be provided with performance indicators which apply for their specific area of responsibility. The process of the actual indicator development is based on a hieratical step-by-step approach obeying the following two rules.

1. Indicators should not be assigned to individuals, which does not have organizational power to enforce, or don not have full impact on the outcome
2. Indicators should not be assigned to individuals, where the employee does not have the professional competencies to influence the outcome.

The indicators would be designed through the organization (top-to-bottom), from strategic objectives into operational indicators, until one of the rules is violated. It is an iterative process, where each indicator is confirmed by the two rules. If one of the rules are violated, the indicator line, are either stopped, or transformed into proxy indicators. In the case where an indicator is split up, there should be a significant reason to so, because the indicator landscape is attempted minimized. The process of continuously repeating the rules, secures that indicators aren't forced to deep in the organization.

The description of the individual indicator plays almost as an important part of the performance system as the structure itself. If indicators aren't described properly, the assessment of these would often become a mess. Therefore it is recommended that the description of the indicators is compatible with some of the receivers of the mandatory reports. If the organizations indicators resample the recipients' structure, it would lighten the data adjustment. In the Danish healthcare sector, the National Indicator Project (NIP) plays a significant role. All Danish hospitals are obligated to construct mandatory report on a biannual basis. The structure of indicators is therefore encouraged to use the same template as NIP. In this way, indicators used internally, could unaltered be used as reporting for NIP or other national agencies.

Testing the approach

The model was tested at the radiology department, and a detailed 3-dimensional indicator landscape was constructed. Based on the hospitals overall strategic plan, a performance matrix vas developed. The strategic plan is a Balanced Scorecard look-a-like, where the four strategic objectives are divided into twelve sub-strategic goals. Each of the departments of the hospital is obligated to follow all twelve goals, which mean they all figure in the matrix. In terms of clarity, only the four strategic objectives are shown, but the underlying level shows each of the twelve sub-strategic goals. In the z-axis, there are three mandatory reports which are to be implemented, board of directors, NIP reports and the report for the Danish Quality model. The report for board of directors is a description of department management, according to the strategic goals. Each department are obligated to conduct an annual report, stating progress on all twelve strategic goals. The Danish National Indicator Project (NIP) measures the quality of care provided by the hospitals to groups of patients with specific medical conditions. These reports are published on a website (www.sundhed.dk) signifying the performance of Danish hospitals. These reports have therefore a significant value in terms of performing well. The Danish

Quality model resembles the Business Excellence model in industrial organizations. The model consists of a series of standards for persistent quality of care in the Danish healthcare sector. During the next years there will be an accreditation of all Danish hospitals, and if they act in accordance with the standards they will become certified. These three reports are for a Danish hospital the foremost important, why we chose these as the z-axis. The y-axis is representing the actual planning levels at the hospital. The Hospital of Southern Jutland is fusion of four independent hospitals. Therefore management is structured as a unified top management, and a head of each department. The radiology department therefore has one head of the department, and four local managers which handles daily operation. That leaves management at the hospital in three steps. The full matrix for the radiology department of southern Jutland is shown in Figure 2.

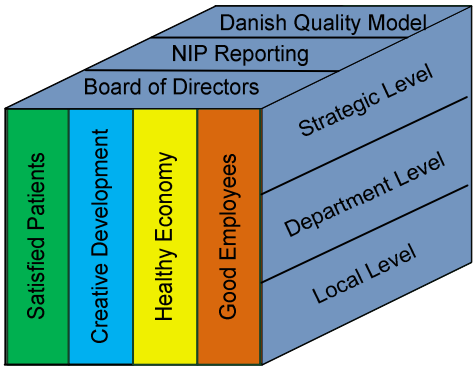

Figure 2: Developed performance matrix, Radiology department of Southern Jutland

By using the two stop-rules in the indicator construction rules only about 40 percent of the indicators reach department level, and only 10 percent of the indicators reach the local management level. Meaning that there were seen a significantly decrease in indicators for local managers. The decrease in indicators is significantly easing the administrative burden of middle managers. Previously middle managers used considerably amount of time reporting on indicators which they didn't have full impact on. With this new structure, the reporting part has been minimized to only encompass the indicators they directly are responsible for. The model therefore gives a more transparent and organization specific structure. The model also provides each organizational layer with the possibility to evaluate its own impact according to the overall strategic objectives.

One of the main objectives for the development of this performance management approach was to make the model useful in a visual environment. Managers which aren't educated in management need to have an intuitive tool, and because many humans are visual oriented, graphics are considered helpful. The model has therefore been built in a web-based environment. By "slicing" through the matrix, indicator sub-levels appear, signifying which measures apply for this particular area. As Figure 3 shows, by opening "Satisfied patients", the sub-goals for this strategic goal become apparent. Furthermore illustrates the right-hand box where the user presently is located in the performance matrix. By "clicking" your way further down web-based model, all indicators through the planning levels becomes present.

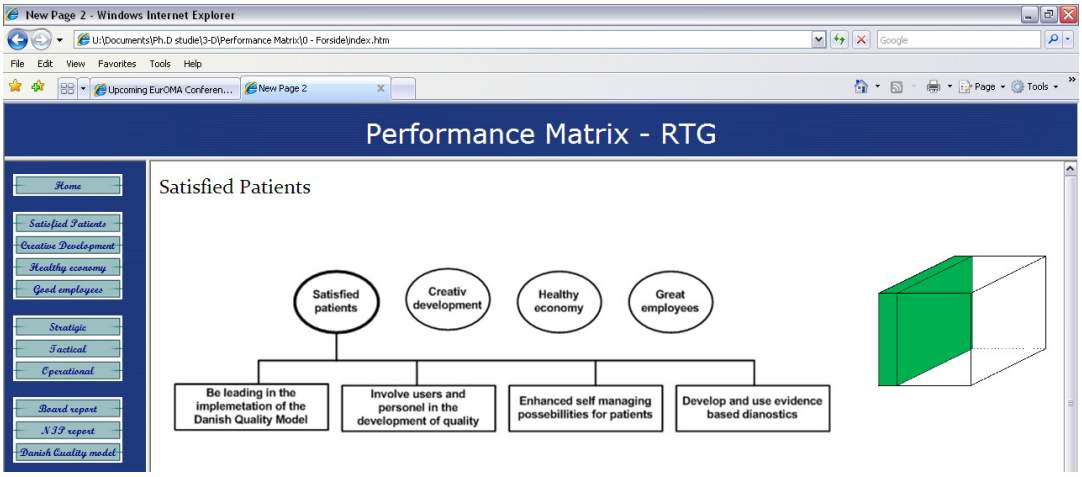

Figure 3 Satisfied Patients

As described each of the indicators resembles the indicator structure from NIP, which mean that indicators are described by following template; Indicator name, Purpose, Responsible, Field of application, Indicator description, Displaying guidance, Data foundation, Indicator goal, Timeframe, Guiding documents, Benchmark and References. As example the indicator "Waiting list, is shown in Figure 4. As for all of the indicators the right-hand side is displaying where the location in the performance matrix. Figure 4 is displaying the strategic use of waiting list in the hospital is according to the strategic goal 1, indicated as a green box. Waiting list is connected to the sub-goal "Be leading in implementation of the Danish Quality Model", which is the reason that the "slice" is narrow.

	Description
Indicator name	Waiting List
Purpose	Continuously monitor the maximal waiting time for a non-acute patient, distributed on modalities
Responsible	Head of department
Field of application	Each four radiology sections of the hospital
Indicator description	Waiting time to the next open examination slot in the booking system for each modality
Displaying guidance	Y-axis: Waiting time in days X-axis: Calendar days 6 month back
Data foundation	Data is collected from RIS (Radiology Information System)
Indicator goal	Waiting time below 20 days, Complying with National Treatment assurance (4 weeks)
Timeframe	At all time
Guiding documents	The Danish Quality model (www.ikas.dk) The National Indicator Project (www.nip.dk)
Benchmark	Monthly benchmarked internally between all four locations Bi-annual the waiting time is benchmarked externally between Danish hospitals
References	The Danish Quality model, Standard 3.1.1- Standard 3.2.1- Standard 3.6.1 - Standard 3.8.1- Standard 3.11.1

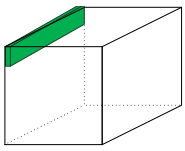

Figure 4: Waiting list indicator, referring to the strategic goal "Satisfied patients".

Besides being part of the Danish quality model, Waiting list also figure in the bi-annual report for board of directors and in the NIP reports. As for all of the indicators in the performance structure, the web based environment is built, and has been tested at the hospital.

Discussion

The increasing demand for reporting on more and more specific key factors is insisting on an even more all-embracing IT architecture in the future. The demand for clinical equipment capable of conducting performance evaluation would be increasing. The need for all hospital information systems to be able to interact with each other would likewise increase in the future, due to the increasing demand for both national and international benchmarking. Therefore more and more information is needed to handle healthcare production systems. This trend is already putting a mark on software providers which are developing software to meet the demand for performance software. Digital Dashboards, as this approach, are increasingly being implemented as a way of interactively displaying organizational performance (Morgan et al. 2008). Furthermore the last decade's growth towards using more mathematical strict process management approach in industrial organizations is likely to be beneficial in healthcare sector as well. The concept of Six Sigma is already gaining acceptance in several healthcare institutions, and an advancement of this method would be likely in the future (Woodward 2006). In this aspect the use of IT based models will continue to be more and more essential, because the models complexity demands computing power to give valuable feedback. But one key issue is that healthcare organizations would experience information overload. The technical capacity is present, technical providers can provide the equipment which can handle this massive amount of data, and exchange these with other facilities. But are the system developers capable of structuring the data so only useful data is communicated? Is there paid enough attention to the limiting of performance information? Our guess is "No".

A satisfying level of information is individual, some want much and some want less. This is why information management is becoming a more and more complicated task. But with this model, information according to performance is both available and transparent. Available so that employees have the opportunity to gather required information, and transparent because they have the opportunity to see in what context the indicator is measured. It's possible to see only the big lines, but the matrix also gives the opportunity of more detailed descriptions. Therefore this approach is seen as a step in the direction of thoroughly selecting which data, for individual needs. The easy task is to provide all data to everybody, but to provide only the necessary and specific data is an art. Managers and employees would neglect the information, cause by the information spamming.

The issues of uniting soft and hard measures, fitted to changing demands from national authorities necessitate extremely flexible performance models. But this is exactly what a future healthcare performance management system has to embrace. Development of new treatments contributes to the ongoing changing environment, and as a consequence patient expectations to quality continue to intensify. More and more hospitals are using strategic development plans which changes every 4-6 years. These aspects are contributing to the demand for extremely versatile performance systems. When developing suitable performance management systems, the task of deducing measures deep in the organization is a key matter. The task of implementing individual or team-based indicators is currently a hot topic at numerous hospitals, and is approached by several scientists all over the world. By using the proposed performance structure, the configuration of the indicators becomes understandable to the user. When a performance problem occurs, it clearly appears which parts of the organizations obligations performance is lacking. By visualizing the present indicators in a matrix form, managers have a tool for identifying unsatisfying performance, and in the light of this call for corrective actions.

Conclusions

The future healthcare sector is demanding continues development of performance management model, where flexibility and transparency should define the models of tomorrow. Standards of quality in care would forever be increasing, and the demand for extensive reporting likewise. Healthcare institutions are required to perform first-class in a range of areas, and to manage the organization towards high class performance, a fine-mesh performance model has to be developed. The development of more holistic oriented systems would become essential an essential challenge for healthcare organizations if they are to cope with the external pressure in the future. Deep cross-organizational evaluation would to a great extend support the organizations to manage performance, and consequently secure high quality of care.

Limitations

It is clear that when the model is developed in cooperation in the same environment where it is tested, it would limit the generalizing potential. To fully prove whether the model is useful, it is necessary to widen the scope of the testing to a broader range of healthcare facilities. Despite these implications, the finding in this study can be a useful basis for more research on the difficulties related to the strategic development process in healthcare organizations.

Acknowledgements

We thank the employees of the Hospital of Southern Jutland for contributing to the scientific work. Especially the management has been of tremendous help to us in the process, including coordinating the interview and workflow sessions.

References

Berler, A., Pavlopoulos, S., & Koutsouris, D. 2005, "Using key performance indicators as knowledge-management tools at a regional health-care authority level", *IEEE Transactions on Information Technology in Biomedicine*, vol. 9, no. 2, pp. 184-192.

Coughlan, P. & Coghlan, D. 2002, "Action research for operations management", *International Journal of Operations & Production Management*, vol. 22, no. 2, pp. 220-240.

Digital Sundhed 2008, *National Strategi for digitalisering af sundhedsvæsenet 2008-2012*, Sammenhængende Digital Sundhed i Danmark.

Griffith, J. R., Alexander, J. A., Jelinek, R. C., Foster, D. A., & Mecklenburg, G. A. 2006, "Is Anybody Managing the Store? National Trends in Hospital Performance", *Journal of Healthcare Management*, vol. 51, no. 6, pp. 392-406.

Kosanke, K. The European approach for an open system architecture for CIM (CIM-OSA)-ESPRIT project 5288 AMICE. Computing & Control Engineering Journal 2[3], 103-108. 1991.
Ref Type: Journal (Full)

Landrum, L. B. & Baker, S. L. 2004, "Managing Complex Systems: Performance Management in Public Health", *Journal of Public Health Management & Practice*, vol. 10, no. 1, p. 13.

Lega, F. & Vendramini, E. 2008, "Budgeting and performance management in the Italian National Health System (INHS): Assessment and constructive criticism", *Journal of Health Organisation and Management*, vol. 22, no. 1, pp. 11-22.

Lemieux-Charles, L., McGuire, W., Champagne, F. o., Barnsley, J., Cole, D., & Sicotte, C. 2003, "The use of multilevel performance indicators in managing performance in health care organizations", *Management Decision*, vol. 41, no. 8, pp. 760-770.

Morgan, M. B., Branstetter, B. F., IV, Lionetti, D. M., Richardson, J. S., & Chang, P. J. 2008, "The Radiology Digital Dashboard: Effects on Report Turnaround Time", *Journal of Digital Imaging*, vol. 21, no. 1, pp. 50-58.

Radnor, Z. & Lovell, B. 2003, "Defining, justifying and implementing the Balanced Scorecard in the National Health Service", *International Journal of Medical Marketing*, vol. 3, no. 3, p. 174.

Veillard, J., Champagne, F., Klazinga, N., Kazandjian, V., Arah, O. A., & Guisset, A. L. 2005, "A performance assessment framework for hospitals: the WHO regional office for Europe PATH project", *International Journal for Quality in Health Care*, vol. 17, no. 6, pp. 487-496.

Winter, R. & Munn-Giddings, C. 2001, *A handbook for action research in Health and social care*, 1 edn, Routledge, New York.

Woodward, T. 2006, "Addressing variation in hospital quality: is six sigma the answer?", *IEEE Engineering Management Review*, vol. 34, no. 1, p. 25.

P2: Benchmarking in healthcare using aggregated indicators

Submitted to: International Journal of Health Planning and Management
ISBN:
Submission date: 15 December, 2009
Acceptance date: 1st July, 2010
Publication date:
Type: Full paper publication
Status: Conditionally accepted, currently through 2nd review phase

Benchmarking in healthcare using aggregated indicators

Authors

Andreas Traberg*
M.Sc., Ph.d. student
Technical University of Denmark
DTU Management Engineering
2800 Kgs. Lyngby
Denmark
Phone: (int+45) 4525 4405
E-Mail: atr@man.dtu.dk

Peter Jacobsen
Associate Professor
Technical University of Denmark
DTU Management Engineering

Nadia Monique Duthiers
MA, GDBA (Management and Organization)
Hospital of Southern Jutland
Radiology department

***Andreas Traberg is the corresponding author**

Summary
Benchmarking has become a fundamental part of modern health care systems, but unfortunately, no benchmarking framework is unanimously accepted for assessing both quality and performance. The aim of this paper is to present a benchmarking model that is able to take different stakeholder perspectives into account. By presenting performance as a function of a patient perspective, an operations management perspective, and an employee perspective a more holistic approach to benchmarking is proposed. By collecting statistical information from several national and regional agencies and internal databases, the model is constructed as a comprehensive hierarchy of indicators. By aggregating the outcome of each indicator, the model is able to benchmark healthcare providing units. By assessing performance deeper in the hierarchy, a more detailed view of performance is obtained. The validity test of the model is performed at a Danish non-profit hospital, where four radiological sites are benchmarked against each other. Because of the multifaceted perspective on performance, the model proved valuable both as a benchmarking tool and as an internal decision support system.

Keywords
Healthcare, Performance Management, Aggregated indicators, Benchmarking

Healthcare performance management
The healthcare sector is one of the fastest growing areas of the economy of most developed countries (Glance et al. 2008). Governments (and taxpayers) invest large amounts of money in it directly or indirectly, and expect a high quality of service from this sector (Purbey, Mukherjee, & Bhar 2007). Demographical developments increase the demand from national and local governments for better quality and higher performance at a lesser cost, and for care catered to different groups of patients (Mohammadi, Mohammadi, & Hedges 2007). The ultimate goal is to manage quality and performance. But you cannot

manage it until you have a way to measure it, and you cannot measure it until you can monitor it (Eagle & Davies 1993). But monitoring quality and performance is a difficult task, which implies that the concepts are well defined and understood before they become measureable. In literature dealing with quality and performance management a semantic confusion has arisen, resulting in both terms being used randomly to describe common ground, but in this paper we will not examine one without implicitly considering the other. In the following 'performance' will be used to describe both terms.

Performance management in healthcare has become as essential a task as it is in the business environment. Neely argues that there are seven main reasons that performance measurement has attracted much attention recently: the changing nature of work; increasing competition; specific improvement initiatives; national and international quality awards; changing organizational roles; changing external demands; and the power of information technology (Neely 1999). Performance measurement provides the basis for an organization to assess how well it is progressing towards its predetermined objectives, helps to identify strengths and weaknesses, and decides on future initiatives, with the goal of improving organizational performance. Performance measurement is not an end in itself, but a tool for more effective management. Results of performance measurement indicate what happened, not why it happened, or what to do about it. In order to make an organization effective, the performance measurement outcomes must be able to make the transition from measurement to management (Purbey, Mukherjee, & Bhar 2007).

In order to measure whether health care provides value for money, indicators are used to measure performance and the results are benchmarked against each other within and across institutions. Benchmarking has become an intrinsic part of most developed health care systems, but unfortunately, health care is still a major industry in which no indicators are unanimously accepted as tools for defining, measuring, and ultimately benchmarking the performance of its services (Ondategui-Parra et al. 2004). Several methodological challenges remain in the field of benchmarking, many of them related to the selection and the quality of indicators used to make comparisons both within and between health care systems (Wait & Nolte 2005). These challenges are largely due to the situation, that the tools each cover specific stakeholder interests, e.g. patient satisfaction, clinical performance, patient safety and waiting times.

The different stakeholders in the health care system all have varying perspectives on how to interpret performance (Loeb 2004). In most countries with public health care systems, the Government as the grant giving authority devises the superior strategic goals and efficiency requirements. Patients also act as stakeholders expecting best possible treatment and safety. In recent years patients have become increasingly involved as partners in care, rather than just being receivers of care. Because of this development patient concerns have been able to affect the design of care, in Denmark resulting primarily in a higher level of information and a focus on reducing waiting times. Employees represent a third stakeholder. As patients are becoming partners in care, employees are changing status from care providers to developers of care. Development of care is among others related to professional competency, technology and teamwork. In an attempt to cover all aspects of health care, indicators representing different perspectives of performance have been developed; resulting in stakeholder dependent viewpoints.

Obtaining a holistic and objective assessment of health care performance useful for health care management is a difficult transition to make, because the individual assessments point in many different directions. Attempts at resolving this complicated task have in

some cases resulted in an overload of indicators with little mutual relation, and small practical value (Geraedts, Schwartze, & Molzahn 2007). Evaluation becomes a series of still lifes, rather than a holistic assessment of performance. Clinical indicators, patient satisfaction surveys, workplace and patient safety evaluations, are as individual models for assessment, indispensable in the evaluation of healthcare. The issue in regards to these frameworks is that they are stand-alone-models which only portray a segment of reality (Tarantino 2003).

An empirical study has shown that the ambiguous information which exists in performance measures used at the hospital department level, maintains the decoupling between clinical activities and management control practices. This decoupling creates management control problems because it hampers the knowledge on the cause-effects of actions, which is important in order to undertake strategic decisions and diagnostic action (Pettersen & Nyland 2006). For health care managers the issue poses a real problem, because without indicators structured in relation to operational context, managers are unable to make informed decisions (Rundall et al. 2007). If health care institutions are to provide high performance, health care managers must be able to make decisions that relate multiple stakeholder interests (Minkman, Ahaus, & Huijsman 2007).

Motivation

The aim of this paper is to present a benchmarking model, which is able to take different stakeholder perspectives into account, and provide a structured and reliable model, which represents performance in a holistic manner. The attempt is to provide a model which, by aggregating indicators, is able to provide a performance overview of the organization through key measures. The reason for using aggregation is to limit the amount of performance indicators and at the same time exploit the huge amount of already registered data. Four local sites were benchmarked against each other, in order to evaluate individual performance. By evaluating whether the organizational role of each department becomes apparent in the result, the validity of the model is tested.

The Case

The case used for this study is a Danish non-profit healthcare institution, the result of a fusion between four former independent hospitals. The hospitals were merged at management level, but the four sites still act as operational parts in the new hospital. The choice of case originates in three issues constituting a challenge to any benchmarking model; 1) Sites are inhomogeneous in size and equipment, 2) hospital and sites represent multilevel management, and 3) sites are assigned different purposes e.g. acute vs. non acute and teaching obligation vs. no teaching obligation. The particular character of this case provides a challenge as well as a possibility for benchmarking for example sites vs. sites and/or department vs. external departments.

The radiology department, which constitutes this case, employs 128 staff members distributed on four sites. The department performs almost any form of radiological examination. The distribution of patients is dependent on the type and amount of equipment and geographic location. The department treats approximately 145.000 patients per year, where about 40% are acute patients.

Methodology

This work is divided into three phases, first a qualitative design phase where the hierarchical indicator model is constructed, second a quantitative test, and third a qualitative validation of the designed model. The construction of the indicator model was

performed as a single case study (Morgan & Morgan 2009), where information was collected from various data sources including workshops (Meyer, Pope, & Mays 2000), analytical interviews (Kreiner & Mouritsen 2006), and informal conversations. Qualitative data were solely collected from employees at the radiological department. The interviews were conducted across all organisational levels. Managers were used as the primary source of data, where medical managers, nursing managers and the project coordinator were interviewed.

The quantitative test included both internal and external data. The collected data stem from various sources such as HR databases, a document management system, and the Radiology Information System/Picture Archiving and Communication System (RIS/PACS). The external data were collected from four federal units and governmental agencies; 1) The Unit of Patient-Perceived Quality's survey of patients' experiences in Danish hospitals, a patient satisfaction survey conducted every two years(The Unit of Patient-Perceived Quality's website 2009). The objective of the survey is to benchmark patient experiences by comparing responses across hospitals over time. The survey includes 30 questions which are answered by about 30.000 patients. In addition, the Danish Quality model, which is a Danish accreditation institution, assesses how well information to patients is distributed (The Danish Institute for Quality and Accreditation in healthcare website 2009). This information is regarded as fundamental for determining the level of patient satisfaction. 2) The Danish Quality Model is an accreditation framework developed by the Danish Institute of Quality and Accreditation in healthcare. The model itself consists of 35 standards related to organizational issues, 54 standards focusing on the continuity of care, and 15 specific disease related standards. All of these standards contain indicators related to different organizational levels. 3) The National Indicator Project has as its purpose to evaluate the treatment of; acute surgery, Chronic Obstructive Pulmonary Disease, Diabetes, heart failure, hip fracture, lung cancer, schizophrenia, and stroke (the National Indicator Project´s website 2009). 4) Patient safety records created by the National Board of Health (The National Board of Health´s website 2009) and the Danish Patient-Safety Database (The Danish Patient Safety Database´s website 2009). It is important to notice that all external data are public, and validated by the federal units and governmental agencies issuing them.

The model construction

The form of the model was chosen, because the aim of this study was to use an index as a common denominator for all included indicators. An example of such an approach is Nakajima´s metrix, introducing the use of aggregated indicators in an Overall-Equipment-Efficiency indicator (OEE), where *Availability*, *Performance*, and *Quality* is combined into one single measure (Nakajima 1986).The OEE combines the indicators into one expression for how "well" equipment, assembly lines or manufacturing lines work. Aggregated indicators rely on mathematical summarization of the outcome of individual measures combined into superior merged indicators.

The guiding principle

The concept of clustering indicators into an OEE measure was adopted and modified to fit health care settings. The merging of indicators provides an index of performance, which does not relate to one single measure, but to a cluster of indicators, each resulting in a high level indicator representing a summation of included lower level indicators, see Figure 1.

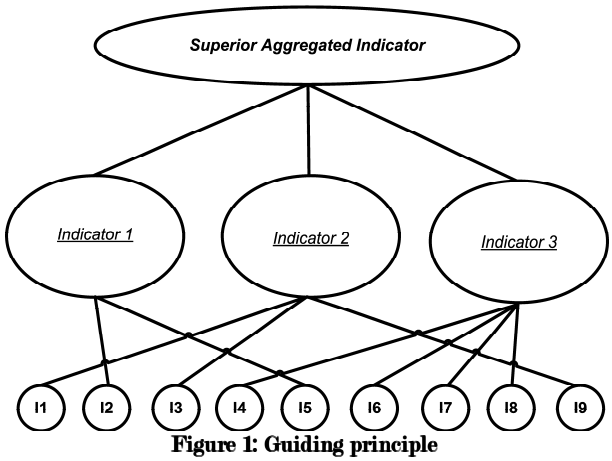

Figure 1: Guiding principle

The summation of lower level indicators in clusters, into a higher level indicator provides the possibility to trace performance both ways. An important reminder when using aggregated indicators is that the indicator outcome in itself only has value in comparison. The aggregated indicator is a fictional number, which represents an estimate of a subsidiary level's outcome. An OEE of 0.85 for example does not provide meaning unless this number can be benchmarked against another's performance, past performance or even an organizational target. The use of aggregated indicators is therefore useful, both in external and internal contexts.

Developing the indicator hierarchy

The qualitative phase of this study has centred on shaping an indicator hierarchy to match health care settings. The case hospital's strategic plan became instrumental in creating the hierarchy's superior structure. The strategic plan was designed to satisfy three main stakeholders; patients, operations management, and employees – a design defining the three superior clusters used in the model. The stakeholders are by the strategic plan defined as equal, and in the model assigned equal mathematical weight. The three main stakeholder groups are in line with what scientific literature refers to as the main stakeholders in modern healthcare (Minkman, Ahaus, & Huijsman 2007). Adopted from manufacturing and service, the patient is regarded as a "customer", which the organization has to treat in competition with other healthcare providers in order to secure market shares (Rochette & Féniies 2008). Securing operational excellence in "production" is common knowledge, as well for manufacturing companies as for healthcare providers (Langabeer 2008). It is widely accepted that poor health among employees, and low job satisfaction affect organizational performance, which has led to a focus on avoiding this (Riedel & Lynch 2006).

The interviews were conducted based on the choice of these stakeholders as a foundation for the model's design. The three stakeholder groups each have different expectations and requirements in order to be satisfied by the performance delivered by a healthcare facility. Therefore the aim was to make an aggregated indicator structure for each of these stakeholder groups. The structure has incorporated measures which affect the individual stakeholder's expectations and requirements of how "well" the system performs. The final benchmark result is calculated as an average of the outcome of the three stakeholder clusters. This aggregation of outcomes creates a relation between the stakeholders at a superior level, which enables representation of performance in a holistic manner. Each indicator was evaluated and placed under the perspective where it was assumed to be most

adequate. In the indicator selection phase it was explicitly important to avoid possible duplication between indicators in the three perspectives. Therefore each indicator was evaluated according to stakeholder association and possible duplication. To limit the mutual impact between the perspectives all probable related indicators, only appears ones in the hierarchies. We are aware of possible relationships between indicators from the three stakeholder perspectives. As described in multiple scientific papers ((Doherty 2008;Eilers 2004;Roelen et al. 2008), the outcome of any performance indicator would be affected by different issues in the organization. It is often possible to make a certain correlation probable, but the degree of this mutual impact is at least very difficult, or even impossible to quantify.

As the model is developed as management information tool and the indicators which are incorporated in the hierarchies are selected and agreed upon by management at the radiology department. As the model supports the decision-making processes, the mangers are chosen as the primary source of data. The interviews conducted with all participants in the management team, where organizational consensus where obtained at two workshops after the interviews.

Patient perspective

Based on the interviews conducted, performance related to a patient perspective can be broken down into safety and satisfaction. Patient safety is constituted by mortality, morbidity, infections and unintended incidents which can occur during hospital stay. The safety related measures are chosen because these four indicators traditionally define patient safety at the hospital. Patient satisfaction is grouped into four clusters derived from the interviews conducted; satisfaction survey, information, complaints, and contact person, see Figure 2.

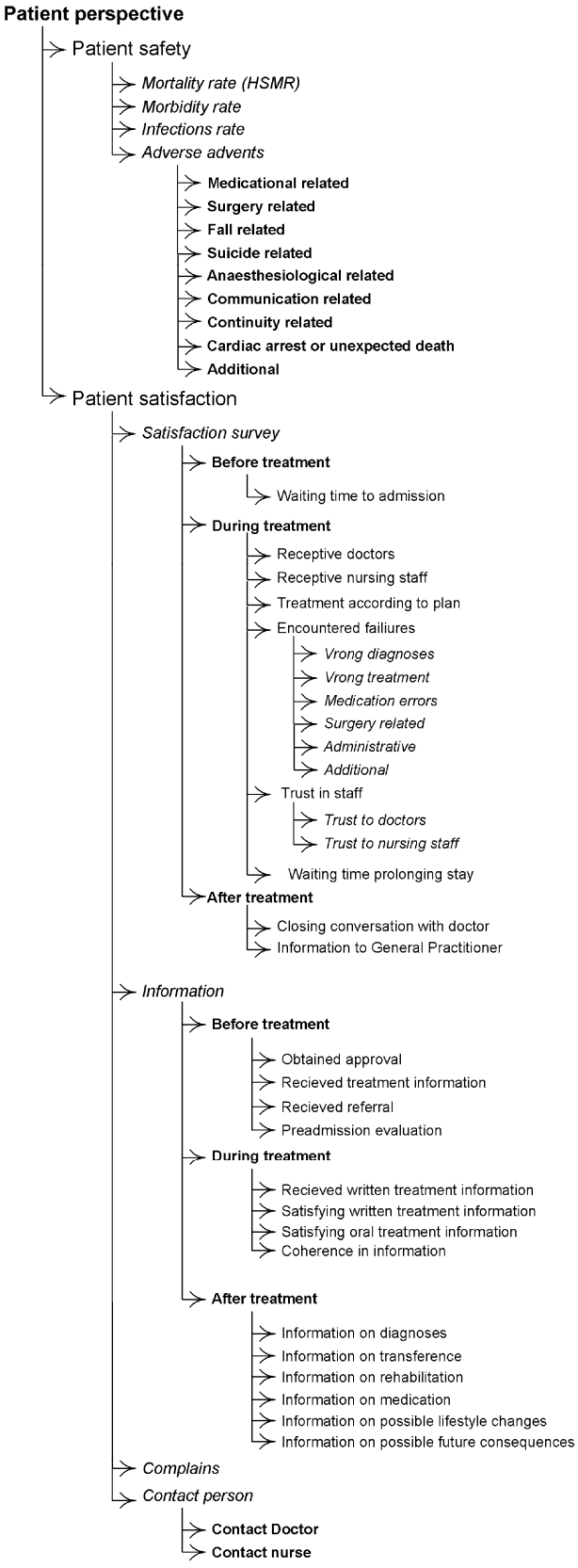

Figure 2: Indicator hierarchy for Patient perspective

To give an example of how the calculations are executed, the way in which patient satisfaction is calculated, is explained. Each cluster in the hierarchy has been weighted in the event that one cluster needs a higher priority than another. In this way the aggregated indicators for patient safety and patient satisfaction constitute the patient perspective indicator, see Equation 1.

$$\text{Patient perspective} = \frac{(W_{Psafe} * \text{Patient safety}) + (W_{Psatis}\text{Patient satisfaction})}{W_{Psafe} + W_{Psatis}}$$

Equation 1 Patient perspective

The same procedure repeats itself for the patient safety indicator, where mortality, morbidity, infection rates, and adverse advents compose the input, and these four clusters are again calculated as a weighted average, see Equation 2.

$$\text{Patient safety} = \frac{(W_{Mt} * \text{Mortality rate}) + (W_{Mb} * \text{Morbidity rate}) + (W_I * \text{Infection rate}) + (W_A * \text{Adverse advents})}{(W_{Mt} + W_{Mb} + W_I + W_A)}$$

Equation 2 Patient safety

Morbidity and infection rates are calculated as standard percentiles, whereas mortality rates are calculated using the Hospital Standardized Mortality Ratio (HSMR), see Equation 3.

$$\text{Mortality rate} = 1 - \frac{\text{HSMR [\%]}}{100}$$

Equation 3 Mortality rate

The rate of unintended incidents is calculated as a weighted average based on nine clusters, see Figure 2, and the outcome is a percentage output of the ratio between the reported number of adverse advents and the total production, see Equation 4.

$$\text{Adverse advents} = \frac{\text{Adverse advents [No.]}}{\text{Total production}}$$

Equation 4 Adverse advents

Operations management perspective

Operations management is grouped into four main clusters, see Figure 3.

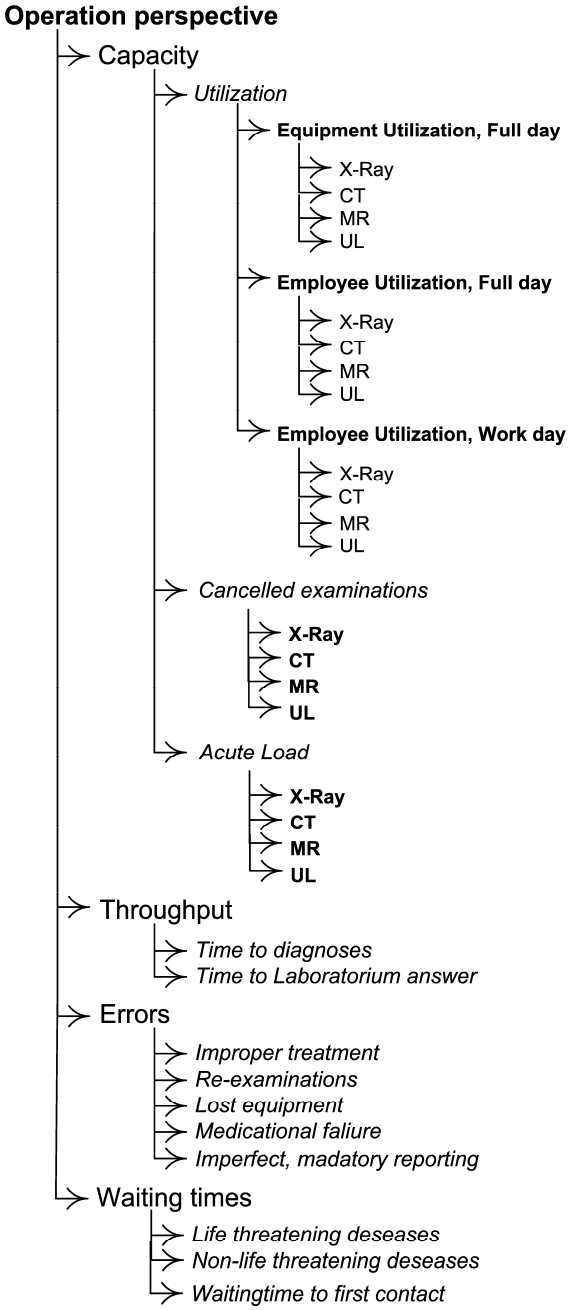

Figure 3: Indicator hierarchy for Operations perspective

The calculations are executed by the same procedure as used in the patient perspective. Because of little consensus among practitioners (Lafond, Brown, & Macintyre 2002), it is important to emphasize the calculations regarding utilization. The important issue concerning the calculation of utilization is this study's inclusion of non-attending patients. During the interview sessions, the amount of non-attending patients was consistently mentioned as being significant. The absence of these patients cause open slots in the planning schedule. The rate of utilization is therefore highly sensitive to non-attending patients. The model compensates for these open slots by adding the number of non-attending patients to the total production, see Equation 5.

$$Utilization_{Equipment} = \frac{(\sum Production + \sum Number\ of\ non\ attending\ patients) * operational\ time}{No.equipment * 24\ Hours * Operational\ time * Efficiency\ factor}$$

Equation 5 Equipment utilization

There are multiple ways of calculating rate of utilization. The model itself is not sensitive to the choice of calculation method, as long as the calculation is performed alike at all sites.

To provide a full picture of how well the capacity is utilized at a hospital department, calculating the utilization rate of equipments is not enough (Lafond, Brown, & Macintyre 2002), due to the situation that the utilization rate of employees is not dependent on the utilization rate of equipment, because employee resources are not restricted to the use of equipment. The utilization rate for employees is therefore calculated separately, see Equation 6.

$$Utilization_{Employees.Fullday} = \frac{((\sum Production + \sum Number\ of\ non\ attending\ patients) * operational\ time}{\left(No.employees * No.shifts * \frac{Hours}{Shift}\right) + Overtime}$$

Equation 6 Employee utilization, Full day

Because of the different purposes assigned to the sites, the employee utilization rate is calculated in two ways. Due to varying "opening hours", some sites have their entire acute load during the day shift, whereas the acute sites have patients coming in during evenings and nights. In order to compensate for this difference, Equation 7 calculates the rate of utilization during the dayshift. The factor F_a relates to the percentile of acute patients arriving during a shift.

$$Utilization_{Employees.day} = \frac{((\sum Planned + F_A * \sum Acute) + \sum Number\ of\ non\ attending\ patients) * operational\ time}{((No.employees\ per\ day\ \ shift * \frac{Hours}{Shift}) + Overtime)}$$

$Site\ 1: F_A = 1/2$
$Site\ 2: F_A = 1$
$Site\ 3: F_A = 2/3$
$Site\ 4: F_A = 1$

Equation 7 Employee utilization, work day

Employee perspective

The employee perspective is constituted by three main clusters, resembling those from the patient perspective, seeFigure 4.

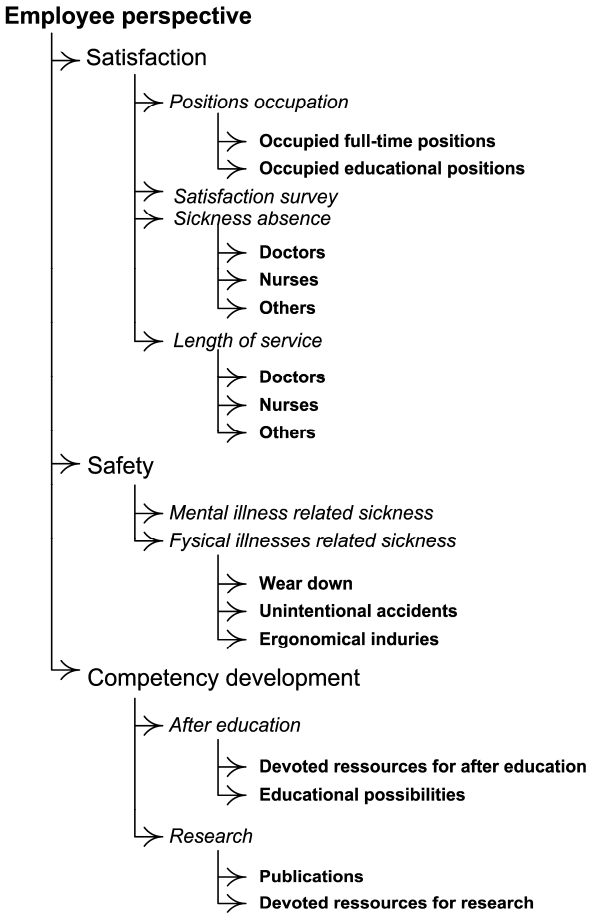

Figure 4: Indicator hierarchy for Employee perspective

The importance of a motivated and satisfied workforce is scientifically documented e.g.(Herzberg 2003), (Riedel & Lynch 2006), (Williams 2008). In scientific literature plus the interviews conducted during this study, the ability to fill open positions, satisfaction surveys, sick leave, and the average length of service, are noted as being valuable indicators in the assessment of employee satisfaction. In some scientific work (e.g. (Franco, Bennett, & Kanfer 2002) and (Dieleman et al. 2003)), salary is regarded as an important factor, which to some extent also applies to Danish healthcare. Healthcare professionals did not mention salary during the interviews or workshops conducted; only when the issue was addressed directly. Wages in the Danish public health care system are set by joint agreements between labour unions and the government, thus limiting the hospitals' ability to negotiate raises. Salary therefore, does not appear as an element in the satisfaction cluster.

The interview sessions revealed that employees are concerned about their workplace, and generally have a desire to influence and develop their professional positions as well as themselves as individuals. Competency development is an issue of great importance for Danish healthcare professionals. Hospitals that prioritize research and development represent more attractive workplaces than those who do not have this priority. Because of its distinctive role in Danish health care, competency development is given a cluster of its own rather than being a subset to "satisfaction". Due to the high rank assigned to competency development, this cluster will have a more significant impact on the outcome of the employee perspective.

The benchmarking procedure

Because indicators are not individually comparative, a benchmarking procedure that can compensate for this lack needs to be applied. In order to compare indicators, benchmarking is performed in three steps by an index of averages (see Table 1);

1. For each indicator an average for all involved locations, specific results are calculated. (e.g. patient satisfaction = 81%)
2. Based on this average, a location specific index is calculated. This index is calculated as a location specific result compared to average (e.g. Location 1 = 1,1)
3. To present the aggregated result for each location, a weighted average of the indexes is calculated (e.g. aggregated result Location 1 = 0,95).

Perspective	Indicator	Loc. 1	Loc. 2	Loc. 3	Average	Formula	Index Loc 1
Patient	*Patient satisfaction*	90%	84%	69%	81%	$Index\ 1 = \frac{90}{81}$	1,1
Operations Management	*Capacity*	0,3	0,5	0,7	0,5		0,6
Employee	*Length of service*	5 years	2 years	6 years	4,3 years		1,16
Aggregated result						$\frac{1,1 + 0,6 + 1,16}{3}$	0,95

Table 1: Benchmark procedure

This procedure is carried out for each of the clusters described in the previous sections, causing aggregated results to appear in several levels of the hierarchy. The multi level results enable the tracking of performance back through the calculations.

By using averages we get a baseline for benchmarking. By adapting the use of averages for benchmarking we have sidestepped the task of having to set target values. The use of target values is just as useful for benchmarking purposes. By transforming all performance measures into average dependent indexes, comparing otherwise incomparable measures becomes possible.

Benchmarking result

By applying the described model with data input from the radiology department, benchmarking of four sites becomes possible, see Figure 5.

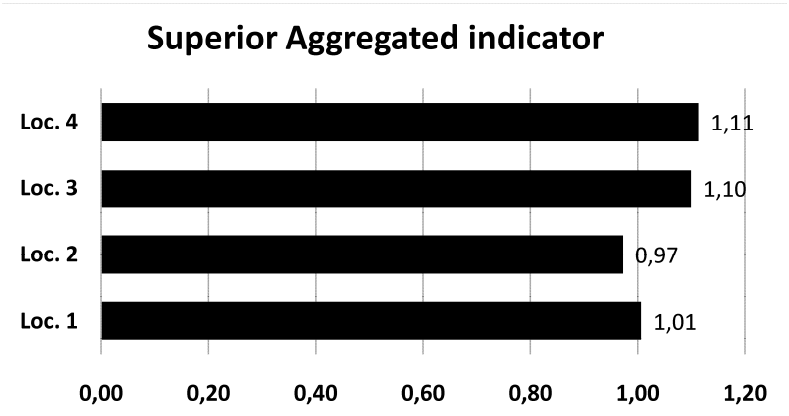

Figure 5 Benchmark result

The overall result in Figure 5 clearly shows that site 3 and 4 perform better than site 1 and 2. By looking at the outcomes of the three perspectives (patients, operations management, employees), the underlying results composing the Superior Aggregated Indicator (SAI) become evident.

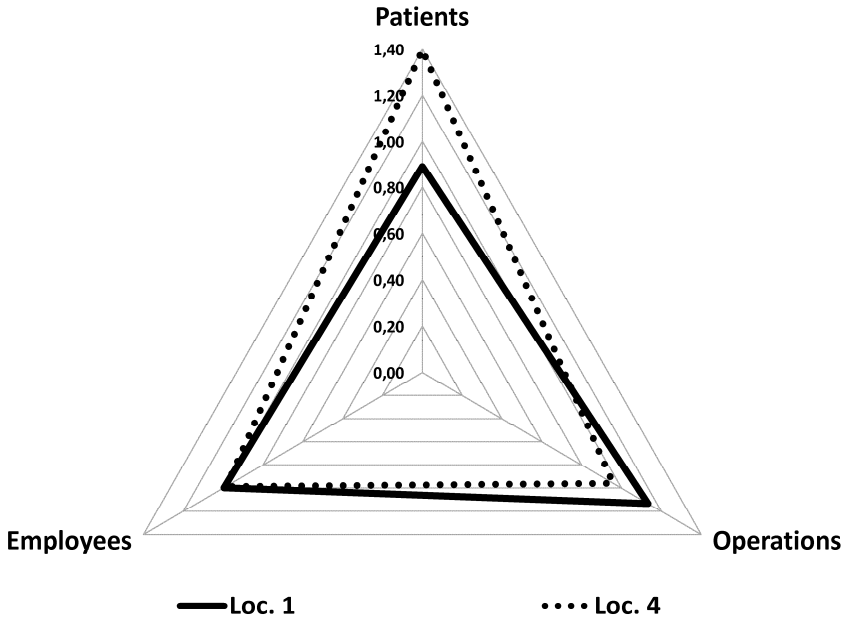

Figure 6 Benchmark result, two sites

Assessing the result for each perspective at sites 1 and 4, it is evident that the sites do not perform equally well in all areas. Site 1 appears to be a "production" unit with an emphasis on examination related processes, whereas site 4 seems to focus more on patient care. The area of the triangle represents SAI; the bigger the area, the better the overall performance. The reason that the productive site (1) comes out with a low SAI, is due to a low outcome in the patient perspective. Movement further down below the three perspectives, will present more detailed information about different outcome levels in the hierarchy, ultimately providing a more detailed description of the SAI.

Validity test

An anonymised presentation of the results, constituted the validity test of this model. During an interview three leading staff members at department level, were asked to assign site names to the results. Because the three leading staff members were able to connect results and known site roles, recognition proved a useful criterion in testing the validity of the model. Recognition proved that the qualitative construction of the hierarchies was in concordance with a perceived reality, because the quantitative results reflect the staff members' perception of performance on the four sites. The notion of a "perceived reality" is important to emphasize here, because there are no absolute values for good or bad performance.

Discussions and Conclusions

The objective of this project was to measure performance at the department level of a hospital, but the tool developed in this study has shown applicability far beyond the case it was developed from. The model acts as a common denominator, by including all aspects

of performance relevant to healthcare managers, at the same time enabling them to go several layers down and examine single sets of indicators related to different stakeholders. In addition to this, superior indicators can be benchmarked providing the option of comparing performance across similar departments. The objective has been to develop an internal information tool for managers at a hospital. Present the model is not suited for governmental surveillance, because the indicator hierarchy is constructed with internal focus. Future work will include research in benchmarking non similar departments on both national and international levels. In this context the model can also serve as a strategic tool.

The performance view put forth in this article is more versatile than a single stakeholder view. By making performance a function of several stakeholder perspectives, the complexity of managing modern healthcare departments is accentuated in the model. Even though the indicator hierarchy emphasizes this complexity, the understanding and viewing of performance is made simple through the aggregation of indicator outcomes. In the future such a simple expression would still stand true even though the number of indicators used to express health care performance can be expected to increase. The use of the model as a measuring tool for performance in healthcare is thus further legitimized by it not being dependent on the number of indicators used, but it is of course dependent on the placing of indicators in relation to the three stakeholder groups. So if necessary, indicators thought to be useful for depicting performance, can easily be incorporated as clusters in the lower levels of the hierarchy.

Having designed a generic model, using it as an external benchmarking tool by comparing performance across departments and hospitals, begs the question. External benchmarking though, poses several challenges such as differences in culture, jurisdiction, organisational structure, etc. In addition, the model in its present structure is not designed for benchmarking only one specific type of department. Because of the emphasis on operations management, the model is considered particularly suitable for use at departments focusing on production, e.g. radiology or orthopaedic surgery. The model needs further adjustment in order to portray performance in departments emphasizing patient care, e.g. oncology or geriatrics. The operations management perspective constitutes the problem, because it includes aspects of productivity that are not directly transferable to care units. Because the present model is build upon a case study, any customized adjustments in the indicator hierarchy would require thoroughly identification of new or revised indicators and datasets.

Although the model stands out as a benchmarking tool, it has an obvious potential as an in-house decision support system. By 'benchmarking' a department against its past performance, the model shows pro- and regress in relation to all stakeholders. A department manager can then easily detect performance problems within his/her areas of responsibility and thus prioritize and allocate resources where and when necessary. If the model is used as a decision support system, the weighting of the indicators takes on particular importance, because weighting becomes a strategic tool for targeting performance in accordance with the organisation's objectives. At the hospital level different weighting profiles can be created according to the purpose of each department, such that a for example a radiological department and orthopaedic surgery do not have the same weighting profiles. Because all quantitative data are an integrated part of measuring performance in the Danish health care sector, and therefore known and accessible, the model can easily be implemented.

The work here has proven it possible to present a benchmarking model, which combines and relates measurements from different stakeholders in one structure, and represents performance in a holistic manner. In addition the model has proven valuable as both decision support system and strategic tool.

References

Dieleman, M., Cuong, P., Anh, L., & Martineau, T. 2003, "Identifying factors for job motivation of rural health workers in North Viet Nam", *Human Resources for Health*, vol. 1, no. 1, p. 10.

Doherty, L. Culture of reporting errors linked to higher satisfaction for patients. Nursing Standard 22[51], 5. 2008.
Ref Type: Magazine Article

Eagle, C. J. & Davies, J. M. 1993, "Current models of "quality": An introduction for anaesthetists", *Canadian Journal of Anaesthesia*, vol. 40, no. 9, pp. 851-862.

Eilers, G. M. 2004, "Improving Patient Satisfaction With Waiting Time", *Journal of American College Health*, vol. 53, no. 1, p. 41.

Franco, L. M., Bennett, S., & Kanfer, R. 2002, "Health sector reform and public sector health worker motivation: a conceptual framework", *Social Science and Medicine*, vol. 54, no. 8, pp. 1255-1266.

Geraedts, M., Schwartze, D., & Molzahn, T. 2007, "Hospital quality reports in Germany: patient and physician opinion of the reported quality indicators", *BMC Health Services Research*, vol. 7, no. 1, p. 157.

Glance, L., Osler, T., Mukamel, D., & Dick, A. 2008, "Impact of the Present-on-Admission Indicator on Hospital Quality Measurement: Experience With the Agency for Healthcare Research and Quality (AHRQ) Inpatient Quality Indicators", *Medical Care*, vol. 46, no. 2, p. 112.

Herzberg, F. 2003, "One More Time: How Do You Motivate Employees?", *Harvard Business Review*, vol. 81, no. 1, p. 86.

Kreiner, K. & Mouritsen, J. 2006, "The analytical interview," in *The Art of Science*, Insight@CBS, pp. 153-176.

Lafond, A. K., Brown, L., & Macintyre, K. 2002, "Mapping capacity in the health sector: a conceptual framework", *International Journal of Health Planning and Management*, vol. 17, no. 1, p. 3.

Langabeer, J. 2008, *Health care Operations Management*, 1 edn, Jones and Bartlett Pulishers, Sudbury.

Loeb, J. M. 2004, "The current state of performance measurement in health care", *International Journal for Quality in Health Care*, vol. 16, no. 1-supplement, p. i5-i9.

Meyer, J., Pope, C., & Mays, N. 2000, "Using qualitative methods in health related action research", *British Medical Journal*, vol. 320, no. 7228, p. 178.

Minkman, M., Ahaus, K., & Huijsman, R. 2007, "Performance improvement based on integrated quality management models: what evidence do we have? A systematic literature review", *International Journal for Quality in Health Care*, vol. 19, no. 2, pp. 90-104.

Mohammadi, S. M., Mohammadi, S. F., & Hedges, J. R. 2007, "Conceptualizing a Quality Plan for Healthcare", *Health Care Analysis*, vol. 15, no. 4, pp. 337-361.

Morgan, D. & Morgan, R. 2009, *Single-Case Research Methods for the behavioral and health sciences*, 1 edn, SAGE, Thousand Oaks.

Nakajima, S. 1986, "TPM - CHALLENGE TO THE IMPROVEMENT OF PRODUCTIVITY BY SMALL GROUP ACTIVITIES", *Maintenance Management International*, vol. 6, no. 2, pp. 73-83.

Neely, A. 1999, "The performance measurement revolution: why now and what next?", *International Journal of Operations & Production Management*, vol. 19, no. 2, pp. 205-228.

Ondategui-Parra, S., Bhagwat, J. G., Gill, I. E., Nathanson, E., Seltzer, S., & Ros, P. R. 2004, "Essential practice performance measurement", *Journal of the American College of Radiology*, vol. 1, no. 8, pp. 559-566.

Pettersen, I. J. & Nyland, K. 2006, "Management and control of public hospitals-the use of performance measures in Norwegian hospitals. A case-study;", *International Journal of Health Planning and Management*, vol. 21, no. 2, p. 133.

Purbey, S., Mukherjee, K., & Bhar, C. 2007, "Performance measurement system for healthcare processes", *International Journal of Productivity and Performance Management*, vol. 56, no. 3, pp. 241-251.

Riedel, J. E. & Lynch, W. D. 2006, "Estimating the Total Costs of Poor Employee Health", *Benefits & Compensation Digest*, vol. 43, no. 12, p. 30.

Rochette, C. & Féniíes, P. "A Framework to link patient satisfaction with customer satisfaction", in *Euroma 2008*, Springer, pp. 1-9.

Roelen, C. A. M., Koopmans, P. C., Notenbomer, A., & Groothoff, J. W. 2008, "Job satisfaction and sickness absence: a questionnaire survey", *Occupational Medicine*, vol. 58, no. 8, pp. 567-571.

Rundall, T., Martelli, P., Arroyo, L., McCurdy, R., Graetz, I., Newwirth, E., Curtis, P., Schmittdiel, J., Gibson, M., & hsu, J. 2007, "The Informed Decisions Toolbox: Tools for Knowledge Transfer and Performance Improvement Practitioner Application -", *Journal of Healthcare Management*, vol. 52, no. 5, p. 325.

Tarantino, D. P. 2003, "Using The Balanced Scorecard as a Performance Management Tool", *Physician Executive*, vol. 29, no. 5, p. 69.

The Danish Institute for Quality and Accreditation in healthcare website. 2009. 10-12-2009.
Ref Type: Internet Communication

The Danish Patient Safety Database´s website. 2009. 10-12-2009.
Ref Type: Internet Communication

The National Board of Health´s website. 2009. 10-12-2009.
Ref Type: Internet Communication

the National Indicator Project´s website. 2009. 10-12-2009.
Ref Type: Internet Communication

The Unit of Patient-Perceived Quality's website. 2009. 10-12-2009.
Ref Type: Internet Communication

Wait, S. & Nolte, E. 2005, "Benchmarking health systems: trends, conceptual issues and future perspectives", *Benchmarking: An International Journal*, vol. 12, no. 5, pp. 436-448.

Williams, M. "Designing system resilience in NHS hospitals", in *Euroma 2008*, Springer.

P3: Operational benchmarking of Japanese and Danish hospitals

Submitted to: Proceedings of the 17th International Annual EurOMA conference, Managing Operations in Service Economies
Paperid: HLT21
ISBN: 978-972-99847-3-0
Submission date: 15th January, 2010
Acceptance date: 15th February, 2010
Publication date: 6th June, 2010
Type: Full conference paper published in proceedings
Status: Published

Operational benchmarking of Japanese and Danish hospitals

Authors

Andreas Traberg*
M.Sc., Ph.d. student
Technical University of Denmark
DTU Management Engineering
2800 Kgs. Lyngby
Denmark
Phone: (int+45) 4525 4405
E-Mail: atr@man.dtu.dk

Kenji Itoh
MA, GDBA (Management and Organization)
Tokyo Institute of Technology
Department of Industrial Engineering and Management
2-12-1 Oh-okayama Meguro-ku 152-8552 Tokyo
Japan

Peter Jacobsen
Associate Professor
Technical University of Denmark
DTU Management Engineering
2800 Kgs. Lyngby
Denmark

Andreas is the corresponding author

Abstract

This benchmarking model is designed as an integration of three organizational dimensions suited for the healthcare sector. The model incorporates posterior operational indicators, and evaluates upon aggregation of performance. The model is tested upon seven cases from Japan and Denmark. Japanese hospitals focus on productivity and reducing errors provide operational benefits, which primarily is achieved by high degree of overwork among staff. Danish hospitals on the contrary pay the price of productivity, with focus on pleasing caring needs of the patient and limiting overwork among employees.

Keywords: Benchmarking, Healthcare, Operational performance

Benchmarking healthcare

Healthcare performance assessment today often incorporate several dimensions such as safety, effectiveness, appropriateness, timeliness and responsiveness of services, along with measures of efficiency and equity (Wait & Nolte 2005). However this variety of dimensions have left the health care industry where no framework is unanimously accepted as tool for measuring and benchmarking the quality and performance of healthcare services (Ondategui-Parra et al. 2004). The disagreement originates from the fact that performance indicators are inherently controversial in healthcare, because they require an operational definition to be measured. As a result the use and development of

new methods for measuring and evaluating healthcare performance has been rapidly evolving, and debated, in the last decades (Folan & Browne 2005). Even though there are disagreement towards the usefulness of benchmarking in healthcare (MacVaugh 2006), most countries have developed comprehensive national strategies for collecting data and evaluating quality and performance within their own healthcare system. Benchmarking healthcare performance has become an intrinsic part within most developed health care systems, which enables politicians, national agencies and hospital managers to survey the delivered services (Purbey, Mukherjee, & Bhar 2007). But these benchmarks rely on national understanding of quality and performance, which often differs greatly from country to country.

Internationally, accreditation agencies like the HQS, ISQua and Joint Commission provide the possibility for hospitals to be certified according to a set of international standards for good quality. But these accreditation frameworks "just" approve hospital procedures they do not mutually compare hospitals. One agency which conducts healthcare comparison in an international context is the Organisation for Economic Co-operation and Development (OECD). The OECD performs a repetitive assessment where healthcare sectors as a whole, are mutually compared. The 2009 comparison announced that Japan uses fewer physicians and fewer nurses than the OECD average, but at the same time has the highest number of acute beds and significantly more diagnostic technology than OECD average (Official OECD webpage 2010). At the same time, Denmark has an average number of doctors and diagnostic equipment but well above average on nurses, and a significantly lower number of acute beds than OECD average. On the basis of this OECD benchmark, the Danish healthcare system seems considerably more inefficient than the Japanese, more people taking care of fewer beds and less equipment.

But these figures are strategic national indicators, which do not tell the story of operational performance at a Danish hospital or at a Japanese hospital. Cultural differences, different legal foundation and varying expectations to delivered care are factors which are not addressed in the OECD benchmark. If international benchmarking should be used in an operational context, and not just in a political context, the indicators need to resemble the versatile performance information managers use to make informed decisions at the hospitals (Dummer 2007;Liyanage & Egbu 2008). Scientists have proposed several frameworks which enable benchmarking across national borders. Many of these benchmarking frameworks relate to the use of financial indicators as a mean to assess whether a hospital has good or bad performance, e.g. (Chen et al. 2006; Evans 2004). But healthcare performance and quality implies much more than economic figures and assessment of medical errors, why versatile performance evaluation tools often are addressed. Griffith (2002) framework uses Balanced Scorecard to assess performance as a multitude of indicators (Griffith, Alexander, & Warden 2002). But the framework does not compile results, and only represent them as singular measures which represent an operational informative disadvantage. Swaminathan (2008) tries to aggregate healthcare performance outcome to present information on progress according to childhood immunization (Swaminathan, Chernew, & Scanlon 2008). But the narrow focus only portrays a fraction of performance for a department, which is limiting the potential as a mean to benchmark hospital performance. The Analytical Hierarchical Process (AHP) used by Dey (2008) is an attempt to provide insight into weaknesses within healthcare organizations (Dey, Hariharan, & Despic 2008). A comprehensive framework which enables benchmarking across countries, struggles with one particular problem namely subjectivity.

Motivation and method

Benchmarking traditionally struggles with trade offs like the ones described and for that reason this paper tries to propose a new way to evaluate hospital performance in two different healthcare sectors. By evaluating Danish and Japanese hospitals upon indicators which are used in managing operational performance in both sectors, country specific differences are accentuated. Not focusing on high level indicators, but emphasis on describing operational performance for departments. By aggregating performance for quantifiable indicators of large scope, the paper tries to compensate for some of the challenges in international benchmarking.

The benchmark is being developed in a comparative research project, where both researchers and clinicians from Denmark and Japan have been involved. The development of the benchmarking model has been performed as a multiple case-study in two consecutive phases. First a qualitative development phase and second a quantitative evaluation phase. The qualitative selection of suitable indicators is performed in close collaboration with clinicians which participated in interviews, and subsequent served as respondents to data questionnaires. Using quantitative data analysis to benchmark the hospitals upon the selected indicators, an insight to performance differences among the Japanese and Danish hospitals are provided.

The case hospitals

For this benchmark seven case departments are included, four Danish and three Japanese. The first Japanese hospital is public hospital belonging to a local municipality, the second and third hospitals are University hospitals, respectively belonging to a national and a private university. The Danish cases are located on four individual sites, belonging to the same public hospital. The Danish hospitals were merged at management level, but the four sites all acts as operational parts in the new hospital, see Table 1

Table 1: Case hospitals for comparison

Hospital	***Located***	***Case department***	***Employees***
Hospital A	Fukuoka, Japan	Dialysis	14
Hospital B	Ibaragi, Japan	Radiology	21
Hospital C	Tokyo, Japan	Dialysis	48
Hospital D	Jutland, Denmark	Radiology	45
Hospital E	Jutland, Denmark	Radiology	16
Hospital F	Jutland, Denmark	Radiology	44
Hospital G	Jutland, Denmark	Radiology	9

The case departments chosen for the benchmark are all "producing" department, thereby no admission or intensive wards. This is a deliberate choice, because wards have different tasks than producing department, which might bias the result.

Two different healthcare systems

Because this paper deals with benchmarking in Japan and in Denmark, two very different healthcare systems, this section would shortly introduce the fundamentals in both.

The Japanese healthcare

After 1945 the governments of the Allied Forces reconstructed the political and social structure of Japan, and new laws and actions were implemented, having a great and positive consequences for health (Suzuki, Gibbs, & Fujisaki 2008). The Japanese social security system has continued to grow, especially the last decade. In fact, social security

has created many job opportunities as demonstrated by the fact that the number of workers in the healthcare and welfare sector almost doubled from about 1.7 million in 2000 to about 3.3 million in 2005 (Kousei Roudou Hakusho 2009). An important characteristic of the Japanese system is that many of the hospitals are privately owned, in particular the smaller hospitals. (Chen, Yamauchi, Kato, Nishimura, & Ito 2006). Japan has about the lowest per capita health care costs is 2.51 (Intl $, 2006;) among the developed nations of the world and its population is one of the healthiest. That is largely due to lifestyle factors, such as low rates of obesity and violence.

Japan would in the future encounter with some severe demographical changes. The population began to decline in 2005, and in the future Japan will face further aging of society with fewer children leading to further population decrease (National Institute of Population and Social Security Research). It is predicted that by 2030, Japan's seniority rate will rise to 31.8%, indicating that one out of three Japanese will be a senior citizen (aged 65 or older), and that the figure will top the 40% in 2055 (Tatara & Okamoto 2009). There is a concern that changes in demographic structure may lead to the decline in the labour force and affect the sustainable development of the Japanese economy and thereby healthcare sector. Another concern is doctor shortages, although the number of doctors increased there still is a severe shortage of doctors in many areas, particularly in a country side.

Danish healthcare

Denmark's healthcare sector is primarily public, covering 98% of all admission beds. The private sector dealing primarily with small clinical procedures, pharmaceutical and dental care (Strandberg-Larsen et al. 2007). Public hospitals in Denmark are part of the municipal administrative structure, which consists of five regions and 98 municipalities. It is the primary task of the regions to manage the health care system, that is the hospitals, psychiatric units and health care insurance system. The total expenditure on health per capita is 3.34 (Intl $, 2006), which compared to Japans 2.51 is relatively high (Official WHO webpage 2010).

The Danish social security system faces like the Japanese some critical challenges over the years to come. As Japan, Denmark faces aging population, which demands nursing and treatment. Due to change in life styles and work characteristics, more and more Danes are struck with life-style diseases like obesity and diabetes. Combined with recruitment difficulties, Danish healthcare are in need for more human resources in the sector, or future comprehensive structural changes. Many Danish hospitals struggle with overcrowding, which constitutes an emerging problem.

The benchmarking procedure

The aim of this benchmark is to present performance as one superior aggregated index which acts as a common denominator for all included indicators in the comparison. This approach has previously been used by Nakajima (1986) which introduced the use of aggregated indicators into an Overall-Equipment-Efficiency indicator (OEE) (Nakajima 1986). The OEE measure included *Availability*, *Performance*, and *Quality* combined into one single measure. In benchmarking context, the aggregated indicator approach has been used by De Toni (2008) to evaluate research institutions (De Toni et al. 2008). The merging of indicators provides indexes of performance, which does not relate to one single measure, but is a representation of all included lower level indicators, see Figure 1.

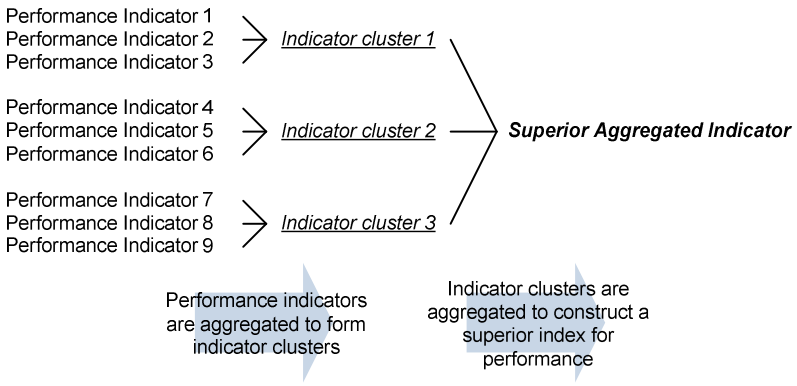

Figure 1: Guiding principle

To be able to compare performance outcome of different performance indicators into one key performance value, normalization of the data is necessary. Normalization serves the purpose of bringing the indicators into a dimensionless quantity, thereby making the indicators comparable (Stapenhurst 2009). As data normalization method the Standard score is chosen, or more commonly referred to as the z-score, see Equation 1. The z-score corresponds to a data point in a normal distribution. It converts all indicators to a common scale and thereby making them comparable regardless of data foundation.

$$z - score = \frac{(Data\ point_n - Mean)}{Standard\ Deviation} \tag{1}$$

An advantage of using the z-score is that it encourages mean scores over high variation, which obey with the primary objective for healthcare facilities of complying with standards for acceptable performance (Lim, Tang, & Jackson 1999). It is more desirable for hospitals to be performing acceptable on all indicators, than perfect in some and poor in others, thereby reducing performance inconsistency in delivered care.

The benchmarking procedure is performed in three consecutive steps, which starts by calculating the mean value and the standard deviation from the collected data. Second, each indicator in the benchmark is transformed into a z-score representing each data value in relation to the standard deviation. Finally, all z-scores are aggregated through summarization. This aggregated result represents the performance level by each particular facility, in a particular cluster of indicators, see Figure 2.

Indicator cluster

	Data-point					Z-score		
	Hosp. 1	Hosp. 2	Hosp. 3	Mean	St. Dev	HF 1	HF 2	HF. 3
Indicator 1	v1	..	..	m1	St.d 1	z1	..	..
Indicator 2	v2	..	..	m2	St.d 2	z2	..	..
Indicator ...	v..	..	..	m..	St.d ..	z..	..	..
Indicator n	vn	..	..	mn	St.d n	zn	..	..
Benchmark result for Indicator cluster No. X						$\sum_{i=1}^{n} zi$	..	..

Step 1 · Step 2 · Step 3

Figure 2: Benchmarking procedure

The procedure repeats itself whenever indicators or clusters are aggregated to a superior level. As a result of this approach the number of included indicators is unimportant, because each indicator equally acts as contributor to a given cluster. Mathematically indicators in "large" clusters have minor weight, than the indicators in

"smaller" clusters. Because all healthcare facilities are evaluated upon a unified hierarchy of indicators, weighting of individual indicators is considered negligible.

Selection of indicators

The next step is the identification of indicators to be implemented in the benchmark. The overall structural frame was decided beforehand, where *Patients*, *Employees* and *Operations* constituted the backbone in the benchmark. The reason why these were chosen is that they constitute the main stakeholders in healthcare. There were set no limitations for how many indicators could be in each dimension. Furthermore it should be mentioned that the number of indicators in each dimension is not an expression for the importance of that cluster, but a result of the availability of comparable data.

To make a fair selection of indicators, clinicians from both Japan and Denmark have been interviewed according to which indicators were, in their opinion, meaningful and useful in daily operations. Based on these interviews, questionnaires were used to collect quantitative data. Based on the interviews it was possible to construct a hierarchy of indicators which represented a common understanding of important performance indicators for management in the seven hospitals, see *Figure 3*.

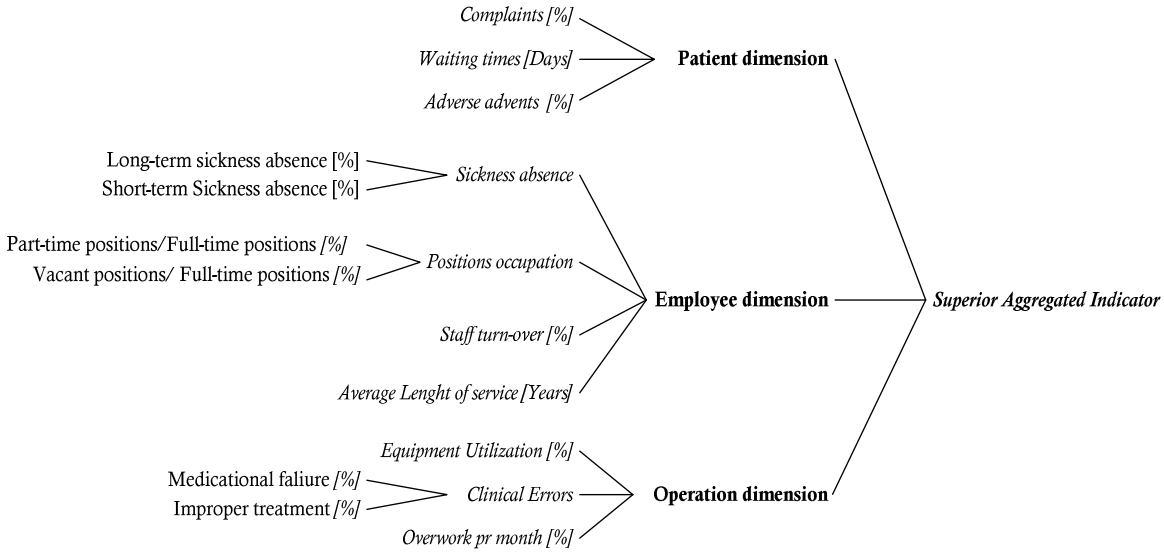

Figure 3: Indicator hierarchy

The reason why overwork is placed in the operations dimension and not the employee dimension is to adjust for the amount of resources put into heightening utilization and decreasing clinical errors. All performance indicators are made generic in the sense that they do not particularly address radiology or dialysis they might as well be applicable for any other producing departments, e.g. orthopaedics, cardiology, etc. Because the indicators in the operations dimension are heavily influenced by "production" indicators the presented hierarchy are assumed maladjusted for admission departments or intensive wards.

Benchmarking performance

The data foundation for the benchmark was collected during spring 2010, and represents the year of 2009. Aggregating the outcome of all three dimensions, the Japanese hospitals have an overall better result than the Danish, see Figure 4. The Japanese hospitals result is all positive indexes, with only one positive Danish hospital.

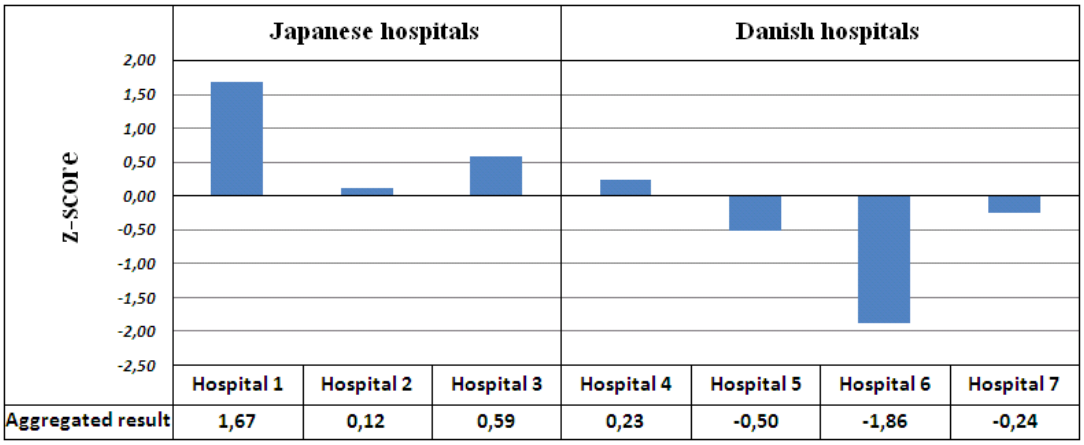

Figure 4: Aggregated benchmark result

To conduct the in-depth analysis of the foundation for the aggregated the result, each of the three dimensions are described in detail in Table 2. All the collected data are anonymous, so they would only appear as z-values in this presentation.

Table 2: Detailed benchmark result

	***Japanese hospitals* (z-values)**			***Danish hospitals* (z-values)**			
	Hospital 1	Hospital 2	Hospital 3	Hospital 4	Hospital 5	Hospital 6	Hospital 7
Complaints	0,47	-0,60	-2,05	0,46	0,47	0,63	0,63
Waiting times	0,79	0,79	0,79	-0,37	-0,06	-2,00	0,06
Adverse advents	1,37	-1,15	0,36	-0,14	0,36	-1,66	0,87
Patients dimension	***0,88***	***-0,32***	***-0,30***	***-0,02***	***0,26***	***-1,01***	***0,52***
Sickness absence	0,74	0,74	0,68	0,00	-0,37	-1,40	-0,40
Position occupation	0,34	0,01	0,44	-0,39	0,30	-0,45	-0,25
Staff turnover	-0,13	0,06	-0,12	0,24	-0,13	-0,25	0,33
Length of service	1,34	0,27	0,70	-0,58	-0,58	-0,58	-0,58
Employees dimension	***0,57***	***0,27***	***0,43***	***-0,18***	***-0,19***	***-0,67***	***-0,22***
Equipment utilization	0,62	1,08	0,42	1,22	-1,11	-0,91	-1,31
Clinical errors	1,02	1,11	1,01	-0,96	-0,72	-1,06	-0,39
Overwork	-0,47	-2,00	-0,52	0,77	0,64	0,83	0,76
Operations dimension	***0,39***	***0,06***	***0,30***	***0,34***	***-0,40***	***-0,38***	***-0,31***

It becomes apparent that the Hospital 1 excels over all other hospitals in the benchmark, having positive performance indexes for 8 out of 10 indicators. Hospital 4 is the best performing Danish hospital, especially because of a very fine operational dimension. As the only hospital in the benchmark, Hospital 4 has managed to keep a high utilization of equipment, with a minimum of overwork. The Japanese hospitals all have high equipment utilization levels, but this is mainly achieved by a high degree of overwork. Another distinct difference between Danish and Japanese hospitals seems to be the employee dimension, where Danish hospitals have lower performance than the Japanese. There are two key explanations for this, sickness absence and length of service. Due to increasing pressure on the Danish healthcare, sickness absence has become a significant factor in several hospitals, plus shifting jobs is considered an career vice

advantage in Denmark. Japanese hospitals excel within waiting times, which constitutes a major problem in Danish healthcare within most specialties.

Discussion

The results of this study accentuate some of the performance differences between Japanese and Danish healthcare. Both countries have what most professionals would acknowledge as high performing healthcare sectors. Though there are big differences in which areas the hospitals have high performance. As suggested in the OECD benchmark Japanese hospitals are very efficient, and this combined with a relative low rate of clinical errors. This benchmark highlight two main explanations for the high efficiency, namely overwork and sickness absence. Both of these factors are contributing to more available human resources for production, enhancing the probability of high utilization. Efficiency and few errors resemble what is known from Japanese industrial production, which has been known for the same virtues.

Danish healthcare has for many years made efforts to satisfy the caring needs of the patient during admission. There are signs that the Danish efforts in the nursing patients have been on the cost of productivity. Because this benchmark does not include satisfaction surveys, due to lacking Japanese data, it is impossible to conclude whether the efforts have paid off in relation to patients. The only indication of patient satisfaction is a low degree of patient complaints at the Danish hospitals. Implementing patient satisfaction in benchmarking across national borders would be a very difficult, because cultural differences presumably would heavily influence the answers. There are indications that Japanese healthcare is striving towards more patient oriented care in the future. Recently, more attention has been paid to patient's rights such as informed consent and shared decision-making (Tatara & Okamoto 2009). This change in focus suggests that Danish healthcare is ahead in terms of patient care, though it is impossible to validate upon the presented results in this benchmark.

The employee dimension is interesting because it stresses that Japanese hospitals outperforms Danish hospitals. But it is important to recognize that the indicators used in the benchmark do not tell about satisfaction with work itself. It only relates to the how well the department is running from management point of view. The ability to hold on to experienced employees, having limited turnover and low sickness absence is considered to be organizational benefits. It has to be mentioned though, that turnover for Japanese doctors is special. Japanese doctors virtually belong to a professor in the medical university. In most cases the professor has authority to send his former student to another hospital every second or third year, independent from their will and wish. For that reason doctor's turnover is rather high in Japan. Turnover rate is therefore not a reliable indicator of doctor's satisfaction in Japan. In Denmark though, this indicator is, among others, a reasonable sign of satisfaction, but because of the differences it is unsuitable as a benchmark for employee satisfaction. The use of part-time workers is also an important factor in the employee dimension, because the part-time/full-time ratio should be as low as possible. The reason is that many part-time workers and vacant positions are considered a flow problem for organizations. As the results shows, Danish hospitals struggles more with keeping a stable workforce than the Japanese hospitals. Related to the employee dimension, overwork constitutes a significant difference between the two countries. Because of shortage of healthcare staff, particularly doctors in Japan, a small number of doctors have to take care of many patients. Therefore they must work longer hours, which is indicated in the benchmark. The amount of working hours in Japan is very seldom in Denmark, due to labour agreement and legislation. Because the overwork indicator is placed in the operations dimension in this benchmark, the outcome is not affecting the employee dimension. Therefore the employee dimension would from

management point of view provide reasonable performance information, but not as a token of good working conditions. It has not during the research period been possible to gather data on work-life satisfaction from Japanese hospitals.

Implementing more indicators into the proposed framework would further clarify some differences between hospitals in different healthcare sectors. The development of this benchmark has shown that Danish managers use significantly more performance information in the management of department than the Japanese managers does. Limited data availability and lack of uniformity of data across countries, limits the scope of most benchmarking initiatives. This has of course also affected this benchmark in the development of the hierarchy, where uniform data were shortage. In particular, Japanese hospitals do not have enough data related to indicators which Danish hospitals hold, and therefore hospital managers or department leaders must in some cases estimate them. This can lead to low reliability of data. A solution towards securing higher reliability in the future work is to question hospital managers about confidence or an estimate of error towards the data. This could be valuable information to the data which is not regular recorded at the hospital. Based on the estimated value and confidence, it would be possible to calculate a value or degree of each indicator with allowance which resembles a confidence interval or limit. Comparison between two countries' hospitals is thereby possible in terms of the indicator value for Danish hospitals, and upper and lower confidence limits as well as the mean value of the indicator for Japanese hospitals.

Another important discussion is regarding weighting of indicators within the framework. The importance of the individual indicators is different between the countries. Because of different healthcare and social systems for instance, sickness absence is not so meaningful in Japanese hospitals while it is a good indicator for employee satisfaction in Denmark. Allocation of weight profiles for the indicator hierarchies may be a way of enhancing the reliability of the model. If data are available from many hospitals, it is possible to apply the Factor Analysis or the Principal Component Analysis (PCA) separately to Danish and Japanese data. Thereby estimate factor loadings to all the indicators for each of dimensions, thereby obtaining two sets of weights: Danish criteria and Japanese criteria. Alternatively the AHP could be applied. These techniques are easy approachable if there are sufficient data available.

Conclusion

The paper discovers some of the differences between Japanese and Danish healthcare. The results point to Japanese hospitals as having better aggregated performance than the Danish hospitals. High equipment utilization and few clinical errors are to some extend achieved by much overwork among Japanese healthcare staff. Danish hospitals pay the price of productivity by focus on pleasing the caring needs of the patient and limiting working hours for employees. The structure of the benchmark is regarded suitable for evaluating operational healthcare performance, because of the possibility of calculating performance in few key indicators, without loosing the strength of deep detailed measures. Though there are seen some difficulties with international benchmarking, primarily caused by cultural and structural differences and availability of data.

References

Chen, X. y., Yamauchi, K., Kato, K., Nishimura, A., & Ito, K. 2006, "Using the balanced scorecard to measure Chinese and Japanese hospital performance", *International Journal of Health Care Quality Assurance*, vol. 19, no. 4, pp. 339-350.

De Toni, A. F., Montagner, M., Tonchia, S., & Zanchetta, A. "Benchmarking Performances of Research institutions: A Measurement Model", in *EurOMA International Conference 2008*.

Dey, P. K., Hariharan, S., & Despic, O. 2008, "Managing healthcare performance in analytical framework", *Benchmarking: An International Journal*, vol. 15, no. 4, pp. 444-468.

Dummer, J. 2007, "Health care performance and accountability", *International Journal of Health Care Quality Assurance*, vol. 20, no. 1, pp. 34-39.
Evans, J. R. 2004, "An exploratory study of performance measurement systems and relationships with performance results", *Journal of Operations Management*, vol. 22, no. 3, pp. 219-232.
Folan, P. & Browne, J. 2005, "A review of performance measurement: Towards performance management", *Computers in Industry*, vol. 56, no. 7, pp. 663-680.
Griffith, J. R., Alexander, J. A., & Warden, G. L. 2002, "Measuring comparative hospital performance / Practitioner response", *Journal of Healthcare Management*, vol. 47, no. 1, pp. 41-57.
Kousei Roudou Hakusho 2009, *Annual Health, Labour and Welfare Report 2007-2008*, Japanese Ministry of Health,Labour and Welfare.
Lim, P. C., Tang, N. K. H., & Jackson, P. M. 1999, "An innovative framework for health care performance measurement", *Managing Service Quality*, vol. 9, no. 6, pp. 423-433.
Liyanage, C. & Egbu, C. 2008, "A performance management framework for healthcare facilities management", *Journal of Facilities Management*, vol. 6, no. 1, pp. 23-36.
MacVaugh, T. 2006, "Benchmarking: Should the art or science be benched?", *Healthcare Purchasing News*, vol. 30, no. 4, p. 13.
Nakajima, S. 1986, "TPM - Challenge to the improvement of productivity by small group activities", *Maintenance Management International*, vol. 6, no. 2, pp. 73-83.
Official OECD webpage, J. Official OECD webpage, Japan. 2010. 10-4-2010.
Ref Type: Internet Communication
Official WHO webpage, D. Official WHO webpage, Denmark. 2010. 7-4-2010.
Ref Type: Internet Communication
Ondategui-Parra, S., Bhagwat, J. G., Gill, I. E., Nathanson, E., Seltzer, S., & Ros, P. R. 2004, "Essential practice performance measurement", *Journal of the American College of Radiology*, vol. 1, no. 8, pp. 559-566.
Purbey, S., Mukherjee, K., & Bhar, C. 2007, "Performance measurement system for healthcare processes", *International Journal of Productivity and Performance Management*, vol. 56, no. 3, pp. 241-251.
Stapenhurst, T. 2009, "Normalization: How to Compare Apples with Pears," in *The Benchmarking Book: A How-to-Guide to Best Practice for Managers and Practitioners*, 1 edn, Oxford, pp. 112-137.
Strandberg-Larsen, M., Nielsen, M. B., Vallgårda, S., Krasnik, A., Vrangbæk, K., & Mossialos, E. 2007, *Japan: Health system review. Health Systems in Transition*, European Observatory on Health Systems and Policies, 9(6).
Suzuki, Y., Gibbs, T., & Fujisaki, K. 2008, "Medical education in Japan: A challenge to the healthcare system", *Medical Teacher*, vol. 30, no. 9-10, p. 846.
Swaminathan, S., Chernew, M., & Scanlon, D. P. 2008, "Persistence of HMO Performance Measures", *Health Services Research*, vol. 43, no. 6, pp. 2033-2049.
Tatara, K. & Okamoto, E. 2009, *Japan: Health system review. Health Systems in transition*, The European Observatory on Health Systems and Policies, 11(5).
Wait, S. & Nolte, E. 2005, "Benchmarking health systems: trends, conceptual issues and future perspectives", *Benchmarking: An International Journal*, vol. 12, no. 5, pp. 436-448.

P4: Rethinking Performance evaluation in Healthcare

Submitted to: Journal of Health Organization and Management
ISBN:
Submission date: 13 January, 2011
Acceptance date: 23th March, 2011
Publication date:
Type: Full paper publication
Status: Accepted, undergoing 2nd minor revision

Rethinking Performance evaluation in Health Care

Authors

Andreas Traberg*
M.Sc., Ph.d. student
Technical University of Denmark
DTU Management Engineering
2800 Kgs. Lyngby
Denmark
Phone: (int+45) 4525 4405
E-Mail: atr@man.dtu.dk

Peter Jacobsen
Associate Professor
Technical University of Denmark
DTU Management Engineering
E-Mail: pj@man.dtu.dk

Nadia Monique Duthiers
MA, GDBA (Management and Organization)
Hospital of Southern Jutland
Radiology department
E-Mail: nadia.duthiers@shs.regionsyddanmark.dk

Abstract

Purpose: The objective is to develop a framework for healthcare performance evaluation, enabling decision-makers to identify areas indicative of corrective actions. The framework should provide information on strategic pro-/regress in an operational context, justifying the need for organizational adjustments.

Design/methodology/approach: The study adopts qualitative methods in constructing the framework, subsequently implementing the framework in a Danish MRI unit. Workshops and interviews are used for the qualitative construction phase, and two internal and five external databases are used for a quantitative analysis.

Findings: By aggregating performance outcomes, collective measures of performance are achieved. This enables easy and intuitive identification of areas not strategically aligned. In general the framework proved helpful to operational decision-makers struggling with extensive amounts of performance information.

Practical implications: The framework's strength lies in the identification of performance problems prior to decision making. The quality of decisions is bound to the individual decision maker. The framework only functions to support these decisions.

Scientific implications: The implementation of the framework on a single case in a public and highly political environment restricts the generalizing potential. The authors acknowledge that there may be more suitable approaches in organizations with different settings.

Originality/value: The study challenges the traditional use of performance reporting by combining strategic weight assignment and performance aggregation in hierarchies. This way, the framework accentuates performance as a function of strategic pro- or regress, which assists decision-makers in placing operational effort in pursuit strategic alignment.

Keywords:
Performance measurement, Decision support, Health Services Sector, Holistic performance, Strategic alignment.

Performance evaluation in healthcare

Managing modern healthcare is becoming increasingly complicated as institutions evolve into integrated health systems comprised of hospitals, outpatient clinics and surgery centre's, nursing homes, and home health services (Curtright, Stolp-Smith, & Edell 2000). Additionally, increasing demands for individualized high performing services, intensified patient inflow and technological innovations, all results in pressure on health expenditures (Strandberg-Larsen et al. 2007;World Health Organization 2008). This development has led to a growing need for more reliable performance evaluation tools to guide the increasingly complex decision-making processes (Swaminathan, Chernew, & Scanlon 2008). But performance of healthcare services are often difficult to quantify, and numerous methods are acknowledged as tools to assess performance and quality of healthcare services (Mohammadi, Mohammadi, & Hedges 2007). Accurate diagnosis and treatment are no longer enough, stakeholders need high performance in all facets of modern healthcare (Elleuch 2008). This signifies that healthcare organizations need to move beyond a narrow medical view and embrace a holistic approach to the concept of healthcare performance. As an attempt to provide holistic performance information, measurement systems consequently have become more wide ranging and their use more widespread (Cheng & Thompson 2006;Lega & Vendramini 2008). Nowadays there is consensus in defining performance in relation to explicit goals reflecting the values and requirements of various stakeholders (such as patients, professions, regulators, etc.). Therefore, performance evaluation have evolved into a multi-faceted concept including, patient load analysis (Mital 2010), work environment (Jones et al. 2009), patient satisfaction (Kutney-Lee, McHugh, & Sloane 2009), mortality rates (Barros 2003), surgical performance (Treasure et al. 2002), incentives structures (Buetow 2008) to name a few examples of the extensive work conducted within specific healthcare performance evaluation. Because of a common acceptance of the strength in both high-level and specific measures, many healthcare facilities are adopting both types within their internal evaluation procedures. Likewise national and international agencies are evaluating the performance of healthcare services to an until now unseen extend. Performance indicators, quality audits and accreditation standards are therefore gradually becoming fundamentals in the vocabulary of healthcare professionals around the world.

Most of these initiatives are well thought through, well documented and well executed, however collectively they poses one significant drawback. This vast selection of self-contained initiatives is limiting the overview for the individual decision-makers. Practitioners experience the cost of this intense focus on evaluation as a heavy administrative, and the feed-back are often overwhelming and confusing. Thus, in many cases the expansion of the administrative burden has not been providing more operational value for healthcare organization - just more work. Managers and a high percentage of operational employees are using large portions of their time on administrative tasks related to reporting on performance and quality initiatives. Contradictory to the initial objective, the expanding quantity of registrations, reports, standards, budgets, etc, has limited the organizations ability to make use of all the information at hand. Decision-makers are constantly faced with a vast selection of indicators which in some cases lead to administrative fatigue and information overload (Bovier & Perneger 2003). Few employees are able to understand and grasp all the information produced at modern healthcare facilities. Unfortunately, this result in several decisions is not based upon

quantitative data, even though the evidence is available. Instead upon more subjective assessments in risk of not being aligned with an organization's strategy (Ormrod 1993). This apprehension continuously affects healthcare facilities, which in these years feel forced into changing their organizational structure (Kocakülâh & Austill 2007). As a way to comprehend all information, more and more employees need to be responsible for sub-parts of the organizational decision-hierarchy. This segregation of tasks into smaller areas of responsibility seems as an obvious structural response (Evans & Weir 1995). But segregation is demanding considerably more from the managing processes within the organization (Walley, Silvester, & Mountford 2006). Decision-making is accordingly moving away from the operational levels of healthcare organization and into the strategic levels of the organization, thus prolonging the ability to make corrective adjustments and thereby delaying necessary changes.

This is a key concern because it can lead to a descending performance spiral (administrative tasks vs. operational productivity) where operational employees utilize their time on administrative task instead of value adding activities. Hence if performance information is to be used as proactive decision support without increasing the administrative burden, the representation of organizational performance needs to be changed. A more intuitive and holistic representation is needed, easing the identification of performance problems throughout the organization.

Motivation and methodology

Assuming that comprehendible performance information along with an intuitive representation is a necessity for modern healthcare organizations to reach their strategic objectives. Thus this paper focuses on the question of constructing holistic aggregated performance information, capable of portraying strategic change based on operational performance measures. The aggregated measures must provide information about current operational performance compared to past performance, in order to represent strategic pro-/regress. The framework needs to justify whether operations are in statistical control or not, according to the strategic objectives of an organization. The advancement of performance measurement systems is considered a key step towards improving the healthcare sectors capabilities. The motivation of this work has been to contribute to this advancement, by focusing on holistic performance measurement. Taking point of departure in strategic objectives and transform this into decision-support information for operational management.

The study is performed as a single case study (Morgan & Morgan 2009;Voss, Tsikriktsis, & Frohlich 2002), with close relations to the staff at a Danish radiology department. The empirical focus of this study is on the Magnetic Resonance Imaging (MRI) modality. The framework has been developed in two steps: A qualitative development phase and a quantitative test phase. The qualitative development of the framework was performed in collaboration with staff members who participated in workshops and interviews (Winter & Munn-Giddings 2001). The justification for using workshops is regarded a possibility for reflection at a higher level than by solely using interviews. As preparation for these workshops, interviews are used as a promoter of discussion topics. All internal process related data were collected from the Radiology Information System/Picture Archiving and Communication System (RIS/PACS), and Human Resources (HR) data were collected from the Hospital Information System (HIS). External data were collected from five federal units and government agencies; 1) The Unit of Patient-Perceived Quality's survey of patients' experiences in Danish hospitals (The Unit of Patient-Perceived Quality's website 2009). 2) The Danish Quality framework (The Danish Institute for Quality and

Accreditation in healthcare website 2010). 3) The National Indicator Project (the National Indicator Project´s website 2009). 4) Patient safety records created by the National Board of Health (The National Board of Health´s website 2009) and the 5) Danish Patient-Safety Database (The Danish Patient Safety Database´s website 2010). It is important to notice that all external data are available to the public, and validated by the federal units and governmental agencies issuing them. The data collection has striven towards basing the performance measures on already collected, validated and published data, with an eye to enhance credibility and validity of the outcomes.

Structural outline

In the construction of any decision support framework, there are two key issues that have to be properly addressed; which indicators to measure (Flapper, Fortuin, & Stoop 1996), and how to evaluate them (Dummer 2007). This particular framework is based on four successive steps, to deal with these two issues;

1) Selection and placement of indicators in hierarchies. The selection of suitable indicators is regarded as being of critical importance, because it establishes the organization's goals and priorities (Neely et al. 1994).
2) A structure for mutually weighing the indicators, assigning mathematical weight in accordance with strategic significance. If the weights are not strategically aligned, the usefulness of the information as decision support is assumed limited.
3) Normalization of outcomes. Normalization assigns indicators a dimensionless quantity; thereby making them comparable regardless of initial value
4) The aggregation procedure calculates an aggregated performance index. This procedure provides information on change according to past performance.

Combined these four steps constitutes the guiding structure which allows for interpretation of performance in relation to strategic objectives. The specifics in each of these steps are adapted from state-of-the-art performance measurement proposals, and fitted to the particular settings constituting modern healthcare.

Step I: Selection and placement of indicators in hierarchies

To fulfil the intention of presenting performance in few key measures characterising overall performance, indicators are structured in a hierarchy, see Figure 1. The aggregated performance index is thereby represented as a common denominator for all included indicators in the framework. The justification for adopting the concept of aggregated measures, is that it can incorporate a vast array of information and at the same time decrease complexity (Jollands, Lermit, & Patterson 2003). This approach has been previously used by (Nakajima 1986) who introduced the use of aggregated indicators in his Overall-Equipment-Efficiency indicator (OEE). The OEE included *Availability*, *Performance*, and *Quality* combined into a single measure. The aggregation of indicators provides an index of performance which is a representation of all included lower level indicators.

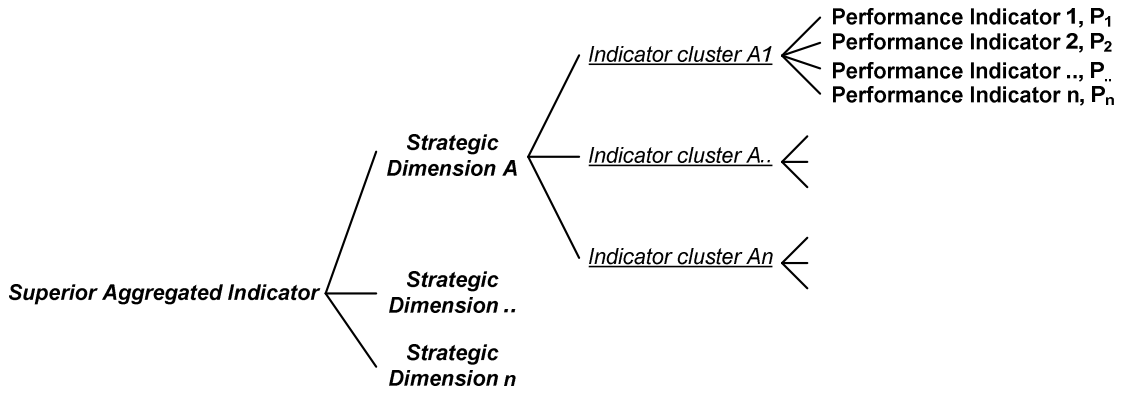

Figure 1: Structural principle of indicator hierarchy

The hierarchy consists of four levels: A superior aggregated indicator, strategic dimensions, indicator clusters, and performance indicators. The selection of suitable indicators for a given department is individual (Evans 2004), e.g. an intensive ward is likely to choose other indicators than radiology. Thus the decision-makers who is to use the framework needs to be an integrated part in the construction of hierarchies. In this process guiding the selection of dimensions, clusters, and indicators related to operational users and stakeholders within the department (Matta & Patterson 2007;Moullin 2004). This particular construction is useful in tracing poor performance backwards through the hierarchy, since poor performance outcomes will be reflected in the upper levels of the hierarchy. By enhancing transparency, the root cause for any overall performance problems is more easily identified.

Step 2: Weighing of indicators

When incorporating several indicators into a performance information system, the indicators will inevitably be of different strategic importance. Without individually assigned weights, indicators in "large" clusters will mathematically have less weight, than indicators in "small" clusters, as long as the comparison is made by simple average. This arrangement constitutes a problem, because some indicators merely support a decision, while others are governing in terms of which decision is made. To compensate for this, the framework adapts the concept of Analytical Hierarchy Process (AHP) (Saaty 1982), to make a quantitative distinction between the indicators, the clusters, and the dimensions within the framework. AHP is a multiple criteria decision-making approach assigning mathematical weights based on either qualitative assumptions or quantitatively underpinned arguments. By applying this approach, it allows subjective as well as objective factors to be considered in a decision-making process (Dey, Hariharan, & Despic 2008).

After the hierarchy of suitable indicators is constructed, the decision-makers conduct a systematic pair-wise comparison of the incorporated indicators, assigning values of relative intensity to each individual indicator within a cluster, see Figure 2. E.g. if indicator P1 is assessed to be of extreme importance (score of 9) in relation to P2, then the intensity (I) of 9 is placed on the I_{12} position and the reciprocal value (score of 1/9) on the I_{21} position. The mathematical construction of the matrix signifies that each indicator´s, comparative intensity within the cluster is stated as the summation of each column.

Scale for pair wise comparison		
Intensity	***Definition***	***Explanation***
1	Equal importance	Two elements contribute equally
3	Moderate importance	Slight favor of one indicator over another
5	Strong importance	Strong favor for one indicator over another
7	Very strong importance	Very strong favor for one indicator over another, demonstrated in practice
9	Extreme importance	Favor for one indicator with highest possible importance
Intensities of 2,3,6 and 8 can be used to express intermediate values		
Reciprocals of above	If performance indicator n has one of the above values assigned to it when compared with performance indicator m, then m has the reciprocal value when compared with n	

Matrix for pair wise comparison				
	P_1	P_2	$P_{..}$	P_{1m}
P_1	1	I_{12}	$I_{1..}$	I_{1n}
P_2	I_{21}	1	$I_{2..}$	I_{2m}
$P_{..}$	$I_{..1}$	$I_{..2}$	1	$I_{..m}$
P_{n1}	I_{n1}	I_{n2}	$I_{n..}$	1
Sum	$\sum_{j=1}^{n} I_{j1}$	$\sum_{j=1}^{n} I_{j2}$	$\sum_{j=1}^{n} I_{j..}$	$\sum_{j=1}^{n} I_{jm}$

Figure 2: Scale and Matrix for indicator comparison, adopted from (Saaty 2008)

To normalize the individual intensities, intensities is divided by the corresponding sum. Then the relative weight of each indicator within the cluster is calculated as the summarization of each row, see Figure 3.

	P_1	P_2	$P_{..}$	P_{1m}	Weight
P_1	$1 \Big/ \sum_{j=1}^{n} I_{j1}$	$I_{12} \Big/ \sum_{j=1}^{n} I_{j2}$	$I_{1..} \Big/ \sum_{j=1}^{n} I_{j..}$	$I_{1m} \Big/ \sum_{j=1}^{n} I_{jm}$	$w_{E1} = \sum_{i=1}^{m} E_{1i}$
P_2	$I_{21} \Big/ \sum_{j=1}^{n} I_{j1}$	$1 \Big/ \sum_{j=1}^{n} I_{j2}$	$I_{2..} \Big/ \sum_{j=1}^{n} I_{j..}$	$I_{2m} \Big/ \sum_{j=1}^{n} I_{jm}$	$w_{E2} = \sum_{i=1}^{m} E_{2i}$
$P_{..}$	$I_{..1} \Big/ \sum_{j=1}^{n} I_{j1}$	$I_{..2} \Big/ \sum_{j=1}^{n} I_{j2}$	$1 \Big/ \sum_{j=1}^{n} I_{j..}$	$I_{..m} \Big/ \sum_{j=1}^{n} I_{jm}$	$w_{En} = \sum_{i=1}^{m} E_{..i}$
P_{n1}	$I_{n1} \Big/ \sum_{j=1}^{n} I_{j1}$	$I_{n2} \Big/ \sum_{j=1}^{n} I$	$I_{n..} \Big/ \sum_{j=1}^{n} I_{j..}$	$1 \Big/ \sum_{j=1}^{n} I_{jm}$	$w_{En} = \sum_{i=1}^{m} E_{ni}$

**Note by construction* $\sum_{j=1}^{n} w_j = 1$

Figure 3: Normalizing matrix

The procedure repeats itself when determining the relative weights in comparing clusters to clusters and likewise dimensions to dimensions. The procedure is independent of the amount of indicators, clusters or dimensions compared. This constitutes strength in relation to the practical usage, where some organization prefers more measures implemented, than others. Further this also signifies that expansion of a hierarchy in no way composes a problem.

Step 3: Normalization procedure
If aggregation of multiple, dissimilar indicators, is to be possible, outcomes need to be normalized, in order to make them comparable. As data normalization method, the Standard score, more commonly referred to as the z-score, is chosen, see Equation 1.

$$z - score(\sigma) = \frac{(Data\ point - Mean\ value)}{Standard\ deviation}$$

Equation 1: Z-score

The z-score corresponds to a data point in a normal distribution, and converts all data into a common scale, making them comparable regardless of initial units (Stapenhurst 2009). A positive z-value indicates performance above mean for a given period of time. The magnitude, positive or negative, indicates how much the value differs from the mean in regards to the standard deviation. When normalizing all performance data according to z-scores, the outcome becomes an index in relation to past performance. The number of data points can be adjusted to the character of the individual indicator depended on the time span which is necessary to provide reasonable comparative values. The normalization procedure is performed for all indicators (P_1 -P_n) in a dimension or cluster, see Figure 4. The mean value and the standard deviation are calculated for each data set containing value points (v_{11} to v_{1m}), and are then transformed into a corresponding z-score.

The justification for choosing the z-score is that healthcare facilities commonly wish to reduce variation in delivered service (Lim, Tang, & Jackson 1999). It is considered more desirable to be performing acceptably in all aspects than perfect on some and weakly in others. By adopting the z-Score, the framework encourages mean scores over high variation, if to secure high overall performance. This premise is further supported in Danish healthcare by the vision of the Danish Quality Framework, which states that; "All patients have the right to the *same homogeneous high quality service*, no matter where they are treated" (Danish Institute for Quality and Accreditation in healthcare 2008).

Step 4: Aggregation procedure
Finally all z-scores are aggregated through weighted averages, thereby presenting a measure relative to performance history, represented by step 4 in Figure 4.

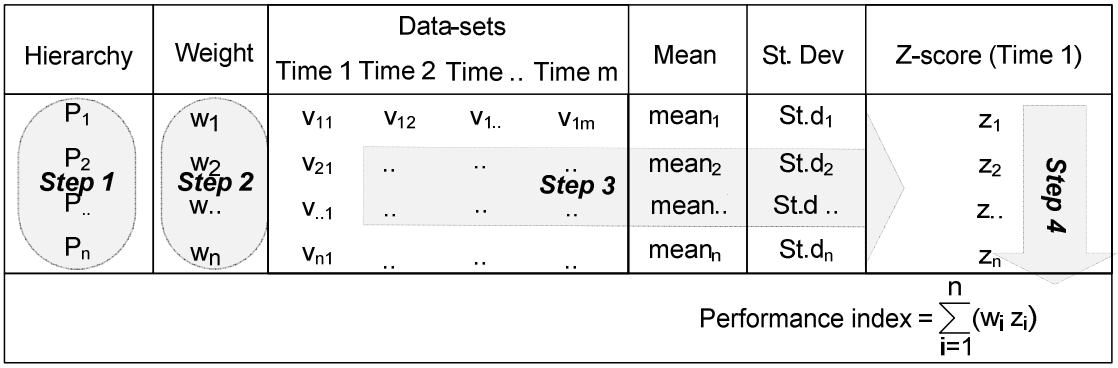

Figure 4: Schematic outline of evaluation framework

As a consequence of this approach the amount of incorporated indicators is case dependent, because each indicator acts as contributor to a given cluster with the assigned weight. The number of data points can be adjusted to the nature of the individual indicator. E.g. In some cases, retrospective data for several years are necessary to incorporate, whereas in other cases a few months' data is suitable. The performance index which is determined as the weighted average now represents performance outcomes as a representation of organizational importance. Low prioritized indicators will not have as much impact as high prioritized indicators. This allows for monitoring of performance as a function of strategic importance.

Implementation at an MRI unit

To test whether the four step procedure is useful as guidance for healthcare decision-makers, the framework is applied to an MRI unit in a Danish hospital. The case is a radiology department at a Danish non-profit hospital, consisting of four individual sites; the result of a fusion between four former independent hospitals. The hospitals were merged at management level, but the four sites still act as separate operational parts in the new hospital. The radiology department employs 128 staff members in total and treats approximately 145 patients on a daily basis. The quantitative test is performed exclusively on the MRI unit on one site, which examines approximately 3500 patients per year, distributed on about 70 different types of MRI scans. The unit receives both acute and planned patients. The unit employs both full- and part-time employees, and students.

The MRI hierarchy

The dimensions used in the construction the hierarchy were based on the strategic plan of the hospital, where *Patients*, *Employees* and *Operations* constitute the backbone. The selection and placement of indicators were determined through a series of workshops with radiologists, technicians, a project manager, and the head of the department. It was desirable to include a large amount of indicators, to provide as complete a performance picture as possible (Curtright, Stolp-Smith, & Edell 2000). Based on the indicator selection and the clustering, the full hierarchy was constructed as shown in Table 1.

Table 1: Indicator hierarchy for the MRI unit

	Dimension	Cluster	Indicator	Formula
Superior indicator	Patients	Safety	Adverse advents	$\frac{\#\ Adverse\ advents}{\#\ Total\ procedures}$
			Incorrect treatment	$\frac{\#\ Incorrect\ treatments}{\#\ Total\ procedures}$
			Re-called patients	$\frac{\#\ Re-called\ patients}{\#\ Total\ procedures}$
		Information	Received written info	$\frac{\#\ Patients\ without\ written\ info}{\#\ Total\ procedures}$
			Satisfaction (written info)	Adopted from external survey
			Satisfaction (oral info)	Adopted from external survey
		Satisfaction	Satisfaction survey	Adopted from external survey
			Waiting time for treatment	Adopted from external Database
			Complaints	$\frac{\#\ Complaints}{\#\ Total\ procedures}$
	Employees	Occupation profile	Part-time employees	$\frac{\#\ Part-time\ employees}{\#\ Full-time\ employees}$
			Available posts	$\frac{\#\ Open\ posts}{\#\ Full-time\ employees}$
			Educational positions	$\frac{\#\ Students}{\#\ Full-time\ employees}$
		Work environment	Overtime	$\frac{\#\ Overwork\ hours}{\#\ Scheduled\ hours}$
			Sick leave	$\frac{Sick\ leave\ [days]}{\#\ Scheduled\ [days]}$
			Turnover rate	$\frac{\#\ Employees\ who\ left}{\#\ Full-time\ employees}$
			Satisfaction survey	Adopted from external survey
		Risk	Reported work hazards	$\frac{\#\ Reported\ hazards}{\#\ Full-time\ employees}$
			Long-term sickness absence	$\frac{Sick\ leave > 15\ days}{Sick\ leave}$
	Operations	Planning	Acute load	$\frac{\#\ Acute\ patients}{\#\ Total\ procedures}$
			Non-Attending patients	$\frac{\#\ Non-attending\ patients}{\#\ Total\ procedures}$
			Cancelled examinations	$\frac{\#\ Cancelled\ examinations}{\#\ Total\ procedures}$
		Efficiency	Operational time	$\frac{\sum Equipment\ producing\ [min]}{\#\ Total\ procedures}$
			% procedures (7-15)	$\frac{\#\ Examinations\ (between\ 7-15)}{\#\ Total\ procedures}$
			Throughput	$\frac{\sum(Description\ end - exam\ start)}{\#\ Total\ procedures}$
		Utilization	Employee utilization rate	$\frac{(\#\ Exams \times Operational\ time) - Overwork}{\sum Working\ hours}$
			Equipment utilization rate	$\frac{(\#\ Exams \times Operational\ time)}{(60 \times 24 \times 365) - Downtime}$

The hierarchy includes 27 performance indicators distributed in 9 clusters. The chosen indicators show overall coherence with other scientific work dealing with healthcare performance measurement, e.g. (Byrne 2006;De Toni, Andrea, & Mattia 2007;Kollberg,

Elg, & Lindmark 2005). The allocation of weights was also conducted as workshops, where clinicians discussed the relative importance of indicators. The data collection was performed by the researchers, and data validity was continuously settled with the workshop attendees. After applying quantitative data from the MRI unit, the aggregated performance profile were calculated, see Table 2. Due to the normalization method, positive values indicate that the MRI unit is performing above average of retrospective data, and negative values indicate that current performance is lower than past performance.

Table 2: Aggregated performance result for the MRI unit

	Dimension *(weight)*	z-score	Cluster *(weight)*	z-score	Indicator *(weight)*	z-score
					Adverse advents *(0,630)*	0,00
			Safety *(0,503)*	0,09	Incorrect treatment *(0,250)*	0,31
					Re-called patients *(0,120)*	0,12
					Received written info *(0,463)*	0,45
	Patient *(0,573)*	0,28	Information *(0,348)*	0,18	Satisfaction (written info) *(0,329)*	0,76
					Satisfaction (oral info) *(0,208)*	-1,33
					Satisfaction survey *(0,586)*	1,88
			Satisfaction *(0,148)*	1,13	Waiting time for treatment *(0,224)*	0,69
					Complaints *(0,190)*	0,00
					Part-time employees *(0,595)*	0,45
			Occupation profile *(0,570)*	0,30	Available posts *(0,277)*	0,13
					Educational positions *(0,129)*	0,00
					Overtime *(0,438)*	-0,95
Result 0,13	Employee *(0,320)*	0,07	Work environment *(0,259)*	-0,36	Sick leave *(0,240)*	0,45
					Turnover rate *(0,202)*	0,20
					Satisfaction survey *(0,120)*	-0,77
			Risk *(0,171)*	-0,08	Reported work hazards *(0,833)*	0,11
					Long-term sickness absence *(0,167)*	-1,00
					Acute load *(0,387)*	-1,06
			Planning *(0,684)*	-0,62	Non-Attending patients *(0,443)*	-0,48
					Cancelled examinations *(0,170)*	0,00
					Operational time *(0,657)*	-0,53
	Operation *(0,107)*	-0,51	Efficiency *(0,244)*	0,06	% procedures (7-15) *(0,207)*	2,14
					Throughput *(0,136)*	-0,27
			Utilization *(0,072)*	-1,38	Employee utilization rate *(0,875)*	-1,37
					Equipment utilization rate *(0,125)*	-1,46

As the aggregated result indicates, overall performance is within control with a small increase on **0.13** in total aggregated performance. The underlying reason for this small

progress is found in the positive result on 0.28 within the patients' dimension. The relatively large positive patient dimension constitutes a significant element in the overall outcome. This of cause due to the significant weight it has been assigned. The operations dimension shows a negative score with negative values in both planning and utilization. The reason the Operations' dimension value does not bring down the overall aggregated result is due to a relative low assigned weight. The negative value in utilization points out another curiosity. Equipment utilization has surprisingly yielded a very poor result, in a period with a fair amount of overtime. The explanation being that the department has not increased the number of planned patients in this period, and has had fewer acute patients. Throughput time has increased during the same period. In a period with an increase in overtime and an increase in throughput time, it is not irrelevant that technicians have been working overtime, and perhaps thereby creating additional work for clinicians without extra resources being allocated to reporting. Data collected at the Radiology department have shown that a high amount of non-attending patients use private sector MRI for diagnostics due to prolonged waiting times in public sector MRI. This relation is enhanced by the increase in use of extra insurance allowing for the use of private health care facilities in Denmark. The correlation is also evident in the hierarchy shown by a decrease in waiting times and a decrease in non-attending patients. In a period with a shortening of waiting times the result is thus fewer non-attending patients.

Discussion

In light of the research objective, the authors believe that the framework possesses the desired strengths, particularly by providing a still life of current performance compared to past performance. The distinctiveness of this framework lies in the combination of normalization according to past performance, and the use of the AHP concept as a method for prioritization and to analyze and represent the outcome in few key measures. This allows for monitoring of performance pro- and regress as a function of strategic importance. The framework has potential to include large amounts of information and at the same time targeting this information for use in decision support in strategic decisions. The hierarchical construction calls for strategic alignment, since indicator structures are deduced through the organization from the strategic objectives. The framework thus provides a clear indication whether an organization is strategically aligned or not. Weak areas of a given organization would be visible, and its significance on overall performance likewise. The framework deals with the reality of facing massive amounts of information and the request for targeted information. Information is presented in a holistic manner, at the same time allowing focus on clustered and target measures. This provides decision-makers in tracing poor performance through the hierarchy, easing the identification of root-causes. Output represents operational performance with a positive or negative value, according to strategic progress or retreat. The strength of specific measures is still apparent, enabling operational decision-makers to correct unacceptable performance.

However, a very structured and methodical development phase is a prerequisite if the above mentioned benefits are to be obtained. The hierarchy must be constructed to reflect the desired aspects at the unit of analysis. The strategic plan of the hospital, in this case, has placed patients and employees at the centre of all activities. The choice of "operations" as a third element was based on a radiological focus. As a result of this, the patients' and employees' dimensions have been assigned the highest weights in the hierarchy. Obliviously a sloppy selection and placement would naturally signify an unusable reporting system. Assigning weights likewise is a difficult process, because usually a hospital's strategic plan does not prioritize dimensions, they commonly state areas in focus. Weight assignment will therefore be a result of subjective assessments made among

workshop attendants. The authors do not ignore the difficulty of assigning strategic weights to indicators, which can be further complicated by unclear strategic objectives (Neely & Al Najjar 2006). At the same time the weight assignment needs to be justifiable within the highly political environment healthcare finds itself. In spite of this, the weighing process is made easier due to the clustering of indicators in the framework. Successively to the implementation of the framework, the Radiology Department has continuously evaluated the weighing of indicators, clusters, and dimensions in order for it to reflect a current focus. In order to maintain overall strategic stability, it is though kept in mind to keep priority changes at a minimum. Derived from this, it is still important to notice that the framework itself does not secure quality in decisions. Quality of decisions is bound to the individual decision-maker. Even though high quality information is provided, the decision process may be influenced by internal pressure, politicians, or even poor judgment. The aim of the study has not been to validate decisions, only to provide valid and transparent information used for decision support.

There are two main issues regarding the validity of the mathematical outcome. First the validity of external input data has been secured by the agencies issuing them, and internal input data by employees at the department. Second the mathematical construction of the framework does not affect data validity. The normalization and aggregation of data do not in itself change the reliability of the data, but it naturally affects the interpretation of outcomes. In order to test whether the presentation of results corresponds to perceived reality, a method of recognition was used. The test proved that the qualitative construction of the hierarchies was in concordance with a perceived reality, because the quantitative result reflected staff members' perception of performance. The framework showed progress and retreat in the organizational areas which were expected. The notion of a "perceived reality" is important to emphasize here, because there are no absolute values for good or bad performance. The weight profiles are somewhat subjective, since the assignment is conducted based on the interviewees' perception of mutual importance. Thereby is the interpretation of performance biased to represent the "perceived reality" of those which have participated in the construction of the hierarchy and designed the weight profile. The advantage though is that the priorities are explicitly formulated, where presently is implicit.

The implementation of the framework on a single case in a public and highly political environment also restricts the generalizing potential. The authors acknowledge that there may be more suitable approaches in organizations with different settings. Despite this consideration, we share a clear conviction that the framework can be easily adapted to different settings.

Conclusion

The main objective has been to cluster large amounts of performance information in order to make it understandable for operational decision-makers. The paper proposes that performance indicators are built in a hierarchical construction, prioritized according to strategic significance, normalized according to past performance and finally aggregated to present an expression for overall performance. The advantage of using such a combination allows for the monitoring of performance pro- and regress as a function of strategic objectives. Placing indicators in a hierarchy provides the possibility of tracing performance from a strategic level through to tactical and operational indicators. What otherwise would have been a qualitative assessment of strategic importance, is now quantified by representing performance as weighed, aggregated measures. The study

challenges the traditional use of performance reporting by combining strategic weight assignment and performance aggregation in hierarchies.

References

Barros, P. P. 2003, "Random Output and Hospital Performance", *Health Care Management Science*, vol. 6, no. 4, pp. 219-227.

Bovier, P. A. & Perneger, T. V. 2003, "Predictors of work satisfaction among physicians", *The European Journal of Public Health*, vol. 13, no. 4, pp. 299-305.

Buetow, S. 2008, "Pay-for-performance in New Zealand primary health care", *Journal of Health Organisation and Management*, vol. 22, no. 1, pp. 36-47.

Byrne, M. 2006, "Implementing Performance Management in the Irish Health Sector", *The Health Care Manager*, vol. 25, no. 2, pp. 114-121.

Cheng, S. M. & Thompson, L. J. 2006, "Cancer Care Ontario and integrated cancer programs: Portrait of a performance management system and lessons learned", *Journal of Health Organisation and Management*, vol. 20, no. 4, pp. 335-343.

Curtright, J. W., Stolp-Smith, S. C., & Edell, E. S. 2000, "Strategic performance management: Development of a performance measurement system at the Mayo Clinic", *Journal of Healthcare Management*, vol. 45, no. 1, pp. 58-68.

Danish Institute for Quality and Accreditation in healthcare 2008, *Den Danske Kvalitetsmodel*, Danish Institute for Quality and Accreditation in Healthcare, AArhus, Denmark.

De Toni, A. F., Andrea, F., & Mattia, M. 2007, "A performance measurement system for facility management: The case study of a medical service authority", *International Journal of Productivity & Performance Management*, vol. 56, no. 5, p. 417.

Dey, P. K., Hariharan, S., & Despic, O. 2008, "Managing healthcare performance in analytical framework", *Benchmarking: An International Journal*, vol. 15, no. 4, pp. 444-468.

Dummer, J. 2007, "Health care performance and accountability", *International Journal of Health Care Quality Assurance*, vol. 20, no. 1, pp. 34-39.

Elleuch, A. 2008, "Patient satisfaction in Japan", *International Journal of Health Care Quality Assurance*, vol. 21, no. 7, pp. 692-705.

Evans, J. R. 2004, "An exploratory study of performance measurement systems and relationships with performance results", *Journal of Operations Management*, vol. 22, no. 3, pp. 219-232.

Evans, J. & Weir, C. 1995, "Decision processes, monitoring, incentives and large firm performance in the UK", *Management Decision*, vol. 33, no. 6, pp. 32-38.

Flapper, S. D., Fortuin, L., & Stoop, P. P. M. 1996, "Towards consistent performance management systems", *International Journal of Operations and Production Management*, vol. 16, no. 7, pp. 27-37.

Jollands, N., Lermit, J., & Patterson, M. 2003, "The usefulness of aggregate indicators in policy making and evaluation: a discussion with application to eco-efficiency indicators in New Zealand", *Ecological Economics, The ISEE Journal*.

Jones, M. K., Jones, R. J., Latreille, P. L., & Sloane, P. J. 2009, "Training, Job Satisfaction, and Workplace Performance in Britain: Evidence from WERS 2004", *LABOUR*, vol. 23, no. s1, pp. 139-175.

Kocakülâh, M. C. & Austill, A. D. 2007, "Balanced Scorecard Application in the Health Care Industry: A Case Study", *Journal of Health Care Finance*, vol. 34, no. 1, pp. 72-99.

Kollberg, B., Elg, M., & Lindmark, J. 2005, "Design and Implementation of a Performance Measurement System in Swedish Health Care Services: A Multiple Case Study of 6 Development Teams", *Quality Management in Health Care*, vol. 14, no. 2, pp. 95-111.

Kutney-Lee, A., McHugh, M. D., & Sloane, D. M. 2009, "Nursing: A Key To Patient Satisfaction", *Health Affairs*, vol. 28, no. 4, p. w669.

Lega, F. & Vendramini, E. 2008, "Budgeting and performance management in the Italian National Health System (INHS): Assessment and constructive criticism", *Journal of Health Organisation and Management*, vol. 22, no. 1, pp. 11-22.

Lim, P. C., Tang, N. K. H., & Jackson, P. M. 1999, "An innovative framework for health care performance measurement", *Managing Service Quality*, vol. 9, no. 6, pp. 423-433.

Matta, M. E. & Patterson, S. S. 2007, "Evaluating multiple performance measures across several dimensions at a multi-facility outpatient center", *Health Care Management Science*, vol. 10, no. 2, pp. 173-194.

Mital, K. M. 2010, "Queuing analysis for outpatient and inpatient services: a case study", *Management Decision*, vol. 48, no. 3, pp. 419-439.

Mohammadi, S. M., Mohammadi, S. F., & Hedges, J. R. 2007, "Conceptualizing a Quality Plan for Healthcare", *Health Care Analysis*, vol. 15, no. 4, pp. 337-361.

Morgan, D. & Morgan, R. 2009, *Single-Case Research Methods for the behavioral and health sciences*, 1 edn, SAGE, Thousand Oaks.

Moullin, M. 2004, "Eight essentials of performance measurement", *International Journal of Health Care Quality Assurance*, vol. 17, no. 3, pp. 110-112.
Nakajima, S. 1986, "TPM - Challenge to the improvement of productivity by small group activities", *Maintenance Management International*, vol. 6, no. 2, pp. 73-83.
Neely, A. & Al Najjar, M. 2006, "Management Learning Not Management Control: The True Role of Performance Measurement", *California Management Review*, vol. 48, no. 3, p. 99.
Neely, A., Mills, J., Platts, K., Gregory, M., & Richards, H. 1994, "Realizing Strategy through Measurement", *International Journal of Operations & Production Management*, vol. 14, no. 3, pp. 140-152.
Ormrod, J. 1993, "Decision making in health service managers", *Management Decision*, vol. 31, no. 7, p. 8.
Saaty, T. L. 2008, "Relative measurement and its generalization in decision making why pairwise comparisons are central in mathematics for the measurement of intangible factors", *Estadística e Investigación Operativa / Statistics and Operations Research*, vol. 102 (2), pp. 251-318.
Saaty, T. L. 1982, "Priority setting in complex problems", *IEEE Transactions on Engineering Management*, vol. EM-30, no. 3, pp. 140-155.
Stapenhurst, T. 2009, "Normalization: How to Compare Apples with Pears," in *The Benchmarking Book: A How-to-Guide to Best Practice for Managers and Practitioners*, 1 edn, Oxford, pp. 112-137.
Strandberg-Larsen, M., Nielsen, M. B., Vallgårda, S., Krasnik, A., Vrangbæk, K., & Mossialos, E. 2007, *Health Systems in Transition*, European Observatory on Health Systems and Policies, 9(6).
Swaminathan, S., Chernew, M., & Scanlon, D. P. 2008, "Persistence of HMO Performance Measures", *Health Services Research*, vol. 43, no. 6, pp. 2033-2049.
The Danish Institute for Quality and Accreditation in healthcare website. 2010. 1-3-2010.
Ref Type: Internet Communication
The Danish Patient Safety Database´s website. 2010. 10-4-2010.
Ref Type: Internet Communication
The National Board of Health´s website. 2009. 10-12-2009.
Ref Type: Internet Communication
the National Indicator Project´s website. 2009. 10-12-2009.
Ref Type: Internet Communication
The Unit of Patient-Perceived Quality's website. 2009. 3-2-2010.
Ref Type: Internet Communication
Treasure, T., Valencia, O., Sherlaw-Johnson, C., & Gallivan, S. 2002, "Surgical Performance Measurement", *Health Care Management Science*, vol. 5, no. 4, pp. 243-248.
Voss, C., Tsikriktsis, N., & Frohlich, M. 2002, "Case research in operations management", *International Journal of Operations & Production Management*, vol. 22, no. 2, pp. 195-219.
Walley, P., Silvester, K., & Mountford, S. 2006, "Health-care process improvement decisions: a systems perspective", *International Journal of Health Care Quality Assurance*, vol. 19, no. 1, pp. 93-104.
Winter, R. & Munn-Giddings, C. 2001, *A handbook for action research in Health and social care*, 1 edn, Routledge, New York.
World Health Organization 2008, *The world health report 2008 : primary health care now more than ever.*, WHO Press.

P5: Finding your Way through the Information Jungle; Using "Performance Accounts" to simplify Performance Information in Healthcare

Submitted to: The European Journal of Public Health
ISBN:
Submission date: 26 April, 2011
Acceptance date:
Publication date:
Type: Full paper publication
Status: Submitted

Finding your Way through the Information Jungle; Using "Performance Accounts" to simplify Performance Information in Healthcare

Authors

Andreas Traberg*
M.Sc., Ph.d. student
Technical University of Denmark
DTU Management Engineering
2800 Kgs. Lyngby
Denmark
Phone: (int+45) 4525 4405
E-Mail: atr@man.dtu.dk

Peter Jacobsen
Associate Professor
Technical University of Denmark
DTU Management Engineering

Nadia Monique Duthiers
MA, GDBA (Management and Organization)
Hospital of Southern Jutland
Radiology department

1. Abstract

Background: As healthcare stakeholders demand accountability in all services, it is evident that healthcare organizations need to move beyond a medical view and embrace a more holistic approach to performance evaluation. Despite this unquestionable fact, several healthcare providing units fail to structure outcomes of existing evaluations, leading to misdirected operational decisions. Instead, national and international benchmarks, quality accreditation audits, satisfaction surveys, etc., flood the desk of clinicians and managers, making it difficult to initiate corrective actions aligned with strategic objectives. Consequently this paper proposes a holistic approach for the structured evaluation of existing data, with the aim of identifying areas suited for corrective actions, balanced against strategic objectives.

Methods: The work has been conducted as a case study at a Danish public hospital. The qualitative development of the framework was performed in close collaboration with staff members who participated in workshops and interviews. As a basis for the quantitative implementation, all internal process related data were collected from internal Hospital Information Systems. Five federal units and government agencies provided external data.

Results: This work presents a novel approach to performance interpretation, taking the form of a "Performance Account", portraying operational performance as a function of strategic progress and regress. By structuring and aggregating normalised performance data according to organisational priorities, the established way of interpreting healthcare performance is challenged.

Conclusions: Applying a performance account to the evaluation of operational performance is argued to improve the ability of clinicians and healthcare managers to initiate corrective actions aligned with strategic objectives.

2. Keywords

Healthcare management, Performance evaluation, Decision Support, Case study

3. Introduction

In pursuit of healthcare accountability, the development and usage of strategic plans is common practice for all healthcare organization, where Balance Scorecard (BSC) "imitations" is the far most utilized performance measurement system (1;2). But few public healthcare organizations have designed a structured practice for evaluating these plans (3;4). A key influence affecting practice is that hospitals operate in highly political environments, where priorities shift rapidly (5;6). The political agenda is highly influenced by economical, medical, technological, and organisational developments, resulting in an influx of urgent initiatives (7). With every new urgency, the demand for evidence is concurrent, in order to document the effect (8-10). As a result vast amount of national and international benchmarks, quality accreditation audits, satisfaction surveys, etc, are applied to provide insight for healthcare stakeholders (11). Most of these initiatives are theoretically sound, well documented and well executed, however collectively they pose the one significant drawback that they function as stand-alone initiatives in practice while the objective is for them to supplement each other (12). The vast selection of self-contained initiatives is limiting the overview for the individual decision-makers (13;14). This constitutes a considerable challenge in pursuit of strategic alignment within most healthcare organizations.

The lack of structured evaluations of strategic plans also makes it difficult to evaluate, whether the organisation is in alignment with strategic objectives (15). Basically a destination (mission) is decided upon, but no map (evaluation) is given. A piecemeal approach to strategic alignment is thus adapted, where the pursuit of stand-alone initiatives dominate decision making (16). E.g. LEAN implementation, introduced as an exterior demand to improve lead times. A project of this magnitude can draw resources away from other initiatives, which in a larger context might show greater organisational benefits. The absence of a holistic view on performance limits the odds of initiating proper corrective actions with long term benefits.

Consequently, the aim of this work is to propose a framework for the structural evaluation of strategic plans, by comparing all strategic areas in relation to organisational priorities. The framework thus provides decision makers with a map of context, serving to point out areas suited for corrective action. The work has striven toward making the framework unaffiliated to departments, hospitals or organizational level, hence constructing it as a generic framework for healthcare performance evaluation. The practical applicability is illustrated by a case implementation.

4. Methods

The work has been conducted as a case study, where a Danish radiology department at have been the focal case. The first case is a radiology department, which employs 128 staff members distributed on four sites. The distribution of patients to the four sites is primarily dependent on the type of equipment available and secondary geographic location. The department treats approximately 145.000 patients per year, where 40% are acute patients. The framework has been developed in two steps: A qualitative development phase and a quantitative test phase. The qualitative development of the framework was performed in collaboration with staff members who participated in workshops and interviews (17). All internal process related data were collected from internal Hospital Information Systems, and five federal units and government agencies provided external data. The data collection has striven towards basing the performance measures on already collected, validated and published data, with an eye to enhancing credibility and validity of the measures.

5. Results

The result is the construction of a performance account based on a three step approach. Healthcare decision-makers are thus provided with at template for evaluating performance according to strategic objectives;

1. Construct indicator hierarchy
2. Assign weight profile to hierarchy
3. Aggregate, normalized performance data throughout hierarchy

Constructing indicator hierarchies

Generally two techniques are used to simplify performance representation; selection and aggregation. The selection process reduces the number of indicators by using statistic methods or qualitative techniques, and aiming at eliminating redundant information. Aggregation implies accumulation of multiple indicators into a lower number but also changes the interpretation of indicators by relating them to each other. E.g. distance and time related to each other can be interpreted as speed.

The justification for not choosing selection as a method is the evident need for large amounts of indicators, due to the complexity of health services. The strength of aggregated measures is that performance indices represent a common denominator for all included indicators. In order to be valuable for interpretation, aggregation needs to be conducted in such a manner that the aggregated outcome contains neither more nor less information than the included indicators. Still, if indicators in a data set duplicate rather than complement each other, there is no sense in including both indicators, since the information held by one can be found in the other. By aggregating indicator outcomes, the entropy of the indicator system is constant. However, the quantity of information is perceived as having decreased. It is important to notice that the selection and relation of indicators is highly contextual (18), e.g. an intensive ward is likely to choose other indicators than radiology (19).

Weight assignment according to organizational importance

The initiation of corrective actions is based on the performance outcome of individual indicators relative to their respective organizational importance. In order to clearly point out the need for corrective actions, indicators need to be prioritized according to organisational importance. Clear cut priorities thereby eliminate the uncertainty in assessing the need for corrective actions. It is therefore necessary to apply a Multi-Criteria Decision Analysis (MCDA) tool to decide upon priorities. Since prioritizing indicators according to strategic importance involves a distinct element of subjectiveness (e.g. politics and ethics) an MCDA tool needs to be able to cope with this qualitative aspect. Organizational priorities are seldom permanent, but vary according to multiple factors along with the context in which they are applied. Because decision making is subject to multiple influences, there are no set laws to characterize in fine detail structures that apply to every decision (20). Consequently the Analytical Hierarchy Process (AHP) is chosen as it involves building multiple choice criteria into a hierarchy, to establish the relative importance of included indicators (21). This method compares indicators pair-wise, assigning them an overall ranking. AHP is useful when the decision-making process is complex, particularly when decisions include alternatives which are not obviously distinctive and/or disparate. Decision makers assign values of relative intensity to the included indicators, according to perceived importance see Table 1.

Intensity	*Definition*	*Explanation*
1	Equal importance	Two elements contribute equally
3	Moderate importance	Slight favor of one indicator over another
5	Strong importance	Strong favor for one indicator over another
7	Very strong importance	Very strong favor for one indicator over another, demonstrated in practice
9	Extreme importance	Favor for one indicator with highest possible importance
Intensities of 2,3,6 and 8 can be used to express intermediate values		
Reciprocals of above	If activity *i* has one of the above nonzero numbers assigned to it when compared with activity *j*, then *j* has the reciprocal value when compared with *i*	

Table 1: Scale for pair wise comparison

The absolute numbers for each pair wise comparison are shown in the matrix below, where inverses are entered in the transpose position. Then approximate the priorities from this matrix by normalizing each column and thereby recover the eigenvector from the system of homogeneous linear equations (20).

$$\begin{array}{c} \\ E_1 \\ \cdot \\ \cdot \\ \cdot \\ E_n \end{array} \begin{array}{c} \begin{array}{cccc} E_1 & \cdot & \cdot & E_n \end{array} \\ \begin{bmatrix} \frac{w_1}{w_1} & \cdot & \cdot & \frac{w_1}{w_n} \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \frac{w_n}{w_1} & \cdot & \cdot & \frac{w_n}{w_n} \end{bmatrix} \end{array} \begin{bmatrix} w_1 \\ \cdot \\ \cdot \\ \cdot \\ w_1 \end{bmatrix} = n \begin{bmatrix} w_1 \\ \cdot \\ \cdot \\ \cdot \\ w_1 \end{bmatrix}$$

$$\underline{\underline{A}} \cdot \underline{\underline{\omega}} = n \cdot \underline{\underline{\omega}}$$

Equation 1: Eigenvalue problem

The eigenvector (ω) hence constitutes a numerical representation of the indicators' relative priorities. Because the assessment of relative importance is based upon subjective judgment, the priorities reflect the way decision makers interpret importance. The calculations are performed throughout the hierarchy, resulting in a profile of priorities (in numerals), representing each indicator's importance to the organization.

Aggregation of performance outcomes

To express performance as a common denominator for all included indicators, aggregation of outcomes is a requirement. Hence it is important to acknowledge that aggregation is only applicable when all the data are expressed in exactly the same unit, or else the final result is equivalent to "comparing apples and oranges". Therefore if aggregation of multiple measures is to be realizable, data needs to be normalized into the same unit. As data normalization method the Standard score, more commonly referred to as the z-score, is chosen.

$$z - score(\sigma) = \frac{(Data\ point - Mean\ value)}{Standard\ deviation}$$

Equation 2: Z-score

The z-score is a statistical measure that quantifies the distance (measured in standard deviations) a given data point places itself from the mean of a data set. It is a dimensionless mass derived by deducting the population mean from a data point and then dividing the difference by the population standard deviation. The magnitude of the z-score represents the number of standard deviations between the data point and the mean. A negative z-score represents that the data point is below the mean, and a positive score represent a data point above the mean. So when normalizing all performance data according to z-scores, the outcome is a benchmark of current performance according to past performance. This procedure transforms the performance outcomes into a comparable unit, which then can be aggregated.

The aggregation process itself is conducted as a weighted average of the indicators within clusters, using the profile of priorities along with the normalized performance data. Finally the framework is assembled in a performance account, where the hierarchy of indicators are portrayed as fiscal posts would be in a financial account, see Figure 1.

Indicator	*Weight*	*Performance level (Z-score)*
Superior performance index		$\text{Performance} = \sum_{l=1}^{m} (w_l z_l)$
- Bottomline Cluster a	w_a	$\text{Performance}_a = \sum_{i=1}^{n} (w_{ai} z_{ai})$
Indicator a1 ⋮ Indicator an	w_{a1} ⋮ w_{an}	z_{a1} ⋮ z_{an}
- Bottomline Cluster m	w_m	$\text{Performance}_m = \sum_{i=1}^{n} (w_{mi} z_{mi})$
Indicator m1 ⋮ Indicator mn	w_{m1} ⋮ w_{mn}	z_{m1} ⋮ z_{mn}

Figure 1: Schematic outline of performance account

The particular design of the Performance Account makes possible to include as many indicators and clusters as desired. As long as there are defined profile of weights and procedure for normalization of performance outcomes, aggregation into superior performance index is possible. Hence the Performance Account is considered to have merits regardless of strategic plan.

Implementation at a radiology department

To prove the practical applicability of the framework, the Performance Account is applied at the Radiology department, at the Hospital of Southern Jutland. The aim is to validate whether the approach advances the holistic interpretation of strategic progress or regress. The practical implementation therefore estimates the overall operational progress, during the strategic plan valid from 2007-2010, see Figure 2. The development of the Performance Account starts with the transformation of the strategic plan into a set of operational indicators, covering key matters of the strategic plan within Radiology. The indicators are subsequently clustered in hierarchies to portray their organizational relationship. The profile of priorities is constructed in workshops, where key decision makers discuss the prioritization within the hierarchy. Finally the performance outcome

is normalized, and aggregation according to hierarchical placement is undertaken, see Figure 2.

Performance Account RTG, 2007-2010		
Indicator	***Weight***	***Score***
Pro-/regress Quality 24/7		**0,11**
Patient Bottom-line	**0,68**	**0,12**
Patient satisfaction	***0,41***	***0,15***
Complaints (Satisfaction)	0,25	-0,01
Patients (tilgang)	0,75	0,20
Patient safety	***0,33***	***0,05***
Clinical Quality	0,41	0,01
- Complaints (Safety)	*0,54*	*-0,01*
- Image optimization	*0,2*	*0,00*
- Unintended occurrences	*0,25*	*0,05*
Equipment hygiene	0,33	0,13
Patient process (forløb)	***0,26***	***0,16***
Waiting list	0,47	0,45
Competences	0,43	0,00
Co-operation	0,1	-0,46
Employee bottom-line	**0,09**	**0,24**
Work environment	***0,4***	***0,00***
Psychological work environment	0,5	0,11
- Employee turnover	*0,34*	*-0,08*
- Sickness absence	*0,21*	*0,18*
- Work satisfaction	*0,45*	*0,22*
Physiological work environment	0,5	-0,11
Recruitment	***0,4***	***0,48***
Non-Danish speaking/ Danish speaking	0,33	1,03
Special employments/regular employments	0,33	-0,07
Students/Full-time employees	0,33	0,49
Radiation hygiene	***0,2***	***0,26***
Economy bottom-line	**0,23**	**0,04**
Effective work processes	***0,41***	***0,11***
Capacity utilization	0,67	0,03
- X-Ray	*0,2*	*-0,12*
- Ultrasonic	*0,2*	*-0,26*
- MR	*0,2*	*0,06*
- Biopsies	*0,2*	*0,05*
- CT	*0,2*	*0,40*
Non- attending patients	0,33	0,27
Production plans	***0,33***	***0,00***
Compliance with budget	***0,26***	***0,00***

Figure 2: Performance account at the radiology department

The overall progress has been positive (0.11), where all three strategic dimensions have progressed according to the organizational objectives. The Patient bottom line (0.11) has been a focus area in the strategic plan, which also has provided decent results over the last four years. Especially the enhancement of the patient processes (0.16) and patient satisfaction (0.15) results have proven admirable. The Employee bottom line (0.24) is excellent, though the organizational priority signifies only limited effect on the overall result. Collectively the progress within the radiology department portrays that radiology have directed themselves according to the organizational obligations laid out for them.

6. Discussion

Since there are no absolute values for good or bad performance (22), there are no way to verify the correctness of the aggregated performance outcome itself. However an important discussion upon the implementation of the Performance Account is whether the normalized performance indices comprise a consistent informational basis. Since the validity of raw performance is not affected by the normalizing and aggregating procedure, reliability thus becomes a pivotal point in this discussion. Reliability lies in the performance account being a reflection of reality; that a negative result is an actual indication that something needs to be corrected. The challenge is that weight profiles are subjectively quantified. As a result the weighted aggregation becomes a reflection of the interviewees' priorities. The interpretation of performance is therefore influenced by the "perceived reality" of those who have constructed the hierarchy and designed the weight profile. Indeed the thoroughness of the prioritization according to strategic objectives becomes a key issue regarding reliability, as it determines the end result.

With a meticulous prioritization and construction according to strategic objectives, the Performance Account confronts the issue of increasing informational complexity. The advantage is that evaluation of strategic plans is becoming more synthetic and manageable. The Performance Account has proven a valuable method for decreasing complexity and enhancing the holistic interpretation of operational performance. This without compromising the valuable information contained within the individual data sets. The experienced decrease in complexity is opposed to the actual quantity of information which in unaffected, as the aggregation includes all lower levels of performance outcomes. Thereby it is possible to take vast amounts of performance information and present it in a simple manner customized to the individual organisation. This approach implies taking a notion of perceived reality without a clear informational basis and representing actual reality in a condensed form.

7. Acknowledgements

The work was supported by The Hospital of Southern Jutland, however the hospital had no role in the design and conduct of the study, the interpretation or analysis of the data, or the approval the manuscript.

8. Conflict of interests

None declared

9. Key points

- The Performance Account brings new insight to the domain of holistic healthcare performance measurement, as it portrays operational performance, as a function of strategic progress and regress

- The Performance Account has merits in combining qualitative prioritization and aggregation of normalized performance data.
- The Performance Account grants the possibility to take vast amounts of performance information and present it in a simple manner customized to the individual organisation.
- In designing the Performance Account, the thoroughness of the prioritization according to strategic objectives becomes a key issue regarding reliability, as it determines the end result.

10. References

(1) Neely A. The evolution of performance measurement research: Developments in the last decade and a research agenda for the next. International Journal of Operations & Production Management 2005;25(12):1264-77.

(2) Curtright JW, Stolp-Smith SC, Edell ES. Strategic performance management: Development of a performance measurement system at the Mayo Clinic. Journal of Healthcare Management 2000;45(1):58-68.

(3) Moullin M. Eight essentials of performance measurement. International Journal of Health Care Quality Assurance 2004;17(3):110-2.

(4) Neely A, Al Najjar M. Management Learning Not Management Control: The True Role of Performance Measurement. California Management Review 2006;48(3):99.

(5) Furnham A. Performance management systems. European Business Journal 2004;16(2):83-94.

(6) Griffith JR, Alexander JA, Jelinek RC, Foster DA, Mecklenburg GA. Is Anybody Managing the Store? National Trends in Hospital Performance. Journal of Healthcare Management 2006;51(6):392-406.

(7) Hauck K, Street A. Do targets matter? A comparison of English and Welsh National Health priorities. Health Economics 2007;16(3):275-90.

(8) Moullin M, Soady J. Outcomes, processes and capability: using the public sector scorecard in public.: Springer; 2008.

(9) Stronks K, Mackenbach JP. Evaluating the effect of policies and interventions to address inequalities in health: lessons from a Dutch programme. The European Journal of Public Health 2006;16(4):346-53.

(10) Drummond MF, Botten G, HÃ¤kkinen U, Pedersen KMl. Assessing the quality of Swedish health economics research. Scandinavian Journal of Public Health 2006;34(6):566-7.

(11) Laamanen R, Ã¨vretveit J, Sundell J, Simonsen-Rehn N, Suominen S, Brommels M. Client perceptions of the performance of public and independent not-for-profit primary healthcare. Scandinavian Journal of Public Health 2006;34(6):598-608.

(12) Geraedts M, Schwartze D, Molzahn T. Hospital quality reports in Germany: patient and physician opinion of the reported quality indicators. BMC Health Services Research 2007;7(1):157.

(13) Norcross L. Building on success [performance management systeml. Manufacturing Engineer 2006;85(3):42.

(14) Neely A. The performance measurement revolution: why now and what next? International Journal of Operations & Production Management 1999;19(2):205-28.

(15) Rundall T, Martelli P, Arroyo L, McCurdy R, Graetz I, Newwirth E, et al. The Informed Decisions Toolbox: Tools for Knowledge Transfer and Performance Improvement Practitioner Application -. Journal of Healthcare Management 2007;52(5):325.

(16) Dummer J. Health care performance and accountability. International Journal of Health Care Quality Assurance 2007;20(1):34-9.

(17) Winter R, Munn-Giddings C. A handbook for action research in Health and social care. 1 ed. New York: Routledge; 2001.

(18) Evans JR. An exploratory study of performance measurement systems and relationships with performance results. Journal of Operations Management 2004;22(3):219-32.

(19) Berler A, Pavlopoulos S, Koutsouris D. Using key performance indicators as knowledge-management tools at a regional health-care authority level. IEEE Transactions on Information Technology in Biomedicine 2005;9(2):184-92.

(20) Saaty TL. Relative measurement and its generalization in decision making why pairwise comparisons are central in mathematics for the measurement of intangible factors. Estadística e Investigación Operativa / Statistics and Operations Research 2008;102 (2):251-318.

(21) Saaty TL. Priority setting in complex problems. IEEE Transactions on Engineering Management 1982;EM-30(3):140-55.

(22) Kollberg B, Dahlgaard JJ, Brehmer PO. Measuring lean initiatives in health care services: issues and findings. International Journal of Productivity and Performance Management 2007;56(1):7-24.

In pursuit of effective healthcare performance evaluation, the development and application of strategic plans, is common practice for all healthcare organizations. But few public healthcare organizations have succeeded in designing a structured practice for evaluating these plans. A key influence affecting this practice is that hospitals operate in highly political environments, where priorities shift rapidly, resulting in an influx of urgent initiatives. Every new initiative entails an increased demand for well-structured performance reporting to measure the effect of each initiative. Consequently, vast numbers of national and international benchmarks, quality accreditation audits, satisfaction surveys, etc, floods the desks of managers and clinicians. Hence management pursues isolated issues, resulting in a piecemeal approach with little strategic alignment throughout the organization. The absence of a holistic view on performance limits the probability of initiating proper corrective actions, thereby hindering consistent strategic advancement in public healthcare.

The thesis focuses on how to design a holistic Management-By-Objectives framework, capable of portraying strategic change in an operational context. The result is a "Performance Account", which enables managers and clinicians to evaluate performance according to organizational priorities. The framework provides decision makers with a map of context, serving to point out areas suited for corrective action with the aim of long term organizational benefits, due to strategic advancement.

ISBN 978-87-92706-07-2

DTU Management Engineering
Department of Management Engineering
Technical University of Denmark

Produktionstorvet
Building 424
DK-2800 Kongens Lyngby
Denmark
Tel. +45 45 25 48 00
Fax +45 45 93 34 35

www.man.dtu.dk