DTU Library

Development of a Generic Performance Measurement Model in an Emergency Department

Sørup, Christian Michel

Publication date:
2015

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

Citation (APA):
Sørup, C. M. (2015). *Development of a Generic Performance Measurement Model in an Emergency Department*. DTU Management Engineering. DTU Management Engineering. PhD thesis No. 5.2015

Development of a Generic Performance Measurement Model in an Emergency Department

PhD thesis 5.2015

DTU Management Engineering

Christian Michel Sørup
March 2015

DTU Management Engineering
Department of Management Engineering

Development of a Generic Performance Measurement Model in an Emergency Department

Christian Michel Sørup

Printed by: Department of Management Engineering, Technical University of Denmark

This document was prepared using the LaTeX document preparation system.

Technical University of Denmark

DEVELOPMENT of a GENERIC PERFORMANCE MEASUREMENT MODEL in an EMERGENCY DEPARTMENT

PhD Thesis

In achievement of the PhD degree
at the Technical University of Denmark
with a public defence on
May 13, 2015
at 1 PM

by

Christian Michel Sørup

born August 27, 1984
Gentofte, Denmark

Main supervisor:	Associate Professor Peter Jacobsen DTU Management Engineering Management Science Technical University of Denmark Kongens Lyngby, Denmark
Co-supervisors:	Head of Department Lisbet Isenberg Ravn Department of Emergency Medicine Herlev Hospital Herlev, Denmark
	Senior Consultant Jakob Lundager Forberg Department of Emergency Medicine Helsingborg General Hospital Helsingborg, Sweden
Board of Commitee:	Professor Dariusz Ceglarek The International Digital Laboratory University of Warwick Coventry, United Kingdom
	Senior Consultant Jens Ole Laursen Department of Emergency Medicine Sønderborg Sygehus Sønderborg, Denmark
	Professor Jesper Larsen *(Committee Chair)* DTU Management Engineering Management Science Technical University of Denmark Kongens Lyngby, Denmark

DANSK RESUME

Med de seneste og kommende års strukturændringer i det danske sundhedsvæsen forventes færre, men større, hospitaler. De nye sygehuse skal i fremtiden kunne håndtere et tiltagende patientflow, hvilket skaber både logistiske og ressourcemæssige udfordringer. I forbindelse med oprettelsen af de nye sygehuse bliver én af de største ændringer oprettelsen af "fælles akutmodtagelser" (FAM). I praksis betyder det en samling af skadestuer, akutmodtagelser og observationsenheder beliggende på flere hospitaler og adskilt på de enkelte sygehuse til én drifts- og funktionsmæssig enhed. En FAM kommer til at dække et optageområde på 200.000- 400.000 borgere, dvs. et 2-3 gange større befolkningsgrundlag end sygehusene i dag typisk dækker.

De danske regioner har indset, at en fælles modtagelse af akutte patienter vil være et vigtigt skridt på vejen i opnåelse af højnet kvalitet i behandling samt effektiv ressourceudnyttelse. For at sikre, at akutmodtagelsernes logistik og ressourcer bliver anvendt bedst muligt, og at de dårligste patienter hurtigt bliver identificeret og behandlet, vil det være et krav, at modtagelsesprocessen formaliseres og evidensbaseres. For at kunne honorere disse krav gennemføres en række tiltag, eksempelvis kompetenceøgning på akutområdet, nye samarbejdsformer, forbedret dokumentation og registrering samt brugen af triage. Disse initiativer er imidlertid nye og ikke velvaliderede til dato. Det vil være hensigtsmæssigt at kunne måle effekten af de enkelte ovenfor nævnte tiltag.

Målet med dette ph.d.-projekt har været at udvikle en model til måling af performance i en FAM. Den nye model er tilpasset således at kun de vigtigste indikatorer analyseres for at give et samlet overblik over afdelingens performance. Yderligere er en grundigere forståelse af de indbyrdes afhængigheder mellem indikatorerne søgt for at opnå dybere indsigt i FAM som system. Mod-

ellen muliggør monitorering af, hvor godt afdelingen performer over tid, herunder hvordan performance ændrer sig ved forskellige nye tiltag. I sidste ende vil en sådan model være et vigtigt redskab til at imødekomme ledelsens vision om at give patienten den bedst mulige behandling under sit ophold i FAM, samt at opnå den højest mulige ressourceudnyttelse.

SUMMARY

Fewer and larger hospitals are expected in the forthcoming years due to the latest and on-going structural changes in the Danish healthcare sector. These large hospitals must be able to handle an increasing flow of patients, thus creating challenges for both logistics and resources. In connection to the establishment of the new hospitals, one of the biggest changes is the new concept of emergency departments (called "FAM"). In practice, the new emergency department (ED) is a merger of the former ED, urgent care unit, and observation unit, where most acute patients are to pass through a joint entrance. From here, patients are either treated completely or transferred to other specialty in-hospital departments. The EDs at the new hospitals can expect to cover a demographical area with 200.000-400.000 inhabitants, equivalent to an area two to three times as large as the current catchment area.

The Danish regions have realised that the establishment of a joint reception of acute patients in the ED will be an important step towards improved quality in treatment and better utilisation of resources. It is a requirement that the reception process is formalised and evidence-based to ensure the logistics and resources at the ED is applied in the best possible manner and the most urgent patients are attended first. To meet such requirements, several initiatives are launched. These are, for instance, improving current competencies, new working procedures, enhanced documentation and registration practices, and the use of triage. All of the mentioned initiatives are new and not well validated to date. It would be desirable to enable measurement of each of the initiative's effects.

The goal of this PhD project was to develop a performance measurement model for EDs. The new model comprises only the most important performance measures that provide an estimate for overall ED performance levels. Further-

more, a thorough analysis of the interdependencies between the included performance measures was conducted in order to gain deeper knowledge of the ED as a system. The model enables monitoring of how well the ED performs over time, including how performance is impacted by the various initiatives. In the end, the developed model will be an important management tool to meet the management's vision of providing the best possible care for the acute patient meanwhile achieving the highest possible utilisation of resources.

ACKNOWLEDGEMENTS

My endeavor as a PhD student is rapidly approaching its final chapter and I am pleased to say that I can look back at three productive and enjoyable years at DTU Management Engineering. Several people deserve a token of my gratitude for having had a positive impact on my progress.

The Emergency Medicine Departments at Nordsjælland- and Herlev Hospital respectively deserves thanks for having made this PhD project possible by funding the project and believing that I was able to produce great results. My co-supervisors Dr. Jakob Lundager Forberg and Dr. Lisbet Isenberg Ravn have both provided much guidance. I am confident that I would never have made it this far without you and your colleagues' help.

A big thank you goes to Dr. Philip Dean Anderson and the employees at the Emergency Medicine Department of Beth Israel Deaconess Medical Center, Boston MA. My research stay here changed the scope and conclusions of the study, mainly due to the discovery of system dynamics and systems thinking. Indeed, my experience in Boston was a tremendous one that I will never forget.

The PhD journey would not have been as enjoyable without the pleasant company of my colleagues in the Operations Management research group at DTU. Both past and present fellow PhD students deserve a well-intentioned thought for all those fun times we have shared. Especially my office mates, Pelle Jørgensen and Diana Feibert, have always been (and are still) great company so showing up on work was never an obstacle – even on those dark and rainy days.

Also, thanks to my family and friends for having tolerated much scientific talk throughout the years. All dependent upon what the future will bring, there is a chance that I will continue my talking and I hope that you will still hear me out from time to time.

On a final note, I would like to express my deepest gratitude for all the guidance that you, Peter Jacobsen, have provided me. We have had many great talks, both work related and personal, and it would never have been such a big pleasure to have completed a PhD thesis without having you as my main supervisor.

Christian Michel Sørup
Kongens Lyngby (March 2015)

LIST OF PAPERS

During my PhD, I have first authored six articles and four posters. Three journal papers have been published within the PhD project's timeframe while one journal paper is currently in review. Only the journal papers were chosen to be included in this thesis. The conference papers were early versions of the subsequent journal papers and were therefore elided.

Included in thesis:

5.1 – CHRISTIAN MICHEL SØRUP, PETER JACOBSEN AND JAKOB LUNDAGER FORBERG (2013), *Evaluation of emergency department performance: A systematic review on recommended performance and quality-in-care measures*, Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine, Vol. 21, No. 62, p. 1-14

5.2 – CHRISTIAN MICHEL SØRUP AND PETER JACOBSEN (2013), *Healthcare performance turned into decision support*, Journal of Health, Organization and Management, Vol. 27, No. 1, p. 64-84.

5.3 – CHRISTIAN MICHEL SØRUP AND PETER JACOBSEN (2014), *Patient Safety and Satisfaction Drivers in Emergency Departments Re-visited - An Empirical Analysis using Structural Equation Modeling*, Health Systems, Vol. 3, p. 105-116.

5.4 – CHRISTIAN MICHEL SØRUP, DANIEL SEPULVEDA ESTAY, PETER JACOBSEN AND PHILIP DEAN ANDERSON (2015), *Balancing patient flow and returning patients: a system dynamics study on emergency department crowding factors*, Health Care Management Science, *in review*.

Excluded from thesis:

1. CHRISTIAN MICHEL SØRUP, PETER JACOBSEN, PHILIP DEAN ANDERSON, LISBET ISENBERG RAVN, AND JAKOB LUNDAGER FORBERG (2014), *An integrated performance measurement model for emergency department assessment*, 6^{th} Danish Emergency Medicine Conference, poster presentation.

2. CHRISTIAN MICHEL SØRUP, DANIEL SEPULVEDA ESTAY, PETER JACOBSEN AND PHILIP DEAN ANDERSON (2014), *Tradeoffs between alleviating emergency department crowding and return visits*, 6^{th} Danish Emergency Medicine Conference, poster presentation.

3. CHRISTIAN MICHEL SØRUP AND PETER JACOBSEN (2013), *Application of Structural Equation Modeling to Determine Emergency Department Patient Satisfaction Drivers*, Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine, suppl. 2,(A18).

4. CHRISTIAN MICHEL SØRUP, PETER JACOBSEN AND JAKOB LUNDAGER FORBERG (2013), *A Literature Review Analysing Endorsed Performance and Quality-In-Care Measures for Emergency Department Assessment*, Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine, suppl. 2,(A12).

5. CHRISTIAN MICHEL SØRUP AND PETER JACOBSEN (2013), *What drives Emergency Department Patient Satisfaction? An Empirical Test using Structural Equation Modeling*, In proceedings of the 20^{th} EurOMA Conference, HOM-24.

6. CHRISTIAN MICHEL SØRUP AND PETER JACOBSEN (2012), *What drives Emergency Department Patient Satisfaction? An Empirical Test using Structural Equation Modeling*, In proceedings of the 4^{th} World P&OM Conference (EurOMA), HEA25.

CONTENTS

List of Tables **2**

List of Figures **3**

1 Introduction **7**
1.1 Transformation of the healthcare sector 7
1.2 The Danish acute care system 9
1.3 Preparing for future challenges 11
1.4 How to define quality and efficiency? 12
1.5 Scientific structure of the thesis 15
1.5.1 "Market basket" or "balanced scorecard" approaches . . . 15
1.5.2 Underlying mechanisms which drive behaviour 16
1.5.3 Decision making effects made visible 18

2 Research design **21**
2.1 Research problems 21
2.2 Stakeholders 22
2.2.1 Scientific community 22
2.2.2 ED decision makers 23
2.2.3 Patients 24
2.2.4 Private companies that offer technological solutions 24
2.3 Research questions 24
2.3.1 Selecting performance measures 25
2.3.2 Dynamics behind the identified performance measures . . 26
2.3.3 How a performance measurement model can be developed 27

3 Scientific approach **29**
3.1 Theory of science . . . 29
3.2 Critical realism . . . 30
3.3 Ontology, epistemology and methodology . . . 31
3.4 Application of critical realism to this study . . . 33
3.5 System dynamics . . . 34
3.5.1 CR in relation to system dynamics . . . 36
3.6 Statistical modelling . . . 38
3.6.1 Hypothesis testing . . . 38
3.6.2 Structural equation modeling . . . 40
3.6.3 Statistical process control . . . 42
3.6.4 CR in relation to statistical modelling . . . 45
3.7 Validity . . . 47

4 Empirical foundation **51**
4.1 Included cases . . . 51
4.1.1 Primary cases . . . 51
4.1.2 Secondary case . . . 53
4.2 Data acquisition . . . 55

5 Incremental development of the model **59**
5.1 P1: Identification of performance measures . . . 60
5.2 P2: Unravelling a single performance measure; sickness absence . . . 64
5.3 P3: ED patient safety and satisfaction drivers . . . 70
5.4 P4: ED patient flow dynamics . . . 76
5.5 The final performance measurement model . . . 82

6 Discussion **95**

7 Conclusion **109**

8 Future research opportunities **113**

Bibliography **119**

Appended papers **130**

LIST OF TABLES

1.1 IOM's proposed dimensions of quality in healthcare 13
1.2 Different definitions of efficiency 14

4.1 Primary databases in use at the two primary case hospitals 57

5.1 Search strings and resulting hits of literature review 61
5.2 EFA factor loadings (N=685) 71
5.3 Stressed model sensitivity test results for protocol trigger values . 79
5.4 Baseline sensitivity test results for selected parameters associated with ED crowding 79
5.5 Number of interviews by profession at the included case EDs . . 83
5.6 Explanation of columns in tab *Measurement Definitions and Data Retrieval* 86
5.7 Explanation of columns in tab *Underlying Measurement Network* . 87

6.1 Concepts of validity in CR with study tactics 107

8.1 Examples of clinical performance measures for disease specific conditions; PCI = percutaneous intervention (Graff *et al.*, 2002; Lindsay *et al.*, 2002; Schull *et al.*, 2011) 115

LIST OF FIGURES

1.1 PhD roadmap of data mining enhanced ED performance measurement . . . 19

3.1 Stratified ontology after Mingers (2004) . . . 32
3.2 Components of a stock- and flow diagram . . . 35
3.3 Signatures . . . 42
3.4 Random C-chart as example . . . 44
3.5 Normal distribution with probabilities . . . 45
3.6 Research process and validity from a CR viewpoint (Johnston and Smith, 2008) . . . 48

4.1 Future hospitals in the Capital Region of Denmark . . . 52
4.2 Funnel of information (Forrester and Senge, 1980) . . . 56

5.1 Flowchart of article selection following PRISMA guidelines . . . 62
5.2 Top 25 % of recommended, discussed, or analysed performance measures in included literature reviews . . . 62
5.3 Overview of factors with a direct influence on employee absence rates . . . 65
5.4 Check list sheet of determinants for employee absence . . . 67
5.5 Determinant's status represented by shaded areas indicating focus area . . . 68
5.6 Theoretical model of hypotheses to be tested . . . 72
5.7 Measurement model (CFA) with appended model fit indices. Error terms have been omitted. LOS = Length of stay . . . 73

5.8 SEM results of tested hypotheses, non-standardised path coefficients reported only. Note: N=685, $P < 0.05$*, $P < 0.01$** 74
5.9 ED census comparison (1^{st} to 3^{rd} of October, 2013) 77
5.10 SD model overview depicting the admission process, the two feedback mechanisms, and the competing inflow of elective patients requiring an inpatient bed. Auxiliary variables and sub-views omitted. 78
5.11 Final Performance Measurement Model; the main menu 85
5.12 Screenshot of *SPC Chart Selection Guide* 85
5.13 Screenshot of tab *Measurement Definitions and Data Retrieval* 88
5.14 Screenshot of tab *Underlying Measurement Network* 89
5.15 The ED Performance Overview 90
5.16 Example of performance measure overview; C-chart of adverse events alongside Pareto diagram depicting severity of incident . . 91

6.1 Challenges of analysing a potentially causal and attributable connection between action and performance measure (adapted from Smith *et al.* (2009)) 99
6.2 Confounding variable with moderating effect 100

CHAPTER 1

INTRODUCTION

1.1 Transformation of the healthcare sector

Western European healthcare systems have continuously been transformed as a result of various reforms since the 1980s (Saltman and Figueras, 1997). These systems differ in composition but do share some common traits in terms of policies. The common traits are tendencies towards centralisation, new managerial forms, and economic incentives to promote healthy competition all of which are examples of new public management (Hood, 1995). Similarities in reforms of various countries can be explained by the need to face a shared healthcare related problem caused by an ageing population, higher expectations on public services, shrinking budgets, and fast scientific technology development which changes best-practice treatment procedures (Blank and Burau, 2004).

In Denmark, fewer and larger hospitals are expected in the forthcoming years due to on-going structural changes in the healthcare sector (Carlsen, 2009). These new hospitals are planned to handle an increasing flow of patients, thus creating challenges for both logistics and resources. According to the Danish Health and Medicine's Authority (*DHMA*), 21 hospitals that provide emergency services will be the result of the new fully deployed healthcare structure scheduled for 2020. This is a reduction from the current 40 public hospitals (DHMA, 2014). With the introduction of the healthcare sector's structure, patients will be required to travel further for qualified treatment. In order to minimise the inconvenience of larger distances, the hospitals are to be built with easy access from highways or rail transportation, some will include helipads. Also, the primary sector is to be strengthened. Ensuring high quality pre-hospital care (i.e.

outside the hospital) and early start of initial treatment interventions for acute patients at the hospital are important focus areas for a well-functioning health-care system (DHMA, 2007). The rollout of the new acute hospital structure is controlled individually by the five regions of Denmark. Together, they have outlined the overall plans, goals, and principles in a timeframe between 2007 and 2012. The emergency department (ED) is today already a focal point of the plan for the handling of acute patients and will continue to be so with the introduction of the new hospitals. Today, most EDs in Denmark are mergers of urgent care units, former concept of emergency departments, and observation units. This is a new concept that ensures a joint point of entry for all acute patients with the exception of a few selected patient groups (e.g., ST-segment elevation, myocardial infarction, obstetrics, and gynaecology). The Danish Health and Medicines Authority have defined the new emergency department concept ("fælles akutmodtagelse" or "FAM") as follows:

> *"The term "fælles akutmodtagelse" (FAM) covers the shared physical location at a hospital in one cadastre where acute patients can receive preadmission evaluation and where imaging diagnostics and treatment are provided by physicians from several specialties, regardless if treatment is provided on an outpatient basis or requires admission to in inpatient department. The FAM is staffed with physicians, nurses, and other healthcare staff."* (freely translated from DHMA (2007)

The FAMs at the new hospitals are expected to cover a demographical area containing 200.000 - 400.000 inhabitants, which is equivalent to an area two to three times as large as the current catchment area (Holm-Petersen, 2010). It is a requirement that the reception process is formalised and evidence-based to ensure that the logistics and resources at the FAM is applied in the best possible manner and the sickest patients are attended first. Some of these requirements are explained in detail in the Danish Quality Model (Hundborg, 2009). Establishment of the FAMs will be based on the recommendations made in the 2007 report by the Danish Health and Medicines Authority entitled *Styrket akutberedskab – planlægningsgrundlag for det regionale sundhedsvæsen* (DHMA, 2007).

Full implementation of the new ED is estimated to take between five and ten years, thus the re-organisation is still ongoing at many hospitals. The burden of treating and diagnosing acute patients will be moved to a larger extent from the inpatient wards to the FAM. It is intended that more acute patients will be fully treated and discharged directly from the FAM without admission to an inpatient ward - this is an important decision due to fewer inpatient beds becoming available. Restricting the number of inpatient beds should be seen in the light of rapidly improved diagnostic options, for instance in telemedicine, and also a general preference for patients returning home instead of being admitted. The new FAM concept creates numerous management challenges which will be resolved through radical changes to the organisational architecture. Throughout the rest of this PhD thesis, the term "ED" will be used consistently instead of FAM to avoid confusion. The reason is that FAM is a purely Danish concept while ED is an internationally recognised term.

1.2 The Danish acute care system

Patient care provided during acute ambulatory visits or admission includes all the activities starting from a patient's first contact with the healthcare system and ends at the point at which the patient is discharged to home. In order for a patient to get in contact with the Danish healthcare system, the primary care sector or the emergency call system (112) is the first stop. The capitol region of Denmark is a special case where a new non-emergency visitation unit, named "Den Præhospitale Virksomhed" (*DPV*) or "1813", was launched in January 2014. The 1813 unit is staffed with nurses and doctors with the purpose of directing the region's patients to the most appropriate care provider (DPV, 2014). Patients are to dial 1813 when their general practitioner (GP) is closed. Although not encouraged, patients may also show up directly at the ED without first having contacted the primary care sector or the telephonic visitation units.

In Denmark, contact can be made to the GP, the 1813 unit, or the emergency call unit (112) dependent upon how ill the patients believe they are. Contact to the primary care system can be directed to either the patient's own primary care physician or a regional primary care physician (*vagtlæge*) operating through an urgent care system all dependent upon what time the call is made. During opening hours, the patient is to call his or her own GP. If help is needed out of the opening hours, contact needs to be directed to the primary care consultation service. Contact can be done by telephone or by showing up in person. The primary care physician may be able to resolve their issue through telephonic consultation, dispatch an on-call physician (*mobil lægevagt*) to their private home, refer them to an emergency department (*skadestue/akutmodtagelse*), or send an ambulance (Pines *et al.*, 2011). Ambulance dispatch is controlled by regional units. The ambulances can provide advanced life support and can, dependent on severity, be supported by physicians specialised within anaesthesiology and nurse anaesthetists who can treat the patient on site. A conclusion to release the patient can be made by the physician on site, provided that adequate treatment was given, making a visit to the ED unnecessary. Should the patient be in need of further hospital treatment, the ambulance will in most cases transport the patient to an ED but can, in rare cases, also deliver the patient directly to an inpatient ward. In the ED, the patient may be admitted to an inpatient ward after having received diagnosis- and initial treatment services or be transferred to another hospital if relevant specialised care cannot be rendered on site. The essential steps that make up emergency care delivery can be summarised chronologically in the list below.

- Triage evaluation
- Emergency stabilisation/resuscitation
- Performance of focused history and physical exam
- Diagnostic studies
- Diagnosis
- Therapeutic interventions
- Pharmacotherapy
- Observation and reassessment
- Consultation and disposition
- Prevention and education
- Documentation

Any admission to an inpatient department can be characterised as either acute or scheduled/elective. The difference between these groups is that the elective patient is scheduled ahead of arrival to the hospital while acute patients must be greeted immediately. Note that an acute patient the hospital may receive information of incoming acute patients. Acute admissions involve all of the above outlined steps in the emergency care system to some extent.

Since the recommendations for the FAM concept is still in the process of being implemented, the scope of work in the ED has not yet changed drastically. As more of the initial management of emergency patients shifts from inpatient wards to the ED, it is expected that EDs will experience an increase in length of stay (LOS) and census and this may lead to ED crowding in the future. According to Lippert *et al.* (2007), around 80 % of intensive care units (ICUs) are on full capacity on a regular basis. Relieving occupancy rates is sought resolved by transferring patients between ICUs in the attempt to balance available resources. Another option is to free capacity through cancellation of planned surgeries but this option will only be considered in very stressed situations (Lippert *et al.*, 2007). Hospital crowding is thus a natural consequence of such high utilisation rates and is a frequent scenario.

1.3 Preparing for future challenges

To cope with the increased patient flow, while making sure high levels of quality and efficiency are maintained, a wide range of initiatives have been launched in the EDs to meet the standards dictated by the Danish government. International experiences from Australia, Canada, and the USA have demonstrated that treatment quality in EDs can be improved by, for instance, the use of triage (Farrohknia *et al.*, 2011; FitzGerald *et al.*, 2010), optimising patient flows through the ED (Wiler *et al.*, 2010), introducing fast-track diagnosing and treatment for low acuity patients in parallel with main patient inflow (Oredsson *et al.*, 2011), and better teamwork (Athlin *et al.*, 2013). Such experiences serve as inspiration for specific adjustments that strengthen the ED.

Another very relevant topic widely perceived as being imperative for a well-functioning ED is having specialty doctors employed fulltime in the ED. Danish attending physicians currently employed in the ED are typically trained in well-established specialties such as general medicine, orthopaedic surgery, surgery (parenchyma), or anaesthesiology (DHMA, 2014). In the USA, another relatively new specialty was established some 60 years ago by the name of emergency medicine *(EM)* (Graves *et al.*, 1986). A doctor specialised within EM has knowledge within three main domains: 1) clinical assessments, 2) leadership and teamwork, and 3) education. Clinical assessment abilities crucial for the emergency physician is first to be able to differentiate between the critically- and less ill patients for timely treatment and second, to consider differential diagnoses ensuring a potential "plan b" in worst case scenarios. Thinking of differential diagnosis should always be done regardless if the patient presents with a prior diagnosis from the primary sector. Management abilities become important for optimal coordination between specialties that are required for treating a patient suffering from multiple illnesses. With a broad general knowledge of which specialties are required for the further treatment of a patient comes higher flow efficiency. High flow efficiency will become realised since inpatient wards, ambulatories, and surgery will to a larger extent be able to focus on those patients that do need their specialised treatment. The ED holds great potential for the education of medical residents. Supervision and education utilising the principles of apprenticeship gives a safe introduction for new doctors to the management of acute patients. While EM as a specialty is not yet implemented in Denmark, it is highly likely that it will be introduced within the upcoming years (Mackenhauer and Petersen, 2014). Most recently, Sweden and Finland introduced the new specialty with enthusiasm (Brabrand and Ekelund, 2010; Hallas *et al.*, 2013).

All these initiatives are launched with the overall purpose of improving the quality and efficiency in the services delivered. What is highly warranted is a performance monitoring tool capable of assessing the impact of these initiatives on selected macro-level measures. These measures should connect to aspects within quality and efficiency while being in line with the strategy of the department.

1.4 How to define quality and efficiency?

Quality in healthcare organisations is no easy concept to grasp. All stakeholders have a general desire to improve healthcare quality but may have diverse understandings regarding how this is achieved. For example, clinicians are mostly interested in evaluating the quality of services provided to individual patients. The government may be more interested in ensuring that treatment provided by the ED lives up to minimal standards while patients prefer best-practice treatment including state of the art technological solutions. Patients though have less medical knowledge and are thus more attuned to whether the behaviour of the healthcare professionals coincides with expectations and whether their symptoms are relieved upon completed treatment. Any attempt to measure quality in healthcare must take the complexity of diverging perspectives into consideration. Most people believe that they can feel its presence or absence when quality is experienced. However, when the question comes to how it is measured, it becomes challenging to define exactly which scale can be used to measure quality. Hence, it is not known how much of it is present at a given point in time or when it is increasing or decreasing. In the highly heterogeneous emergency care system where most acute patients are seen, quality is certainly a multi-faceted variable.

In the ED, the whole spectrum of disease types is represented. Easily tracked significant outcome variables rarely happen, for example mortality. The sickest patients that hold the highest potential for measuring the effects of treatment interventions are usually transferred to in-hospital departments or other treatment facilities further down the line of the hospital care system. The extent to which diagnostic evaluations and subsequent treatment provided in the ED is difficult to determine when there is no option to assess the continuity of care.

Some definitions have been offered by healthcare organisations in an attempt to capture the essence of quality-in-care. These however only sketch the broad lines of the concept and are therefore highly abstract. One such definition was put forward by the Institute of Medicine *(IOM)* in 2001:

> *"Quality of care is the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge"* (IOM, 2001b)

The above quotation provides a qualitative sense of the nature of healthcare quality. However, it does not provide concrete parameters that can be applied for measuring its presence or absence. The definition of quality may be anything anyone wishes it to be albeit it must reflect values and goals of the healthcare system and the society in which it exists (Donabedian, 2005). The IOM specified several areas that are commonly referred to as comprising a medical care system of high quality (IOM, 2001b):

Area	**Definition**
Safe	No harm to patients from the care that is supposed to help them
Effective	Provide the best possible evidence-based services to all patients who could benefit and refrain from providing services to those not likely to benefit (avoid under- and overuse)
Patient-centered	Provide respectful and responsive care in compliance to individual patient preferences, needs, and values
Timely	Reduce waits and potentially harmful delays for both patients and care givers
Efficient	Avoid waste, especially waste of medical equipment, supplies, ideas, and energy
Equitable	Provide continuous high quality of care regardless patients' personal characteristics (i.e. gender, ethnicity, geographic location, and socioeconomic status)

Table 1.1: IOM's proposed dimensions of quality in healthcare

Concrete measures for efficiency were not addressed in the subsequent National Healthcare Quality Report in 2001 because of the "considerable methodological and measurement issues involved" (IOM, 2001a). Since then, there has been a lot of work on conceptualising efficiency measures and major progress has been done. Measuring efficiency adds another perspective to quality measurement and specifically highlights trade-offs in quality improvement. Hence, efficiency measures plays an important role in tracking the effects of health policy initiatives (Gerdtham *et al.*, 1999). Despite the importance of this value, a variety of definitions for efficiency exists, just as this is the case for quality. The diversity in definitions results in confusion amongst stakeholders about adequacy or desirability of alternative measures of efficiency. To demonstrate the variety in wording, a non-exhaustive list of definitions is provided in Table 1.2:

Author	**Definition of efficiency**
IOM (2001a)	Avoid waste, especially waste of medical equipment, supplies, ideas, and energy
MPAC (2006)	Using fewer inputs to get the same or better outcomes. Efficiency combines concepts of resource use and quality
Palmer and Torgerson (1999)	Health care resources are being used to get the best value for money

Table 1.2: Different definitions of efficiency

Despite the common characteristics in these definitions, they differ enough to cause confusion when it comes to measuring levels of efficiency (Burgess, 2012). Efficiency in this study is defined as being the relation between the resources utilised to handle a given patient and the output of the healthcare system. This definition is broad enough to comprise different in- and outputs while allowing for flexibility of what to evaluate under the term "efficiency" and whether available data and methods support the construct.

As outlined, there are many methodological challenges to overcome in connection with the measurement of healthcare quality and efficiency. All stakeholders, be it caregivers, patients, regulators, or private companies, are interested in knowing the payoff resulting from spending increasing amounts of funds on healthcare improvements. A clear perception of whether governmental subsidies increase the value of healthcare services provided is still warranted today. There has to date not been developed a thorough systematic approach capable of determining whether healthcare quality and efficiency is on the rise in the ED. In Denmark, great opportunities for performance measurement do exist and these have also been exploited in the past. Clinical registries, national patient satisfaction surveys, national registries of adverse events and patient complaints, and the Danish Health Care Quality Programme (for accreditation) were all means that were utilised when the *National Indicator Project* containing performance measures for several different diseases was launched in 2000 (Mainz *et al.*, 2009). This project specified quality standards with appertaining indicators for six diseases being 1) stroke, 2) hip fractures, 3) schizophrenia, 4) acute gastrointestinal surgery, 5) heart failure, and 6) lung cancer (Mainz, 2003).

Performance measurement in connection to quality monitoring is thus a topic that has received a lot of attention on a national strategic level but not to the same extent on departmental levels. What is generally accepted though is that insights into departmental performance levels indeed are important to pursue. Since no generic approach to measure performance exist, individual healthcare organisations perform ad hoc measurements according to whatever measures they find most relevant.

This PhD thesis develops a generic performance measurement model that consists of several performance measures mostly relevant for the clinical director but also for other stakeholders indirectly. Special emphasis is put in understanding the complex interactions that naturally occur when introducing the many diverse performance enhancing initiatives. All acquired knowledge throughout the study is condensed in a software platform that provides the clinical director with in depth insight into the department's levels of performance over time.

1.5 Scientific structure of the thesis

1.5.1 "Market basket" or "balanced scorecard" approaches

Providing healthcare is highly complex in a variety of domains. Patients present themselves based on symptoms rather than a prior diagnosis. Clinical decision making contains several difficult steps such as what medication needs to be administered, correct treatment pathway, and likely diagnosis. These are all based on an incomplete decision foundation. The staff in the ED experience varying degrees of stress which depend upon patient arrivals and changes that occur in the condition of existing patients. Caregivers have multiple tasks to administer and must balance these according to acuity (e.g. spend time with current patient versus see other urgent patients). Lots of activities are ongoing simultaneously in an ED and many of these are recorded either manually or automatically in various databases, creating a wealth of available data.

Because of the high degrees of complexity, there is a need to handpick a set of performance measures that encompass this complexity. These performance measures need to be defined and commonly understood while being in compliance with the seven IOM recommendations in order to cover the diverse nature of acute healthcare delivery. Today, many healthcare managers are required to spend excessive hours analysing archival data in order to gain new knowledge about the system – time that could be used much more productively. A collection of such performance measures can be regarded as a "market basket" or "balanced scorecard" capable of painting a broad picture of an overall population based on a carefully sampled subset (Sorian, 2006). The warranted "market basket" or "balanced scorecard" must reflect the broad patient population presented to the ED and the wide scope of treatment services provided by the ED system to ensure a proper depiction of ED performance.

Discussing which performance measures to include is greatly debated worldwide. Some believe only a handful of performance measures are sufficient while others are inclined to include a larger amount for higher levels of detail. Some existing quality measurement frameworks for emergency care include representative disease entities of varying acuity to serve as intermediary measures for how care is provided across the spectrum of patient diseases and acuity in the ED. Examples of diseases that involve tracking of specific measures are asthma, pneumonia, and acute myocardial infarction (Lindsay *et al.*, 2002). Whatever choice is made in regard to which specific patient diseases are of interest (if any) then one should make sure that these represent the majority of patients treated in the given ED, that they are valid across age groups, and that evidence supports a certain best-practice treatment that has high impact on recovery.

Establishment of a performance measurement model requires data gathering, filtering, and processing. Longitudinal internal benchmarking is made highly difficult or downright impossible if data is gathered in a non-standardised manner. This problem appears as a result not only because registration practices change over time but also from having multiple data handling systems that are not connected to a shared data storage source for easy access. Adding to the problem is the issue of cumbersome documentation protocols where acute patients are transferred from the ED to external healthcare providers or, initially, are receiving patients from ambulance services. For instance, a patient may arrive by ambulance to the ED, be admitted to an inpatient ward after initial treatment, and be transferred to another hospital's inpatient ward, before finally being discharged to his/her home. Here, there is a great need for standardised exchange of information between transitions. Hence, a precise measurement of performance puts great demands on the supporting information infrastructure, preferably with a minimum of human interference.

1.5.2 Underlying mechanisms which drive behaviour

Causal mechanisms behind the processes that reflect social patterns play a major part of this study. The ED can be regarded as a social construct involving heaps of interrelated simultaneous processes between multiple professions. Enhanced knowledge of the ED as a complex system can be obtained if these processes are understood in regard to how they impact with each other. Attention to causation is a returning hot topic for many social researchers (Imai *et al.*, 2011). The causation of mechanisms have been studied intensely in many different sciences. Having a general definition that covers all of these domains does not exist. For instance, disciplines within health sciences look at neurological networks while social scientists study more diffuse phenomena (Hedström and Ylikoski, 2010). However, some shared definitions are generally accepted.

Firstly, mechanisms are identified based on the outcome they produce. Characterising this effect must be done with care (see section 6 for an explanation). Secondly, mechanisms are micro-level causal notions that cannot be reduced. This can be a great challenge to overcome as socially constructed processes may blur effects through unmeasured mediating variables. Interventions may also not have sufficient effect for their impact to be traceable. A more straightforward and highly visible causal network compared to a social construct would be that of a car's engine (provided that the hood is open) (Mayntz, 2004). Here, the engine's components being fan, electrical wiring, battery, etc. can be physically assessed. Such leisure is not possible when analysing social constructs. Thirdly, mechanisms are comprised of structures. These particular structures contain the driving components of the mechanism's behaviour. This is analogous to an airplane's "black box" where the researcher is interested in making the box transparent in the sense that properties and relations causing the effect under analysis become clear. Finally, mechanisms must be regarded as hierarchical in structure. In practice, mechanisms can be broken into sub-mechanisms that ultimately relate to each other. There will at some point be an exhaustive level of detail that can be regarded as fundamental. However, they may nonetheless be relative to other mechanisms across sciences (Hedström and Ylikoski, 2010).

It seems apt to apply a mixed-research approach to cover the major processes' interconnections relevant for the emergency care system. Qualitative techniques cover a concise high-level business process mapping and a mapping of ED employee representatives' perception of how the processes relate. Statistical approaches are applied to elucidate if peer-reviewed literature conclusions upon hypothetical causal links can be seen in empirical data material gathered at the PhD project's case hospitals. What statistical methods do not offer is insight into the dynamics behind the causal mechanisms. To analyse such dynamics, computer simulation comes in handy, especially that of system dynamics (Forrester, 1994). This study has produced a total of three journal papers upon the subject of causal interrelations.

1.5.3 Decision making effects made visible

Making decisions in a complex setting like the ED often puts the employee's cognitive abilities to a test. The employee may have a good understanding of how the system works based upon individual interactions between variables. However, the challenge grows when trying to predict how external policy alterations will impair system behaviour. What will the result be when specialty physicians are introduced within emergency medicine? Will finishing treatment in the ED result in better quality? Many variables are needed to determine plausible answers for such questions. Subtle system interactions will come into play making the total outcome of such interventions hard to interpret. A clinical decision maker's intuitive judgement and subsequent decision making is impaired by complexity and stress. Since the precision of the decisions made in an ED are very important, there is a need for support to compensate for natural human deficiencies. An array of techniques originating from information science and management science *(MS)* make it possible to set up performance measurement models *(PMS)* in the shape of computer software, either as stand-alone tools or imbedded within existing computer environments. Note that the overall goal of a performance measurement system is to present the system-wide effects of decisions that have already been made and implemented. This is not to be confused with a different type of program typically referred to as "information decision support systems" *(IDSS)* as this is defined as *"an interactive computer-based information system that is designed to support solutions on decision problems"* (Lee and Huh, 2006). The difference between the two is that while the PMS presents retrospective performance data, the IDSS can be applied in real-time for concurrent decision making.

Variation in data needs to be assessed in order to determine the effects of various initiatives launched in the ED at a given date. This can be done through the use of process control charts that can assess the statistical probability of seeing special cause variations as a symptom of change in the system's behaviour (Hart and Hart, 2002). A complete performance measurement model consisting mostly of control chart windows allows the ED decision maker instant insight into the data trends. A high-level roadmap presenting 1) the overall idea of the PhD thesis and 2) the different phases which the study has gone through is presented in Figure 1.1.

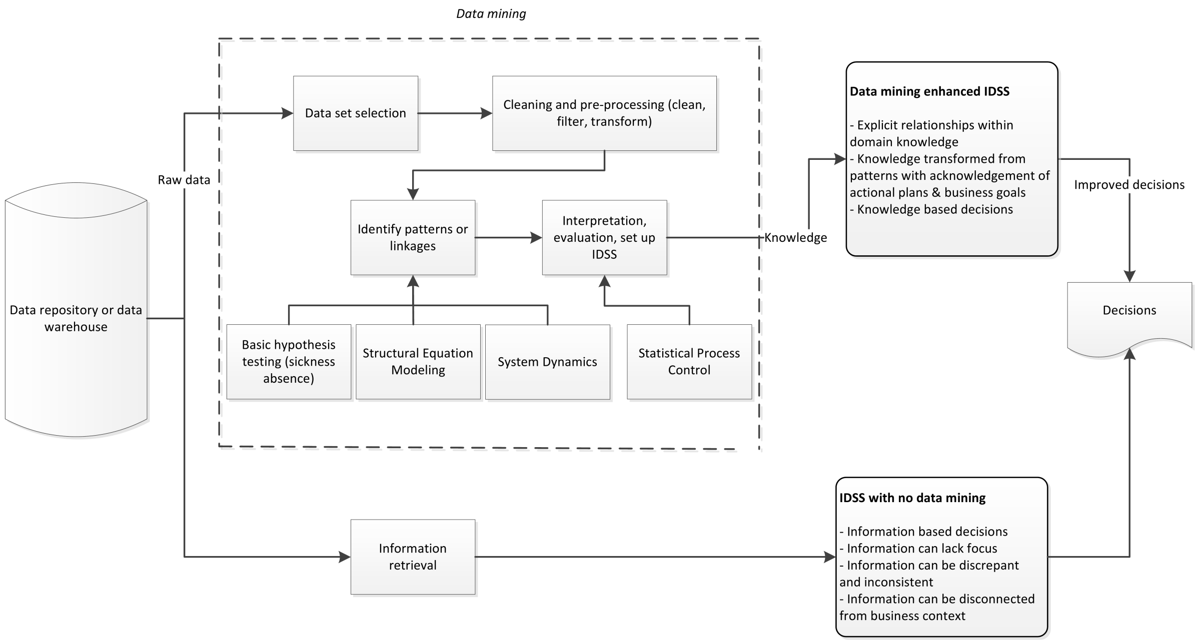


Figure 1.1: PhD roadmap of data mining enhanced ED performance measurement

The cylinder in Figure 1.1 symbolises the hospital's data warehouse which comprises smaller in-hospital databases containing differentiated performance measures. If the bottom path is followed, information retrieval happens with a minimum or no prior data mining effort. System-wide effects become impossible to assess if evaluation is made based upon fragmented data elements in isolation. Therefore, the upper path is what is desired. All the phases this study has gone through are enclosed within the broken lined rectangle. To start off, the performance measures are selected and retrieved from the data warehouse. A cleaning, filtering, and processing of this data is required before any further analysis can be conducted. Three attempts to understand the underlying patterns behind selected indicators were conducted. The first attempt was made using basic statistical hypotheses testing of underlying drivers for sickness absence rates. The second attempt applied more advanced statistics through the use of structural equation modeling, testing more hypotheses connected to patient satisfaction simultaneously. The last attempt applied computer simulation of the admission processes at a major hospital in Boston, MA. All knowledge of the underlying patterns behind the selected performance measures was encapsulated in a computer program. The computer program presents a guided user interface that presents an intuitive and quick overview of the ED's performance by extracting and processing data automatically from the data warehouse.

To summarise, the outcome of this study is a data driven performance measurement model capable of assessing system performance in retrospect while explicitly mapping the underlying mechanisms for causal explanation.

Each of the phases in the study is elaborated in much more detail in chapter 5.

CHAPTER 2 RESEARCH DESIGN

The following chapter will present the research questions and the scientific foundation upon which this thesis was built. Firstly, a brief introduction to the overall research problem is given. Next, the primary- and secondary stakeholders are introduced together with their main interests. Identification of stakeholders is paramount to specify research aims and to maximise the value of the practical outcomes. The meta-question is then stated and afterwards split into three sub-questions that have a guiding purpose to ensure comprehensive treatment of the task at hand.

2.1 Research problems

Clinical decision makers require a collection of comprehensible performance measures to ensure a well-functioning department. It is of great importance that these performance measures contain the core strategic values and are presented in an intuitive way in order to allow a quick yet comprehensible overview. Therefore, the paramount question for this research is phrased as follows;

> **How can a holistic performance measurement model be developed which, using selected indicators reflecting quality and efficiency, measure whether the patient receives better treatment by the introduction of organisational changes?**

Making the performance measurement model holistic implies transparency of how the performance measures relate. As a main focus, this thesis' main concern is focussed upon how to present performance information based on carefully selected measures seen in a broad operational context. The practical outcome is a performance measurement model capable of holistically tracing the effects of performance variations. Two different methods of indicator use are usually mentioned in relevant literature which relate to the locus of control and type of resulting action (Willis *et al.*, 2007). External indicator systems are of particular relevance for political, commercial or community groups as key indicators which are monitored for performance verification purposes. On the other hand, internal systems focus primarily on audit-feedback loops that allow monitoring of departmental performance levels. This study strives to develop an internal performance assessment tool for use by ED decision makers. The research at hand intends to fulfil two goals; the practical performance measurement model holds special value for the emergency medicine community as it strives to solve the given research problem while it also strives to carry out theoretical advancements within the current performance measurement literature.

2.2 Stakeholders

The stakeholders that take special interest into this study can be split in two primary- and two secondary groups. The primary beneficiaries are the scientific community, mainly from management science, and the ED decision makers. The secondary stakeholders have an indirect interest in the work as successful implementation and future possibilities are of value. The acute patients that go through the ED constitute a key client who benefit from high fidelity performance enhancing initiatives. Companies offering technological solutions, being either of clinical importance or IT support, may see a potential in this study to extend their existing businesses.

The four stakeholders make up the common players of modern healthcare both as internal active players and external influencing actors. All stakeholders hold different interests implying a varying degree of benefit they can gain from this study.

2.2.1 Scientific community

This PhD project's primary academic intention is to advance the understanding performance measurements in the healthcare community. This goal can be achieved through the adaption of state of the art scientific measurement techniques grounded in systems thinking (elaborated upon in chapter 3). The work provides new knowledge in terms of a novel performance measurement approach and is demonstrated on the complex system of EDs. This progress in theoretical development has future practical implications of interest to the scientific community.

Contributions to the ongoing scientific debate have been done through the publications of journal papers targeted both at emergency medicine- and management science related journals. Furthermore, a number of presentations at international conferences have also been conducted in order to gain quality feedback so that the conference papers could be altered into subsequent journal papers. By acquiring comprehensive knowledge of recent high impact publications, this study's publications are based upon current trends and hot topics in performance measurement literature. All contributions aim to reach a broad audience that entails a critical reaction in terms of theory building by fellow researchers.

The present work consists of four journal papers and six conference contributions. All journal papers are submitted to internationally peer-reviewed journals, three of which have been published at the time of writing. The conference contributions come in the form of two conference papers and four poster presentations, that were all abstracted in *Scandinavian Journal of Trauma, Resuscitation, and Emergency Medicine*. Both conference papers were presented at the annual *EurOMA* conference in the years 2012 and 2013. All posters were presented at the annual *Danish Emergency Medicine Conference* (DEMC) in the years 2013 and 2014. The basic publication strategy was to cover a broad spectrum of journals to demonstrate that the scientific relevance is interdisciplinary with the scientific communities. Submission to either journals or conferences has been done as successive steps moving towards the study's final outcome.

2.2.2 ED decision makers

Adjusting the final model to consist of a practical hands-on IT solution is a natural consequence of the close collaboration with clinicians and ED decision makers. Consequently, the ED employees are obvious stakeholders in this PhD project. The EDs at both Herlev- and Nordsjælland have taken a special interest in the development of the project in their role as co-supervisors and sponsors. Direct observations and continuous interviews have been conducted at both departments which have had major influence upon the final outcome. Therefore, the suggested performance measurement model is tailored for optimum usability in EDs. Inspiration to develop this performance measurement model was not only sought from the primary collaborators. Perspectives and thoughts from personnel at other EDs have also been taken into careful consideration.

Such perspectives have not been restricted to national borders but have also contained active participation from a major Northern American ED. Inclusion of many EDs increases the potential for widespread use and it can reasonably be assumed that other EDs will be able to benefit from the results of this study. The potential for general use of these results is addressed in detail in the validity section (see section 6).

2.2.3 Patients

With improved system performance insight for the ED decision makers, a natural beneficiary would be the patients. Today, patients themselves play a major role in what is to be expected from the healthcare sector due to a shift in paradigms in modern healthcare (Pellegrino, 1999). Expectations are higher than ever which makes it imperative for the ED decision makers to launch improvements of highest possible impact. A performance measurement model that can aid in targeting areas of performance that hold great value for patients is a means of meeting these expectations. Hence, there is every reason to believe that an improved understanding of how efficient and high quality care can be given in the ED will positively affect the patients.

2.2.4 Private companies that offer technological solutions

The rapid evolution of technological solutions has the potential of changing modern healthcare to the better. Many functions in healthcare are highly dependent upon advanced medical equipment to not only aid clinicians in delivering the best possible clinical treatment but also manage the progressively complex processes within the hospital setting. Healthcare organisations are becoming increasingly more dependent upon the suppliers' ability to provide customised solutions fitting their urgent needs. At the same time, the pace at which new technological solutions are presented to the market is very fast which therefore makes it important for the suppliers to know exactly the demand.

If this study's recommendations are implemented in the ED's existing performance monitoring dashboard, it will become more clear for suppliers where to target future technological equipment and solutions with a greater degree of scientific evidence.

2.3 Research questions

The overall research problem was split into three consecutive research questions. These research questions were formulated to ensure a natural progression throughout the study. Three independently defined stages framed the outline for the most crucial steps involved in developing the final performance measurement model. Formulating the research questions has been under continuous revision throughout the research period. Learning new concepts and techniques has had the biggest impact upon the research focus and the research question has been re-phrased accordingly to secure consistency. Dividing the overall goal of the study into separate research questions is based upon the current pool of scientific management- and emergency medicine literature. Identified key scientific challenges, practical deficiencies, and methodological concerns identified in the current body of literature all constitute the empirical basis for framing the research problem. Also, the practitioners' frustrations and difficulties add to a more practical aspect of the research problem. Based upon the theoretical- and practical considerations just mentioned, three research questions were formulated.

The first research question deals with which of the many suggested performance measures should be included in the performance measurement model to-be developed. The second research question concerns a deeper understanding of the underlying mechanisms that connect the selected performance measures. Finally, the third research questions wraps up the complete study by suggesting a generalizable performance measurement model for assessment of ED performance. Each of the three research questions are elaborated further in the following.

2.3.1 Selecting performance measures

The reason for collecting and monitoring various performance measures varies according to perspective. Patients, clinicians, or policy makers have different agendas thus they differ in opinions on what should be measured and which priorities these measures should be given, in part because they are used for different purposes (Cameron *et al.*, 2011). Despite having narrowed down the primary stakeholders, the debate amongst the international community of emergency physicians on what should be measured is still ongoing. Consensus on a set of performance measures hampers the possibility of a standard evaluation of overall ED performance levels.

Attempts to establish consensus upon a set of performance measures has been sought; the latest was at the *Second Performance Measures and Benchmarking Summit* on February 24, 2010 where 32 US emergency medicine leaders met to discuss and debate key terminology and measures (Welch *et al.*, 2011). Other studies from, for instance, Canada and the United Kingdom have set up Delphi panels asking national emergency medicine experts to rate and debate selected performance measures stemming from literature recommendations (e.g. Coleman and Nicholl (2010); Schull *et al.* (2011)). If the outcomes of such studies are compared, the apparent result is a diverging set of recommended performance measures that also differ in their respective level of abstraction. The first research question is defined as follows:

RQ1: *Which performance measures should be used in order to assess quality and efficiency in an emergency department?*

To answer this question to satisfaction, a comprehensive systematic literature review is required. Using PRISMA guidelines, contemporary meta-studies on recommended, discussed, or analysed performance measures were analysed and recorded (Liberati *et al.*, 2009). All identified performance measures were afterwards subject to a survey amongst ED staff at five different Danish public hospitals with the purpose of ensuring that no important performance measures had been disregarded. The included studies in the literature differ in their levels of detail. This means that some studies set up performance measures for major patient groups, such as acute myocardial infarction (AMI), pneumatic-, and asthma patients, while others rely on a broader set of measures independent of patient type (Graff *et al.*, 2002). A finer grained selection of performance measures does indeed enable comprehensive and in depth tracking of various aspects of ED patient care and efficiency. However, the question is whether such knowledge is deemed worthwhile with regards to the greater expense of the effort required to make sense of such a vast amount of data? Is there an appropriate compromise between level of detail and swift insight into current performance levels that can be sought? How can such data be hierarchically presented in an intuitive manner?

Such questions were the primary focus of the initial phase of the study in order to establish a standardised, consensus-agreed set of performance measures which reflect ED performance and a. Or put differently, establishing the very backbone of the future performance measurement model.

2.3.2 Dynamics behind the identified performance measures

Knowledge about which components will constitute the performance measurement model is the next and most challenging phase of the study. The selected performance measures are to be linked according to, if possible, causes and effects. The situation is somewhat similar to knowing all the bones, nerves, blood vessels etc. in the human body but without knowing about the cardiovascular system, pulmonary system, joints, and ligaments. To investigate such patterns is no easy matter as the various performance measures identified in the first phase are highly heterogeneous, non-linear in behaviour and may exhibit delayed feedback upon impact. However, the goal is definitely worth pursuing as knowledge about how these interconnections interact will allow better control over the system when it is possible to predict the consequences of the interventions that have yet to be performed. Hence, the next research question is stated:

RQ2: *What are the patterns or linkages between the selected performance indicators?*

Acquisition of the ability to recognise and develop causal relationships is crucial for reasoning. It forms the very basis for learning to manoeuvre intelligently within a complex system. Great effort has been invested into examining causation, beginning with the Ancient Greeks, by philosophers, mathematicians, cognitive scientists, and computer scientists, and many others (Jonassen and Ionas, 2006). Everyone operates on a common sense belief that causality does exist. However, a more theoretical debate amongst scientists and philosophers about whether causality can be acknowledged at all has long been an almost quarrelsome debate (Mazlack, 2007). Perhaps a complete understanding of all potential factors may open for a crystal clear answer of whether an effect will occur or not? Such a question is unlikely answered as it is not possible to fully know if all elements of relevance have been included. However, not having complete insight should not be discouraging, since actions are carried out based upon individual perceptions.

Any kind of analysis on causation will have to deal with imprecision, uncertainty, and incomplete knowledge (Mazlack, 2009). An approach which mixes initial interviews with employees in the ED with subsequent quantitative analysis using statistics and computer simulation served as a means to answering the second research question.

2.3.3 How a performance measurement model can be developed

The third and final research question is a natural development from the two previous research questions. The selected performance measures along with their interconnections must be presented in a practical IT solution that preferably can be integrated into an existing IT-based ED data repository. The developed performance measurement model must contain several features. For instance, data is presented in an intuitive manner, early warnings are visible, it must be flexible and able to allow for future alterations, and tracking the systemic effects of interventions must be possible. All causal interconnections are encapsulated in a guided user interface (GUI), allowing the user easy navigation while hiding the underlying network complexity. How to develop the final model is central for the third research question:

RQ3: *How can a generalised performance measurement model be developed that can monitor quality and efficiency on the basis of knowing about a series of performance measures and their relations?*

At a first glance, this task resemble a classic consultancy task but the study's main focus remains an academic one, which is to draw generalizable conclusions rather than giving specific case-based recommendations and implementing these (Voss *et al.*, 2002). Data applied to demonstrate the model's capabilities is retrieved from three separate EDs adding to the generalising potential of the study.

Comparison of performance on selected measures becomes possible but it is not intended in this study. Rather, the model should be applied for internal longitudinal performance assessment only.

CHAPTER 3

SCIENTIFIC APPROACH

There can be a wealth of potential answers to the stated research questions if the scientific approach is not addressed explicitly. The scientific approach serves as the main guiding principle for both reader and researcher to understand how the empirical material is gathered and how this should be processed. In the following section, the choice of methodology that frames the entire research conducted in this study is presented. All justifications made are intended to inform the reader of the rationale behind both the approach and the subsequent interpretation of results. Consistency, structure, and scientific validity can be addressed when following a rigorous scientific approach. This chapter addresses the scientific potential and limitations that have gradually appeared throughout the progress of the study. Elaborative explanations are also included regarding the philosophy of science, and which methods have been selected and how the question of validity is addressed.

3.1 Theory of science

How did the researcher perceive reality? This question is indeed an important one and therefore critical to highlight in order for the reader to fully comprehend the scientific approach. When conducting research, the actions of the researchers is driven by the systems of belief by which they generate and understand knowledge and their claims upon reality. The systems of beliefs, or paradigms, can be characterised by their answers to three categories of questions; ontology, epistemology, and methodology (Guba, 1990). Usually, a scientific hypothesis is claimed to be a testable preliminary assumption (unlike a proposition) that can explain a phenomenon. What differ are the requirements for how the hypothesis can be formulated, tested, and subsequent results interpreted. This depends upon the chosen theory of science. Theory as a concept covers a broad range of aspects starting from the very abstract (almost philosophic) to concrete hypotheses of particular phenomena. A theory is a contemplation that describes specific aspects of a phenomenon apart from other aspects that also may impact the phenomenon (Danermark *et al.*, 2002). Ontology is theory on what exists in the world and how it exists. Epistemology is the theory of knowledge; that is what can be known in the world and how this can be known. The question if something really exists belongs to the ontological field while questions regarding methods and objectivity are addressed in the epistemological field (Buch-Hansen and Nielsen, 2005). With methodology, knowledge is about the methods applied to generate scientific knowledge. Any methodology section includes a number of methods alongside how to use these and in what situations these become applicable.

In brief, ontology concerns overall assumptions about the nature of reality; epistemology refers to the assessment and justification of proposed knowledge claims; and methodology deals with the processes through which the knowledge claims are created (Orlikowski and Baroudi, 1991). Theory of science is thus an overall theory about both science and research.

This study follows the philosophical theory of critical realism as founded by Roy Bhaskar in the philosophy of science and later extended into the social sciences by authors such as Archer *et al.* (1998) (Bhaskar, 1979). The paradigm is put in relation to both theoretical applicability and practical benefits whilst highly connected systems thinking (Mingers, 2014).

3.2 Critical realism

Critical rationalism emerged as a theory of science in 1975 when Roy Bhaskar, a British philosopher, first developed the concept of transcendental realism in his book, *A Realist Theory of Science* (Bhaskar, 1975). Bhaskar later developed the perspective of critical naturalism, which was then merged with transcendental realism and finally into the umbrella term critical realism (*CR*). CR originates from the scientific theoretical discussions of natural sciences. However, having originally been advanced as an alternative to positivism, critical realism later began to highlight other perspectives such as sociology (Archer *et al.*, 1998), economics (Lawson, 1997) and also business disciplines such as management (Ackroyd and Fleetwood, 2000) and marketing (Sobh and Perry, 2006). CR can be regarded a more holistic substitute to both postmodernist and positivist approaches.

CR claims that a reality exists beyond the understanding of human perception. Only some aspects of the world can objectively be perceived by our senses (Johnston and Smith, 2008). Sometimes our senses are not fully reliable. Illusions and sense data can fool the senses, causing a misinterpretation. When an event is misperceived, the incidence and properties of the event are separate from our understanding which means that the mechanisms driving the event operate beyond our awareness of its activity.

3.3 Ontology, epistemology and methodology

Bhaskar splits knowledge objects into either transitive and intransitive objects (Bhaskar, 1975). Transitive objects are of intangible character and include paradigms, methods, and theories. These are of subjective character and will only exist alongside human activity. Intransitive objects, on the other hand, are all real things, structures, processes, mechanisms, and events of the world. The distinctions between what is happening and what is perceived of that and between an event and what underlying mechanism that caused the event are the very essence of critical realism. In CR, these distinctions form three nested domains of reality being 1) the empirical domain, 2) the actual domain, and 3) the real domain and together these are called a *stratification of reality* (see Figure 3.1).

The empirical domain is a sample of the actual domain. The empirical domain consists of events that can be experienced through measurement and perception (Wynn and Williams, 2012). The second domain, the actual, is a subset of the real which includes all events that are initiated by causal interactions between underlying structure and entities, regardless if these events are observed by humans. Finally, the real domain contains all entities and structures that exist in the world together with their built-in causal powers. The nesting of domains is done in such a way that an event in the actual domain, caused by an activation of an underlying mechanism, may not be recognised as a registered experience in the empirical domain. Mechanisms may or may not be activated in the real domain. They may even be counter-acted, thus not producing any noticeable change in the actual domain. The overarching goal of CR is to make an attempt to apply knowledge of experiences in a certain situation to analyse assumptions on how the world works in terms of structures and mechanisms that must exist in order for some outcome to have occurred (Mingers, 2004).

CR has an ontological view that the world exist independently of humans' perception of it. Unlike positivists, CR does not assume knowledge to be final and precise. This should be understood as a prerequisite for true realisation

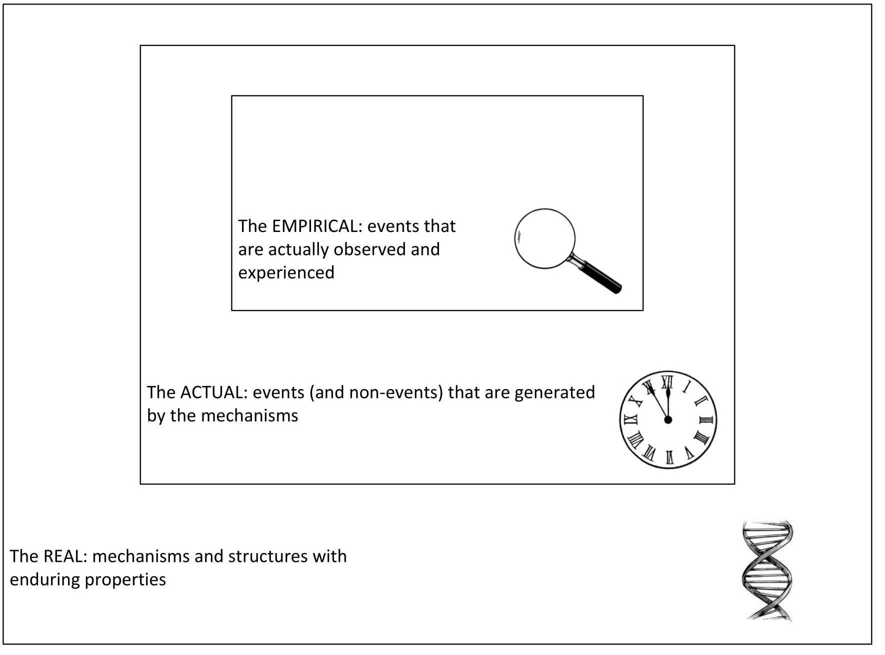


Figure 3.1: Stratified ontology after Mingers (2004)

of what knowledge is capable of explaining – not to be mistaken from scepticism. New knowledge originates in the social sphere which is constantly extending the episteme, i.e. the knowledge base. All new knowledge is building on existing knowledge through extensions, reformulations, or reinterpretations. Knowledge is therefore a social phenomenon produced by humans in a historical context. This is quite important when CR is applied as a research method due to the ability to perform analyses that are built from a theory base grounded in a common understanding, because the objects under investigation will have to be theoretically defined prior to analysis (Danermark *et al.*, 2002). Any analysis starts with an exploratory phase that serves the purpose of reaching beyond the existing knowledge base thus bringing in new knowledge on the issue under investigation.

CR is highly apt for a case study approach, giving the researcher numerous degrees of freedom in terms of choosing between methods (Ackroyd, 2010; Easton, 2010; Miles and Huberman, 1994). No definite research method is recommended by Bhaskar – as long as the method of choice can shed light on a deeper understanding of reality, it complies with CR (Mingers *et al.*, 2013). Numerous strategies for CR based research have been promoted including the identification of certain mechanisms, explanations to how the mechanisms interact within context, and descriptions of the context within which the mechanisms are working (Ackroyd, 2010). A central point of departure in the methodology of CR is that understanding of reality can never be achieved without the use of a theoretical language. The purpose of the chosen theories is to sort, explain, and predict. The more accurate the theories can explain and predict certain phenomenon, the higher the validity. The results obtained when the theories are tested against observations are quantitative connections and regularities.

A case study is an empirical enquiry with the goal of investigating a contemporary phenomenon from within its original context and is especially useful when the boarders between context and phenomenon are not difficult to determine (Yin, 2009). Usually, a case study involves only few social entities or events where data is collected from several sources (Easton, 2010).

> *"The case study inquiry copes with the technically distinctive situation in which there will be many more variables of interest than data points, and… relies on multiple sources of evidence, with data needing to converge in a triangulating fashion."* (Yin, 2009)

Triangulation means undertaking different approaches to illuminate a particular problem. In this study, three EDs participated constituting three cases in total. Inclusion of these cases had the power of promoting generalizable results at the expense of limiting the time to go in depth with each of the cases. Such considerations are elaborated upon in chapter 4.

3.4 Application of critical realism to this study

Now that the basics of CR have been explained, attention is turned towards why this philosophy is particularly well suited for management science *(MS)* studies.

Mingers provides three arguments for CR's suitability in relation to MS: 1) CR enables the researcher to take a realist stance whilst accepting the major points of critique towards naïve realism, 2) CR addresses both natural- and social sciences covering both hard- and soft approaches, and 3) CR potentially fits quite well with the perceived reality as understood by MS researchers (Mingers, 2009). Point three can be argued for taking into consideration either the theory of MS or by examining empirical examples of its practice. The latter may be more difficult to argue for as theory falls short. Focus should then be on practical examples demonstrating an array of techniques accompanied with some methodology for its use. Many examples of practical interventions can be found which constitutes another problem. Two MS approaches have been applied in this study; *systems dynamics* and *statistical modelling*. Both approaches will be presented in brief and explored relative to the overall philosophical frame.

3.5 System dynamics

System dynamics *(SD)* is a computer-aided simulation technique for business and policy simulation modelling based on feedback systems thinking. SD can be utilised on any complex dynamic system characterised by interdependent factors, information feedback, and circular causal mechanisms. SD was invented in the late 1950s by Professor Jay Forrester who has made numerous valuable contributions to the SD society. Most famous is the controversial book *Urban Dynamics* in which Forrester modelled the higher level balancing forces that control population, housing, and industry in a city (Forrester, 1969). The model predicted the fidelity of suggested remedies to the city's social problems and became a source of inspiration for many governmental policy decisions. More publications followed that promoted systems thinking to a broader audience. The most known works include *Limits to Growth* (Meadows *et al.*, 1972), *The Fifth Discipline* (Senge, 2006), and *Business Dynamics – Systems Thinking and Modeling for a Complex World* (Sterman, 2000).

The general SD approach starts with framing problems dynamically and then goes on through iterative mapping and modelling stages, and then to finally to building confidence in the model's behaviour and its implications on present policies. Analytically, the underlying structure of any SD model consists of an intertwined system of nonlinear, first-order differential- or integral equations,

$$\frac{d}{dt}x(t) = f(x, p) \tag{3.1}$$

where x is a vector of stocks or levels, p is a set of parameters related to change in the stocks or levels, and finally f being a nonlinear vector -valued function (Richardson, 2013). Simulation runs are split into discrete time intervals of time steps dt (which is user defined). Conceptually, systems are in SD modelled as closed loops of causal feedback being of either balancing or reinforcing character. These loops are dominant to varying degrees, thus holding key information to feedback understanding. Another fundamental aspect of SD modelling is the notion of endogenous change. Changes in exogenous variables can trigger system behaviour. The causes to why the system behaves as it does are contained within the system itself. All systems can be argued to be open systems in reality but this would be very hard to simulate as no model boundaries could be specified.

An example of a small system is one of a pendulum where a displacement corresponds to an exogenous change (you pull the pendulum to the left or the right) whereas the swinging of the pendulum upon release is caused by the system (the pendulum's new position, momentum, and oscillations). Theory building and policy analysis are highly impaired by this perspective. Undertaking an endogenous view eliminates the tendency, especially in social systems, to blame others for the failure of policy initiatives. What makes closed loop causality and feedback special is that it can be delayed, deceitful, and ambiguous. Any SD model strives to make explicit the system behaviour that comes from within the system itself.

The main building blocks in an SD model are stocks (levels) and flows (rates). Components in simulation software are depicted as pipes (flows) with directions and boxes as stocks (see Figure 3.2). A typical analogy to this is one of a bathtub (as stock) with a faucet (inflow) and a drain (outflow). A constant inflow gives a linearly rising stock whereas a linearly rising inflow results in the stock rising in a parabolic manner etc.

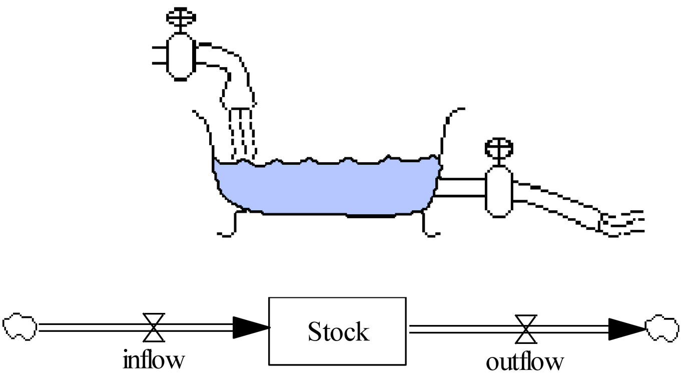


Figure 3.2: Components of a stock- and flow diagram

Noting the causal mechanisms and feedback loops does not provide the full picture of the dynamic system behaviour. To gain better knowledge of a system of interest, the modeller will, in practice, start by performing a qualitative assessment of the system, formulate a dynamic problem definition, then ensure inclusion of all relevant factors that affect system behaviour, and set up a causal loop diagram depicting the either balancing or reinforcing feedback loops.

The diagram is altered into a stock and flow diagram in computer simulation software, such as Vensim^{TM}, where each component is analytically specified. It is in the later stages when the model is able to recreate historical behaviour and the robustness of the model has been tested that deeper insights are enabled. Even though a discrete view on single events and decisions is consistent with an endogenous feedback perspective, system dynamics is generally a continuous view perspective. Applying such a perspective means looking beyond events and dig deeper into the system structure and patterns behind these. Insights into how the connections interact between system structure and behaviour are the ultimate goal for a system dynamics researcher and stem from applying the overall continuous perspective.

3.5.1 CR in relation to system dynamics

The basic paradigm of a system dynamics researcher corresponds well to critical realism's distinction between the real and the actual. All events that are experienced and where a possible explanation is desired are causally generated by the system's structure.

Another similarity between CR and SD is the emphasis on feedback structure. A clear distinction between the overt system behaviour and the underlying pattern of interconnections that is responsible for the behaviour was emphasised by both Sterman and Senge. They said that since ”behaviour follows structure”, different people are bound to exhibit the same kind of behaviour when put into the same system. Such a claim is remarkably similar to Bhaskar's distinction between the actual domain where events can be observed and the domain of the enduring mechanism that cause the events (Mingers, 2009). In a SD model, actual behaviour of the system is dependent on the strength and balance of included feedback loops. Simulation of plausible events may or may not reveal predictable changes in parameter values. Hence, a SD model can be applied to explain both the occurrence and the absence of an event. When running a SD model, causal relationships interact with each other (the real) thus creating different patterns of behaviour (the actual). Some of these patterns can be further analysed by the modeller (the empirical) by opening various graphs and tables within the modelling software and these patterns needs explanation based upon the underlying system structure (Mingers, 2009). Generative mechanisms in CR describe a causal force that is the result from a configuration of system structure. In SD, Senge describes four basic structures of SD models that correspond to the before mentioned generative mechanisms (Senge, 2006, chapter 6). These basic models are coined ”system archetypes”; generic models that contain a specific structure that can be seen in many situations and produce a predictable behaviour when active.

When it comes to the retroductive approach in CR, the explicit modelling approach in SD is a great fit. A general approach starts by identifying a problematic or unwanted behaviour in a complex system (Martinez-Moyano and Richardson, 2013; Sterman, 2000). Such behaviour is depicted by the use of charts and graphs; these are referred to as *reference modes* (Sterman, 2000, page 90). Reference modes will depict descriptive data that show the development of a particular problem over time. Hypotheses as to why this undesirable behaviour exists are then made and linked to the type of system structure that may be responsible for its generation. The initial approaches is remarkably similar to that of retroduction in CR. Factors, of relevance to the problem statement, are listed and mapped using causal loop diagrams at first (Binder *et al.*, 2004). How these factors connect can be based on multiple sources of information. First and foremost, they can be based on what Forrester calls "mental models" (Forrester and Senge, 1980). These mental models are maps of how the system behaves and are contained in the minds of the people that have their daily routines inside the system. The mental maps connect the components of the system to an overall working engine. When a causal loop diagram is in place, the modeller will have to translate it into a stock- and flow model that can be simulated. If the model is able to replicate the desired behaviour, there is better reason to believe that the model captures the essential causal traits at work. Is the model then validated? A SD model is recognised to be unprovable with reality. Attempts to eliminate alternative explanations to model behaviour, is the main approach to enhance trust in the robustness of the model. As Forrester and Senge wrote:

> *"We believe confidence is the proper criterion because there can be no proof of the absolute correctness with which a model represents reality... one tests a systems dynamics model against a diversity of empirical evidence, seeks disproofs, and develops confidence as the model withstands tests."* (Forrester and Senge, 1980)

A validated model is therefore a model that contains important elements of particular parts of a system, exhibit reasonable behaviour according to reference modes and understanding, and can withstand various tests of extreme conditions. Such a model can then be used to improve the understanding of system behaviour when changing particular triggers. SD models therefore strive to make the users aware of the causal mechanisms that dominate system behaviour more than trying to predict what will happen. This fact goes hand in hand with Bhaskar's diagnosis, explanation, and action methodology for introducing change.

3.6 Statistical modelling

Statistical analysis is a different type of modelling which recognises explicitly the uncertainty in a data set. The implicit theory of statistical modelling does inherently not follow well the principles of CR. Statistical modelling is founded on a conception, from Hume, of concurrent events, assuming that events happen on a regular basis and these can be quantified over time resulting in a data set of relevant variables (Hume, 1967).

A set of mathematical equations can be formed on the basis of the variables which represents generalizable laws. Such a worldview is in essence empiricist and goes in contrast to a critical realist understanding of the world (Mingers, 2003). Despite the criticism towards statistical modelling, critical realists do acknowledge its value, especially when it comes to assisting the researcher in explaining and learning from findings rather than to predict (Mingers *et al.*, 2013). This is true in particular for more advanced statistical techniques that go beyond basic hypotheses testing and strive to test causal mechanisms behind several variables. Critical realists applaud methods such as factor- and path analysis that explores latent constructs and causal interconnections. Three statistical techniques have been applied in this study, namely *hypothesis testing*, *structural equation modeling* (SEM) and *statistical process control* (SPC).

3.6.1 Hypothesis testing

A statistical hypothesis test is a technique of statistical inference applied when statistically testing a hypothesis. Suppose there is an interest in testing if two variables are related. What is desired is to statistically test whether this relationship can be justified on the basis of a data sample. Statistical inference refers to one main question: "What is the probability that the relationship observed in sample data could come from a population in which there is no relationship between the two variables?" (Knoke *et al.*, 2002, page 88). If there is a high probability of obtaining a relationship even though none exists (called a type 1 error), one should be very hesitant to conclude a general relationship between the two variables in a population. On the contrary, if the chance for making a type 1 error is very small, the hypothesis stated in the beginning should be accepted.

There are several hypothesis tests to choose from. What determines the appropriate technique is determined by which scale the two variables that are to be compared are measured in. Before setting up a hypothesis between two variables, which of the two is the dependent- and independent variable must be stated. Variables can be 1) dichotomous, 2) categorical or 3) numerical. A dichotomous (i.e. binary) variable can be one of two options. Therefore, examples of such could be gender or any kind of yes/no questions. Categorical variables contain several fixed options from a limited number of possibilities. Instances of categorical variables include blood type (A, B, AB or 0) and a patient's country of origin. Last, numerical variables can be measured on either a continuous or discrete scale. Such variables answer questions that relate to frequencies, i.e. ″how many″ or ″how much″.

Of relevance to this study was the method of statistical hypothesis testing between two variables that were both measured on numerical scales. Pearson correlation analysis (or merely ″correlation analysis″) is then an appropriate technique to utilise for testing the hypotheses (Soyemi, 2012). Correlation procedures assume a linear relationship between two variables together with the independent variable being normally distributed along the dependent variable. Should the data violate the normality condition, the methods are still sufficiently robust to yield acceptable results in terms statistical significance (Knoke *et al.*, 2002, page 169).

Visualisation of the data elements can be done by plotting all data elements in a scatter plot. This visual representation will depict the relation between any two numerical (continuous) bivariate variables. The closeness of the scatter to a straight line can be derived numerically by calculating the correlation coefficient (Miller *et al.*, 2005, chapter 11). This correlation coefficient is given by:

$$r = \frac{1}{n-1}\sum_{i=1}^{n}(\frac{x_i - \bar{x}}{s_x})(\frac{y_i - \bar{y}}{s_y}) \quad (3.2)$$

where r is the sum of products of the standardised variables divided by a factor of n–1. The correlation coefficient can take a value between -1 through 0 to +1. A positive correlation coefficient means that the two variables linearly follow each other's development. On the contrary, a negative correlation coefficient implies an inverse linear relation between the two variables.

Having calculated the correlation coefficient, a statistical significance test can be performed to determine the p-value. The null hypothesis states that the population correlation coefficient from which the data sample stem is zero (Sedgwick, 2012). The alternative hypothesis states that the correlation coefficient is different from zero. It is two-tailed which means that the correlation coefficient can be either < 0 or > 0. In order to determine if the obtained correlation coefficient is not due to chance, one must consult a Pearson correlation significance table to see if the two variables correlate sufficiently compared to sample size. Such tables are readily available on the internet and in statistical textbooks.

3.6.2 Structural equation modeling

Since SEM is comprised from several related statistical techniques, there is no single source of its origin. It dates back to the early 20^{th} century with the first attempts to what is now known as exploratory factor analysis, first explained by Charles Spearman in 1904 (Fabrigar *et al.*, 1999). Along came the basics of path analysis by bio geneticist Sewall Wright in 1918, where observed covariances was first demonstrated to relate to parameters of a model that could show causal effects among a set of variables. Wright can furthermore be attributed the path diagrams which are graphical representations of direct and indirect causal mechanisms that are still being applied today (Wright, 1921). Jöreskog, Keesling, and Wiley came up with an integrative approach combining factor analysis and path analysis in the early 1970s (Jöreskog, 1969; Keesling, 1972; Wiley, 1973). This approach was later conceptualised by Bentler in 1980 in a framework (at that time called the *JWK model*) that later provided the basis for the computer software for model analysis; a program named LISREL (Jöreskog and Sörbrom, 1982). The JWK model has undergone several updates since then and is now known strictly as SEM.

A SEM model can serve both a confirmatory and an exploratory purpose. Confirmatory modelling is the usual scope of current SEM analyses and begins with a number of hypotheses with limited impact in a causal model. The hypotheses need to be operationalised by importing measurement data before running the analysis on how well the proposed model fits the data. Causal assumptions made prior to model testing can then be falsified (Bollen, 1989). An example of SEM's use can be found in Hölzel *et al.* where a causal model was developed to investigate factors on patient involvement in clinical decision-making in connection to conflicts between patients and clinical staff.

Several pieces of information are required by the computer simulation software before a SEM analysis can be conducted. Which variables to include, their proposed impact on each other, and their respective directionalities need specification prior to analysis. The a priori specifications represent the hypotheses that altogether make up the model. In other words, a model is given from the beginning of the analysis and the main question will be whether the empirical data is supportive. It can occur that the data does not provide sufficient support for the model; hence the hypotheses need re-specification or the model may be disregarded completely. Another important aspect is to test *alternative models* (sometimes referred to as *equivalent models*) that may fit the data equally well, sometimes even better, and/or be more parsimonious (Shah and Goldstein, 2006). Alternative models refer to competing models that contain other hypotheses but with the same variables included in the base model, which, in practice, implies a change in direction of a particular hypothesis. Changing directions of hypotheses requires adequate support from literature and while the model that provides the best results is retained, alternative models are rejected. Variables are grouped in two categories being *observed* (manifest) or *latent*. Observed variables are the measured data elements acquired for instance through a questionnaire. For latent variables, these must be measured on a continuous scale. In SEM, these latent variables are theoretical constructs that are made up from a number of observed variables that share a common aspect. An example of such could be *intelligence* because there are no single measures that rightfully can cover the whole meaning of the term. Plausible observed variables must then measure a broad range of factors related to intelligence.

The difference between SEM and other statistical techniques is that SEM is capable of analysing both observed and latent variables. In comparison, multiple regression and ANOVA analyses are only able to consider observed variables (Tabachnick and Fidell, 2013).

Sample size is important to consider since SEM is a large sample technique. If the data set imported into the SEM analysis is too small, making the model ”under- identified”, standard errors may become inaccurate and the probability of obtaining wrong statistical estimates is increased. Naturally, one may ask ”what is then a sufficiently large sample for a SEM analysis?” To answer this question, several aspects that impact model complexity needs to be taken into consideration. For instance, large samples are needed if a highly complex model is to be correctly analysed because it will inevitably contain numerous parameters compared to that of a simpler model. Additionally, the estimation type can put more demanding requirements to sample size (several estimation options exist in current SEM software). Maximum likelihood is often chosen as a default estimator in SEM (Kline, 2011, chapter 7). Distributional characteristics need to follow a near normal shape imposing yet another requirement. If a low complexity model is to be analysed using a maximum likelihood estimator, a rule of thumb says that 200 cases is the minimum number of cases to include (Shah and Goldstein, 2006). With less than 100 cases, any SEM model (unless very simple) will not yield reasonable results.

A number of statistical tests need to be conducted to evaluate the fit of a SEM model to a particular data set. These tests must comply with threshold limits specified for each statistical figure, all of which have been subject to tightening in the past years due to a general lack of methodological rigour in previously published studies (Schreiber, 2008; Shook *et al.*, 2004). With this said, there are however four main reasons as to why statistical significance is of less importance in SEM compared to that of other statistical approaches such as multiple regression and ANOVA. First, SEM considers the evaluation of complete models, which implies a broader level perspective. Of course, statistical figures of separate effects can be assessed but ultimately the decision is whether to reject/modify the whole model. Second, remember that SEM is a large sample technique. Therefore, generated results have a higher possibility to be highly statistically significant; something which is quite trivial when the sample is large. In other words, Hays stated that if sample size is large, a statistical significant result can be extracted from practically any study (Hays, 1981, page 293). Third, latent variables are *estimates* that are subject to change if the estimation technique is changed or sometimes if other software solutions are used. The latter would bring about minor changes but these can however impact the p values enough to affect hypotheses testing (ex. p = *0.052* vs. p = *0.048* for like effects when testing at level of significance $\alpha = 0.05$). Fourth, more emphasis is put on estimating the magnitude of effects compared to the outcomes of the various statistical tests especially when it comes to research in behavioural sciences.

In this study, statistics has been applied not only to empirically test relations between factors but also to monitor process data and determine if changes to the ED system has resulted in an improvement. The latter technique is called statistical process control *(SPC)*.

3.6.3 Statistical process control

Walther Shewhart, a physicist, is commonly referred to as being the inventor of SPC in the early 1920s. He was employed at Bell Laboratories, USA at that time and was responsible for improving the quality of telephones (Shewhart, 1928). During his employment, he developed a theory on variation that provided the foundation for modern SPC. Essentially, the theory can be demonstrated by looking at five signatures spelling "CMS", written by the same person under like conditions (see Figure 3.3).

Figure 3.3: Signatures

Two observations are of importance: 1) even though the signature has been written under the same conditions and by the same person, the signatures do show some variation, and 2) the variation lies within certain limits. One could justifiably claim that the process of writing looks stable if nothing is known about the underlying process. What understanding of the signatures would a traditional statistical approach allow? All five signatures would have to be compared to a standard. Some of the signatures would not live up to the standard. If a statistical test was applied, one may find a signature that is much different from the standard. The essential feature of Shewhart's original theory on variation is that it can categorise variation according to the action needed to change it.

Classical statistical analysis techniques allow for natural variation but needs an aggregation of time series measurements that may impair decision making (Benneyan *et al.*, 2003). SPC is a branch of statistics that applies time series analysis methods together with a graphical depiction of data. The graphical representation can yield rapid and valuable insights in an intuitive fashion to lay decision makers. Researchers and practitioners can use SPC and its various tools (most importantly the control charts) to determine and communicate the impact of healthcare improvement efforts.

A control chart is a graphical presentation of time series data that is able to differentiate between common cause variation and special cause variation. Common cause variation can be found in any process data and is caused by chance. Special cause variation differs from common cause in the fact that variation in the process can be assigned to a cause outside the process. The control chart is an easy way to monitor a process over time while determining if special cause variation is present. Therefore, the technique is useful to determine if a process is within statistical control and, if so, if interventions to the process have had a significant effect. Note that if any special cause variation is present in the data, this should be analysed and removed before the work on improving the process can start. Also, control charts can be applied to highlight if a change to the process has resulted in an improvement.

Any control chart is comprised from some basic components that are depicted in Figure 3.4. A center line represents the average (or sometimes median) of the data sample corresponding to a completely stable process. Two horisontal lines, the upper control limit *(UCL)* and the lower control limit *(LCL)*, are also plotted. Data points that happen to fall outside the control limits indicate a special cause variation worth of further investigation. Special cause variation can also be indicated by the trend in which the process has developed over time. Such indication is when a predefined number of consecutive data points are decreasing- or increasing - in slope or if the center line is crossed a number of times. As goes for any kind of statistical tool, SPC can at times be misleading. There is the possibility for both type 1 errors (identifying special cause that is

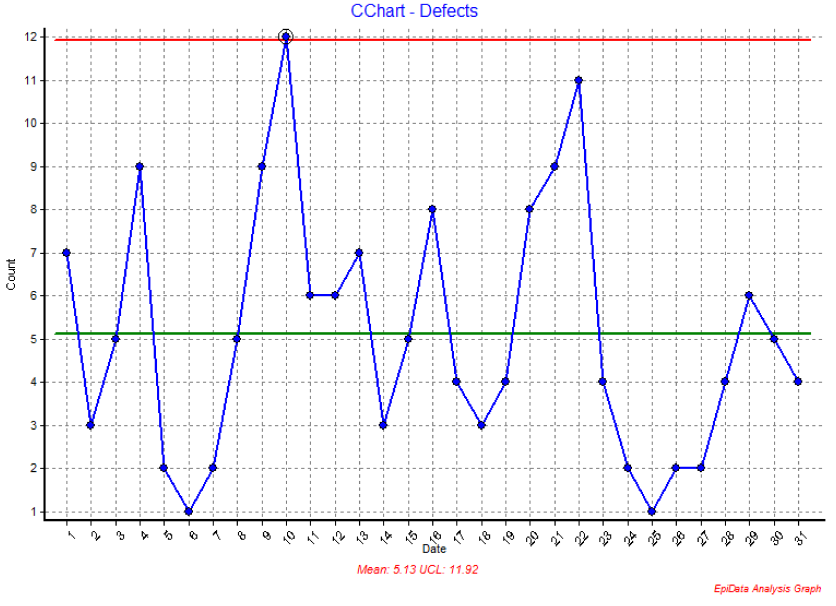


Figure 3.4: Random C-chart as example

merely common cause variation) and type 2 errors (special cause that is mistaken for common cause variation). Hence, Shewhart defined the control limits at three times the standard deviation from the center line. If the data follows a normal distribution, the span of the standard deviations in which the data points fall is depicted in figure 3.5. 68 % of all data points would fall within one standard deviation, 95 % within two standard deviations, and finally 99.73 % within three times the standard deviation (Hart and Hart, 2002).

In the healthcare setting, many performance measures are of the lower-the-better or higher-the-better character. Examples of important performance measures are laboratory turnaround times, patient satisfaction scores, medication errors, infection rates, number of patient falls, door-to-doctor times, counts of adverse events, and many more. One of many examples on the use of SPC in healthcare can be found in Morton *et al.* (2008), where monitoring infection rates at a hospital in Queensland, Australia lead to more efficient cleaning procedures and introduced early warning indications of new outbreaks.

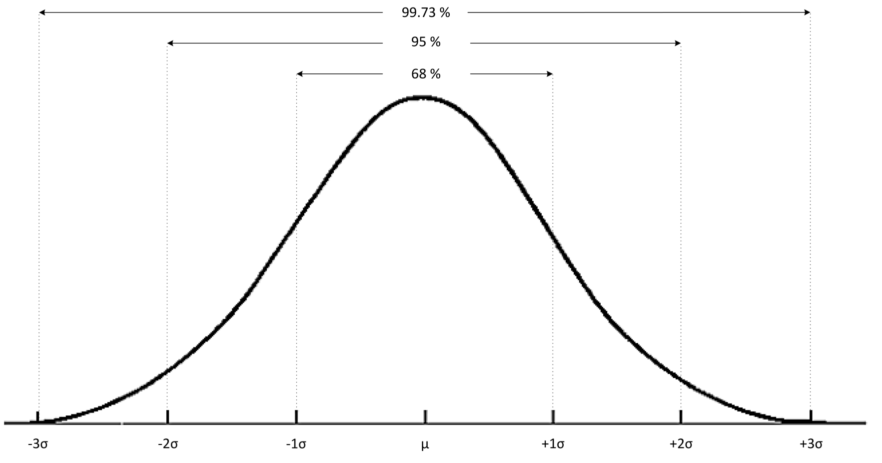


Figure 3.5: Normal distribution with probabilities

3.6.4 CR in relation to statistical modelling

When considering statistical modelling methods, this approach leans, to varying degrees, more towards an empiricist worldview than that of CR. Several contradictions to a critical realist view exists, some more severe than others. First of all, causality as a term is rather crude being basically a Humean notion of a constant series of events (Mingers, 2003). Quantified material is extracted, a mathematical procedure followed, and the most sensible links are promoted on the basis of a proposed model. The generated results are often used afterwards for predictive purposes. Such use is a problem seen from a critical realist point of view because one remain focussed on what can be empirically tracked and do not make the effort in understanding the underlying mechanisms that may explain the behaviour. Therefore, the researcher is left with poor grounds on which to conclude if correlations are indeed genuine or bogus.

Stability of calculated regression coefficients alongside corresponding statistical level of significance are automatically assumed to include only the factors that have major impact on the system being modelled. Implicitly, this means that any other factor not included in the model has minor to zero effect on the system's behaviour. In reality, these assumptions may prove to be wrong since there is no way of knowing what the impact might be if the number of factors under analysis is expanded. Another quite unrealistic trait of regression is the assumption of normal distributed data, unidirectional causality, linearity, etc. These assumptions are commonly known not to be representative of real empirical data. As explained in section 3.5, the real world contains very complex interactions showing counterintuitive non-linear behaviour (Forrester, 1971; Hunt *et al.*, 2012).

Therefore, the original idea of testing models for statistical significance is as follows: If a hypothesised model is assumed true (null hypothesis) then the alpha level is corresponding to the probability of obtaining observed values within a specified range (and reject the null hypothesis on the wrong basis, i.e. a type 1 error). Bearing the latter statement in mind, agreeing that a basic statistical model is "true" is deemed nonsense from a CR perspective, since the base assumptions are far from reality (Carlsson, 2010, chapter 15). All of the mentioned statements make it possible for competing models to fit equally well to a given dataset. Although guidelines to determine best-fits in a stepwise manner have been developed, models are made up from intuition, experience, and usefulness for the end-user (King, 2005). All these problems that now have been presented may point in the direction that a critical realist would completely abandon any kind of statistical testing, especially since empirical verification does not give cause to concerns when causal effects may be present but not reflected in empirical measurements. Some statistical modelling is however seen as powerful in CR but the purpose of such exercises must be re-considered.

Of main interest to the critical realist is the understanding of underlying mechanisms that drive behaviour in a system. Some statistical analyses can aid in this task, because these can be applied in an early exploratory phase, for instance in identifying specific patterns in larger datasets. Testing models that contain several hypotheses can provide results that point towards a number of constraints that are responsible for generating an underlying structure with a certain degree of non-randomness in the empirical data. The reader should however keep in mind that the results could be biased by the data treatment process. Still, detection of such patterns in datasets is quite difficult. Techniques such as exploratory factor analysis, cluster analysis, and regression are useful for provision of a starting point for further in depth investigations. Factor analysis and path analysis can take one step further in understanding patterns since these two combine the identification of common factors making up a construct and the application of differential equations that formulate the constructs analytically (Porpora, 1998). By regarding the problem as an artificially closed system where normally exogenous variables are controlled for, multivariate analysis can be very useful.

3.7 Validity

It is paramount for the researcher to address the question of whether the conclusions made are valid and reliable. Had this study been similar to a clinical experiment, validity could be sought by repeating a study to see if the same results were obtained. In organisational research, such experiments are no option since the system in which these experiments were to be conducted would be in constant change. To argue that the conclusions from this study are valid, thus making a scientific contribution implies a workaround that the system under analysis is not a closed one. Since the critical realist acknowledges that reality cannot be described to perfection, the aim is to add to further understanding of what is already known about a given phenomenon. Because identification of underlying generative mechanisms is the main focus of the study, the concepts behind establishment of validity can be described through an extension of Figure 3.6.

Identification and suggestion of solutions belongs to a scientific way of thought and the techniques mentioned in this chapter are not a contradiction of CR. It is the author's belief that the approach applied to answer the stated research questions comply to the frame of CR. Some points that provide arguments to this are:

- An array of different applied research methods (direct observations, semi-structured interviews, case studies, system dynamics modelling, and advanced statistical modelling) emphasises the coherence with CR due to the fact that the research problem determines which methods are suitable and not vice versa.

- The study can be regarded as somewhat interdisciplinary as theory from management science, knowledge management, emergency medicine, and philosophy are all included within the study. Digging into deeper layers of understanding through more fundamental philosophical reflection or even on a biological level is beyond the scope of this study.

- Approaches that are advocated by CR were utilised in order to answer the research questions adequately. Such approaches encompass interrelated induction- deduction, abstraction, and retroduction (Buch-Hansen and Nielsen, 2005).

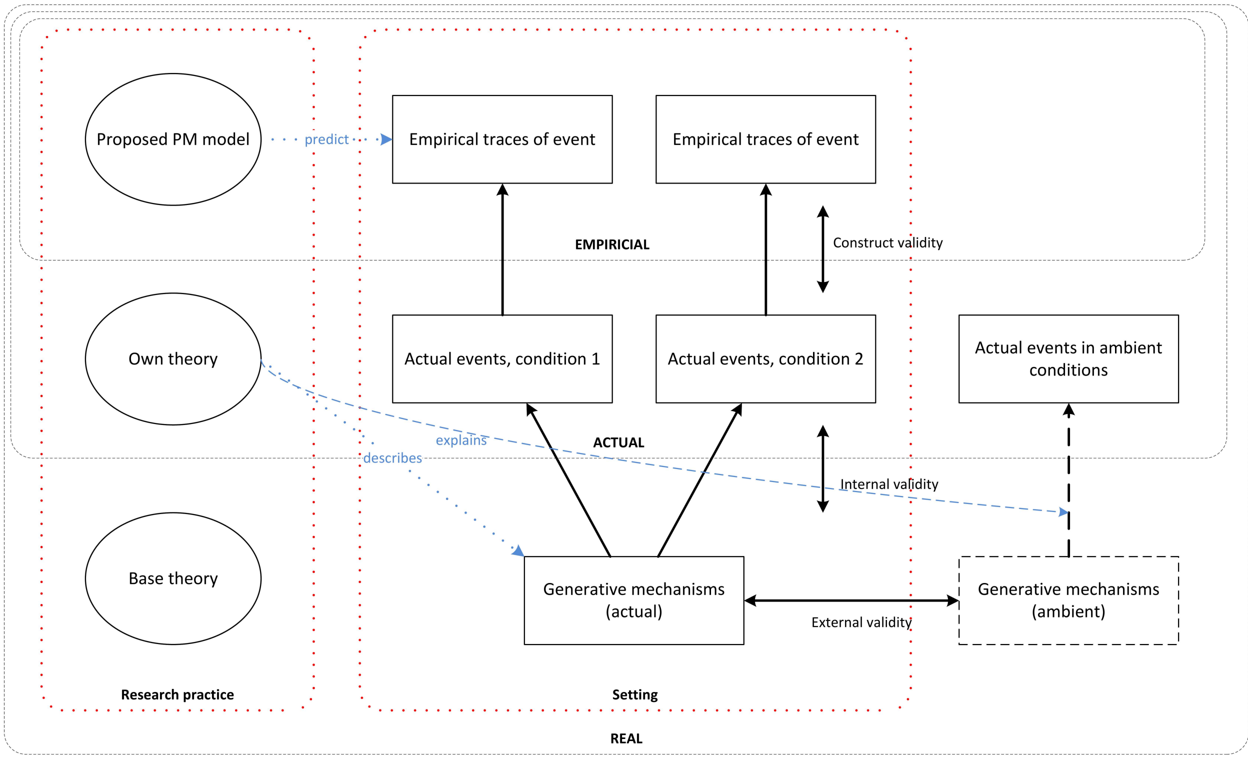


Figure 3.6: Research process and validity from a CR viewpoint (Johnston and Smith, 2008)

- Fundamentally, critical realists acknowledge that all knowledge is fallible while a separated reality exists. Such a perspective is revealed through the manner conclusions are built throughout the research that has been conducted. The main focus has been to find conditional learning points leaning towards truth, since it is recognised that establishing absolute truth is an utopian premise.

- Clarity of concepts is promoted as being extremely important by Bhaskar, especially when working with social factors (Easton, 2010). Research that is not built solely upon mathematical formulations definitely needs thorough explanation so that applied concepts and terms are defined in an unambiguous manner. More fertile grounds for future research probabilities comes with clearer definitions.

The above bullets underline a clear distinction from a positivist worldview which see reality as fixed, directly measurable, knowable, and where there is but a single truth (Rubin and Rubin, 2012). Even though some statistics have been applied to explore relationships between isolated areas of interest, the main focus remained an investigation of the generative mechanisms which are responsible for certain phenomenon to surface. On the contrary, the research is also very much different from traditional social constructivism, since no real attention has been given to explaining how technology or knowledge has been shaped by social factors. The research done in this study falls somewhere in between as it implicitly reflects the assumption that it is an independent yet interwoven natural and social reality that is under analysis.

CHAPTER 4 EMPIRICAL FOUNDATION

This chapter will present the cases that have been included in the study. First, an introduction to each of the cases will be given. Explained next are the details of how the qualitative data was acquired. How the quantitative data was obtained from each of the cases wraps up the chapter.

4.1 Included cases

Two EDs have financially supported this study, thus making them primary cases. These are the EDs at Herlev and Nordsjælland Hospital respectively, both situated in the Capital Region of Denmark. Two co-supervisors have taken an active interest into the study in their roles as senior consultants at the beforementioned EDs. Both co-supervisors have introduced the author to the daily routines, provided assistance in data retrieval, set up interview appointments with key personnel, and provided valuable feedback on both conference- and journal contributions. Additionally, the Department of Emergency Medicine at Beth Israel Deaconess Medical Center *(BIDMC)* in Boston, MA constituted a secondary case where a part of the study was conducted. The author collaborated closely with a senior attending physician who, like the two co-supervisors, provided assistance in retrieving data, critically reviewed the author's work and allowed for direct observations during various shifts in the ED.

4.1.1 Primary cases

In the Capital Region of Denmark, the EDs at Nyt Hospital Herlev- and Nordsjælland respectively make up the main cases of the study (see Figure 4.1 for geographical location). Both hospitals are undergoing major structural changes as dictated in the national hospital reform published in January, 2007 (DHMA, 2007). On a national basis, many hospitals will no longer provide emergency care treatment. However, this does not impact the two primary cases.

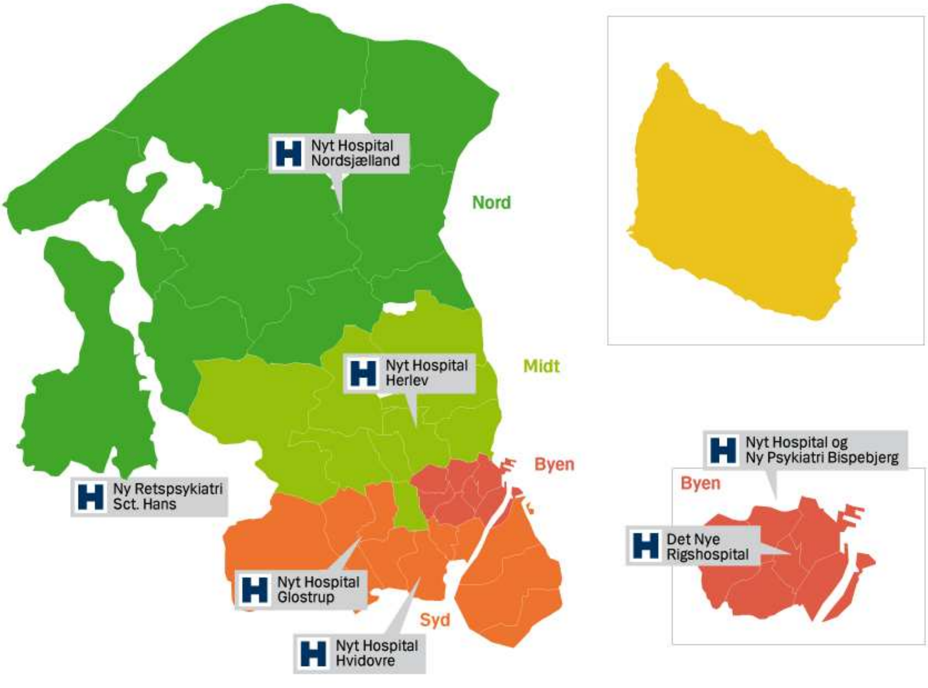


Figure 4.1: Future hospitals in the Capital Region of Denmark

Nordsjælland Hospital consists of three facilities that organisationally function as one unit. In 2020, Nordsjælland Hospital will move from its existing buildings to a single new facility in Hillerød, where hospital treatment for approximately 310.000 citizens is to be provided. Today, the main ED at Nordsjælland Hospital is located in Hillerød and is a joint unit consisting of an urgent care clinic, trauma bays, observational unit, secretariat, and reception; that is identical to the FAM concept explained in chapter 1. Two urgent care clinics are also connected to the hospital. These are situated at Frederikssund Hospital and Sundhedshuset in Elsinore. The ED staff at the main site in Hillerød currently consists of five specialty physicians of which three are specialised within emergency medicine. These physicians are responsible for coordinating and supervising not only all acute patient procedures but they also act as mentors for medical residents undergoing their education. Complex clinical decision making is done in collaboration with the hospital's other specialty physicians.

Herlev Hospital is to be greatly expanded in order to cope with more patients, especially within the field of cancer treatment (REGH, 2009). Today, Herlev Hospital covers an area of 425.000 citizens but this number will increase by 50 % when the expansion is completed in 2018. The ED is currently structured slightly differently than at Nordsjælland Hospital. The interdisciplinary teamwork is grouped in three teams denoted 1) medical, 2) surgical, and 3) urgent care. The medical team primarily treats patients suffering from illnesses related to oncology, haematology, neurology, and other medical conditions such as infections, impaired metabolism, and lung diseases. The surgical team focuses on gastro surgery, gastro medical conditions, and urology. Finally, the urgent care team treats both smaller and major traumas and illnesses connected to cardiology and orthopaedic surgery. Similar to the ED in Nordsjælland, specialty physicians are summoned when more complex medical decision making is to be conducted. An example of this is when ED patients are admitted to an inpatient ward.

Both primary cases treat approximately the same amount of patients each year (despite the catchment areas being different), refer to the same region, and follow the Danish Emergency Process Triage *(DEPT)* scale (Skriver *et al.*, 2011). Although differently organised, both primary cases were willing to participate as sponsoring partners of this study. They share an interest in expanding their understanding of how to assess their department's performance and to learn more about how the effects of potential interventions can be tracked. The author has been introduced at both EDs to the daily routines of all staff professions (within the ED and supporting departments), conducted semi-structured interviews, obtained data from diverse databases. The author has also had regular status meetings with both co-supervisors.

4.1.2 Secondary case

Contact was established with the Department of Emergency Medicine at BIDMC located in the Longwood Medical Area of Boston, MA. The visit to BIDMC was conducted in a timespan from October 1^{st}, 2013 to April 15^{th}, 2014. BIDMC is rated amongst the best hospitals in the United States and is one of several major tertiary level one teaching hospitals situated in Boston affiliated with Harvard Medical School. The ED at BIDMC differs from the two other cases in various ways. First of all, emergency medicine as a specialty is well implemented and recognised among hospital staff. Having specialised ED physicians is an important asset to BIDMC as more patients can be discharged upon completed treatment in the ED compared to otherwise (Anderson *et al.*, 2006; Scott *et al.*, 2009). In its position as a level one hospital, more complicated patients are transferred from other communal hospitals in Boston and its vicinity. An observation made by the author was that the staff to patient ratio was about one to two. Therefore, it is hypothesised that more attention to the individual patient can be given; something that improves patient perceived experience. Having great human resources is costly and should be seen in light of the US health care system being primarily pay-for-performance, whereas the Danish is primarily a tax paid system. However, the patients that were admitted to BIDMC were often complex cases that required more resources before a diagnosis could be concluded and treatment finished. The ED employees are organised in three teams consisting of medical residents, a senior physician, nurses, and technicians. The teams are multidisciplinary and can therefore provide treatment regardless of the patient's chief complaints. From a performance measurement perspective, the ED at BIDMC is an interesting example as they have developed a customised real-time monitoring tool consisting of several performance measures. Data is registered automatically and the IT platform is integrated to enable full patient anamnesis, containing information about, for instance, laboratory tests on the spot. Data is thus measured with high degrees of precision, ensuring construct validity. In time, imaging results were to be integrated as well.

During the stay at BIDMC, focus was put on answering the second sub-question of how the various interconnections between the measures interact. It was realised that the attention should be moved from looking at performance measures to the underlying process which the performance measures reflect. Therefore, focus was focused upon how simulation, more precisely SD, could provide answers about these interconnections that are influenced by time delays and complex non-linear behaviour. A prior problem statement needs to be formulated to set up the boundaries of a SD model. ED crowding was chosen as the main topic, since this problem is worldwide and is also of great interest to the case hospital. BIDMC may in hindsight not be the optimal hospital to investigate ED crowding effects as this is not perceived there as being a big problem. Reversing this viewpoint, it is interesting to identify what it is that BIDMC does differently compared to elsewhere. The SD model comprises an analysis of the admission process at BIDMC with special emphasis upon an emergency protocol named "Code Help". This protocol can be activated when predefined threshold values for ED census are met. Activation of the protocol alerts other inpatient wards of ED congestion, resulting in either a relocation- or more rapid discharge of current inpatients. Patient flow is improved and ED congestion may be resolved. The outcome of the external stay became two conference contributions and a journal paper.

4.2 Data acquisition

Data from the included cases has been obtained in different formats. These can be grouped into being either of *qualitative* or *quantitative* character.

Qualitative data

In this study, non-numeric data comprises *direct observations* and *interviews*. Both categories are highly important to be able to create a more nuanced picture of the research problem at hand as all information cannot be transferred into numeric form.

Direct on-site observations served as a means to gain an initial thorough understanding of what goes on in the ED. These observations was used to map the various generic processes and to identify which actors are present together with their corresponding responsibilities. Observations were done of all included cases as the organisations and procedures are not identical.

Interviews were conducted as semi-structured with questions posed in a deductive manner and as more loose discourse. Through interviews, empirical material can be collected with as little interference as possible (Tong *et al.*, 2007). A semi-structured interview allows the interviewee to comment upon given questions, thus resembling a more open conversation as opposed to a controlled interview where precise answers are sought. Semi-structured interviews were chosen for different purposes. These included identification of relevant performance measures, their interconnections, and validation of incremental findings.

Applying a model thinker's mindset, information available to the researcher about a given system exists primarily in the form of qualitative data being conceptions of the system and descriptive written information (Luna-Reyes and Andersen, 2003). As mentioned in section 3.5, Forrester refers to these data as mental models that exist in the minds of the actors in the system (Forrester and Senge, 1980). Such information is invaluable especially when trying to connect quantitative measures according to causal influences. The funnel of information by Forrester and Senge demonstrates the value and richness of information in the quest for understanding system behaviour (see Figure 4.2).

An approximately evenly distributed number of interviews were conducted with all of the included cases with key representatives from each profession. By doing so, a representative picture of the system seen from multiple angles was obtained. The duration of the interviews was time restricted as not to keep the interviewee from her original tasks. At times though, interviewees allowed for more detailed explanations.

Quantitative data A number of different data repositories were applied to carry out the study's quantitative analyses. Through testing the simulation models and hypotheses on real life data, validity was added to the findings. Data was extracted from both a number of *internal databases* at the included cases and an *external survey* from a research center in the Capital Region of Denmark. An outline of the acquired quantitative data follows in the following section.

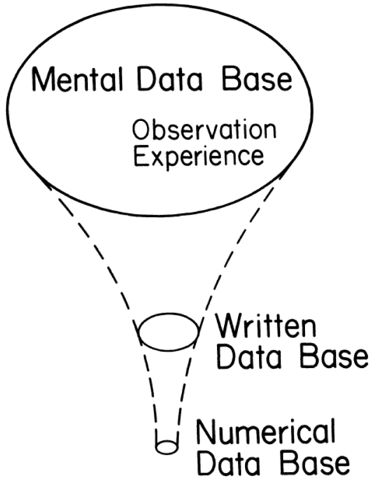


Figure 4.2: Funnel of information (Forrester and Senge, 1980)

Internal databases comprise several systems that each contains needed information on the selected performance measures. Unfortunately, these systems are not integrated to an encompassing whole at the primary cases. The primary data sources therefore include a total of five separate systems containing information on the performance measures to be included in the final performance measurement model. As Herlev- and Nordsjællands Hospital are both situated in the Capital Region of Denmark, it was observed that their data retrieval systems were identical. The databases are listed in Table 4.1.

Besides these databases, internal patient- and employee satisfaction survey results were obtained (only for Nordsjælland Hospital). Patient surveys were conducted as semi-structured interviews with a small sample of patients during a one-week period. The employee satisfaction survey was conducted by an external consultancy company. The most recent employee satisfaction survey was distributed during the summer of 2014 and obtained a response rate of 85 %. It contained 65 different questions regarding both physical- and psychological work factors.

Data was stored more centrally at the secondary case hospital. This made it easier to extract exactly the performance measures of interest for a user specified time interval. Two systems allowed for easy retrieval of many of the performance measures. These were the *ED dashboard* and the *ED time log*. The latter is

Database	Description
OPUS Arbejdsplads	A patient administrative system similar to the electronic patient journal
GS!Åben	A support module for OPUS Arbejdsplads
Labka	Contains information on laboratory tests, blood work and more
CETREA	ED clinical logistics overview usually used in real-time
RIS/PACS	PACS stores data from medical imaging devices as images while RIS contains information on patient demographics, examination orders, and scheduling.

Table 4.1: Primary databases in use at the two primary case hospitals

an integrated part of the ED dashboard. The ED dashboard is a real-time patient data tracking system used for managing workflows. Clinical-, registration-, and time-stamp data are all imported into this system that can be readily monitored by all employees at the hospital with large monitors and tablets. Time-stamps are registered automatically in the ED time log, for instance by the implementation of an automatic recording algorithm behind a standard protocol. This could be the triage time containing a start- and finishing time-stamp. Here, all measurements on vital parameters for a given patient are registered directly in a spreadsheet. Opening a new spreadsheet records the starting time, whereas closing it records the finishing time.

External databases include one major patient satisfaction survey for all EDs in the Capital Region of Denmark (Rimdal and Soerensen, 2012). This survey, specifically targeting patients in the ED, was developed by the *Center for Patient Experience and Evaluation* (formerly *Unit of Patient-Perceived Quality*) in 2013. The survey was the first of its kind, involving only EDs, and reached out to 1940 patients by telephone in the period between February 20^{th} and March 4^{th}, 2013. A total of 23 questions were posed concerning information levels, guidance and parking, reception, waiting times, communication with staff, examination and treatment, physical settings, and general contentment. 14 of the questions were used as empirical foundation for the statistical analysis presented in paper three (see section 5.3).

CHAPTER 5

INCREMENTAL DEVELOPMENT OF THE MODEL

This chapter will guide the reader through the step-wise development of the final recommendations. Four journal articles serve as documentation for the progress throughout the study. Not all published work is presented in this thesis since some of these are earlier works that were altered into journal format at a later stage. For a full list of all published work, please revisit the list of papers. All articles, whether targeted for journal- or conference publication, have been submitted to internationally known scientific publishers. The papers are presented as to chronologically guide the reader through the maze of considerations and incremental work towards the final performance measurement model. Main contributions of the separate articles are promoted to construct a logical row of arguments leading towards the final recommendations.

The first article examines recent review articles on recommended, analysed, and debated performance measures for ED overall evaluation. The outcome of the literature review was an extract of performance measures that provided the very foundation for further analysis of how these were connected. The second article opens up for a single performance measure to analyse underlying leading- and lagging factors in depth. Sickness absence was the chosen factor due to its intangibility and complexity. Statistical hypothesis testing was applied to explore which measures should be tracked thoroughly to hint at future sickness absence levels. The third article broadens the scope of analysis to undertake the domain of patient satisfaction through a more advanced statistical approach, namely SEM. The fourth article abandons the statistical approach in favour of simulation. Here, a SD model is developed on a North American case to examine ED crowding, especially in relation to return visits. All the articles provide a joint foundation which results in the proposed performance measurement model.

5.1 P1: Identification of performance measures

Background
Concerns on how to measure overall performance in EDs are widespread. This comes in natural prolongation of the many restructuring initiatives that are being implemented in EDs worldwide. The new organisation of EDs must be seen in light of an aging population, typically suffering from multiple illnesses. Such patients are highly complex and need thorough examination before an initial diagnosis can be concluded. Much emergency medicine literature addresses the issue of performance measurement seen from an operational viewpoint with special emphasis on ED waiting times and crowding (Kelman and Friedman, 2009). Initiatives targeted to reduce waiting times in EDs have mostly been neglecting other facets of ED quality of care. A one-sided focus on reducing waiting times will most likely leave other aspects of ED quality in care to deteriorate. Thus, applying a balanced performance measurement approach that captures multiple aspects of broad ED performance is warranted. Several calls for performance monitoring frameworks to insure quality and efficiency in healthcare delivery have been made (Schull *et al.*, 2011). Recommendations on which performance measures to monitor not only diverges but also suggest larger spans of measures that compromise overview.

This study's aim was to condense concurrent emergency medicine literature on recommended performance measures for ED macro-level assessment.

Methods
A systematic literature review of internationally peer-reviewed review articles in the databases of PubMed, Cochrane Library, and Web of Science was carried out in the period of April through July, 2012. PRISMA guidelines were applied (Liberati *et al.*, 2009).

Results
1314 titles was the total gross outcome of all searches performed in the mentioned databases. 46 of the results were unique. These were subject to an inclusion/exclusion criteria run-through after having read titles and abstracts. The search strings and resulting hits are presented in Table 5.1. PubMed differs between the singular and plural forms which explain the distinction made in parentheses.

# Search	Variable search string	# Hits
1	performance measure/(performance measures)	13/(46)
2	performance indicator/(performance indicators)	13/(36)
3	quality measure/(quality measures)	24/(47)
4	quality indicator/(quality indicators)	19/(200)
5	quality assessment	67
6	quality evaluation	2
7	performance assessment	15
8	performance evaluation	13
9	quality assurance	657
10	quality improvement	9

Table 5.1: Search strings and resulting hits of literature review

Articles would be included in the final review if the following criteria were fulfilled: 1) the article's main purpose was to discuss, promote, or analyse performance measures that best reflect ED performance, 2) the article was a review article, and 3) the article included macro-level performance measures for overall ED performance assessment. Exclusion criteria included 1) reference to specific patient groups or illnesses, 2) a different setting than the ED, 3) not touching upon performance measurement, 4) primary goal to investigate evidence behind selected indicators, 5) description of how performance measures is to be applied now and in the future, 6) criticism towards particular performance measures, and 7) a language different from English. A flowchart of the filtering of articles can be found in Figure 5.1.

14 articles met the inclusion criteria. These were read in full to note all performance measures mentioned. How many times a performance measure was mentioned indicated a degree of consensus that the particular measure was important for ED assessment. The top 25 % of the recommendations is presented in Figure 5.2. All identified performance measures were afterwards sorted by the authors in three groups being 1) patients, 2) employees, and 3) operations (Traberg *et al.*, 2014).

Discussion

55 performance measures were identified in the included articles. The articles differed in their approaches and recommended tracking of everything between four to 44 performance measures. Most in focus were the operational timestamps indicating great attention to timeliness of care. Patient centeredness has gained ground as patient satisfaction also proved to be of great importance, especially in the American studies where healthcare typically is privatised. Employee related performance measures were mentioned in only two of the 14 ar-

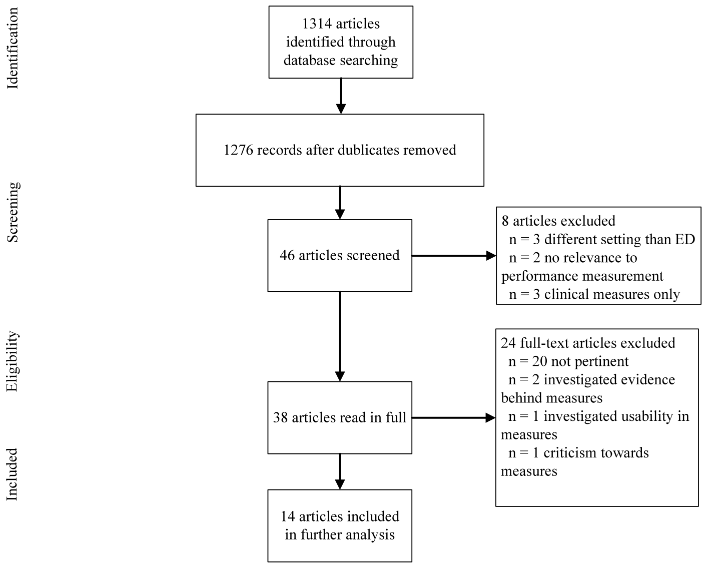


Figure 5.1: Flowchart of article selection following PRISMA guidelines

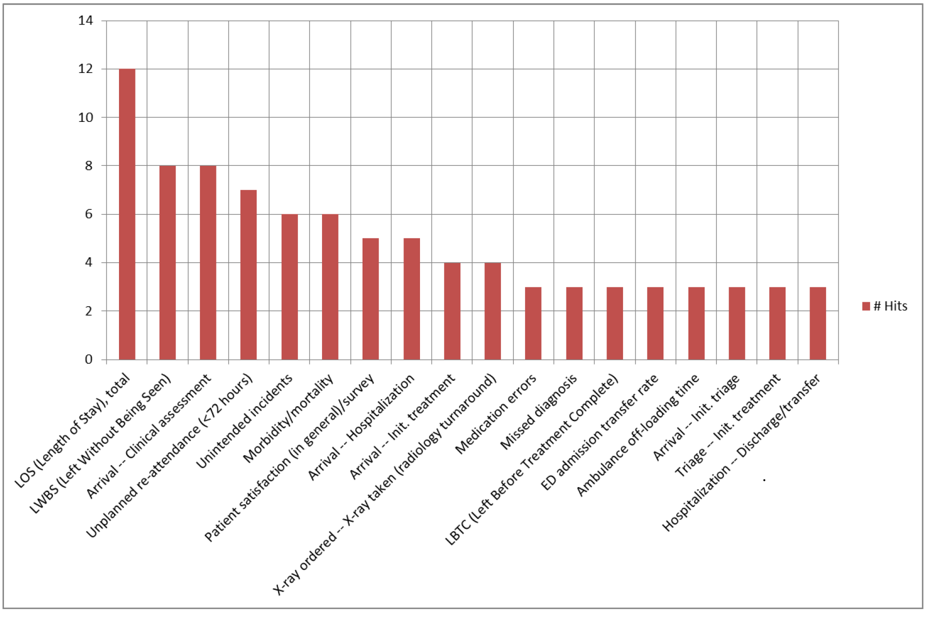


Figure 5.2: Top 25 % of recommended, discussed, or analysed performance measures in included literature reviews

ticles, suggesting employee relations being of less importance. Retaining experienced staff in the ED should be a top priority issue, since such resources are crucial cogs for any ED to function. Great amounts of data are being recorded so there are plenty of performance measures to choose from. However, which performance measures appear most rich in information is still up for debate. A deliberate decision not to include condition specific performance measures was made early in the study. Such a decision is based on the objective of wanting to measure general ED performance regardless patient groups. On the other hand, it must be assumed that inclusion of specific performance measures targeted for monitoring aspects in care of representative patient groups could promote deeper understanding of the EDs performance. Concurrently, establishing consensus on the various performance measures definitions is another key step towards data-driven management. Definitions are at times changed making time-series analysis difficult, thus compromising construct validity.

Conclusion

This systematic literature review identified which performance measures are analysed, discussed, or recommended in recent emergency medicine review articles for overall ED performance evaluation. Consensus on which performance measures should be closely monitored remains a challenging task. Besides a precise definition of each performance measure, decisions on when and how these are recorded should be standardised.

Experience gained

The systematic literature review revealed which performance measures should be included as the backbone of the final performance measurement model. All performance measures were categorised into three aggregated clusters being related to 1) patients, 2) employees, and 3) operations. Agreement on which performance measures to use for overall ED performance assessment is currently a hot topic. Clear definitions on each measure are still to be agreed upon, making the challenge of benchmarking impossible. Hence, the final performance measurement model is to be applied for internal process monitoring exclusively.

Having identified the relevant performance measures that are essential for ED performance evaluation, the next phase of the study can commence. As the ED is a highly complex and dynamic system, understanding the effects of performance enhancing initiatives are dependent on knowing how the performance measures interact.

5.2 P2: Unravelling a single performance measure; sickness absence

Background
The healthcare sector has been experiencing several rounds of budget cuts worldwide since the economic recession began in 2008. A large expenditure of hospital budgets is for salary, implying that dismissing clinical staff was inevitable. The mantra ”doing more with less” had taken first place since the ration between employees and patients increased disproportionally. How then is a well-functioning ED secured? One answer could be through employee retention. An important, yet complex, indicator for employee well-being is *sickness absence*. Lowering current sickness absence levels to a theoretical minimum is an important step in high utilisation of available human resources. But what are the main determinants for sickness absence? How can it be improved?

This article travels deeper into the very structure of a single employee related performance measure, namely that of sickness absence. Management of sickness absence rates is sought conceptualised through the development of a proactive decision support tool directly linked to employee satisfaction survey results. With such a decision support model, the ED decision maker can target interventions for lowering sickness absence levels with higher fidelity.

Methods
A literature study of contemporary journal articles concerning factors directly hypothesised to influence sickness absence was performed. The literature study was supported by 1) three semi-formal interviews with human relations experts at three case organisations in order to capture potential drivers for sickness absence not mentioned in literature and 2) conclusions from the ASUSI project (a large national project, comprising many smaller projects, on sickness absence factors in Denmark) (ASUSI, 2008). All factors identified were presented in an encompassing overview that categorises factors into either voluntary- or involuntary absence. The authors introduced the term *employee absence* to rightfully cover absence caused for motivational reasons. Statistical hypothesis testing was conducted based upon past employee satisfaction survey scores and corresponding employee absence levels. Furthermore, the level of social capital was quantitatively assessed. Empirical data material was provided by four Danish public hospitals and one private pension fund (for evaluation of generalising potential). Positively identified factors had to 1) show a statistically significant Pearson correlation level in connection to employee absence rates and 2) be in agreement with included literature conclusions. Any measure that complied

with both conditions was included in the proposed decision support tool.

Results

The literature survey with supporting semi-structured interviews revealed a gross portfolio of drivers for sickness absence rates. Categorisation of the drivers was split into voluntary- and involuntary absence reasons (as presented in Figure 5.3).

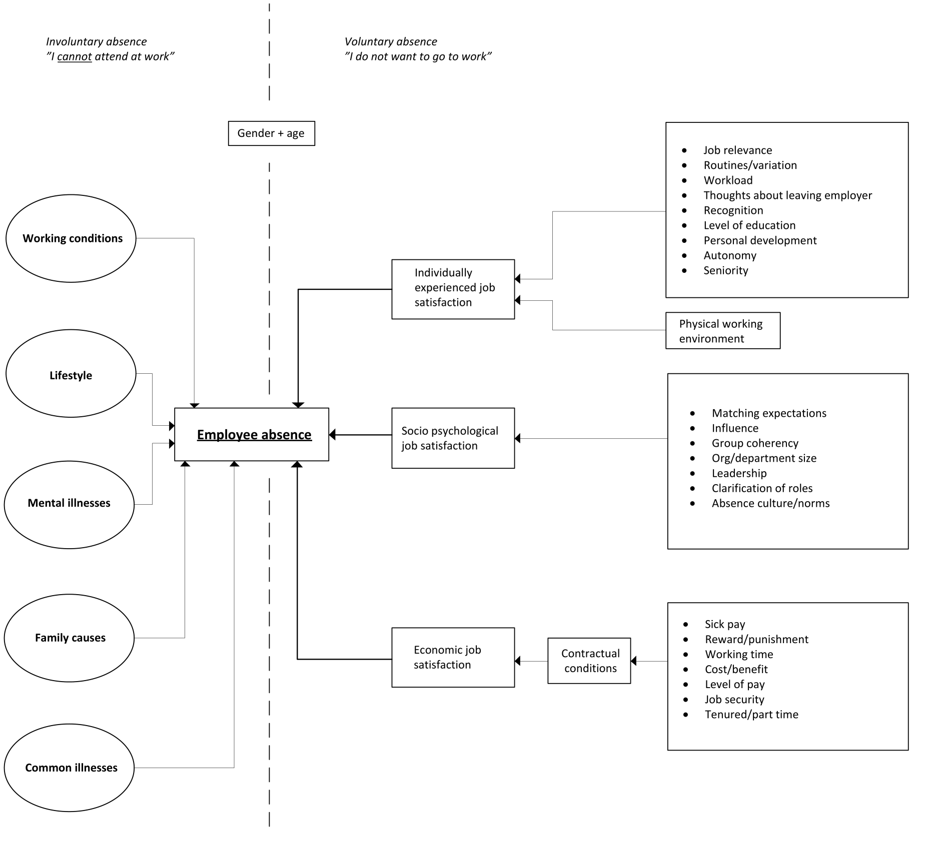


Figure 5.3: Overview of factors with a direct influence on employee absence rates

The most significant results determined by calculated Pearson correlation coefficients while in compliance with literature conclusions were 1) *fairness*, 2) *job security*, 3) *total level of employee satisfaction*, and 4) *thoughts of leaving current position*. Calculations of social capital levels were found to have an inverse relation to employee absence rates. Altogether, four overall constructs obtained sufficient statistical power and was in compliance with literature conclusions. Three of these constructs fit in the concept of "social capital", whereas the fourth concerns general employee satisfaction. The latter construct must address the 1) satisfaction with working conditions, 2) feeling of job security, and 3) whether the employee has thoughts of applying for a job elsewhere. By addressing these elements in employee satisfaction surveys, the scores can be transferred into a decision support model to visually indicate critical levels that may result, or explain, current absence levels. Figure 5.4 presents the four categories, underlying factors, explicit questions with appertaining scores, the calculated average scores that determines the current level (being either high, intermediate, or low), and lastly a colour code for the level.

The determinants' levels can be quickly assessed in the "employee absence level diamond" presented in Figure 5.5 which automatically transforms the scores from the check list in Figure 5.4. The diamond form in the center of the model shows the categories' status. The rectangles in the periphery make up the levels for the separate factors. The colour in each corner of the diamond shape is determined by the average scores of the underlying factors.

22.02.2012		Projectname:								
		Made by:		Low degree	- Some extent -			High degree		
Category	Factor	Colour	Note	1	2	3	4	5	Score	Average
Element 1, social capital	Fairness		Are conflicts solved in a just manner?					x	5	
			Are you acknowledged for a job well done?					x	5	
			Do the management always treat suggestions seriously?					x	5	
			Do work tasks get fairly distributed between employees?				x		4	4,8
Element 2, social capital	Reliance		As employee, are you able to express your opinion?			x			3	
			Do you have confidence in the way the top management manages the work place?			x			3	
			Can you trust the announcements made by the management?		x				2	
			Do the managers bear confidence in their employees?			x			3	2,8
Element 3, social capital	Cooperation		Is there a great working relationship between your colleagues?					x	5	
			Do you experience a good cross sectional teamwork?				x		4	
			Do you have a good professional relationship in your department?					x	5	
			Do you get aid and support from your colleagues when needed?			x			3	4,3
General satisfaction	Job security		Do you feel safe regarding your work-related future at the current employer?	x					1	
			Do you feel secure in your current employment?		x				2	1,5
	Working conditions		Are you content in terms of ergonomics? (lifting, working postures etc.)?		x				2	
			Are you content with the ergonomic tools used to lift heavy loads?		x				2	
			Are you content regarding the safety procedures (i.e. handling of needles, chemicals etc.)?				x		4	
			Are you satisfied with your working conditions in general?			x			3	2,8
	New job?		Do you plan on staying in your present job?	x					1	1,0

Figure 5.4: Check list sheet of determinants for employee absence

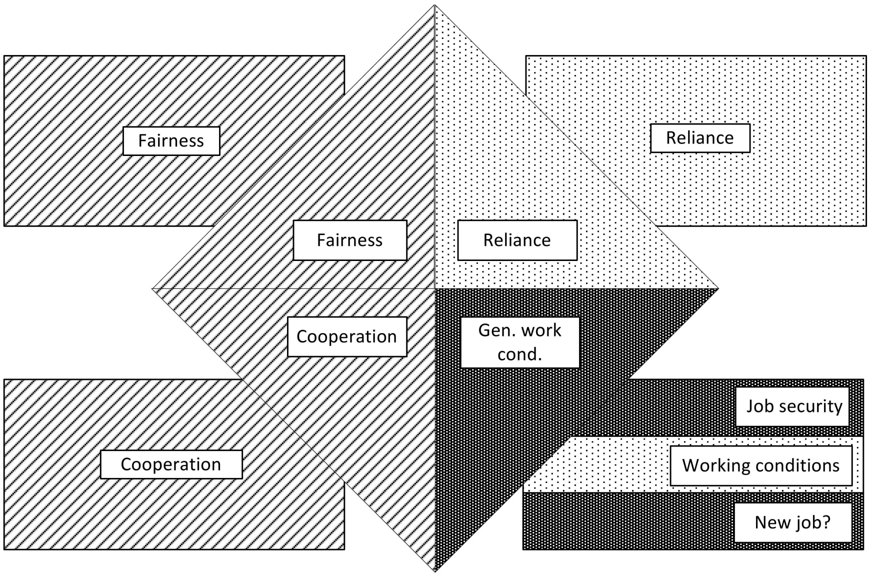


Figure 5.5: Determinant's status represented by shaded areas indicating focus area

Discussion

Questionnaires possess valuable information that can, with some data mining effort, be applied for more accurate managerial interventions. The proposed decision support tool is not a stand-alone solution, since it has not yet been tested in practice and may miss other important drivers. Other factors that may be rooted in an organisation's culture can reasonably well be hypothesised to have a strong impact on what is considered "acceptable" levels of absence. An example is mirroring your nearest manager's absence patterns (Schreuder *et al.*, 2011). If the manager is often absent, it can be hypothesised that so will the employees under him. With the decision support tool suggested, the manager will have support to articulate given areas of focus that the employees see as potentially pressing issues and is responsible for their absence rates. Interestingly, performance improvements and lowered employee absence rates was reported at Frederiksberg Hospital in 2011 in the wake of a major round of layoffs. (Drachmann, 2011). This was however a short- term effect that eventually tipped towards an increase in employee absence, possibly due to high workloads. Such an effect suggests that employee absence behaviour is indeed a complex phenomenon that displays non-linear behaviour. With a more sound empirical foundation, multivariate statistical analysis is enabled, making it possible to extract even more information from the data material at hand.

Conclusion

A comprehensive overview containing drivers for employee absence was developed through a literature survey with supporting semi-structured interviews. Quantitative assessment through statistical hypothesis testing was conducted to reveal main determinants for employee absence. The positively tested determinants were formulated into a decision support model for pinning down where to launch high impact interventions that lowers current employee absence levels.

Experience gained

Opening up for one indicator in the performance measurement model revealed a large sub-network of interconnected factors that can impact system behaviour. This study was a first attempt to apply basic statistics so that determinants for employee absence rates becomes clear. Three elements that together constitute the term "social capital", together with all-round employee satisfaction, were found to follow employee absence patterns. By extracting selected scores from past employee satisfaction surveys and comparing these with concurrent absence rates, the management team is enabled some insight into early warning symptoms of future absence rates. These findings were presented in a decision support model that not only holds valuable information for staff retention but can also be intuitively presented at staff meetings. A key conclusion was that a precise identification of small subset of determinants can ease the overview for the manager. The granular level of detail is comprehensive for many human behavioural factors; here demonstrated with employee absence at scope. The findings made in this study are naturally implemented as an incremental sub-module in the overall performance measurement model.

Analysing each and every performance measure in a similar manner would be extremely time-consuming. Testing more hypothetical causal connections in a single model can be done through the use of more advanced statistical methods such as causal modelling, or more precisely SEM (Kline, 2011). With more complex quantitative models comes the demand of comprehensive empirical material. With a proper data foundation, more robust conclusions on causal mechanisms can be drawn.

5.3 P3: ED patient safety and satisfaction drivers

Background

Few will argue against the meaningfulness of monitoring patient satisfaction as patients are indeed a main stakeholder in the healthcare sector. Despite the general agreement, which aspects are considered most relevant to promote ED patient satisfaction is still an unsolved puzzle. The typical ED patient has no prior diagnosis to his visit and will present himself with unique subjective descriptions of primary complaints. To which degree his expectations of service level is met will determine his final opinion of the total ED visit. Practitioners and academics have proposed many hypotheses of factors being paramount drivers in patient satisfaction – hypotheses based on qualitative research approaches with very little supplementary quantitative testing.

Traditional hypotheses testing specify default models and assume no measurement error. Advancing to multivariate statistical techniques, like structural equation modeling *(SEM)*, enables testing of an a priori model consisting of observed variables and latent constructs grounded in literature. Furthermore, measurement error is accounted for. The model developed can shed light on the power of the relations between the variables if specified correctly. This study develops an integrated framework that includes four constructs reflecting key aspects of patient care in the ED.

Methods

SEM is a statistical method comprised of path analysis, factor analysis, and multiple regression. The path analysis step displays directionalities between the selected variables, i.e. the hypothesised causal relations. An exploratory factor analysis *(EFA)* is then utilised to make explicit the hidden structure (dimensions) of a set of variables. It reduces the attribute space from a larger number of variables to a smaller number of aggregated constructs and as such is a "non-dependent" procedure (that is, it does not assume a dependent variable is specified). The confirmatory factor analysis *(CFA)* can be initiated after the EFA and serves the purpose of testing whether the empirical data fit a hypothesised measurement model. This hypothesised model is based on a priori hypotheses from included literature. In a CFA, several statistical key figures are used to determine how well the model fits to the data. A good fit between the model and the data does not imply a ”correct” model, or even that it explains a large proportion of the covariance. A ”good model fit” only indicates that the model is statistically plausible. The final SEM step derives unbiased estimates, regarding error terms, for the relations between the latent constructs. MPlus version 6.12

was applied for conducting the analysis (Muthen and Muthen, 2007).

This study applies secondary empirical data; a comprehensive patient satisfaction survey involving 1.940 ED patients at 11 Danish EDs, all situated in the Capital Region of Denmark (Rimdal and Soerensen, 2012).

Results

14 questions in the patient satisfaction survey were found applicable for the SEM analysis. Eligible respondents were only those that had responded to all of the 14 questions (listwise deletion), returning a useful total of 685 responses after filtering out potential biased respondents, i.e. responses made by guardians or relatives on behalf of the patient. The EFA revealed a total of four latent constructs (see Table 5.2). These four constructs were named according to what overall topic the questionnaire items covered. The constructs were named 1) safety and satisfaction, 2) waiting times, 3) information dispersion, and 4) infrastructure. The left column in Table 5.2 states the questionnaire item number identical to the numbering in the survey from which it has been extracted.

Question	Description	Construct			
		1	2	3	4
q7	Intuitive signpost				0.883
q8	Parking options				0.867
q13a	Wait times, actual				
q14	Wait times, arrive to treat		0.871		
q15	Info, reason for wait			0.887	
q16	Info, progress in wait			0.894	
q17	Wait times, perceived total stay in ED		0.767		
q20	Safety, correct treatment	0.739			
q25	Safety, symptom awareness	0.803			
q26	Safety, contact person	0.818			
q27	Safety, discharge	0.799			
q28	Facilities, overall assessment	0.315	0.331		
q29	Facilities, cleanliness	0.519			
q31	Total impression of ED visit	0.705	0.373		

Table 5.2: EFA factor loadings (N=685)

Four hypotheses on causal mechanisms between the four latent constructs were posed based on literature conclusions. These are represented in a theoretical model shown in Figure 5.6.

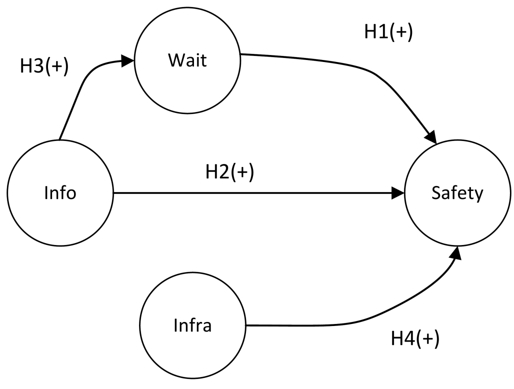
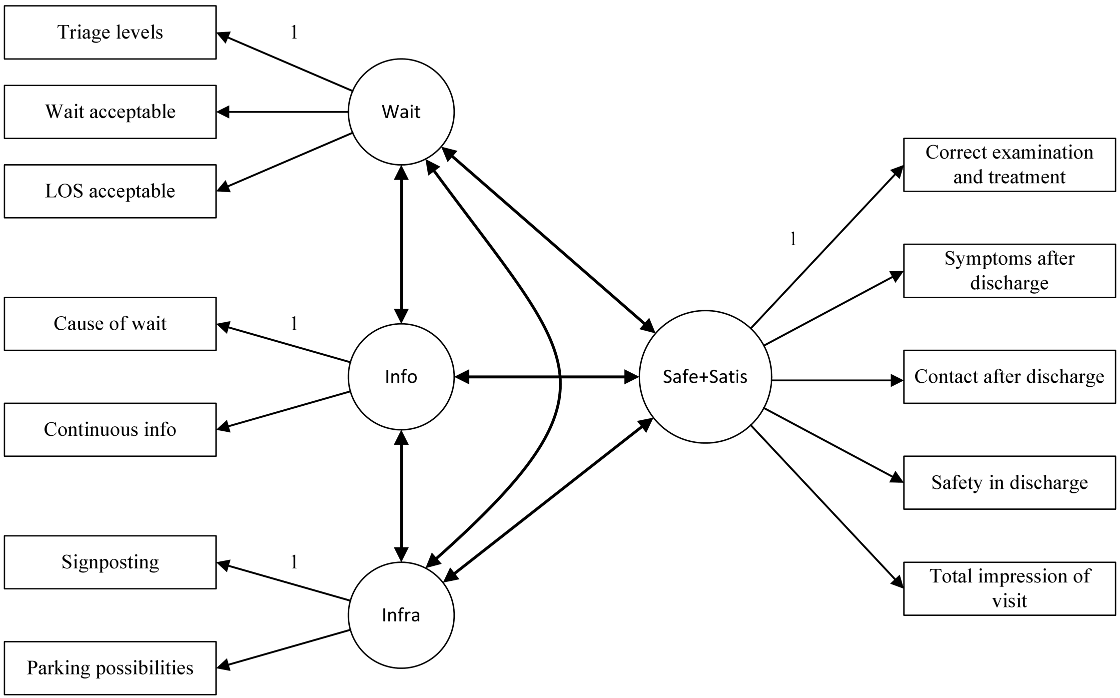


Figure 5.6: Theoretical model of hypotheses to be tested

In brief, the positive sign in parenthesis indicates a positive effect on one factor upon the other, specified by the arrow's direction. The "safety + satisfaction" variable becomes the dependent variable, while the rest are modelled as independent. With the theoretical model in place, the CFA/measurement model was specified and tested against the empirical data (Figure 5.7). The model fit is adequate if the statistical key figures are within certain generally acknowledged ranges (Kline, 2011).

In the CFA, all latent constructs are allowed to covary. The hypotheses depicted in Figure 5.6 were tested in Mplus version 6.12, alongside competing models, and regression coefficients could be determined (Muthen and Muthen, 2007). Results for each hypothesis is reported in Figure 5.8.



Chi square (χ2)	152.126
degrees of freedom	48
p	< 0.001
Root Mean Square Error of Approximation (RMSEA)	0.056
90 % confidence interval	
high	0.066
low	0.046
Comparative Fit Index (CFI)	0.963
Standardised Root Mean Square Residual (SRMR)	0.049

Figure 5.7: Measurement model (CFA) with appended model fit indices. Error terms have been omitted. LOS = Length of stay

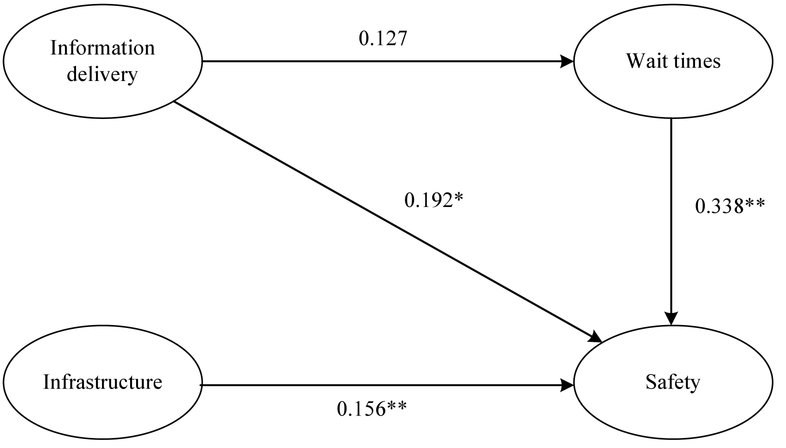


Figure 5.8: SEM results of tested hypotheses, non-standardised path coefficients reported only. Note: N=685, P < 0.05*, P < 0.01**

Discussion

Based on a large empirical sample from all 11 EDs in the Capital Region of Denmark, acceptable waiting times were found to be of greatest importance to the patients, especially the time to initial assessment (also called the *greeter time*). A close second was information dispersion, making the patient feel at ease and left with the feeling that value-adding activities are ongoing. Three of four tested hypotheses were statistically significant but small in sizes. ED patient satisfaction is an important measure that reflects both efficiency and quality in the ED, yet is difficult to handle due to unique requirements and expectations from the individual patient.

The model examined in this study does not take all relevant ED patient satisfaction aspects into account. Several other studies have examined other aspects of patient satisfaction for instance the health care provider's communication- and technical skills (Berg *et al.*, 2012) and patient involvement (Hölzel *et al.*, 2013). There is a potential to combine these studies with the present study to develop a more complete patient satisfaction SEM model. However, an extension of the model would add to the complexity and thus require an even larger empirical sample. The benefit of testing more hypotheses related to ED patient behaviour simultaneously is to shed more light on what aspects ED patients value the most.

Conclusion

A structural equation model was developed to analyse four key constructs driving overall ED patient satisfaction. Three out of four hypotheses were found to be statistically significant based on a comprehensive empirical data sample. An acceptable initial waiting time to see a health care provider was found to be most important, followed by information dispersion between patient and employee, and finally good infrastructure. ED decision makers can now rank the included factors' levels of importance to insure the highest possible payoff for resources invested in promoting ED patient satisfaction levels.

Experience gained

This study demonstrated that it was possible to test more hypotheses in one model through the use of multivariate statistics. Patient satisfaction, another highly complex performance measure of great importance to all the of study's stakeholders, constitutes a higher level aggregated cluster in the final performance measurement model. SEM was a useful technique to quantitatively test causal relations between latent constructs. It would however prove significantly more difficult to apply SEM on a larger scale model that is to consider a more heterogeneous group of performance measures. Thus, a completely different approach to understanding how the performance measures connect between the patient-, employee-, and operations related categories must be considered.

An important insight in the study is made at this stage. The searchlight is directed from looking solely at performance measures to gaining a better understanding of the dynamics behind the processes (of which the performance measures reflect parts of). Simulation techniques can potentially show how processes in an ED are interrelated.

5.4 P4: ED patient flow dynamics

Background

EDs worldwide are facing an increasing problem of having a supply and demand mismatch when it comes to patient intake. More and more patients seek treatment in the ED while capacity is continuously being downsized. The typical ED patient suffers from multiple conditions. Such complexity makes it a difficult and time-consuming task for the ED physician to conclude on which treatment should be prioritised. A natural consequence is a clustering of patients in the ED, resulting in a blurred overview of critical patients, compromising timely treatment, and ultimately lower quality of care. The problem is widely acknowledged and the Institute of Medicine, a recognised American health institution, updated their position statement on ED crowding with a number of recommendations for mitigating the problem (ACEP, 2006). However, the problem has resisted the policy changes and remains an issue which, to make matters worse, seem to increase. The causative mechanisms that appear counter-intuitive are yet to be identified. Simulation, more precisely *system dynamics* (SD), is a technique that can be exploited to understand better the complex non-linear behaviour that ED crowding exhibits (Sterman, 2000).

The study examines two situations in the admission processes at a major tertiary level one hospital in North America; 1) a baseline "business as usual" model and 2) a stressed model with a sudden increased patient inflow. A local protocol *(Code Help)* is investigated for its system-wide effect . In more detail, this goes in particular for the potential risk of having patients return due to premature discharge as a delayed effect upon protocol activation. Additionally, four more generic parameters are examined in the baseline model being 1) inpatient capacity, 2) bed assignment time, 3) transfer time, and 4) number of incoming elective patients.

Methods

A SD model that shows the admission process at the case hospital was developed. In more detail, the model depicts two feedback mechanisms; one balancing the available inpatient capacity and one modelling the reinforcing effect of returning patients after initiating the local protocol. Furthermore, the competition for inpatient beds is included through a parallel inflow of elective patients. The model is built based upon direct observations, semi-structured interviews, archival data from October, 2013, and literature stemming from emergency medicine and healthcare management. The overall model with supplementary object descriptions is shown in Figure 5.10. The model is supported by two subviews that can be assessed in the model documentation. Full model documentation can be acquired with the author. Iterative qualitative assessments were made to ensure a correct understanding of the process modelled. Quantitative testing of the model's output compared to archival data with additional assessment of system behaviour under extreme conditions strengthened belief in the model's robustness and validity.

Results

The SD model was first evaluated based on how well the simulated output resembled acquired data on ED census throughout the month of October, 2013. As can be seen on Figure 5.9, the simulation output returns a good resemblance of actual system behaviour. The jagged graph shows archival data while smooth graph shows the simulation output.

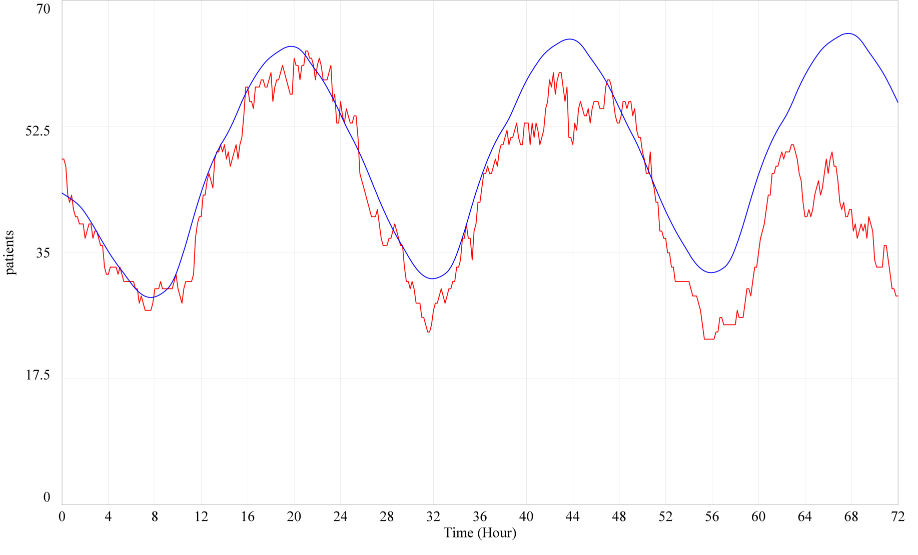


Figure 5.9: ED census comparison (1^{st} to 3^{rd} of October, 2013)

With a model that adequately can replicate system behaviour and display intuitive behaviour under extreme conditions, the next step is to vary parameter settings to identify levers of high system impact. Reported results are split into their relation to either 1) the baseline model or 2) the stressed model with sudden temporary increase in patient arrival.

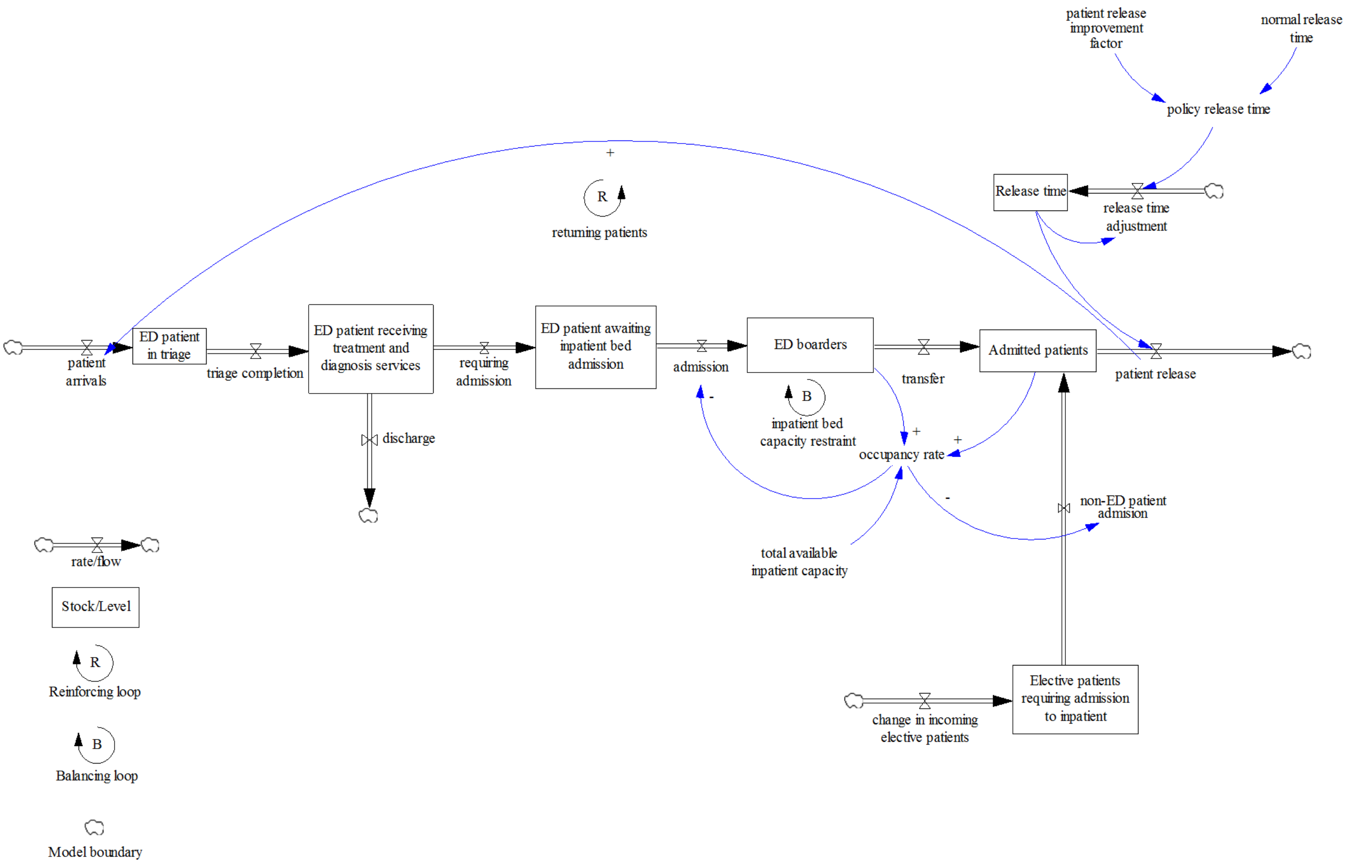


Figure 5.10: SD model overview depicting the admission process, the two feedback mechanisms, and the competing inflow of elective patients requiring an inpatient bed. Auxiliary variables and sub-views omitted.

	trigger, ED boarders			trigger, ED census		
	Base	7	13	Base	48	60
Occupancy	0.9205	0.8971 (2.54)	0.926 (0.59)			
time, max diff	51					
ED census	61.02	64.34 (5.47)	60.06 (1.57)			
time, max diff	67.5					
ED boarders	6.153	6.503 (5.69)	6.052 (1.64)			
time, max diff	70.125					
Admitted patients	453.7	442 (2.58)	456.6 (0.64)			
time, max diff	51					

Table 5.3: Stressed model sensitivity test results for protocol trigger values

	tot avail inpatient beds			bed assign time			transfer time			mean elective patients		
	Base	400	900	Base	1.8	4	Base	0.5	2.5	Base	100	200
Occupancy	0.9167	0.9185 (0.20)	0.7891 (13.92)	0.833	0.89 (0.79)	0.865 (2.04)	0.909	0.905 (0.43)	0.912 (0.37)	0.836	0.689 (17.59)	0.917 (9.75)
time, max diff	56			24			32			16		
ED census	59.96	59.35 (1.02)	60.07 (0.18)	59.53	56.72 (4.72)	62.03 (4.20)	59.53	56.65 (4.84)	61.82 (3.85)	59.96	59.14 (1.37)	59.98 (0.03)
time, max diff	139.5			19.625			19.5			139.5		
ED boarders	6.031	5.967 (1.06)	6.043 (0.20)	5.869	5.403 (7.94)	6.238 (6.29)	5.959	2.784 (53.28)	8.628 (44.79)	6.031	5.945 (1.42)	6.033 (0.03)
time, max diff	142.5			24			23.125			142.25		
Admit patients	455.6	364.7 (19.95)	479.3 (5.20)	438.6	440.3 (0.39)	437 (0.36)	455.7	458.6 (0.64)	453.2 (0.55)	414.7	340.8 (17.82)	455 (9.72)
time, max diff	100			24			48			16		

Table 5.4: Baseline sensitivity test results for selected parameters associated with ED crowding

Tables 5.3 and 5.4 presents the sensitivity analyses indicating baseline, minimum, and maximum values for the given parameter. In parentheses, the leverage potential is given in percent. The values reported have been extracted to the time where the three graphs (base, maximum, and minimum) show the biggest variance. Most leverage potential is found in the largest percentages deviation. Note that there is no effect on varying the ED census trigger due to the threshold being frequently exceeded.

Discussion

A SD model that specifically models the dynamics behind the case hospital's admission process was developed. First, the local protocol's trigger values were analysed based on four outcome parameters. Changing the ED census trigger had no effect as it is exceeded on a daily basis. The ED boarders trigger was identified to be the main lever of alleviating patient flow. If lowered to 7 boarded patients (compared to a base value of 10), the payoff becomes a decreased occupancy rate and number of admitted patients with 2.54 % and 2.58 % respectively, while ED census increases with 5.47 %. Some assumptions as to how many patients return when the protocol is activated, the delay to which protocol is activated until the effect can be measured, and average inpatient length of stay must be taken into account when interpreting the results. With more accurate measures, the simulation model would become more robust, however the dynamics would remain the same. Second, selected generic parameters connected to the admission process that are generally seen as levers for ED crowding were analysed. Getting ED boarders transported to the stem ward faster than normal was found to be the main lever that could ease ED congestion by 4.84 % while raising the number of admitted patients with 0.64 %. In connection, a lowered bed assignment time enhances patient flow with slightly less than the transfer time (ED census lowered by 4.72 %) with a subsequent small elevation in number of admitted patients (0.39 %).

An enhanced knowledgebase would be obtained by focussing on the transition phase between departments. In that way, aspects from the receiving stem ward may change system behaviour, identify other levers of interest, and alter conclusions made from this study. In this study, SD has been demonstrated as a learning platform for ED personnel to better understand how processes are interrelated and how to illuminate where to target potential interventions within current span of control.

Conclusion

This study presents a SD model which examines an ED's admission process at a major tertiary level one hospital in North America. In particular, the effects of changing the trigger values for a local protocol for ED crowding alleviation were examined. Additionally, several generic parameters were included in the analysis as a side objective. For the stressed model with protocol in scope, changing the ED boarder trigger impacted the system the most. Of the generic parameters, bed assignment- and transfer times were found to be of main importance to alleviate crowding situations. Non-linear complex systems, like an ED, are well suited for analysis through the use of SD that is built from differential- and integral equations. In this study, the strength of SD is demonstrated by providing ED decision makers a safe testing environment in which the effects of desired policy changes can instantaneously be assessed.

Experience gained

By switching the attention from performance measures to processes, a deeper understanding of the complex plays between actions and results is enabled. SD was demonstrated to be a learning platform that support decision making, especially where past interventions do not show the expected results – or maybe even contradict intuition. The human mind is capable of grasping immediate effects of particular actions; something that is a lot more difficult when time delays come into play, when effects are non-linear and may be impaired by indirect parameters that are not acknowledged. SD combines mental models and differential- and integral equations, both of which are imperative for understanding complex system structure. For instance, if an intervention's effect is absent and the time delay is not realised, the effects of the intervention is prone to be falsely ascribed to other causes. The combination of human understanding of causal mechanisms and computer simulation software holds great potential in widening the general understanding of how interventions will impact a complex system. An understanding of how the processes link must not only be sought in written- or numerical data. It is all in the minds of the experts who, in this case, are the employees that work in the ED. If such information can go from a tacit state to explicit through semi-structured interviews, the processes can be conceptualised, compared, and in the end simulated after acquisition of relevant data material.

Drawing on the knowledge of mental models, the final step of the study is to make an attempt of connecting the identified performance measures and present this in a guided user interface that can be linked to existing ED data repositories. Presentation of the individual performance measures will primarily be done by using statistical control charts to assess data variability.

5.5 The final performance measurement model

Background

On a global basis, EDs are required to manage increasing numbers of acute patients due to a re-structuring of the healthcare sector and an aging population. To be able to cope with these future requirements, there is a need to re-think how acute treatment will be offered to ensure an acceptable patient flow while maintaining high levels of quality and efficiency in care. Various initiatives have already been launched in an attempt to optimise the ED patient flow. Examples of such initiatives are formalised triage procedures and the launch of emergency medicine as a separate specialty. However, a thorough understanding of how these initiatives impact the ED performance as a whole remains unknown.

The final stage of the study concerns the development of a practical performance measurement model capable of presenting current performance levels in an ED. Knowledge of 1) what to measure and 2) how the identified performance measures relate constitute the novelty of the proposed model. Despite the study being of an academic character, the proposed model is developed as a prototype containing an intuitive guided user interface *(GUI)* which draws on several statistical quality control techniques to present variation in data. Focussing on data variation makes it possible to assess if quality and efficiency, reflected in selected performance measures, have been statistically influenced by an intervention.

Methods

Several methods were applied in the development of the final performance measurement model. Each of these is chronologically described in the following.

Semi-structured interviews

Establishment of the performance measures' connections was based on semi-structured interviews at three case EDs, two Danish and one North American. 39 interviews were conducted with key representatives across professions in the ED. These representatives were senior physicians, medical residents, secretaries, and nurses with varying responsibilities. Table 5.5 presents the number of interviews conducted based on the interviewees' profession and affiliation.

The purpose and structure of all interviews was identical. The interviews were carried out in the given ED with one interviewee at a time. Due to time restraints regarding the interviewee's work schedule, all interviews had a duration of maximum 30 minutes. The questions asked followed a funnel model starting with open questions and then proceeded into more detailed questions. Initially, the researcher made a brief introduction describing the purpose of the

Profession	Nordsjælland	Herlev	BIDMC
Senior physician	4	2	2
Medical resident/ junior doctor	0	2	5
Nurse practitioner	1	2	0
Nurse	3	6	8
ED unit coordinator	1	0	1
Secretary	1	1	0
Total	10	13	16

Table 5.5: Number of interviews by profession at the included case EDs

entire study and why the interview had been set up. The interviewee was presented to the collection of highly analysed, discussed, or promoted performance measures which had been identified in the prior systematic literature review. An A3 paper presenting the hierarchical framework of performance measures served as a basis for the interview.

- **"Which of the presented performance measures do you believe best reflect quality in care within the ED?"**

The interviewee primarily highlighted the most visited measures (those that are consulted on a daily basis). Performance measures consulted less often but on a weekly basis were highlighted in orange.

Having highlighted the commonly used performance measures seen from the interviewee's perspective, the interviewee was then asked to think of the ED when it was under pressure (e.g. in situations with high workload and many patients). Bearing this situation in mind, connections between relevant performance measures were sketched on the A3 sheet.

- **"In situations where the ED is put under pressure, where can the effects of the highlighted measures potential variations be seen in the rest of the measures?"**

The interviewee was encouraged to support these propositions/hypotheses (based on what is measured/not measured) with narratives from their own experiences. Direct links would first be drawn but potential indirect links could evolve as the interview progressed.

Data acquisition and filtering

Data on the identified performance measures was acquired at all included case hospitals. Several staff members from all case hospitals and across departments provided assistance in collecting raw data. The data obtained was blinded to eliminate patient sensitive information. All data elements needed to be within the same time frame in order to make comparisons across performance measures possible. The author strived to collect three months' worth of data from the three included EDs. Extensive data preparation was necessary before presenting the data in the final performance measurement model.

Setting up the GUI

For most of the performance measures, data is presented in an appropriate control chart to quickly assess special- and common cause variations. Other statistical quality techniques, such as Pareto diagrams, support the control charts where applicable (Montgomery, 2005). EpiData was utilised to generate the control charts (Lauritsen and Packness, 2010). Microsoft Excel was chosen as the standard developing platform due to its common use. Microsoft Visual Basics programming language was applied to set up the interface, thus allowing the user to easily navigate between windows. Supporting material was implemented in the model so that the user can consult or update 1) definitions, 2) choice of SPC charts and 3) connections between the performance measures. Each of the three supporting windows are placed as shortcuts in a "main menu" where also the performance overview can be accessed.

Results

When the GUI is opened, the user starts from a main menu (see Figure 5.11). Four options are available: 1) the main ED performance overview, 2) an SPC chart selection guide, 3) complete definitions and data retrieval of included performance measures and 4) an overview of connections with appended causal explanations. The ED performance overview is of main interest while the three other tabs serve as supporting material. In the following, the supportive tabs are explained first.

Starting with the second tab from the main menu, the *SPC Chart Selection Guide* presents a decision hierarchy to assist in which control chart is appropriate for a given data set (see Figure 5.12). While the guide provides a quick overview, some prior knowledge of SPC is recommended.

The primary charts used in the model are the $\bar{X}$&S for the different time intervals and the U chart. The samples are largely above 10 and the sample variance in the attributes data is high due to the ED census daily dynamics. One control chart, named the G-chart, is used for attribute data that depict rare events but is however not listed in the SPC chart guide. As a side note, all tabs in the model includes a *back* function which is conveniently placed in the

Figure 5.11: Final Performance Measurement Model; the main menu

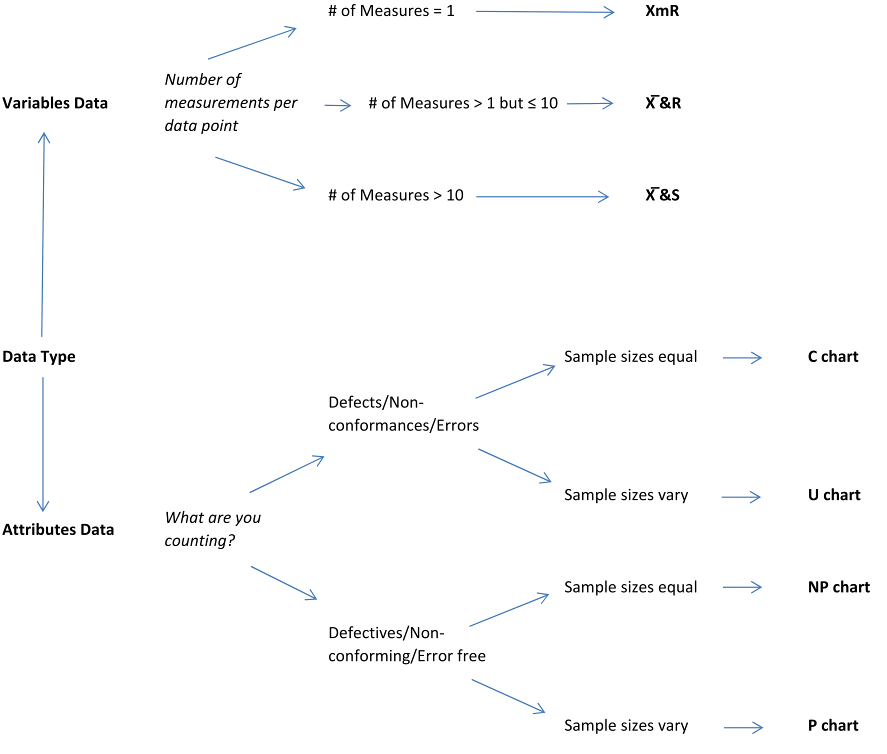


Figure 5.12: Screenshot of *SPC Chart Selection Guide*

top left corner. Information on how each performance measure is defined and recorded can be found in the tab named *Measurement Definitions and Data Retrieval* (see Figure 5.13). This tab contains ten columns and 34 rows, one for each performance measure. Starting from the left, a brief description of the columns is presented in Table 5.6.

Column	**Description**
Category	The aggregate category being either 1) patients, 2) employees or 3) operations
Cluster	A lower level cluster that branches the aggregate category into appropriate themes under which the performance measures relate
Potential responsible for data extract	Indicates the person who will be responsible for extracting data on the given performance measure
Observed variable name	States the name of the performance measure as referred to in peer-reviewed emergency medicine literature
Explanation	A brief statement of what the performance measurement includes
Desired data	This category comes in continuation of the explanation category and specifies the requested data which best possibly reflects the performance measurement mentioned. The data precision may vary according to what precision is officially required by the global emergency medicine society
Data source	From which database the data can be extracted
Definition	The official definition as identified in peer-reviewed literature
Scale	Denotes if the data is of 1) numerical, 2) ordinal or 3) binary character
Diagram (SPC)	Provides an explanation regarding which control chart was selected. Note that more than a single type of control chart can be correct

Table 5.6: Explanation of columns in tab *Measurement Definitions and Data Retrieval*

It is imperative to register any changes in data registration procedures to be able to maintain the relevance of the performance measurement model. The third and last supporting tab (*Underlying Measurement Network*) contains all findings from the 39 interviews (see Figure 5.14). A similar construction as the previous tab applies with some exceptions. There are six columns in total. The first three columns all hierarchically denote the variables starting with 1) the overall category, 2) the lower level cluster and 3) the observed variable. The next three columns are listed and commented in Table 5.7.

Column	Description
Potentially connected to	For each of the performance measures (observed variables), a number of other performance measures are hypothetically linked. A drop down menu is enabled, listing each of the performance measures that were identified to connect through the interviews rounds
Hypotheses stated	Every potential connection between performance measures is explicitly addressed in a brief description
Interview strength	Indicates the number of times the given connection had been emphasised during the interviews. A higher number thus reflects an intense focus or concern

Table 5.7: Explanation of columns in tab *Underlying Measurement Network*

The interconnections may change over time as processes and procedures are developed. As for the rest of the performance measurement model, the connections must be maintained to ensure the correctness of the causal interconnection that is sought highlighted.

The *ED Performance Overview* represents the last tab (see Figure 5.15). This tab completes the information on the selected performance measures' levels which can be accessed from here. When applicable, the performance measures are compared to an index in order to specify data trends. The trend will instantly bring potential areas of low performance to the user's attention. From the performance overview window, the user can click any performance measure (that contains data) of interest to investigate this further. In Figure 5.16, an example is shown to demonstrate how the user is presented to the trends in a single performance measure (here *Adverse Events*). The performance measurements left empty are due to no data being possible to retrieve.

Back to Main Menu

Category	Cluster	Potential responsible for data extract	Observed variable name	Explanation	Desired data	Data source (if applicable)
Patient	Safety	Nanna	Adverse events	Amount and score after review per day	List of all (reported) incidents with score and date	RL Solutions (CSC)
			Missed diagnoses			
		Allan	ED Returns (< 72 hours)	Number of patients returning within 72 hours upon discharge (inpatient ward or ED)	List of all patients returning within 72 hours. Use date for first contact. Date and time for first and following contact along with diagnosis for both courses	OPUS/GS
		Allan	Mortality	Mortality rate	List of all patients arriving through the ED that 1) expired in the ED or 2) expired during hospitalization. Date and time for arrival and death.	OPUS/GS
			Morbidity	Has been omitted from the model due to an assessment of all incoming ED patients		

Figure 5.13: Screenshot of tab *Measurement Definitions and Data Retrieval*

Back to Main Menu

Category	Cluster	Observed variable	Potentially connected to (nb: based on interviews)	Hypotheses stated	Interview (strength)
			check drop down list for options		High number = more strength
Patients	Safety	Adverse events	1st encounter with healthcare personnel - initial diagnosis ready	Risk of more adverse events if long time delay between first encounter with physician and initial diagnosis. This typically relevant for high acuity patients	2
		Missed diagnoses	1st encounter with healthcare personnel - initial diagnosis ready	Potential connection between the time of consultation and missed diagnoses	2
		ED returns (< 72 hours)	1st encounter with healthcare personnel - initial diagnosis ready	Delayed greeter time increases the likelihood of more ED return visits (delayed effect)	2
		Mortality	1st encounter with healthcare personnel - initial diagnosis ready	A delayed greeter time may inflict loss of lives, hence increasing mortality rate (low impact hypothesised)	1
		Morbidity			

Figure 5.14: Screenshot of tab *Underlying Measurement Network*

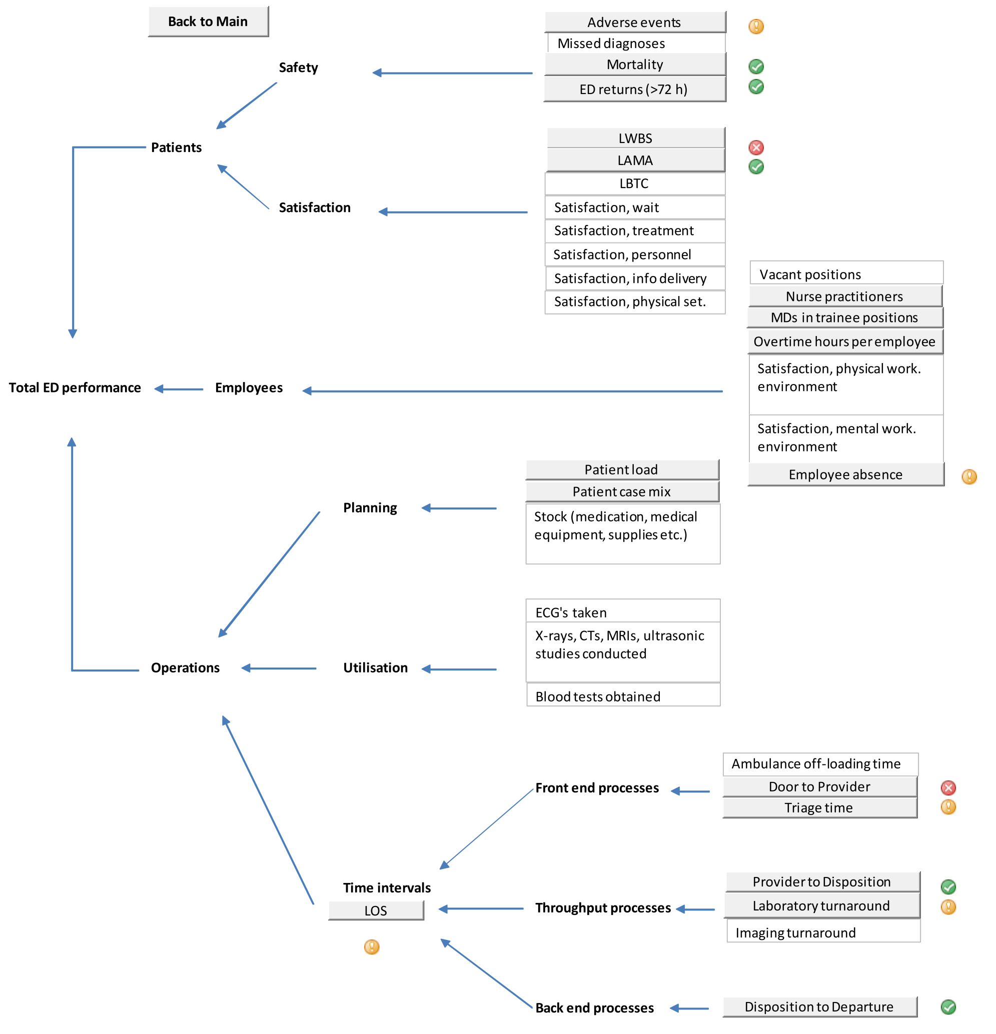
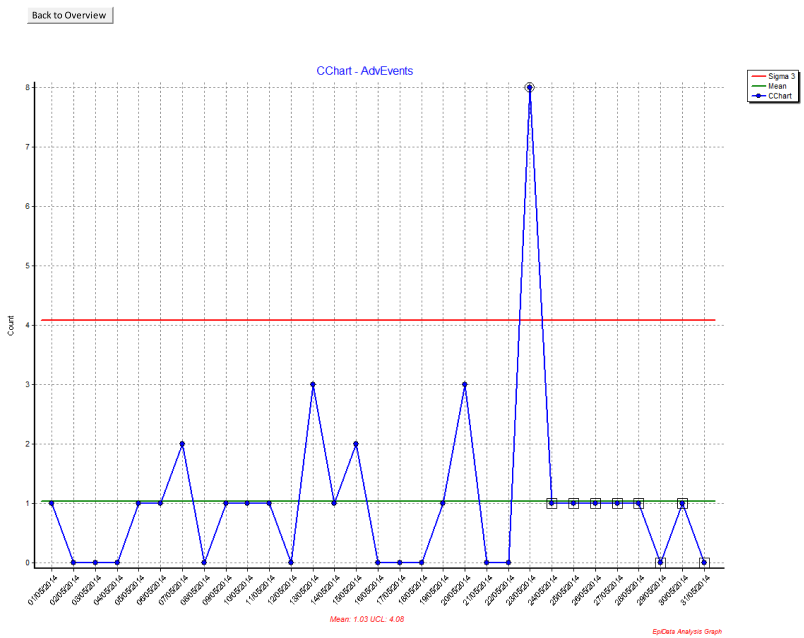
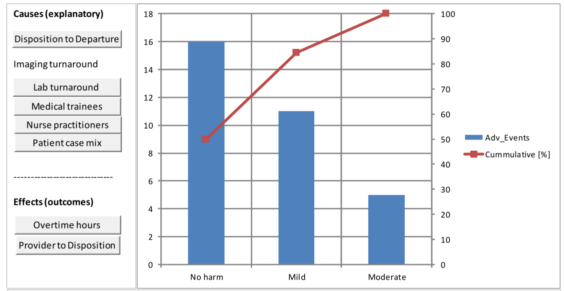


Figure 5.15: The ED Performance Overview





Narrative:

In general, one adverse event is reported during May 2014. One outlier is identified on the 23rd where 8 reports were made. This was done due to an increasing back log that had not been entered in the database. The last eight measurements are all either on the average limit or below suggesting a change in the process.

From the above Pareto diagram, the severity of the events are presented. Five of the events were reported as being "Moderate":
These were reported on the following dates:

13th: Treatment error (misunderstanding)
15th: Insufficient diagnostic examination
23rd: Wrong medication prescribed
23rd: Patient arrival during shifts
25th: Missing information during transition between departments

For more information, see **RL solutions** through contact person (referenced in *Metric Definitions*).

Figure 5.16: Example of performance measure overview; C-chart of adverse events alongside Pareto diagram depicting severity of incident

Discussion

When this study was completed, the final performance measurement model was launched as a prototype. Therefore, there are still elements that are work in progress. Data is required to be manually processed before it is entered in the appropriate windows. The performance measures that can be visually sketched in control charts can be imported in a standardised format and subsequently processed in Epidata with minimal effort. With other data types, some more processing work will be required before analysis can be enabled. This is especially relevant for the raw data which arrives from external parties such as satisfaction surveys. The proposed model provides the ED decision maker with a quick overview of key departmental performance levels. As performance measures merely indicate that a process or outcome may need attention, time must be allowed to track down the causes for changes in system performance. Causal tracing is made possible through the implementation of the underlying network structure. However, the selected performance measures cannot display all possible dynamics of the ED system and it therefore will be necessary to search for causal explanations of system behaviour through other means as well.

The initial motivation for developing a performance measurement model was not alone to visualise the intertwined impact of interventions to the ED system but to also be able to enable tracking of performance in retrospect. Through the use of statistical quality control methods, variations in processes are assessed to determine if these are indeed caused by a deliberate intervention. What remains to be analysed more is the delay during which this effect kicks in. The SD study on ED crowding factors clearly showed complex non-linear system behaviour; something that must not be forgotten when trying to comprehend how a system truly functions. A further investigation of the time delays and magnitude of the interventions' effects on the main processes in the ED is definitely worth pursuing.

The proposed performance measurement model can be streamlined further to create more value. What is required to strengthen the model is, first of all, a supportive data infrastructure. At the two Danish cases, data is stored in multiple databases that are to be integrated into one shared platform to be completed within the next couple of years (REGH, 2014). A common data warehouse should make data extraction easier in the future, preferably without any manual input involved. Another important prerequisite is to allow the necessary time to analyse the data trends. The model gauges many different key aspects of ED performance that will require the user to carefully evaluate the data trends of interest. As such, the model is not a standalone tool where the user can put management on autopilot. Other extensions can be added to include even more insights. A cost module is desirable to estimate expenses and benefits of the various initiatives in the ED. Putting a price tag on processes and procedures is no easy task, especially due to each patient requiring a unique set of treatment interventions. However, assessing value for money, especially seen from a top management viewpoint, will inevitably be a topic that must be examined as budget restrictions and political pressure are still ombrageant.

Conclusion

A performance measurement model that presents the data trends of carefully selected performance measures reflecting all-round ED performance was developed. The performance measures are connected based upon the relationship between causes and effects, which permits causal tracing of changes to the ED system. ED decision makers can draw benefit from the model to visualise both 1) the general status of the department and 2) assess if initiatives have had the desired impact.

The model was set up as a prototype which requires manual input and processing of data. To automate this initial data preparation will in the near future be considerably easier when an integrated healthcare platform is launched and data can be stored in a shared data warehouse. Time must be allocated to properly understand what the data reveals and consequently which decisions are needed to be made. Furthermore, continuous updates to maintain the purpose of the model are important. Maintenance is a required consequence of frequent model use. While intended as a prototype, the model serves as an inspiration towards how a future performance measurement model could be implemented.

Experience gained

Performance measurement is imperative for the ED decision maker to be able to gain insight into system performance and steer the department towards continuous improvements. A step towards a more data-driven performance approach was attempted by demonstrating how an IT-software solution could look in the future. Although still a prototype, the user is provided with a tool to visualise how well the department is performing over time alongside knowledge on how the processes are related. It is likely that system developers may find it interesting to adopt a similar approach for future off-the-shelf software solutions across domains.

Understanding causal interactions is indeed a complex matter. In the developed performance measurement model, the connections are made based upon the interviewees' mental models. A greater understanding of how these are connected in terms of magnitude of impact and time delays is desirable. This may in the future be investigated through simulation studies, such as the one demonstrated on ED crowding factors. The synergistic effects of applying a combination of exploratory statistics and simulation are promising in the task to navigate the ED safely through the challenges that lie ahead.

Even though the collection of performance measures included has been discussed in contemporary peer-reviewed emergency medicine literature, there is a need to continuously evaluate the appropriateness of each performance measure. Some performance measures may be replaced with other more locally relevant ones. Such replacement is possible as the performance measurement model comes with a higher of flexibility. Additional modules to understand the quality of treatment for disease specific patients should be pursued in the future, all dependent on which patient type is more greatly represented in the given ED. Suggestions regarding which diseases that should be closely monitored include sepsis, asthma, acute myocardial infarction, and pneumonia (Graff *et al.*, 2002; Schull *et al.*, 2011). The developed performance measurement model could with great benefit be implemented in already existing IT support systems which can allow automatic data extraction. A closer analysis of the quality of data must be conducted as the performance measurement model will only display the quality that which is reflected by what the input data permits. If data is registered manually, technological solutions that automatically register diverse performance measures should be considered. The less administrative work burden on the clinical staffs' agenda, the more value adding time to treat patients becomes available.

It is the author's hope that the suggested model will eventually be put into practical use at the included case hospitals. Uncovering the true value of the proposed solution can only be achieved through pilot testing. The pilot testing of the model would have been the next step in the study if the required time to do so had been allocated. The author hopes to pilot test the developed performance measurement model in the future.

CHAPTER 6

DISCUSSION

The research design and empirical standpoint frames the researcher's play pen. Conclusions are likely to differ if another research design and empirical lens is chosen. Therefore, it is important to elaborate upon how the research problem was addressed and to pose oneself the question if similar results would have been obtained, provided a different empirical foundation. A critical look at the three research sub-questions is done in relation to the chosen methodology and the empirical material. Furthermore, all incremental conclusions throughout the study are compared to the overall goal, namely the development of a generic performance measurement model for macro-level ED evaluation.

RQ1: Identification of key performance measures for emergency department assessment

Which performance measures should be used in order to assess quality and efficiency in an emergency department?

The starting point for the study was to investigate which elements which would constitute the backbone of the performance measurement model. There are many data elements to choose from as data recording at hospitals has become fairly cheap and accessible. Some would argue that all data elements should be analysed to give the best possible insight into the condition of the ED system. Analysis of such vast amounts of data would be extremely time consuming and would lead to what Pettigrew coined "death by data asphyxiation" (Pettigrew, 1990). Therefore, it was important to identify a minimum data set that could provide a sufficiently comprehensive picture of the current state of an ED without compromising the level of detail. In other words, to strike a balance between having a manageable number of performance measures the capability of tracking all-round system performance.

Any potential performance measure should not only be of relevance to the ED population (defined by patient type, diagnosis, and acuity) and basic process elements of emergency care. Additional considerations for healthcare decision makers include a series of other criteria that have been put forward by the Organization for Economic Cooperation and Development (OECD) and the National Quality Measures Clearinghouse (NQMC) amongst others (NQMC, 2014; OECD, 2013). The lists of requirements are elaborate and there are common traits to be found. Despite some minor differences in these lists, they emphasise a general scientific soundness of the given measure (validity, reliability, and evidence), whether the measure is feasible to obtain (already in use, potential for benchmark, cost/burden to extract), and its susceptibility of impacting the healthcare system. With the guidelines from OECD and NQMC, it should be clearer to select performance measures that are rich on information. However, the assessment of ED quality of care remains hampered by a lack of consensus on appropriate measures.

A systematic literature review regarding which performance measures were debated, recommended, or analysed was conducted. Several meta-studies were identified which contained a similar goal of seeking consensus on a set of performance measures for use in overall ED performance assessment. Interestingly, these studies' conclusions diverged. This divergence of conclusions made it necessary to complete a "meta-study of meta-studies" that ultimately became the study's first journal article. Great experience on health quality measurement systems was found in peer-reviewed literature studies stemming from the United States, Canada, Sweden, Australia, and the United Kingdom. In terms of disease epidemiology profiles and socio-economics, these countries are comparable to Denmark. Also, each of the countries have either developed and specified complete performance measure data sets relevant to emergency care delivery, have set up models for gathering emergency care performance information, or both. 55 macro-level performance measures reflecting aspects of emergency quality care were identified. A striking lack of performance measures relevant to the employees' well-being became apparent. It is hypothesised that with a political agenda that stresses efficient and high-performing healthcare systems, there has been a tendency to neglect the physical working environment of the staff. Furthermore, measurement of employee satisfaction may dependend upon the organisation's culture. During the external stay at BIDMC in Boston, MA, no official employee satisfaction surveys were performed. This may partly be explained by an underlying cultural difference where employees take great pride in being a member of a successful and highly regarded hospital. It goes without doubt that the work in an ED is stressful and highly demanding. The ED as a working place has become increasingly challenging as budget constraints limit the number of hands available to take care of an increasing number of patients, many of whom suffer from multiple diseases, thus adding to the complexity of the task. The healthcare provider is a key asset to the ED. Hence, additional performance measures relevant to the employees' well-being should be added to ensure a more complete assessment of the ED.

The 55 identified performance measures in the systematic literature review were all presented to a selection of ED employees at five Danish hospitals being Nordsjælland, Herlev, Hvidovre, Bispebjerg, and Aabenraa. 82 employees (out of 246) participated in an online survey that sought to critically assess the importance of tracking each performance measures. The underlying reason for this exercise was to retain or discard the identified performance measures while encouraging the consideration of potentially missing performance measures relevant to Danish healthcare. A handful of employee related performance measures were suggested and added to the final performance measurement model. It was chosen to develop the online survey in Danish, implying a translation from English. Even though it was attempted not to lose any original meaning of the original performance measures, the translation into Danish did cause some confusion for a few of the respondents. A report presenting the results of the online survey was later distributed to every survey participant. The report (in Danish) can be acquired upon request. In general, the respondents all agreed on the importance of the included performance measures. One performance measure though was commented upon by several respondents. According to the included literature, the rate of patients that leave without being seen by a healthcare provider (LWBS) is an important performance measure that reflects both patient satisfaction and safety. A comment from a respondent was as follows:

> *"Patients that leave before being seen by a healthcare provider often do not require urgent treatment. Some [patients] are very impatient, partly due to substance abuse problems and do not completely appreciate other patients having a higher priority."* (Nurse, anonymous)

While personal experiences control a statement, the author had to remain critical towards potentially discarding performance measures. In this case, not all patients that leave prematurely are trivial cases and could potentially create scenarios where the patient's health is at risk, meaning that the performance measure must be retained (Welch *et al.*, 2006).

Measuring all-round performance in an ED remains a subjective matter that probably will continue to be a highly debated topic. With national initiatives such as *Sundhedsplatformen* in the Capital Region of Denmark and the Centers for Medicare and Medicaid in the United States, there is a reason to believe that a set of performance measures needs to be rolled out on a broader level which will be supported by an adequate data infrastructure (CMM, 2014; REGH, 2014). Of paramount importance is a careful attention on defining the individual performance measures and maintaining these in a systematic way to ensure a sound scientific foundation for performance measurement efforts.

RQ2: The relations behind the identified performance measures

What are the patterns or linkages between the selected performance measures?

Now that the pieces of the puzzle have been gathered, the next phase of the study concerns how these fit together. The objective of science aiming at understanding mechanism-based phenomena across domains has for many years been a highly discussed topic. However, only during recent decades has this idea been systematically investigated (Bechtel, 2008). Several interpretations as to what constitutes a mechanism-based explanation or causal relationship have been proposed. The fundamental idea is rather straight-forward and can be described as follows:

> *"... At its core, it [a mechanism-based explanation] implies that proper explanations should detail the cogs and wheels of the causal process through which the outcome to be explained was brought about."* (Hedström and Ylikoski, 2010)

Examples of cause-effect relationships from health sciences among measured variables of interest concern treatments, exposures, outcomes, and preconditions (Pearl, 2001). Health science research has had a preference for statistical analysis for estimating correlations between variables. What is underlined by critics is the danger of mistaking a high correlation or covariance for a causal relationship. An example of this danger is the correlation between patients' coughing and fever. A causal mistake would be to conclude that patients are coughing because they have a fever to a causally perceived high degree (or vice versa). The only thing a high correlation will tell is that, for some reason, certain events generally happen simultaneously. Statistical analysis must therefore be supported by a qualitative assessment if more robust conclusions on causal relationships are to be made.

Indeed, knowledge of why actions have a certain effect on the ED as a system will give the decision maker an important edge in the attempt of improving quality-in-care. Therefore, the second sub-question in this study has been thoroughly investigated through multiple approaches, two of which apply statistics and one which uses computer simulation on processes. All three approaches have been documented in the form of journal articles. What makes causal analysis difficult can be illustrated in Figure 6.1. Here, the change in the system (denoted intervention/action) is hypothesised to create a visible effect in a given performance measure. The effect is impaired to an unknown degree by multiple factors. This phenomenon, where information gets "lost in translation", is similar to that of a garden hose with holes. A garden hose with no holes would maintain optimum pressure. If water leaks, there is an apparent waste – equivalent to losing information.

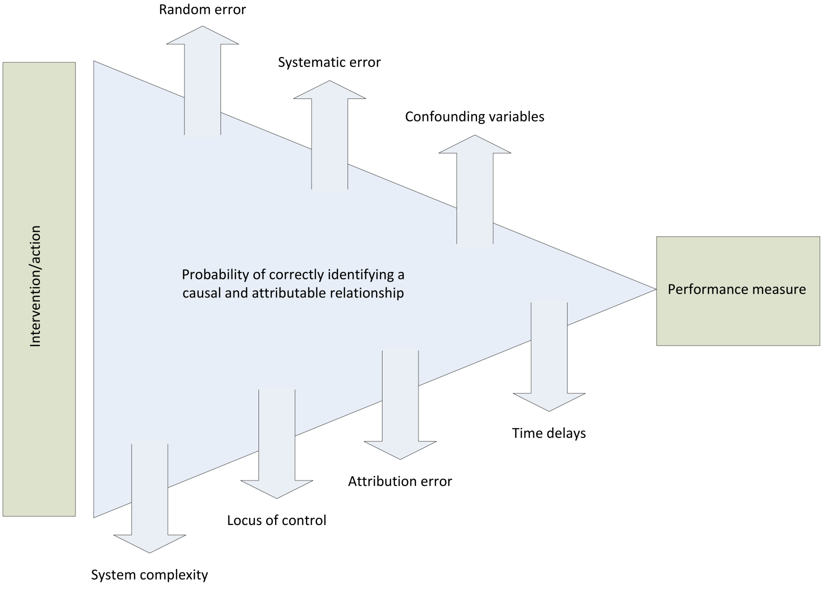


Figure 6.1: Challenges of analysing a potentially causal and attributable connection between action and performance measure (adapted from Smith *et al.* (2009))

Starting from the top, *random error* is caused by chance and will be present in any type of quantitative data. When calculating statistical estimates for a hypothesised causal relationship, two types of error must be considered: type I (false positive rate) and type II (false negative rate). The acceptable rate (α level) of a type 1 error in a statistical set is usually defined as being 0.05 or 0.01 (called the level of significance or p value). This means that there is a five in a hundred or one in a hundred chance of identifying two variables' (intervention/action on performance measure) relationship when in fact no relationship exists. Hence, there is a need to recognise that a statistical test may be misleading. There is also the potential of missing a relationship (type II error). A false negative result is generally easier to accept compared to false positive ones, which is reflected in a standard β level at either 0.1 or 0.2. With a β level equal to 0.1, there is a 10 % chance that the statistical test will show no relationship present when in fact there is one. Variance in data needs to be assessed because if this is due to a great variance present in the data set occurring from chance, it becomes more challenging to draw accurate conclusions as to whether a relation between two variables is present. Hence, frequent checks of how the data is registered alongside both completeness and reliability serve to minimise random error.

Whereas error can happen by chance, it can also happen as a misspecification in registration procedures. This type of error is denoted as *systematic error*. Inaccuracy and bias are introduced when error happens systematically. Evaluation of both size and type of two variables' relationship are obscured with systematic error, as bias can distort the relationship into becoming more or less powerful in size compared to the actual association. Systematic error can be reduced with clear definitions of what needs to be measured and the data registration protocols that are designed to consider potential sources of bias.

Even if a relationship between two variables does exist having assessed both random- and systematic error in the data sets, can we then be confident of the result? *Confounding variables* (or simply *confounders*) constitute yet another pitfall that needs to be considered. The researcher can be deceived by the confounder's ability to indirectly moderate the relationship which is under analysis.

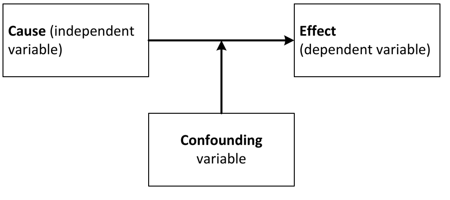


Figure 6.2: Confounding variable with moderating effect

A confounder can be a primary cause (or one of many causes) connected to both the effect and the cause. The possibility of a faulty conclusion regarding a statistically significant effect between cause and effect, is a pitfall, while the true relationship is controlled by the confounding variable. A practical, yet simplified, example is a high correlation between sales of ice cream and deaths by drowning. The wrong conclusion here would be to claim that selling more ice creams inflict more drowned people. A confounder that must be added to the equation is the weather on the given day, as sunny days attracts more people to the beach while being good for the ice cream refreshment business. If the confounder is not acknowledged prior to the testing, conclusions are likely to be biased. Multivariate statistics is able to control for confounders if these are identified. In paper 3 on patient satisfaction and safety (explained in section 5.3), one confounder was included in the level of contentment with waiting times being a moderator between 1) satisfaction with information delivery and 2) general satisfaction and safety.

There can be many possible confounding variables internally as well as externally in a social system, such as the ED, because it indeed is very *complex*. All of these variables have varying degrees of impact upon system behaviour making it a baffling challenge for the researcher to attribute observed processes or clinical outcomes to an action/intervention. For instance, confounders must be sought in patient characteristics, current ED policies, system features, and healthcare resources. The ED has high degrees of socio-technical complexity, which makes it increasingly more likely to miss confounders of importance. Thus, the main inhibitor for inclusion of confounders is the lack of knowledge and acknowledgement of their presence. Complexity was systematised by hierarchically categorising the identified performance measures into clusters according to relevance. A hierarchical model approach may be able to differentiate between the variation in data in a subset of the performance measurement model (Moore *et al.*, 2010).

Even though an intervention is thought to have a certain impact reflected in a performance measure, there may be health policies that dampen or even kill the desired effect. Any healthcare provider's actions are heavily influenced by the infrastructure and policies in the ED, also called the provider's *locus of control*. Sometimes, such policies are deemed likely to encourage a deviance from recommended clinical pathways for instance due to patient preferences. The effect from health policies and practices on interventions to the ED system must therefore be acknowledged and assessed when of establishing causal links between performance measures.

By now, it should be clear that determining causal relationships is not a straightforward matter. A bias that does not involve statistical estimates has been suggested by social psychologists is called *attribution error* (Rouse and Serban, 2011). This bias stems from a difference in perspectives when estimating the causal effects from events that involved the participants. Over-emphasis of circumstantial factors (outside locus of control) is often the case when trying to explain outcomes that were related to own actions, whereas under-emphasising circumstantial factors when looking at others. An example from healthcare can be found in a study by Golomb *et al.* (2007) where doctors had prescribed a certain drug with following side-effects. The doctors proved reluctant to ascribe these effects to the drug and instead pointed to patient health behaviours as being the true cause. In the attempt of developing a performance measurement model for ED assessment, internal physicians can be hypothesised to look more towards external factors as potential confounders. Outsiders such as academics and healthcare staff from other departments may be more prone to seek explanations within the system itself. Whether attribution error is occurring, scientific rigour is underlined by the involvement of both internal and external stakeholders for a more nuanced view on the ED system.

The last pitfall to be aware of is *time delays* between the change in the system and when the effect can be measured. This pitfall relates to the confounding variables and system complexity that can be ascribed to non-linear system behaviour. If the outcome of an action can be seen or measured immediately, nobody will question the cause and effect. An example would be putting your hand on a hot stove; you instantly feel the pain, realise what has happened and a reflex takes care of removing your hand from the stove. The non-linear behaviour of the ED was not realised from the study's beginning. System dynamics is apt for modelling such behaviour through setting up an analytical model that can show these interactions which the human mind is not able to process intuitively.

With all the obstacles on the road to establishing causal relationships, the challenge may seem overwhelming at a first glance. However, the study of causal mechanisms is too important to abandon. No generic method can be applied to cover all of the above mentioned topics. Various approaches must be undertaken to cover some of these gaps. This has been demonstrated in three of the four journal articles included in this thesis. These journal articles each demonstrate how different aspects of the system can be examined through the use of management science theory in the form of statistical analysis and computer simulation. The major take-home lesson learned was to look into how the processes relate rather than focussing on the performance measures exclusively. Correlating performance measures was possible but conclusions are acknowledged to be of exploratory nature compared to actual confirmation of the relationships. With clearer definitions of the performance measures alongside sticking to these over time, more robust analysis of the causal mechanisms is enabled and should be supported with a thorough understanding of the underlying processes that drive the system, for instance visualised by computer simulation. The final performance measurement model follows Forrester's assumption on the employees' mental models being most rich on information when it comes to understanding how the causal mechanisms work (see Figure 4.2).

R3: How a performance measurement model can be developed

How can a generalised performance measurement model be developed that can monitor quality and efficiency on the basis of knowing about a series of performance measures and their relations?

This study presented a final performance measurement model that utilises all of the acquired knowledge. The solution was set up as a prototype GUI utilising a carefully collected and processed data sample. In practice, it utilises the previously developed hierarchical structure that encapsulates the underlying knowledge on connectivity between the included performance measures. Current ED performance can be holistically assessed, in a user specified time window, through intuitive indication of upwards, downwards, or stagnating trends in each performance measure. With such information, the ED decision maker will be informed of areas that may have been affected by system changes and to which degree this is the case. But can the interventions really be assessed with confidence? Attention must be brought to the discussion of why causal mechanisms are difficult to guarantee. The included performance measures have been connected but it would take further investigation to better understand the *how* and *to what extent* these interact. Moreover, initiatives that change the ED working procedures are sometimes launched in parallel making it problematic to distinguish between which initiative had what effect. Bearing this in mind, it is important to state that the developed performance measurement model is no substitute for logical reasoning or instinct, since not all explanations can, and will never, be found reflected in numerical data elements. Statistical process control was deemed an apt technique to assess variation in as many performance measures as possible, while being easy to explain to outsiders. The technique has for many years been applied in industry with healthcare being a late adopter. Information on selected performance measures of interest can therefore be dispersed at morning meetings, via intranet, or maybe internal newsletters.

In terms of the generalizability potential, several cases were included in different stages throughout the study. All of the included cases shared their views on which performance measures were of greatest value for broad ED assessment. These views were based on personal experiences and what is recorded and analysed at the given site. Small adjustments were necessary when switching between a Danish and an American setting. Examples of performance measures that were irrelevant in the US were nurse practitioners *(behandlersygeplejersker)*, since such a position simply did not exist at the case site. Instead medical residents were used. Surprisingly, no employee satisfaction surveys were routinely reported. No apparent explanation for this lack was provided but what struck the author was the particular the pride exhibited by the employees regarding their work at the case site. Although merely a hypothesis, taking such pride in being a member of an organisation could partly explain why no official employee satisfaction surveys were conducted. Indeed, the three included cases have had the greatest impact on the study's final conclusions. However, it is the author's opinion that despite the minor differences in what need to be measured, the developed performance measurement model is of most joint benefit to many EDs worldwide.

The performance measurement model developed in this study is of course tailored for ED use. However, the issue of correctly measuring a carefully selected set of measures while understanding the underlying complex interactions that drive performance is of great interest to other organisations across domains. The approach utilised in this study can be extended to other domains that seek deeper knowledge of how their system works under the surface.

Before the proposals can become successful in practice

As a word of caution, the road to a successful implementation of the developed performance measurement model is not yet clear. Many performance measurement initiatives fall short despite having identified a seemingly right set of performance measures. Implementation can become obstructed due to three main reasons which are *infrastructural*, *political* and *focus* (Neely and Bourne, 2000). Overcoming these hurdles will significantly improve the potential for a successful implementation of the study's propositions. In the following, each of the three reasons are explained.

A lack of an appropriate infrastructure is the main sinner for cumbersome performance measurement. The problem is that data is spread across the hospital based on content in databases that are incompatible for easy retrieval. For instance, operational data exists in one format and laboratory results in another format. Satisfaction survey scores, whether it is patients or employees, is usually the property of an external third party. If excessive time, effort, and resources are required for processing and understanding the performance data, chances are that focus will eventually slip and the purpose of the developed model fall behind. Often individuals responsible for building and implementing an encompassing and solid infrastructure drown in the attempt because it simply takes too long. Building such a solid foundation is not something that can be quickly implemented. Therefore, it requires great persistence to happen. In a highly dynamic environment such as the ED, there is a real need to maintain sufficient energy levels in the pursuit of completing this task. Should the foundation of attention start to crack, frustrations with the implementation, clarifying definitions of the performance measures etc. will be bound to appear, implying that people will get annoyed with the process.

In terms of political issues, many people link performance measures to unpleasant staff monitoring methods. Many organisations, unfortunately, have applied performance systems as a whip. Employees, regardless of rank and domain, can remember examples of performance measure misuse which point out so called poor performers. In such organisations, where a culture of scapegoating exists, nobody likes performance measurement data to be accessible. There exist the opportunity to *game* with data, especially if this is manually registered. Gaming will occur whenever staff seek to embellish measured performance by knowingly changing variables aside from clinical quality (Goddard *et al.*, 2002). If social influences on outcome are not sufficiently accounted for, healthcare personnel may indeed become reluctant to work in areas with disadvantaged social circumstances. To promote accurate measurement, the leadership in the ED must introduce the performance measurement system in a way to eliminate doubts of its purpose. It must eliminate or restrict its potential for misuse (Lilford *et al.*, 2004).

Even when the organisation has been through the tough journey of implementing the performance measurement system with a supportive infrastructure, staff members will have to be selected to manage the measurement data. The ED decision makers need to allow themselves sufficient time to analyse the performance data, otherwise the effort is wasted. It is not enough to produce the charts and reports. Time must be allocated to analyse the data trends in order to decide what is going to be done differently during the following months or years. The mantra in the 1990s was that we were measuring the wrong things; this has now changed to measuring too much. Focus has to be on how to extract value from the data gathered.

Validity and Reliability

Assessment of the study's outcome boils down to whether the conclusions are reliable and valid. Reliability refers to the demonstration of the operations of a study can be replicated with similar results (Yin, 2009). Validity is closely related to reliability and must be assessed from multiple angles. Since this study applied the philosophical perspective of CR, some of the concepts of validity are somewhat different compared to that of the traditional empiricist. Several types of validity in CR have to be considered, being *measurement validity*, *internal validity*, *ecological validity* and *external validity*. Each of these aspects is described in brief in the following.

For the critical realist, empirical evidence is important when it provides valuable information on the actual events happening in the research setting. Thus, validity is not primarily reflected in the data but rather in a logical evaluation of the link between event and data. Claiming adequate measurement validity involves setting up a chain of evidence on the quality of the information that the measurements enables. Measurement validity does not only concern empirical problems, like biased data, but also includes conceptual problems, such as a measure's ability to depict traces of an event. Cook and Campbell (1979) put forward their highly cited definition of internal validity in 1979 which connects to the empiricist's perspective:

> *"[Internal validity] refers to the approximate validity with which we infer that a relationship between two variables is causal or that the absence of a relationship implies the absence of cause."* (Cook and Campbell, 1979)

In the viewpoint of the critical realist, the generative mechanisms are split ontologically from the event patterns, meaning that internal validity concerns more than events and their concurrences. Here, the main focus is on the identification of natural structures, mechanisms, or processes (beyond the researcher's scope) explained by using events and measurements (within the researcher's scope) (Archer *et al.*, 1998). With high internal validity, it is possible to see the difference between events and the operation of these structures, mechanisms, or processes that cause the events. Ecological validity refers to whether a finding meaningfully reflects an event or process in the world (Kelly *et al.*, 2010). More precisely, it is the strength of the evidence that a particular mechanism that causes events in a well-defined domain is identical to the theorised mechanism for the given phenomenon. Therefore, the ecological validity relates to internal validity. Promoting ecological validity will positively affect the internal validity but does not necessarily imply the opposite. Lastly, external validity can be a tricky concept to grasp as it seeks to assess how a presumed causal mechanism can be extrapolated to other similar cases or even across domains. A study with no external validity can only claim the conclusions to be valid for the particular case that was included. Usually, external validity is concluded based on statistical biases in sample descriptive parameters. However, concluding that a particular relationship is generalizable because of no biases does not convince a critical realist of its correctness. The reason for the scepticism is due to the statistics merging empirical traces with an event, and event with a causal mechanism. For a critical realist, the concern for external validity is more in the explanation as to why an identified causal mechanism would hold outside the research setting. External validity in CR therefore deals with the generative mechanism's activity in settings outside the study while ecological validity looks at how the generative mechanism was activated through a particular change. Multiple tests were necessary to account for all four types of validity. In Table 6.1, a summary of how the different types of validity were approached is presented.

Validity type	**Definition in CR**	**Study tactic**
Measurement	Evidence that the measurement provided valuable information of the events in focus	Multiple sources of evidence, (data collection and assessment)
Internal	Evidence of a specific mechanism having triggered the event registered	Multiple sources of evidence (quantitative testing and qualitative confirmation)
Ecological	Evidence that the hypothesised mechanism causing an effect in the practice domain was activated in the research setting in a way that is reflective of practice	Computer simulation and consensus among study participants
External	Evidence that the mechanism causing the events in the research setting are valid in other cases not included in the study	Inclusion of multiple cases with adherence to study protocol

Table 6.1: Concepts of validity in CR with study tactics

The various types of validity have been assessed in all published articles, although not mentioned explicitly. Some issues within the ED setting deserve mentioning as these will have an evident effect on the validity of the study's propositions. First of all, the ED is a highly dynamic environment that, at the time of writing, is subject to many ongoing changes, some of these even being launched in parallel. With a research setting in constant movement, new working processes may surface making it necessary to re-evaluate the causal mechanism foundation. The performance measurement model in itself is deemed unlikely to lose its relevance since it has been made flexible as to add or discard performance measures of choice. Another, potential danger for validity is the fact that definitions on the various performance measures are still up for debate. When local definitions and sampling techniques of data exist, it compromises *instrumentation validity* making it difficult to compare performance over time, especially external benchmarking (Gardner and Wright, 2009). The study's proposed performance measurement model has changed over time due to expert opinion upon what to measure. This fact can decrease validity. Incremental changes were however restricted making the effect on validity small in impact. Thus, the proposals put forward were considered as valid in terms of these vulnerabilities.

Both clinicians and the scientific community involved in the research process have supported the study's findings. There was a general agreement that the performance measures identified, alongside their interconnections, was in line with their perception of how the ED system functions. There is reason to believe that the developed performance measurement model is of benefit to many EDs that are in need of a comprehensive and intuitive performance monitoring tool. The scientific community (i.e. journals and scientific societies) have shown their interest by accepting conference contributions and the publishing the study's results as journal articles. Both primary stakeholders acknowledge the final proposals, thus adding to validity in context of EDs. Testing the robustness of these proposals can be sought by introducing more diverse cases which may serve to identify validity problems that did not arise in the included cases.

CHAPTER 7

CONCLUSION

In the midst of a rapidly changing healthcare sector, the call for accurate performance measurement is louder than ever. There is now a worldwide focus on how to exploit management science methods to enhance the understanding of healthcare systems and a need to manage this understanding to their own advantage. In Denmark, many changes are especially happening to the ED concept which currently lacks a sound solution to evaluate if these changes are for the better or worse. Therefore, this study has a great interest both in terms of the scientific potential and practical problem solving. Currently, the cases included in this study are applying minimal data sets of locally selected performance measures in an attempt of determining whether their departments are impacted by performance enhancing initiatives. Sporadic retrieval of performance data inhibits a holistic evaluation of departmental functioning as a systematic approach is not followed.

The study first considered which performance measures can provide a generic, broad and comprehensive overview. A systematic literature study of published meta-analyses supported by expert opinions revealed 35 performance measures that were subsequently categorised into aggregated clusters entitled *patients*, *employees* and *operations*. The novelty of the study consists of investigating interconnections between the various performance measures, both within and outside the clusters. Three approaches were utilised, all of which were in compliance with the philosophical view of CR. First, a single complex performance measure was analysed in depth using statistical hypothesis testing in order to identify drivers for employee absence rates. The findings were presented in a decision support tool that could direct the decision maker's attention to areas in need of improvement to lower future absence rates. Second, the aggregated cluster of patient satisfaction was modelled and analysed through the use of SEM. This statistical technique was useful for testing multiple hypotheses simultaneously while accounting for measurement error. This advances the knowledge base since basic statistical hypothesis testing does not assume any measurement error. Third, a SD case study was conducted at a major tertiary level one hospital in Boston, MA. Here, the relevant worldwide issue of ED crowding was examined with a specific primary focus on understanding the complex non-linear feedback mechanisms of initiating a local help protocol with system-wide effect. The focus was therefore moved from looking at performance measures exclusively to the dominant processes of which these performance measures reflect. Exploiting simulation to understand system complexity was a main lesson learned in the study and is something that should be pursued more in the future. All knowledge was exploited in a practical IT-software solution that applied carefully gathered performance data from the included cases.

Some obstacles need solving before a full implementation of the study's proposals is advised. A supportive infrastructure that allows automatic retrieval of relevant parameter values in a user determined time window must be established. Furthermore, the performance measurement model must not be misused for control purposes, making the pitch of the proposal to the staff critical for success. Lastly, time for analysing the data generated through model use should be allowed to gain the necessary understanding of how the system is performing.

The study leaves some open ended questions worth investigating in the future. While several possibilities will be explained in chapter 8, resolving and agreeing on concrete definitions of the identified performance measures is a first step to create a solid foundation for subsequent data analysis. The performance measurement model can then be altered into a model that can be simulated, allowing for an enhanced understanding of system performance. This study has generated results of scientific value based upon publications in internationally peer-reviewed journals and presentations at conferences for both the management science and emergency medicine societies. A rigorous scientific approach to iteratively test and verify the study's sub-conclusions served to secure that the developed performance measurement model is of high validity and reliability. Therefore, the study's proposals are of value to EDs facing similar issues as the cases included. Other organisations can with benefit apply the approach undertaken in the study to gain valuable insights into their system's performance.

The study was launched because of a general desire to enable measurement of the effects from performance enhancing initiatives in the ED setting. The results have proven that it is challenging yet possible to create a scientific model with the capability to measure and connect a selection of essential performance measures. Performance measurement initiatives together with a boosted system complexity understanding are becoming paramount for any organisation, regardless of domain, to stay competitive and target high fidelity improvements. For EDs, the developed performance measurement model is tailored to assess key aspects of departmental performance and as such plays a vital role in supporting decision makers in their task of continuous improvements.

CHAPTER 8

FUTURE RESEARCH OPPORTUNITIES

Several opportunities for elaborating upon the results presented in this PhD thesis exist. First and foremost, the final outcome of the thesis is a prototype that has not yet proven its full potential. Implementation at the case hospitals would strengthen the reliability and validity of the final performance measurement model. It is hoped that the performance measurement model can assist the primary users (i.e. clinical directors) in assessing given initiatives and thus to navigate their department towards continuous improvements. Some hurdles need passing before full implementation is possible. Importing data from the existing data warehouse at the given hospital needs to be automated. Setting up algorithms within the software solution capable of doing so would be easiest if the relevant data elements are located in a single database. None of the included cases had a single data warehouse were all data could be accessed. Furthermore, it was concluded that not all data elements were recorded despite these being recommended in international peer-reviewed literature. Full implementation thus requires careful sampling of the missing performance measures over time. As the ED system is in constant change, the suggested prototype is made flexible so that it can allow adding- or deleting performance measures in the framework. This flexibility means that the system must be maintained and occasionally updated so that it will not lose its value. Any performance measure added to the model must comply with the requirements for relevance to the stakeholders, scientific soundness, feasibility in retrieving precise data, and relevance to the overall ED's strategy. Additionally, which other performance measures new measures may connect to must be taken into consideration.

Assessing data validity

Paramount to accurate performance assessment is high precision data. Much of the data included in the performance measurement model can be derived from different ED data repositories. Investigating the accuracy of this data is an apparent research area in order to establish trustworthy reporting procedures for quality monitoring programs. For instance, if timestamps applied to measure departmental success rates is inaccurately measured, the conclusions drawn about current performance levels will be highly biased. With unreliable data, performance measurement becomes inaccurate and can at worse be misleading (Smith, 1994). Indeed, there is a call for rigorous ED performance measurement, notably an understanding of how the data tracking systems record the various measurements. Two overall types of electronic ED tracking systems are available; *active* and *passive* (Gordon *et al.*, 2008). An active tracking system needs manual registration by the ED staff so that the system's information to be up to date. In contrast, passive tracking systems are automated registration points that, for instance, can trace the physical position of patients and employees by the use of wireless sensor technology. A passive tracking system may also be known as a "real-time location system". Accuracy of this recorded data is deemed highly correlated with the distinction of the two data recording systems. It would be interesting to analyse the two data tracking systems and their impact on data variation and accuracy. This could be achieved through a comparative study of selected identically defined performance measures at two or more case EDs.

Condition specific performance measures

Adding extra modules to the proposed performance measurement model adds a more nuanced view of departmental performance. Various performance measures for highly represented patient groups are of special interest to the primary stakeholder, i.e. the ED decision maker. Some disease specific conditions that are common in the ED are pneumonia, strokes, acute myocardial infarctions, STEMI, sepsis, and asthma. Individual performance measures specific for such patient groups provides more detailed information of clinical quality of care. For example, a selection of recommended clinical measures for the three first mentioned diseases is presented in Table 8.1.

Each of these specific conditions deserves to be examined in depth according to which performance measures should be recorded and how. Such measurements could further allow for national benchmarking amongst hospitals if a consensus agreed set of performance measures is determined. While Table

Disease	Measure
Pneumonia	Patient oxygeniation assessment Time from patient arrival to initial antibiotic administration Were correct antibiotics administered? Blood cultures obtained prior to antibiotic administration?
Stroke	Time from patient arrival to stroke team notified Time from stroke team notified to response Time from lab tests ordered to completed tests and interpreted Time from neurosurgery evaluation to start evaluation
AMI	% administered aspirin on arrival % administered beta blockers on arrival % thrombolytics initiated within 30 minutes after arrival % PCI within 90 minutes after arrival

Table 8.1: Examples of clinical performance measures for disease specific conditions; PCI = percutaneous intervention (Graff *et al.*, 2002; Lindsay *et al.*, 2002; Schull *et al.*, 2011)

8.1 presents a number of clinical performance measures, it is worth noting that one should not fall short on linking these to patient life quality. According to Michael Porter of Harvard Business School, importance to the patients' quality of life after having received treatment must be emphasised (Reiermann *et al.*, 2014). Porter provides an example of German hospitals rendering high class prostate cancer treatment with only 5 % post treatment mortality after five years, regardless which German hospital conducts the surgery. However, patients are also concerned about how well they function after treatment, i.e. issues with incontinence and erection. Here, there are great differences between the success rates of the hospitals. Measuring such outcomes gives a more holistic overview of how delivered care works and provides a more qualified statement about healthcare quality.

Economical assessment

The developed performance measurement model does not include financial viability and can therefore only be used to evaluate an ED's operational performance in connection with the quality of outcomes. A cost efficiency analysis of implementing the various changes is necessary to determine if an intervention is to be concluded as a success seen from a hospital manager's viewpoint. The reason for leaving out economical measures at first was a primary concern of measuring the operational impact of initiatives. Inclusion of financial performance would move the model towards a more classic Balanced Scorecard model which is believed to be of higher value to top-level hospital managers compared to that of clinical decision makers. Indeed, it is challenging to put price tags on the many procedures carried out in the ED, mostly because each patient's continuity of care is unique. On the contrary, it is deemed possible to make estimates based on the total expenditure of treating a patient within a certain patient group, possibly specified within triage levels. It is the author's opinion that assessing the return on investments makes most sense when rigorous measurement techniques are installed and maintained for some time before specifying economical measures and when the stakeholder perspective of hospital managers is undertaken.

Even though the cost side of the initiatives is not included explicitly, the effects of launching these can be assessed on a series of performance measures that all relate to operational performance. No generally accepted set of economical ED performance measures exist (to the author's knowledge) but there may be a possibility to develop such variables without actually measuring these in reality. Computer simulation offers this possibility and is definitely seen as an important area that should be exploited in the future. For this purpose, a discrete event simulation model is currently regarded the best choice as this simulation technique is usually applied for point prediction. A hybrid simulation technique that combines the strengths of discrete event simulation and system dynamics is another possibility for healthcare assessment. The hybrid technique is fairly new and not many studies have investigated its benefits (Cernohorsky and Voracek, 2012; Djanatliev and German, 2013).

Towards a decision support tool

The final performance measurement model has the potential of being reformulated into a simulation model as demonstrated in section 5.4. The qualitative aspect of SD consisting of a causal loop diagramming was utilised to set up the links between the included performance measures in the final performance measurement model. The possibility to extend this network into a quantitative SD model, framing a particular problem of importance, would be of great value. It must be emphasised that one should always model a problem and not an entire system, because with no boundaries there would be basis for saying "there is no need to include that" when potential variables are suggested. A SD model aims to develop a test environment where system policy changes can be assessed instantaneously with no costs associated. The greatest payoff is the ability to determine time delays and complex non-linear relations between variables – something that holds great value to any decision maker regardless domain. Visualising performance levels of selected performance measures meanwhile assessing variations in processes is indeed a valuable goal worth pursuing. Extending insights into the very structure behind the processes responsible for system behaviour kicks the current knowledge base to a higher level. Thorough understanding of potential interventions' effects would cause what to happen in the system before launching these in reality would result in better performing systems with higher value for money. When it comes to choosing a simulation technique, SD is usually preferred to model problems at a higher strategic level in order to get insights into the relations between various parts of a complex system. Discrete event simulation is applied to model tactical- or operational level systems for point prediction. An example where a discrete event simulation model would be appropriate in healthcare would be to identify resource bottlenecks (Brailsford and Hilton, 2001).

Patient handoffs between departments or sectors

Currently, Danish patients in need of urgent treatment must first contact a GP, the 1813 system (if in the Capital Region of Denmark), or the 112 emergency call system before being referred to treatment at a hospital, first the ED and maybe also an inpatient ward. Therefore, there is a lot of information being handed over not only between sectors but also in the hospital. A "transition" or "handoff" is a commonly used term used when two or more healthcare personnel make an exchange of responsibility, authority of an operation, or specific clinical information (Cheung *et al.*, 2010). Clear and effective communication between the emergency physician and other healthcare personnel needs rigorous research as information tends to "get lost in translation" (Knutsen and Fredriksen, 2013). Most research looking at patient handoffs are qualitative and exploratory in study approach with a focus on, for instance, patient safety factors (Siemsen, 2011). Standardised procedures and checklists have recently been proved to have a positive impact on information exchange without increasing process times (Dubosh *et al.*, 2014). From an operational management viewpoint, there is a potential for applying simulation to either quantify the improvements of implementing handoff protocols through discrete event simulations or investigate the dynamics behind information exchanges between sectors using SD. Regardless of the choice of simulation technique, broadening the scope to include either the in-hospital patient handoffs or handoffs from one sector to another would allow for a more holistic view of emergency care delivery.

BIBLIOGRAPHY

ACEP (2006), *Crowding*, Annals of Emergency Medicine **47**(6), 585.

ACKROYD, S. (2010), *Critical Realism, Organization Theory, Methodology, and the Emerging Science of Reconfiguration*, in *Elements of a Philosophy of Management and Organization*, pp. 47–77, Springer-Verlag, Berlin, DE.

ACKROYD, S. AND FLEETWOOD, S. (2000), *Realist perspectives on management and organisations*, Routledge, London, UK.

ANDERSON, P.D., PETRINO, R., HALPERN, P., AND TINTINALLI, J. (2006), *The globalization of emergency medicine and its importance for public health*, Bulletin of the World Health Organization **84**(10), 835–839.

ARCHER, M., BHASKAR, R., COLLIER, A., LAWSOM, T., AND NORRIE, A. (1998), *Critical Realism: Essential Readings*, Routledge, London, GB.

ASUSI (2008), *ASUSI-projektet - slutrapport*, Technical Report September, Center for Forskning vedr. Sygefravær og Udstødning (CFSU).

ATHLIN, A.A., VON THIELE SCHWARZ, U., AND FARROHKNIA, N. (2013), *Effects of multidisciplinary teamwork on lead times and patient flow in the emergency department: a longitudinal interventional cohort study.*, Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine **21**(1), 76.

BECHTEL, W. (2008), *Mental Mechanisms: Philosophical Perspectives on Cognitive Neuroscience*, Taylor & Francis, New York, NY.

BENNEYAN, J.C., LLOYD, R.C., AND PLSEK, P.E. (2003), *Statistical process control as a tool for research and healthcare improvement*, Quality & Safety in Health Care **12**(6), 458–464.

BENTLER, P.M. (1980), *Multivariate Analysis with Latent Variables: Causal Modeling*, Annual Review of Psychology **31**(1), 419–456.

BERG, G.M., CROWE, R.E., NYBERG, S., AND BURDSAL, C. (2012), *Trauma patient satisfaction with physician assistants: Testing a structural equation model*, Journal of the American Academy of Physician Assistants **25**(5), 42–51.

BHASKAR, R. (1975), *A Realist Theory of Science*, Harvester Press, Hemel Hempstead, second edition.

BHASKAR, R. (1979), *The Possibility of Naturalism*, Harvester Press, Sussex, UK.

BINDER, T., VOX, A., BELYAZID, S., HARALDSSON, H., AND SVENSSON, M. (2004), *Developing System Dynamics Models from Causal Loop Diagrams*, in *22nd International Conference of the System Dynamics Society*, Oxford, UK.

BLANK, R.H. AND BURAU, V.D. (2004), *Comparative Health Policy*, Palgrave MacMillan, Houndmills, UK.

BOLLEN, K.A. (1989), *Structural Equations with Latent Variables*, Wiley, New York, NY.

BRABRAND, M. AND EKELUND, U. (2010), *Emergency medicine in Scandinavia - an outstanding opportunity for research*, Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine **18**(1), 5.

BRAILSFORD, S. AND HILTON, N. (2001), *A Comparison of Discrete Event Simulation and System Dynamics for Modelling Healthcare Systems*, in RILEY, J (editor), *Planning for the Future: Health Service Quality and Emergency Accessibility. Operational Research Applied to Health Services (ORAHS)*, pp. 1–17, Glasgow Caledonian University, Southampton, UK, first edition.

BUCH-HANSEN, H. AND NIELSEN, P. (2005), *Kritisk Realisme*, Roskilde Universitetsforlag, Frederiksberg, DK, first edition.

BURGESS, J.F. (2012), *Innovation and efficiency in health care: does anyone really know what they mean?*, Health Systems **1**(1), 7–12.

CAMERON, P.A., SCHULL, M.J., AND COOKE, M.W. (2011), *A framework for measuring quality in the emergency department*, Emergency Medicine Journal **28**(9), 735–740.

CARLSEN, C. (2009), *New hospital structure – investments in new hospitals*, Technical report, Danish Regions, Kongens Vænge, DK.

CARLSSON, S.A. (2010), *Design Science Research in Information Systems: A Critical Realist Approach*, in *Design Research in Information Systems SE - 15*, volume 22 of *Integrated Series in Information Systems*, pp. 209–233, Springer US.

CERNOHORSKY, P. AND VORACEK, J. (2012), *Towards Public Health Policy Formulation*, in *International Proceedings of Economics Development & Research 52*, pp. 126–130.

CHEUNG, D.S., KELLY, J.J., BEACH, C., BERKELEY, R.P., BITTERMAN, R.A., BROIDA, R.I., DALSEY, W.C., FARLEY, H.L., FULLER, D.C., GARVEY, D.J., KLAUER, K.M., MCCULLOUGH, L.B., PATTERSON, E.S., PHAM, J.C., PHELAN, M.P., PINES, J.M., SCHENKEL, S.M., TOMOLO, A., TURBIAK, T.W.,

VOZENILEK, J.A., WEARS, R.L., AND WHITE, M.L. (2010), *Improving handoffs in the emergency department*, Annals of Emergency Medicine **55**(2), 171–180.

CMM (2014), *Quality Measures*, https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityMeasures/.

COLEMAN, P. AND NICHOLL, J. (2010), *Consensus methods to identify a set of potential performance indicators for systems of emergency and urgent care*, Journal of Health Services Research & Policy **15**(2), 12–18.

COOK, T.D. AND CAMPBELL, D. (1979), *Quasi-experimentation: Design & analysis issues for field settings*, Rand McNally, Chicago, IL.

DANERMARK, B., EKSTRÖM, M., JAKOBSEN, L., AND KARLSSON, J.C. (2002), *Explaining society - Critical realism in the social sciences*, Routledge, London, UK.

DHMA (2007), *Styrket akutberedskab - planlægningsgrundlag for det regionale sundhedsvæsen*, Technical report, Danish Health and Medicines Authority, Copenhagen, DK.

DHMA (2014), *Faglig gennemgang af akutmodtagelserne*, Technical report, Danish Health and Medicines Authority, Copenhagen, DK.

DJANATLIEV, A. AND GERMAN, R. (2013), *Large Scale Healthcare Modeling by Hybrid Simulation Techniques using AnyLogic Categories and Subject Descriptors*, in *Proceedings of the 6th International ICST Conference on Simulation Tools and Techniques*, pp. 248–257, Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, Cannes, FR.

DONABEDIAN, A. (2005), *Evaluating the Quality of Medical Care*, The Milbank Quarterly **83**(4), 691–729.

DPV (2014), *Helpline 1813*, Technical report, Den Præhospitale Virksomhed.

DRACHMANN, H. (2011), *Sparekniven ramte rigtigt*, Politiken **3**.

DUBOSH, N.M., CARNEY, D., FISHER, J., AND TIBBLES, C.D. (2014), *Implementation of an emergency department sign-out checklist improves transfer of information at shift change*, The Journal of Emergency Medicine **47**(5), 580–585.

EASTON, G. (2010), *Industrial Marketing Management Critical Realism in Case Study Research*, Industrial Marketing Management **39**(1), 118–128.

FABRIGAR, L.R., WEGENER, D.T., MACCALLUM, R.C., AND STRAHAN, E.J. (1999), *Evaluating the Use of Exploratory Factor Analysis in Psychological Research*, Psychological Methods **4**(3), 272–299.

FARROHKNIA, N., CASTREN, M., EHRENBERG, A., LIND, L., OREDSSON, S., JONSSON, H., ASPLUND, K., AND GORANSSON, K.E. (2011), *Emergency Department Triage Scales and Their Components: A Systematic Review of the Scientific Evidence*, Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine **19**(1), 42.

FitzGerald, G., Jelinek, G.A:, Scott, D., and Gerdtz, M.F. (2010), *Republished paper: Emergency department triage revisited*, Postgraduate Medical Journal **86**(1018), 502–508.

Forrester, J.W. (1969), *Urban Dynamics*, MIT Press, Boston, MA.

Forrester, J.W. (1971), *Counterintuitive behavior of social systems*, Technological Forecasting and Social Change **3**(3), 1–22.

Forrester, J.W. (1994), *System Dynamics, Systems Thinking, and Soft OR*, System Dynamics Review **10**(2), 1–14.

Forrester, J.W. and Senge, P.M. (1980), *Tests for Building Confidence in System Dynamics Models*, TIMS Studies in the Management Science **14**, 209–228.

Gardner, T.M. and Wright, P.M. (2009), *Implicit human resource management theory: a potential threat to the internal validity of human resource practice measures*, The International Journal of Human Resource Management **20**(1), 57–74.

Gerdtham, U.G., Rehnberg, C., and Tambour, M. (1999), *The impact of internal markets on health care efficiency: evidence from health care reforms in Sweden*, Applied Economics **31**(8), 935–945.

Goddard, M., Davies, H.T.O., Dawson, D., Mannion, R., and McInnes, F. (2002), *Clinical performance measurement: part 2–avoiding the pitfalls*, Journal of the Royal Society of Medicine **95**(11), 549–551.

Golomb, B.A., McGraw, J.J., Evans, M.A., and Dimsdale, J.E. (2007), *Physician Response to Patient Reports of Adverse Drug Effects*, Drug Safety **30**(8), 669–675.

Gordon, B.D., Flottemesch, T.J., and Asplin, B.R. (2008), *Accuracy of Staff-Initiated Emergency Department Tracking System Timestamps in Identifying Actual Event Times*, Annals of Emergency Medicine **52**(5), 504–511.

Graff, L., Stevens, C., Spaite, D., and Foody, J. (2002), *Measuring and improving quality in emergency medicine*, Academic Emergency Medicine **9**(11), 1091–1107.

Graves, H.B., Mills, J.D., and Wiegenstein, J.C. (1986), *History of Emergency Medicine*, Annals of Emergency Medicine **15**(4), 493.

Guba, E.C. (1990), *The Alternative Paradigm Dialog*, in Guba, E.C. (editor), *The Paradigm Dialog*, pp. 17–30, Sage Publications, Newbury Park, CA.

Hallas, P., Ekelund, U., Bjornsen, L.P., and Brabrand, M. (2013), *Hoping for a domino effect: a new specialty in Sweden is a breath of fresh air for the development of Scandinavian emergency medicine*, Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine **21**(1), 26.

Hart, M.K. and Hart, R.F. (2002), *Statistical Process Control for Health Care*, Duxbury Applied Series/Thomson Learning, Pacific Grove, CA.

HAYS, L.W. (1981), *Statistics*, Holt, Rinehart and Winston, New York, NY, third edition.

HEDSTRÖM, P. AND YLIKOSKI, P. (2010), *Causal Mechanisms in the Social Sciences*, Annual Review of Sociology **36**(1), 49–67.

HÖLZEL, L.P., KRISTON, L., AND HÄRTER, M. (2013), *Patient preference for involvement, experienced involvement, decisional conflict, and satisfaction with physician: a structural equation model test*, BMC Health Services Research **13**(231), 1–10.

HOLM-PETERSEN, C. (2010), *Fremtidens arbejdstilrettelæggelse for speciallæger påFælles Akut Modtagelser*, Technical report, Dansk Sundhedsinstitut, Copenhagen, DK.

HOOD, C. (1995), *The "new public management" in the 1980s: Variations on a theme*, Accounting, Organizations and Society **20**(2-3), 93–109.

HUME, D. (1967), *Enquiries Concerning Human Understanding and the Principles of Morals*, Clarendon Press, Oxford, UK.

HUNDBORG, K. (2009), *The Danish Quality Model*, Ugeskrift for Laeger **171**(20), 1660–1663.

HUNT, J.C.R., TIMOSHKINA, Y., BAUDAINS, P. J., AND BISHOP, S.R. (2012), *System Dynamics Applied to Operations and Policy Decisions*, European Review **20**(03), 324–342.

IMAI, K., KEELE, L., TINGLEY, D., AND YAMAMOTO, T. (2011), *Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies*, American Political Science Review **105**(04), 765–789.

IOM (2001a), *Crossing the Quality Chasm: A New Health System for the 21st Century*, National Academy Press, Washington, DC.

IOM (2001b), *Envisioning the National Health Care Quality Report*, Technical report, National Academies Press, Washington DC.

JOHNSTON, R. AND SMITH, S. (2008), *How critical realism clarifies validity issues in theory-testing research: analysis and case*, in HART, D. AND GREGOR, S. (editors), *Information Systems Foundations: The Role of Design Science*, chapter 2, pp. 1–318, ANU E Press, Canberra, AU.

JONASSEN, D.H. AND IONAS, I.G. (2006), *Designing effective supports for causal reasoning*, Educational Technology Research and Development **56**(3), 287–308.

JÖRESKOG, K.G. (1969), *A general approach to confirmatory factor analysis*, Psychometrika **34**(2), 183–202.

JÖRESKOG, K.G. AND SÖRBROM, D. (1982), *Recent Developments in Structural Equation Modeling*, Journal of Marketing Research **19**(4), 404 – 416.

KEESLING, J. (1972), *Maximum Likelihood Approaches to Causal Flow Analysis*, Unpublished phd thesis, University of Chicago, IL.

KELLY, D., FU, X., AND SHAH, C. (2010), *Effects of position and number of relevant documents retrieved on users' evaluations of system performance*, ACM Transactions on Information Systems **28**(2), 1–29.

KELMAN, S. AND FRIEDMAN, J.N. (2009), *Performance Improvement and Performance Dysfunction: An Empirical Examination of Distortionary Impacts of the Emergency Room Wait-Time Target in the English National Health Service*, Journal of Public Administration Research and Theory **19**(4), 917–946.

KING, G. (2005), *The Dangers of Extreme Counterfactuals*, Political Analysis **14**(2), 131–159.

KLINE, R.B. (2011), *Principles and Practice of Structural Equation Modeling*, volume 3, Guilford Press, New York, NY.

KNOKE, D., BOHRNSTEDT, G.W., AND MEE, A.P. (2002), *Statistics for Social Data Analysis*, Wadsworth, CENGAGE Learning, Belmont, CA, fourth edition.

KNUTSEN, G.O. AND FREDRIKSEN, K. (2013), *Usage of documented pre-hospital observations in secondary care: a questionnaire study and retrospective comparison of records*, Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine **21**(1), 13.

LAURITSEN, J. AND PACKNESS, AA. (2010), *Statistisk Proces Kontrol, Introduktion og håndbog. SPC diagrammer og analyser med brug af EpiData*, Odense Universitetshospital, Odense, DK, first edition.

LAWSON, T. (1997), *Economics and Reality*, Routledge, London, UK.

LEE, K. AND HUH, S. (2006), *A model-solver integration framework for autonomous and intelligent model solution*, Decision Support Systems **42**(2), 926–944.

LIBERATI, A., ALTMAN, D.G., TETZLAFF, J., MULROW, C., GØ TZSCHE, P.C., AND JOHN, P.A. (2009), *The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration*, British Medicine Journal **339**, b2700.

LILFORD, R., MOHAMMED, M.A., SPIEGELHALTER, D., AND THOMSON, R. (2004), *Use and misuse of process and outcome data in managing performance of acute medical care : avoiding institutional stigma*, Lancet **363**, 1147–1154.

LINDSAY, P., SCHULL, M., BRONSKILL, S., AND ANDERSON, G. (2002), *The development of indicators to measure the quality of clinical care in emergency departments following a Modified-Delphi approach*, Academic Emergency Medicine **9**(11), 1131–1139.

LIPPERT, A., ESPERSEN, K., ANTONSEN, K., JOENSEN, H., WALDAU, T., AND LARSEN, K.M. (2007), *Kapaciteten pådanske intensivafdelinger. En national survey af kapacitet, aflysninger og overflytninger af kritisk syge patienter*, Ugeskrift for Laeger **169**(8), 712 – 716.

LUNA-REYES, L.F. AND ANDERSEN, D.L. (2003), *Collecting and analyzing qualitative data for system dynamics: methods and models*, System Dynamics Review **19**(4), 271–296.

MACKENHAUER, J. AND PETERSEN, D.B. (2014), *Akutafdelingerne mangler de rette speciallæger*, Jyllandsposten pp. 1–2.

MAINZ, J. (2003), *Defining and classifying clinical indicators for quality improvement*, International Journal for Quality in Health Care **15**(6), 523–530.

MAINZ, J., HANSEN, A., PALSHOF, T., AND BARTELS, P.D. (2009), *National quality measurement using clinical indicators: the Danish National Indicator Project*, Journal of Surgical Oncology **99**, 500–504.

MARTINEZ-MOYANO, I.J. AND RICHARDSON, G.P. (2013), *Best practices in system dynamics modeling* †, System Dynamics Review **29**(2), 102–123.

MAYNTZ, R. (2004), *Mechanisms in the Analysis of Social Macro-Phenomena*, Philosophy of the Social Sciences **34**(2), 237–259.

MAZLACK, L.J. (2007), *General imprecise causal reasoning representations*, in *2007 IEEE International Conference on Systems, Man and Cybernetics*, pp. 480–485, IEEE, Montreal, CA.

MAZLACK, L.J. (2009), *General Causal Representations in the Medical Domain*, in *2009 2nd International Conference on Biomedical Engineering and Informatics*, pp. 1–5, IEEE, Tianjin, CO.

MEADOWS, D. H., MEADOWS, D. L., RANDERS, J., AND BEHRENS III, W. W. (1972), *The Limits to Growth*, Earth Island, UK.

MILES, M.B. AND HUBERMAN, A.M. (1994), *Qualitative Data Analysis*, Sage Publications, London, UK.

MILLER, I., FREUND, J., AND JOHNSON, R. (2005), *Miller & Freund's probability and statistics for engineers*, Pearson Higher Education, New Jersey, NJ.

MINGERS, J. (2003), *A Critique of Statistical Modelling from a Critical Realist Perspective*, in *European Conference on Information Systems*, pp. 1–13, Naples, IT.

MINGERS, J. (2004), *Real-izing information systems: critical realism as an underpinning philosophy for information systems*, Information and Organization **14**(2), 87–103.

MINGERS, J. (2009), *The Contribution of Critical Realism as an Underpinning Philosophy for OR / MS and Systems*, Journal of Operational Research Society **51**(11), 1256–1270.

MINGERS, J. (2014), *Systems Thinking, Critical Realism and Philosophy*, Routledge, London, UK.

MINGERS, J., MUTCH, A., AND WILLCOCKS, L. (2013), *Critical Realism in Information Systems Research*, MIS Quarterly **37**(3), 795–802.

MONTGOMERY, D.C. (2005), *Introduction to Statistical Quality Control*, Wiley, Hoboken, NJ.

MOORE, L., HANLEY, J.A., TURGEON, A.F., AND LAVOIE, A. (2010), *Evaluating the performance of trauma centers: hierarchical modeling should be used*, The Journal of Trauma **69**(5), 1132–1137.

MORTON, A.P., CLEMENTS, A.C., DOIDGE, S.R., STACKELROTH, J., CURTIS, M., AND WHITBY, M. (2008), *Surveillance of healthcare-acquired infections in Queensland, Australia: data and lessons from the first 5 years*, Infection Control and Hospital Epidemiology **29**(8), 695–701.

MPAC (2006), *Report to the Congress: Increasing the Value of MEDICARE*, Technical report, MedPAC, Washington DC.

MUTHEN, L.K. AND MUTHEN, B.O. (2007), *Mplus User's Guide*, volume 5, Statmodel, Los Angeles, CA.

NEELY, A. AND BOURNE, M. (2000), *Why Measurement Initiatives Fail*, Measuring Business Excellence **4**(4), 3–7.

NQMC (2014), *National Quality Measures Clearinghouse - Selecting Measures*, http://www.qualitymeasures.ahrq.gov/tutorial/selecting.aspx.

OECD (2013), *Health at a Glance 2013: OECD Indicators*, OECD Publishing, Paris.

OREDSSON, S., JONSSON, H., ROGNES, J., LIND, L., GÖRANSSON, K.E., EHRENBERG, A., ASPLUND, K., CASTRÉN, M., AND FARROHKNIA, N. (2011), *A systematic review of triage-related interventions to improve patient flow in emergency departments.*, Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine **19**(1), 43.

ORLIKOWSKI, W.J. AND BAROUDI, J.J. (1991), *Studying Information Technology in Organizations: Research Approaches and Assumptions*, Information Systems Research **2**(1), 1–28.

PALMER, S. AND TORGERSON, D.J. (1999), *Economics notes - Definitions of efficiency*, British Medicine Journal **318**(7191).

PEARL, J. (2001), *Causal Inference in the Health Sciences : A Conceptual Introduction*, Health Services & Outcomes Research Methodology **2**(3-4), 189–220.

PELLEGRINO, E.D. (1999), *The commodification of medical and health care: the moral consequences of a paradigm shift from a professional to a market ethic*, The Journal of Medicine and Philosophy **24**(3), 243–266.

PETTIGREW, A.M. (1990), *Longitudinal field research on change: theory and practice*, Organization Science **1**(3), 267–292.

PINES, J.M., HILTON, J.A., WEBER, E.J., ALKEMADE, A.J., SHABANAH, H.A., ANDERSON, P.D., BERNHARD, M., BERTINI, A., GRIES, A., FERRANDIZ, S., KUMAR, V.A., AND HARJOLA, V.P. (2011), *International Perspectives on Emergency Department Crowding*, Academic Emergency Medicine **18**(12), 1358–1370.

PORPORA, D. (1998), *Do realists run regressions?*, in *2nd International Centre for Critical Realism Conference*, University of Essex, Colchester, UK.

REGH (2009), *Generalplan: fremtidens fysiske rammer for patienter i Region Hovedstaden*, RegH, Copenhagen, DK.

REGH (2014), *Sundhedsplatformen*, www.sundhedsplatform.dk.

REIERMANN, J., MØLLER, M., AND RASMUSSEN, B. (2014), *Michael Porter: Danmarks offentlige sektor måler sine resultater helt forkert*, Mandag Morgen pp. 1–6.

RICHARDSON, G.P. (2013), *System Dynamics*, in *Encyclopedia of Operations Research and Management Science*, p. 1641, Imprint: Springer.

RIMDAL, T.O. AND SOERENSEN, P.V. (2012), *Patienters oplevelser i Region Hovedstadens akutmodtagelser og akutklinikker*, Technical report, Unit of Patient-Perceived Quality, Copenhagen, DK.

ROUSE, W.B. AND SERBAN, N. (2011), *Understanding change in complex socio-technical systems*, Information Knowledge Systems Management **10**, 25–49.

RUBIN, H.J. AND RUBIN, I.S. (2012), *Qualitative interviewing: the art of hearing data*, Sage Publications, London, UK, third edition.

SALTMAN, R.B. AND FIGUERAS, J. (1997), *European health care reform: analysis of current strategies*, World Health Organization, Regional Office for Europe, London, UK.

SCHREIBER, J.B. (2008), *Core reporting practices in structural equation modeling*, Research in Social and Administrative Pharmacy **4**(2), 83–97.

SCHREUDER, J.A.H., ROELEN, C.A.M., VAN ZWEEDEN, N.F., JONGSMA, D., VAN DER KLINK, J.L.J., AND GROOTHOFF, J.W. (2011), *Leadership styles of nurse managers and registered sickness absence among their nursing staff*, Health Care Management Review **36**(1), 58–66.

SCHULL, M.J., GUTTMANN, A., LEAVER, C.A., VERMEULEN, M., HATCHER, C.M., ROWE, B.H., ZWARENSTEIN, M., AND ANDERSON, G.M. (2011), *Prioritizing performance measurement for emergency department care: consensus on evidence-based quality of care indicators*, Canadian Journal of Emergency Medicine **13**(5), 300–309.

SCOTT, I., VAUGHAN, L., AND BELL, D. (2009), *Effectiveness of acute medical units in hospitals: a systematic review*, International Journal for Quality in Health Care **21**(6), 397–407.

SEDGWICK, P. (2012), *Pearson's correlation coefficient*, British Medicine Journal **345**, e4483.

SENGE, P.M. (2006), *The Fifth Discipline : The Art and Practice of the Learning Organization*, Random House Business, London, UK.

SHAH, R. AND GOLDSTEIN, S.M. (2006), *Use of structural equation modeling in operations management research: Looking back and forward*, Journal of Operations Management **24**(2), 148–169.

SHEWHART, W.A. (1928), *Economic aspects of engineering applications of statistical methods*, Journal of the Franklin Institute **205**, 395–405.

SHOOK, C.L., KETCHEN JR., D.J., HULT, G.T., AND KACMAR, K.M. (2004), *An assessment of the use of structural equation modeling in strategic management research*, Strategic Management Journal **25**(4), 397–404.

SIEMSEN, I.M.D. (2011), *Patientovergange - et eksplorativt studie af faktorer der påvirker sikkerheden af patientovergange*, Phd thesis, Technical University of Denmark, Kongens Lyngby, DK.

SKRIVER, C., LAURITZEN, M.M.P., FORBERG, J.L., GAARDBOE-POULSEN, O.B., MOGENSEN, C.B., HANSEN, C.L., AND BERLAC, P.A. (2011), *Triage medfører hurtigere behandling af de mest syge*, Ugeskrift for Laeger **173**, 2490–2493.

SMITH, A.J. (1994), *The Need for Measured Data in Computer System Performance Analysis or Garbage In, Garbage Out*, in *Proceedings Eighteenth Annual International Computer Software and Applications Conference (COMPSAC 94)*, pp. 426–431.

SMITH, P.C., MOSSIALOS, E., PAPANICOLAS, I., AND LEATHERMAN, S. (2009), *Performance Measurement for Health System Improvement - Experiences, Challenges and Prospects*, Cambridge University Press, New York, NY, third edition.

SOBH, R. AND PERRY, C. (2006), *Research design and data analysis in realism research*, European Journal of Marketing **40**(11/12), 1194–1209.

SORIAN, R. (2006), *Measuring, reporting, and rewarding performance in health care*, Technical Report 912, National Commitee of Quality Assurance.

SOYEMI, K. (2012), *Choosing the right statistical test*, Pediatrics in Review **33**(5), e38–44.

SØRUP, C.M. AND JACOBSEN, P. (2013), *Healthcare performance turned into decision support*, Journal of Health Organization and Management **27**(1), 64–84.

SØRUP, C.M. AND JACOBSEN, P. (2014), *Patient safety and satisfaction drivers in emergency departments re-visited - an empirical analysis using structural equation modeling*, Health Systems **3**(2), 105–116.

SØRUP, C.M., JACOBSEN, P., AND FORBERG, J.L. (2013), *Evaluation of emergency department performance - a systematic review on recommended performance and quality-in-care measures.*, Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine **21**, 62.

STERMAN, J.D. (2000), *Business Dynamics - Systems Thinking and Modeling for a Complex World*, Irwin/McGraw-Hill, Boston, MA, first edition.

TABACHNICK, B.G. AND FIDELL, L.S. (2013), *Using Multivariate Statistics*, volume 6, Pearson Education, Allyn & Bacon, Inc. Needham Heights, MA.

TONG, A., SAINSBURY, P., AND CRAIG, J. (2007), *Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups*, International Journal for Quality in Health Care **19**(6), 349–357.

TRABERG, A., JACOBSEN, P., AND DUTHIERS, N.M. (2014), *Advancing the use of performance evaluation in health care*, Journal of Health Organization and Management **28**(3), 422–436.

VOSS, C., TSIKRIKTSIS, N., AND FROHLICH, M. (2002), *Case research in operations management*, International Journal of Operations & Production Management **22**(2), 195–219.

WELCH, S.J., ASPLIN, B.R., STONE-GRIFFITH, S., DAVIDSON, S.J., AUGUSTINE, J., AND SCHUUR, J. (2011), *Emergency Department Operational Metrics, Measures and Definitions: Results of the Second Performance Measures and Benchmarking Summit*, Annals of Emergency Medicine **58**(1), 33–40.

WELCH, S., AUGUSTINE, J., CAMARGO, C.A., AND REESE, C. (2006), *Emergency department performance measures and benchmarking summit*, Academic Emergency Medicine **13**(10), 1074–1080.

WILER, J.L., GENTLE, C., HALFPENNY, J.M., HEINS, A., MEHROTRA, A., MIKHAIL, M.G., AND FITE, D. (2010), *Optimizing Emergency Department Front-End Operations*, Annals of Emergency Medicine **55**(2), 142–160.e1.

WILEY, D. (1973), *The Identification Problem for Structural Equation Models with Unmeasured Variables*, in GOLDBERGER, A.S. AND DUNCAN, O.D. (editors), *Structural Equation Models in the Social Sciences*, pp. 69–83, Academic Press, New York, NY.

WILLIS, C.D., GABBE, B.J., AND CAMERON, P.A. (2007), *Measuring quality in trauma care*, Injury **38**(5), 527–537.

WRIGHT, S. (1921), *Correlation and causation*, Journal of Agricultural Research **20**(7), 557–585.

WYNN, D. AND WILLIAMS, C.K. (2012), *Principles for conducting critical realist case study research in information systems*, MIS Quarterly **36**(3), 787–810.

YIN, R.K. (2009), *Case Study Research: Design and Methods*, volume 5 of *Applied Social Research Methods Series*, Sage Publications, Thousand Oaks, CA.

APPENDED PAPERS

The four journal papers which this thesis is built upon are put in this final chapter of the thesis. The first three of these were published by the time of writing. The final journal paper was currently undergoing its first review at Health Care Management Science.

PAPER 1

SCANDINAVIAN JOURNAL OF
trauma, resuscitation
& emergency medicine

REVIEW **Open Access**

Evaluation of emergency department performance – a systematic review on recommended performance and quality-in-care measures

Christian Michel Sørup[1*], Peter Jacobsen[1] and Jakob Lundager Forberg[2]

Abstract

Background: Evaluation of emergency department (ED) performance remains a difficult task due to the lack of consensus on performance measures that reflects high quality, efficiency, and sustainability.

Aim: To describe, map, and critically evaluate which performance measures that the published literature regard as being most relevant in assessing overall ED performance.

Methods: Following the PRISMA guidelines, a systematic literature review of review articles reporting accentuated ED performance measures was conducted in the databases of PubMed, Cochrane Library, and Web of Science. Study eligibility criteria includes: 1) the main purpose was to discuss, analyse, or promote performance measures best reflecting ED performance, 2) the article was a review article, and 3) the article reported macro-level performance measures, thus reflecting an overall departmental performance level.

Results: A number of articles addresses this study's objective (n = 14 of 46 unique hits). Time intervals and patient-related measures were dominant in the identified performance measures in review articles from US, UK, Sweden and Canada. Length of stay (LOS), time between patient arrival to initial clinical assessment, and time between patient arrivals to admission were highlighted by the majority of articles. Concurrently, "patients left without being seen" (LWBS), unplanned re-attendance within a maximum of 72 hours, mortality/morbidity, and number of unintended incidents were the most highlighted performance measures that related directly to the patient. Performance measures related to employees were only stated in two of the 14 included articles.

Conclusions: A total of 55 ED performance measures were identified. ED time intervals were the most recommended performance measures followed by patient centeredness and safety performance measures. ED employee related performance measures were rarely mentioned in the investigated literature. The study's results allow for advancement towards improved performance measurement and standardised assessment across EDs.

Keywords: Performance, Measures, Indicators, Emergency department, Quality improvement, Quality

* Correspondence: cmiso@dtu.dk
[1]DTU Management Engineering, Technical University of Denmark, Produktionstorvet, building 424, 2800, Kongens Lyngby, Denmark
Full list of author information is available at the end of the article

Background

In Europe, many EDs have undergone organisational changes [1,2]. Hospitals receiving acute patients are increasingly merged to larger organizations. Continuous high expertise in the EDs is promoted through the presence of relevant resources, medical specialties and experienced staff [2]. In Denmark, the new concept behind EDs consists of merging all acute admission units and observatory units into one joint ED. The rationale for re-structuring is first and foremost to cope with an increased amount of patients while securing delivery of high quality and efficiency, concurrently with decreased overall hospital capacity [3]. Promotion of interdisciplinary teamwork and earlier senior physician involvement are examples of means to deliver timely and high quality treatment to patients within the EDs, which is essential for early diagnosis and provision of effective treatment of the increasing number of patients with comorbidities [4,5]. Other prevalent changes include introducing emergency medicine as a separate specialty [6] and formalised use of triage systems [7]. Many different ways of organising the ED is evolving and the costs and effects are being debated [8]. A way of assessing the effect on the re-organisation and the many local initiatives is highly warranted.

Inspired by the private service sector's way of monitoring and evaluating work processes, health care decision makers have seen the importance of adopting a similar view on management [8]. Hence, an increasing number of quality- and performance measurement initiatives have been integrated within the core operations. Performance measurement is a broad topic, which is rarely defined in detail. Most commonly, it is referred to as the process of quantifying actions, where measurement is a process of quantification and following action leads to performance [9]. Individual performance measures are defined as metrics that reflect effectiveness and/or efficiency of an action. A selection of such performance measures thus comprises of a performance measurement system which enables a more comprehensive evaluation of performance. Widely acknowledged performance measurement frameworks such as the Balanced Scorecard [10] and Business Excellence [11] have been implemented in health care to assure strategy alignment with operations. Even though a high percentage of performance measurements initiatives fail, mainly due to either being poorly designed or too difficult to implement in practice [12], successful implementation and use has been reported [13,14].

The fundamental idea of quality assurance in health care was originally to pass accreditations, whereas the healthcare sector now strives to converge quality improvements wherever possible. Many EDs have accepted the Institute of Medicine's (IOM) report from 2001 called "Crossing the Quality Chasm" [15]. In this report, six quality domains are endorsed. These are *safety*, *effectiveness*, *patient-centeredness*, *timeliness*, *efficiency* and *fairness* (*equity*). The terms *efficiency* and *effectiveness* are often used interchangeably. Efficiency refers to the effectiveness of specific procedures whereas effectiveness regards the total outcome [15].

Different initiatives are continuously being presented in EDs in response to the IOM domains. In the United Kingdom (UK), crowded EDs were sought resolved by the introduction of the four hour target as a primary performance measure [16]. This means that only 98% of the patients may stay within the ED for more than four hours.

Focus on a single time-related measure does not necessarily correspond to high levels of quality and can potentially lead to dysfunctional behaviour [17]. Other important performance areas become unmonitored when focussing only on few ultimate measures. As an example, patients are without adequate treatment transferred to other wards more rapidly to keep length of stay in the ED within the accepted upper threshold limits. This can lead to reduced quality, increased costs and difficulties in retaining staff (sustainability). The outcome of the measure would be great yet the obtained quality would be poor.

Asking the clinicians in UK EDs about the subsequent effects of the four hour target resulted in a governmental report in which a total of eight performance measures to best represent quality were suggested by the Department of Health [18]. Eight performance measures were chosen on the basis of best possible evidence, formulated by lay-representatives and are weighted equally (in theory).

The UK EDs are not alone in the dilemma of determining how to evaluate new initiatives on key performance measures aligned with department visions. Similar problems such as crowding and scarce resources are struggled with elsewhere in the world. The selection of which performance measures to highlight also differs according to stakeholder perspective [19]. A clinician's perspective on highly important performance measures is distinct compared to that of a patient, policy maker, or administrator, mainly due to the use of the measures for varying purposes. The entities may be subject to alteration over time depending on evolving clinical evidence, new practices and procedures, public opinions, and health system dynamics. Whereas a policy maker's chief concern involves public accountability or a measurement framework reflecting 'pay for performance', the clinicians will demand procedural improvements for the benefit of enhanced treatment outcomes and clinical safety. From a patient's perspective, the main focus will be on patient centeredness considered excellent medical treatment is delivered.

Consensus is still lacking on which measures are considered to be most accurate, extensive, clearly defined, and based on evidence [20,21]. Working towards a consensus of performance measures that reflect the general performance of an ED and whether or not certain quality improvement initiatives prove efficient is clearly warranted. A shared understanding of performance measures will enable continuous quality improvements and benchmarking opportunities both internally and externally over time.

The aim of this article is to present an overview of the highlighted performance measures suggested in internationally peer-reviewed review articles through the application of PRISMA guidelines.

Methods

Literature search strategy

This review gathers information on published results of review articles highlighting performance measures suitable for overall ED assessments. Identification of such articles was done through a systematic search in the databases of PubMed, Cochrane Library, and Web of Science conducted in the period of April to July, 2012. For all searches performed, the term "emergency department" or "ED" was used as fixed standard phrases. A selection of combined searches was conducted using the following text strings: emergency department, ED, performance indicator(s), performance measure(s), quality assessment, quality assurance, and quality improvement.

To investigate synonyms to the variable search terms, MeSH headings and wildcards were applied. The searches, with the variable search terms, and resulting number of hits are presented in Table 1.

PubMed differs between the singular and plural forms of phrases, hence the distinction shown in parenthesis. For example, the term "performance measures" is recognised as a different keyword compared to "performance measure" although nearly half of the articles reoccurred. A performed search is based on wordings in both titles and abstracts. The searches were performed separately using Boolean operators: "emergency department" OR "ED" AND the given search term. All search hits were filtered to only include reviews and structured reviews in terms of article type. The searches were performed according to the PRISMA guidelines [22].

Table 1 Search strings and resulting hits

Search	Variable search string	# Hits
1	performance measure/(performance measures)	13/(46)
2	performance indicator/(performance indicators)	13/(36)
3	quality measure/(quality measures)	24/(47)
4	quality indicator/(quality indicators)	19/(200)
5	quality assessment	67
6	quality evaluation	2
7	performance assessment	15
8	performance evaluation	13
9	quality assurance	657
10	quality improvement	9

Inclusion/exclusion criteria

Articles were included in the systematic review if they managed to fulfil all of the following stated criteria: 1) the main purpose was to discuss, analyse, or promote performance measures that best reflect ED performance, 2) the articles were review articles, and 3) the articles reports macro-level performance measures, thus reflecting an overall departmental performance level.

Articles were excluded if 1) they referred to a specific patient group or illness, 2) the setting was different than EDs, 3) they did not touch upon performance measurement, 4) they investigated evidence behind selected indicators, 5) they described how measures were used or should be used in the future, 6) they directed criticism towards vaguely defined performance measures, and 7) the language was different from English.

Selection

Selection of articles was performed independently by two of the authors (CMS and PJ) by reviewing titles and abstracts. If any doubts arose, the entire article was assessed. Afterwards, a decision about possible inclusion was made on the basis of a discussion between the two authors (CMS and PJ).

Synthesis of results

According to the Traberg categorisation, all recommended performance measures can be allocated into the three categories; 1) patients, 2) employees, and 3) operations [23]. Traberg's framework was chosen due to its sensible division of performance measures into clusters seen from a clinician's viewpoint.

Results

Study selection

A total of 1314 titles were identified from the applied databases. 46 of these were unique. The unique titles were scanned on the basis of both title and abstract. Then, the inclusion/exclusion criteria were applied leaving 38 articles to be read in full extent. Of the 38 articles, 14 of these met the eligibility criteria and were included for further analysis. A flowchart presenting the article selection process is shown in Figure 1.

By accumulating all search hits presented in Table 1, 38 redundant articles were marked and excluded. A total of 1276 articles' titles and abstracts were afterwards screened based on the inclusion/exclusion criteria. Any article title

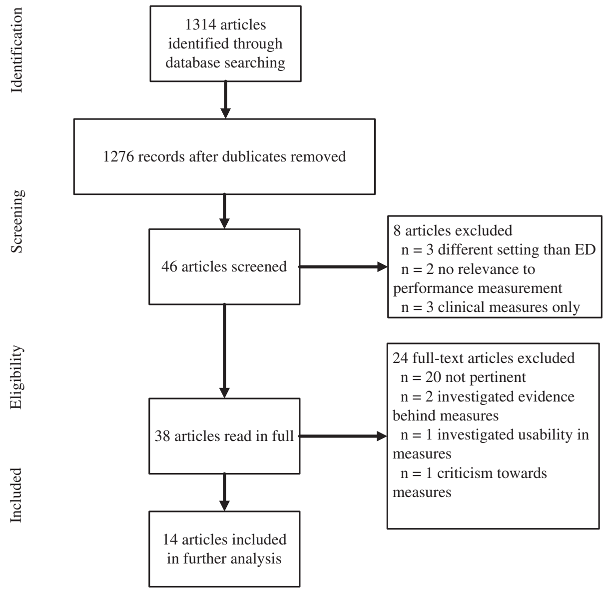


Figure 1 Flow-chart of article selection.

or abstract not deemed relevant by the authors were discarded (n = 1230). These articles were discarded due to 1) being conducted in a setting different from emergency departments or, 2) not relating to performance measurement, or 3) focusing on a specific clinical condition measure (for instance percentage of paediatric asthma patients prescribed anti-inflammatory medication or time to antibiotics given to patients suffering from pneumonia [8]).

The initial filtering returned a total of 46 articles in which all abstracts were read. Eight of these articles adopted an approach which was not in compliance with this study's inclusion criteria. As an example of an excluded article, Persell et al., 2011 implemented and evaluated a quality improvement intervention, which included several clinical condition specific measures [24].

24 articles were not included in the final review due to 1) being non-pertinent, 2) investigated the evidence behind certain indicators [25,26], 3) described how measures were used and will be used in the future [27], or 4) directed criticism towards vaguely defined quality indicators [28]. 14 articles remained to be analysed and compared.

The reference lists of the 14 final articles were browsed for possible relevant articles, yet none met the criteria for inclusion.

Characteristics of the included studies

A comparison of the included articles is presented in Table 2. The two last columns in Table 2 indicate first the preliminary pool of performance measures analysed and second the actual recommended performance measures.

No literature older than ten years that reviews overall ED performance measures was identified. The included articles formulate their primary objective differently but ultimately come up with a list of performance measures which reflect key performance- and quality-in-care measures in EDs. All these performance measures relate to a macro-level aspect, implying that these are generally applicable. In terms of the articles' settings, USA and Canada have had the greatest focus on how to assess performance in EDs based on publications. The UK, Sweden and Australia have now also published their view on what performance measures to report. As units of analysis, paediatric EDs and general EDs were both eligible for analysis in this article since there is no difference in the generally applicable performance measures highlighted when referring to patient age. A differentiation between age and gender would be advisable if the performance measures were matched to specific clinical conditions.

With regards to the chosen approach, most of the articles apply a survey based approach consisting of two or

Table 2 Presentation of included literature (* Focus indicates whether the suggested indicators are more generally applicable or refers to clinical conditions (e.g. indicators related to specific ailments))

Corresponding author	Year	Objective	Focus*	Setting	Method	*Gross indicator portfolio*	*Recommended indicators*
McClelland et al. [33]	2012	Examination of practical aspects in collecting time-based ED measures	Time-related measures only (7)	American, EDs	Structured interviews and few data comparisons	7	7
Beniuk, Boyle & Clarkson [29]	2012	To prioritise quantified crowding measures to assess current ED status	Overall (8)	International EDs (USA, UK, Canada, Australia, Netherlands and Hong Kong)	Standard three round Delphi study	27	8
Alessandrini et al. [15]	2011	Proposition of a measurement framework specific for PEC practitioners and administrators	Overall (13) and condition specific (1)	American, PEDs	Point of departure in IOM recommendations. Alteration into Donabedian's structure, process, outcome categorisation	120	14
Ekelund et al. [34]	2011	1) To assess feasibility in gathering benchmark data in Swedish EDs and 2) to evaluate patient throughput times and inflow patterns	Overall (4)	Sweden, EDs	Comparison of variables reflecting quality measures	4	4
Heyworth [35]	2011	1) Benefits and drawbacks associated with a single time-related measure and 2) proposed quality indicators to assess timeliness, quality, and safety	Overall (8)	United Kingdom, EDs	Description of current state in the UK; reflection on the quality indicators proposed by the Department of Health	8	8
Schull et al. [21]	2011	Seeks consensus on a set of parsimonious quality-of-care indicators for an ED	Overall (11) and condition specific (2)	Canada, EDs	Modified Delphi panel technique, three rounds	170	13
Welch et al. [32]	2011	Consensus of a standard set of performance measures in EDs related to patient flow	Overall (44)	American, North American Benchmark Summit (367 EDs)	Survey and audit	44	44
Coleman & Nicholl [16]	2010	Identification of a indicators usable for PCT commissioners and NHS decision makers to monitor performance	Overall (16)	United Kingdom, EDs and Urgent Care Units	Standard three round Delphi study	70	16
Hung & Chalut [30]	2008	1) Presents which indicators are deemed most useful to assess PEC and 2) which measures are currently being recorded	Overall (15)	Canada, PEDs	2-part questionnaire including a novel ranking formula to prioritize indicators	67	15
Guttmann et al. [31]	2006	Development of measures relevant for paediatric emergency care (children < 19)	Overall (6) and condition specific (8)	American, PEDs	Structured panel process with underlying literature review	109	14
Sibbritt, Isbister & Walker [36]	2006	Provision of a recommended list of performance indicators from routinely collected data in EDs	Overall (9)	Australia, EDs	Data collection and following SPC analysis	9	9
Solberg et al. [3]	2003	Identification of measures in EDs relevant for managing crowding	Overall (38)	American, EDs	Expert consensus on 113 measures; 10 investigators refined the measures to a total of 38	113	38

Table 2 Presentation of included literature (* Focus indicates whether the suggested indicators are more generally applicable or refers to clinical conditions (e.g. indicators related to specific ailments)) *(Continued)*

Graff et al. [8]	2002	How to critically evaluate quality in an ED	Overall (9) and condition specific (29)	American, EDs	Summary. Point of departure in IOM recommendations. Afterwards alteration into Donabedian's structure, process, outcome categorisation	38	38
Lindsay et al. [20]	2002	A systematic approach to identify valid and relevant measures in an ED	Overall (8) and condition specific (13)	Canada, EDs	Modified Delphi panel technique, two rounds	104	21

ED Emergency Department, *IOM* Institute of Medicine, *NHS* National Health Services, *PCT* Primary Care Trust, *PEC* Paediatric Emergency Care, *PED* Paediatric Emergency Department, *SPC* Statistical Process Control.

more rounds of questioning panel members (commonly designated as the Delphi technique) [3,16,20,21,29-31]. This approach serves the purpose of finding consensus for a given topic by filtering responses through every stage. Two review papers report on interviews or audits [32-34]. A single article refers to a British governmental report [35]. Two articles elaborate on the IOM guidelines [8,15] and a single article includes performance measurement tracking by the application of statistical process control (SPC) [36].

A difference exists between the number of performance measures ultimately recommended and the gross pool of indicators investigated for several of the articles. These to amounts are listed as the two last columns of Table 2.

Duplicate performance measures were filtered out only if the wording differed slightly. The authors included core measures if the level of detail was deemed too specific. An example of a low abstraction level can be found in Welch et al. 2011 where the performance measure LOS is divided for admitted-, discharged-, observational-, and behavioural health patients [32].

All recommended performance measures are presented in Tables 3, 4 and 5 in compliance with Traberg's three overall categories.

Some of the suggested performance measures connected to patients needs to be more precisely defined before use. Often indirect measures have to be used instead and that explains the common use of unplanned re-attendance as a performance measure, because it indirectly reflects a missed diagnosis or inadequate treatment. One may argue that the safety measures rather reflect operational efficiency. Such measures' both quantitative and qualitative character explains why they are a part of the proposed framework. For instance, the number of unintended incidents is without value if not accompanied by a qualitative description to pinpoint what went wrong.

Most highlighted in the cluster *patient centeredness* was the outcome measure LWBS (patients who leave before being seen by a physician). This measure can be hypothesised to be related to patient satisfactory levels, as it is often related to crowding and extensive waiting times.

LWBS is also regarded as an important measure due to documented increased risks and adverse outcome in patients leaving before being treated [19,21,29].

A high rate in LWBS points toward potential systemic obstacles in patient reception or triage.

In terms of patient centeredness and satisfaction surveys, none of the included articles elaborates on which latent constructs must be recorded to reflect overall patient satisfaction levels. Much other literature addresses this issue using diverse approaches [37-39]. Employee related performance measures are presented in Table 4.

Only two performance measures connected to employees are suggested by Sibbritt et al. 2006 [36] and Welch et al. 2011 [32] (see Table 4). Insertion of the employee perspective in quality improvements has only recently been suggested by Crouch & Cooke in 2011 [40], entailing that in the future there may be a change in the demeanour of ED performance measurement. Employee related performance measures provide a pointer to which degree the current performance is sustainable. Sustainable performance is also linked to measures such as sickness absence rates, educational programme outcomes, and the amount of staff having the necessary competencies to fulfil their respective job descriptions [19]. In contrast to the employee related performance measures, operational performance measures have harvested more interest. These are presented in Table 5.

The operational performance measures deal primarily with effectiveness mainly related to time based registrations. Changes in working procedures serve a two-fold purpose; 1) timely treatment and 2) improving quality-in-care.

The main ED tasks are fast recognition and treatment of time-dependent critical conditions plus fast disposition to adequate level of care. Therefore, the great focus on time intervals is not a surprising result. LOS is far the most used time interval. LOS is an indirect overall measure of the efficiency of the whole ED stay. Keeping

Table 3 Patient related measures

		Alessandrini 2011	Beniuk 2012	Coleman 2010	Ekelund 2011	Guttmann 2006	Graff 2002	Heyworth 2011	Hung 2008	Lindsay 2002	McClelland 2012	Schull 2011	Sibbritt 2006	Solberg 2003	Welch 2011
Patients															
Safety	Unintended incidents	x				x	x		x		x	x			
	Medication errors	x					x		x						
	Treatment errors						x					x			
	Missed diagnosis					x	x					x			
	Morbidity/mortality			x	x	x			x	x			x		
	Unplanned re-attendance (<72 hours)	x				x		x	x	x		x	x		
Patient centeredness	Complaints	x													x
	Patients Who Left Before Supposed To (PWLBST)														x
	LWBS (Left Without Being Seen)	x	x				x	x	x			x	x		x
	LBTC (Left Before Treatment Complete)		x											x	x
	LAMA (Left Against Medical Advice)														x
Satisfaction	Satisfaction (in general)/ survey	x			x		x	x				x			

Table 4 Employee related performance measures

		Alessandrini 2011	Beniuk 2012	Coleman 2010	Ekelund 2011	Guttmann 2006	Graff 2002	Heyworth 2011	Hung 2008	Lindsay 2002	McClelland 2012	Schull 2011	Sibbritt 2006	Solberg 2003	Welch 2011
Employees															
Occupation profile	Educational positions												x		x
Work environment	Employee complaint ratio														x

Table 5 Operational performance measures; Note: double dashes between factors indicates a time interval

		Alessandrini 2011	Beniuk 2012	Coleman 2010	Ekelund 2011	Guttmann 2006	Graff 2002	Heyworth 2011	Hung 2008	Lindsay 2002	McClelland 2012	Schull 2011	Sibbritt 2006	Solberg 2003	Welch 2011
Operations															
Planning	Acute load													x	
	Bed occupancy rate												x	x	
	Boarding burden		x											x	
Utilization	Utilization rate (lab equipment)	x				x									
	Number of ECG's taken									x					x
	Number of plain radiographic studies														x
	Number of CT studies														x
	Number of MRI studies														x
	Number of ultrasonic studies														x
	Number of laboratory studies														x
	Overall medication usage														x
	Number of behavioural health consultations														x
	Number of specialty consultations														x
	Utilization rate (employees)									x					
Efficiency	Throughput													x	
	ED admission transfer rate		x	x										x	
Time intervals	LOS (Length of Stay), total		x		x	x	x	x	x	x	x	x	x	x	x
	Ambulance off-loading time		x									x			x
	Arrival -- Registration														x
	Arrival -- Treatment space														x
	Arrival -- Clinical assessment	x	x		x		x	x	x			x			x
	Arrival -- Hospitalization	x		x					x			x		x	
	Arrival -- Init. triage	x	x										x		
	Arrival -- Init. treatment			x				x			x	x			
	Registration -- Init. triage	x							x						
	Registration -- Discharge/ transfer								x						
	Triage -- Triage completed														x

Table 5 Operational performance measures; Note: double dashes between factors indicates a time interval *(Continued)*

	Triage -- Init. treatment	x			x			x		
	Admit decision -- Discharge								x	x
	Treatment space -- init. encounter									x
	Init. encounter -- Init. treatment		x							x
	Init. encounter -- Hospitalization						x			
	Init. encounter -- Clinical decision		x		x					
	Init. encounter -- Discharge/ transfer						x			
	Disposition decision -- Discharge									x
	Hospitalization -- Discharge/ transfer				x	x	x			
(diagnostic imaging)	Registration -- X-ray ordered									x
	X-ray ordered -- X-ray taken (radiology turnaround)			x		x			x	x
	Data ready -- Disposition decision time									x
(laboratory)	Blood sample ordered -- Blood sample result (lab turnaround)	x								x
(bed logistics)	Bed ordered -- Bed assigned									x

LOS short also means reducing crowding and keeping an efficient patient flow. Despite that timely treatment is one of the main performance goals for an ED, it is notable that time to treatment is only the ninth most highlighted performance measure (see Figure 2). LOS is often an easy parameter to retrieve from the ED computer system and is relatively easy to define. Time to treatment is more difficult to define and often not as easy to register in a standardised manner. In addition one could argue that time to treatment should be divided into treatment time related to triage category. Thus, data availability and easily defined measures could influence the choice of measures. Indeed, other stakeholders than clinicians contribute to the focus of timely treatment, especially in the wake of crowding. Such stakeholders are patient associations, politicians, and the media.

Presented in Figure 2 is the top 25% of the hits identified in the included literature.

Discussion

The investigated articles differ in their approach, yet share their primary objective which is to analyse, discuss, or promote a series of performance measures that reflect key performance metrics and quality-in-care in emergency departments.

No literature older than ten years that reviews overall ED performance measures was found. During the recent five years, there has been an intensified debate on ED performance measurement. This comes in response to a previous low prioritisation of the emergency medicine area and an increase in ED patient volume over recent years.

A qualitative approach in choosing performance measures seems dominant. Especially the Delphi technique seeking consensus through either audits or questionnaires serves as a means to filter suggestions into core performance measures best suitable for ED assessments.

A total of 55 different performance measures are highlighted in the investigated literature. The level of abstraction in the included papers differs from four to 44 performance measures in total. Most of the suggested performance measures are independent on patient specific indicators and thus serve to reflect overall ED performance levels.

Patients

Patient *safety* is challenging in the highly complex and time critical environment with undifferentiated patients in the ED. Thus, it is an absolutely essential measure which is confirmed by several recommended measures and is being suggested in most of the relevant literature. Tracking the conclusive outcome, mortality and morbidity, seems highly warranted but can be difficult to obtain, except for some of the well-developed countries that register much health statistics, for instance in Scandinavia. Especially mortality reviews engage clinicians and serve as a means for continuous quality improvements [19].

Employees

As the most apparent stakeholder, the patient must remain as the paramount focus and all internal procedures must strive to yield as much value as possible to the quality-in-care. In the periphery of performance measurement focus, treatment services are performed by the employee, who is an essential resource for maintaining the daily operations. High quality treatment and optimal patient flow correlates with a high level of employee contentment, low turnover, and great seniority [41].

In the included literature, the employee aspect has to date not been given a high priority in the assessment of ED performance [21].

Operations

Welch et al. raises the question of how and when to define that the actual patient progress has begun [32]. Does it start when the patient arrives at the ED or when the patient is registered in the ED administration system? Ideally, the registration begins when the patient enters the ED facility but in practice this is difficult to obtain. Therefore, the starting point often is at patient registration. Local circumstances from patient arrival to registration become a factor to include when benchmarking externally.

How many performance measures to include?

Many emergency departments register large amounts of data. Probably, not all registered data is being used. As an ED decision-maker, it is impossible to investigate causes and effects from all registered data. Therefore, it is a necessity to determine which registrations appear most rich in information. It is important to find equilibrium of the required number of performance measures and invested work in collecting data. An extensive amount of performance measures may enable detailed analysis on the expense of extended time consumption. Few performance measures have the advantage of quick overview and thus lack the ability to take multiple aspects of performance into account. As can be read from Table 2, the amount of recommended performance measures varies greatly as a result of desired levels of detail.

Criticism towards performance measurement in EDs

In parallel to the literature recommending certain performance measures, it is important to take notice of the literature which adopts a more critical perspective towards the focus on performance measures [26]. In this

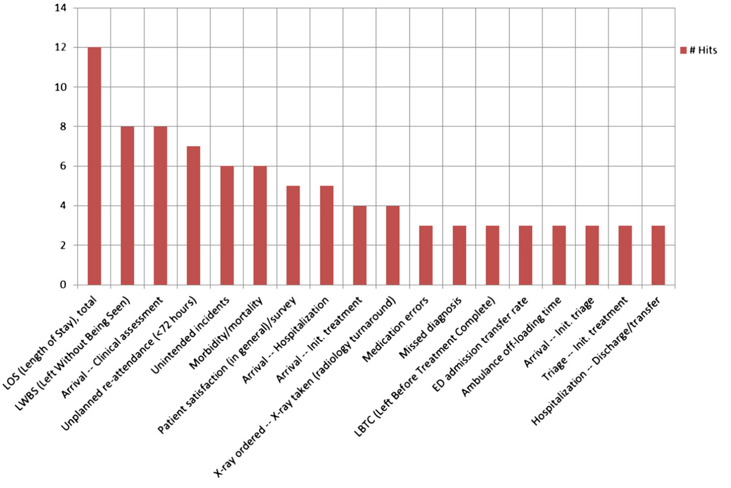


Figure 2 Top 25% of highlighted performance measures in included literature.

literature, evidence and formalisation of ED performance measures is questioned.

Evaluation of what is actually being measured, how to provide evidence for the choices of performance measures, and what the consequences are in implementation of any performance measurement framework is essential.

The authors acknowledge that once a set of performance measures is agreed upon, these should preferably be maintained over time to obtain sufficient data to add statistical strength, validity and reliability to each measure. It is then, ED decision-makers are provided the basis to decide whether to keep or discard given performance measures.

Sibbritt et al. suggests using statistical process control (SPC) when monitoring the department's performance over time [36]. SPC is used to filter common cause variations from special cause variations. Application of SPC, either control- or run-charts, makes it possible to track alterations' effects on key performance measures and is increasingly used in International Health Institute related projects [42].

Data validity questionable

Gordon, Flottemesch, and Asplin report systematic errors and non-normal distributions in ED timestamps which weakens the foundation on which managers make their decisions [43]. Outcomes may also be prone to alterations if employees are given the opportunity to report better status than what is evident [44].

Once a set of performance measures are selected, validation should include a longitudinal study of the retained set of performance measures to ensure construct validity and that clinical processes are driven in the wanted direction [19].

Perspective

Future challenges include a consensus on which performance measures should be in current focus to grasp crucial aspects of performance and contemporarily defining how these should be measured. Some performance measures may only be useful on a local level. However, comparing essential performance measures between EDs could promote learning that supports further quality improvements. It is imperative to agree upon definitions on key terms and measures to promote comparability between ED efficiency and effectiveness.

A joint set of identically defined performance measures across EDs would be beneficial in terms of benchmarking and ultimately continuous quality improvements.

Further studies should investigate the interconnectivity between the selected performance measures. Insight into the performance measures' mutual impact allows for better understanding of ED performance. Furthermore, the use of SPC is deemed a highly important tool in data-driven leadership, since it is useful in measuring if initiatives have the intended effects or variations occur due to common causes.

Conclusions

A structured literature review served the purpose of identifying which performance measures were analysed, discussed or recommended to assess ED performance on a macro-level. The most emphasised performance measures were time intervals and patient-related measures. Only few articles referred to the measurement of employee relevant measures.

In order to monitor the effect on different ED organisations and initiatives, consensus on a shared set of performance measures is needed. Consensus should include agreement on *how* and *when* the data registrations are gathered. These questions are crucial to address for streamlining performance measurement, which could allow for comparability between similar departments.

Moreover, investigation of the interconnectivity between the performance measures and how to measure if launched initiatives have the wanted effects is a sensible future research area.

Competing interests
The authors have no competing interests to declare.

Authors' contributions
CMS has written the manuscript as well as conducted the literature review. PJ has critically reviewed the manuscript. JLF has formulated the study objective and critically reviewed the manuscript. All authors read and approved the final manuscript.

Author details
[1]DTU Management Engineering, Technical University of Denmark, Produktionstorvet, building 424, 2800, Kongens Lyngby, Denmark. [2]The Emergency Department, Nordsjællands Hospital, Dyrehavevej 29, 3400, Hillerød, Denmark.

Received: 23 January 2013 Accepted: 5 August 2013
Published: 9 August 2013

References
1. Alberti G: *Transforming emergency care in England*; 2005. http://aace.org.uk/wp-content/uploads/2011/11/Transforming-Emergency-Care-in-England.pdf
2. Danish Health and Medicines Authority: *Styrket akutberedskab - planlægningsgrundlag for det regionale sundhedsvæsen*; 2007. http://www.sst.dk/publ/publ2007/plan/akutberedskab/styrketakutberedskab.pdf
3. Solberg LI, Asplin BR, Weinick RM, Magid DJ: **Emergency department crowding: consensus development of potential measures.** *Ann Emerg Med* 2003, **42**(6):824–834.
4. Walley P, Silvester K, Mountford S: **Health-care process improvement decisions: a systems perspective.** *Int J Health Care Qual Assur* 2006, **19**(1):93–104.
5. Shetty A, Gunja N, Byth K, Vukasovic M: **Senior streaming assessment further evaluation after Triage zone: a novel model of care encompassing various emergency department throughput measures.** *Emerg Med Australas* 2012, **24**(4):374–382.
6. Hallas P, Ekelund U, Bjornsen LP, Brabrand M: **Hoping for a domino effect: a new specialty in Sweden is a breath of fresh air for the development of Scandinavian emergency medicine.** *SJTREM* 2013, **21**(1):26.
7. Göransson KE, Ehrenberg A, Ehnfors M: **Triage in emergency departments: national survey.** *J Clin Nurs* 2005, **14**(9):1067–1074.
8. Graff L, Stevens C, Spaite D, Foody J: **Measuring and improving quality in emergency medicine.** *Acad Emerg Med* 2002, **9**(11):1091–1107.
9. Neely A, Gregory M, Platts K: **Performance measurement system design: a literature review and research agenda.** *IJOPM* 2005, **25**(12):1228–1263.
10. Kaplan RS, Norton DP: **The balanced scorecard: measures that drive performance.** *Harv Bus Rev (USA)* 2005, **83**(7):172–180.
11. Nicolay CR, Purkayastha S, Greenhalgh A, Benn J, Chaturvedi S, Phillips N, Darzi A: **Systematic review of the application of quality improvement methodologies from the manufacturing industry to surgical healthcare.** *Br J Surg* 2012, **99**(3):324–335.
12. Neely A, Bourne M: **Why measurement initatives fail.** *MBE* 2000, **4**(4):3–7.
13. Huang S, Chen P, Yang M, Chang W, Lee H: **Using a balanced scorecard to improve the performance of an emergency department.** *Nurs Econ* 2004, **22**(3):140–146.
14. Samani DZ, Kohan MZ: **Framework for implementing balanced scorecard in hospitals.** *JIBED* 2011, **5**(4):362–368.
15. Alessandrini EA, Knapp J: **Measuring quality in pediatric emergency care.** *Clin Ped Emerg Med* 2011, **12**(2):102–112.
16. Coleman P, Nicholl J: **Consensus methods to identify a set of potential performance indicators for systems of emergency and urgent care.** *J Health Serv Res Policy* 2010, **15**:12–18.
17. Kelman S, Friedman JN: **Performance improvement and performance dysfunction: an empirical examination of distortionary impacts of the emergency room wait-time target in the English National Health Service.** *JPART* 2009, **19**(4):917–946.
18. Department of Health: *A&E Clinical Quality Indicators - data definitions*; 2011. http://www.fadelibrary.org.uk/wp/wp-content/uploads/downloads/2011/09/AE-clinical-quality-indicators-data-definitions.pdf
19. Cameron PA, Schull MJ, Cooke MW: **A framework for measuring quality in the emergency department.** *Emerg Med J* 2011, **28**(9):735–740.
20. Lindsay P, Schull M, Bronskill S, Anderson G: **The development of indicators to measure the quality of clinical care in emergency departments following a modified-Delphi approach.** *Acad Emerg Med* 2002, **9**(11):1131–1139.
21. Schull MJ, Guttmann A, Leaver CA, Vermeulen M, Hatcher CM, Rowe BH, Zwarenstein M, Anderson GM: **Prioritizing performance measurement for emergency department care: consensus on evidence-based quality of care indicators.** *CJEM* 2011, **13**(5):300–309.
22. Moher D, Liberati A, Tetzlaff J, Altman DG: **Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement.** *Int J Surg* 2010, **8**(5):336–341.
23. Traberg A: *Management-By-Objectives in Healthcare*, PhD Thesis. Technical University of Denmark: DTU Management Engineering; 2011.
24. Persell SD, Kaiser D, Dolan NC, Andrews B, Levi S, Khandekar J, Gavagan T, Thompson JA, Friesema EM, Baker DW: **Changes in performance after implementation of a multifaceted electronic-health-record-based quality improvement system.** *Med Care* 2011, **49**(2):117–125.
25. Stelfox HT, Bobranska-Artiuch B, Nathens A, Straus SE: **A systematic review of quality indicators for evaluating pediatric trauma care.** *Crit Care Med* 2010, **38**(4):1187–1196.
26. Stelfox HT, Straus SE, Nathens A, Bobranska-Artiuch B: **Evidence for quality indicators to evaluate adult trauma care: a systematic review.** *Crit Care Med* 2011, **39**(4):846–859.
27. Pines JM, Fermann GJ, Ferroggiaro AA, Irvin CB, Mazer M, Frank Peacock W, Schuur JD, Weber EJ, Pollack CV: **The role of the society for academic emergency medicine in the development of guidelines and performance measures.** *Acad Emerg Med* 2010, **17**(11):e130–e140.
28. Baker DW, Qaseem A: **Evidence-based performance measures: preventing unintended consequences of quality measurement.** *Ann Intern Med* 2011, **155**(9):638–640.
29. Beniuk K, Boyle AA, Clarkson PJ: **Emergency department crowding: prioritising quantified crowding measures using a Delphi study.** *Emerg Med J* 2012, **29**(11):868–871.
30. Hung GR, Chalut D: **A consensus-established set of important indicators of pediatric emergency department performance.** *Pediatr Emerg Care* 2008, **24**(1):9–15.
31. Guttmann A, Razzaq A, Lindsay P, Zagorski B, Anderson G: **Development of measures of the quality of emergency department care for children using a structured panel process.** *Pediatrics* 2006, **118**(1):114–123.
32. Welch SJ, Asplin BR, Stone-Griffith S, Davidson SJ, Augustine J, Schuur J: **Emergency department operational metrics, measures and definitions: results of the second performance measures and benchmarking summit.** *Ann Emerg Med* 2011, **58**(1):33–40.
33. McClelland MS, Jones K, Siegel B, Pines JM: **A field test of time-based emergency department quality measures.** *Ann Emerg Med* 2012, **59**(1):1–10. e2.
34. Ekelund U, Kurland L, Eklund F, Torkki P, Letterstål A, Lindmarker P, Castrén M: **Patient throughput times and inflow patterns in Swedish emergency departments. A basis for ANSWER, A National SWedish emergency registry.** *SJTREM* 2011, **19**(1):19–37.
35. Heyworth J: **Emergency medicine - quality indicators: the United Kingdom perspective.** *Acad Emerg Med* 2011, **18**(12):1239–1241.

36. Sibbritt D, Isbister GK, Walker R: **Emergency department performance indicators that encompass the patient journey.** *Qual Manag Health Care* 2006, **15**(1):27–38.
37. Boudreaux ED, O'Hea EL: **Patient satisfaction in the emergency department: a review of the literature and implications for practice.** *J Emerg Med* 2004, **26**(1):13–26.
38. Muntlin Å, Gunningberg L, Carlsson M: **Patients' perceptions of quality of care at an emergency department and identification of areas for quality improvement.** *J Clin Nurs* 2006, **15**(8):1045–1056.
39. Göransson KE, von Rosen A: **Patient experience of the triage encounter in a Swedish emergency department.** *Int Emerg Nurs* 2010, **18**(1):36–40.
40. Crouch R, Cooke M: **Making care count: introducing the emergency department quality indicators.** *Int Emerg Nurs* 2011, **19**(2):67–68.
41. Lin BYJ, Wan TTH, Hsu CC, Hung F, Juan C, Lin C: **Relationships of hospital-based emergency department culture to work satisfaction and intent to leave of emergency physicians and nurses.** *Health Serv Manage Res* 2012, **25**(2):68–77.
42. Benneyan JC, Lloyd RC, Plsek PE: **Statistical process control as a tool for research and healthcare improvement.** *Qual Saf Health Care* 2003, **12**(6):458–464.
43. Gordon BD, Flottemesch TJ, Asplin BR: **Accuracy of staff-initiated emergency department tracking system timestamps in identifying actual event times.** *Ann Emerg Med* 2008, **52**(5):504–511.
44. Jones P, Schimanski K: **The four hour target to reduce emergency department 'waiting time': a systematic review of clinical outcomes.** *Emerg Med Australas* 2010, **22**(5):391–398.

doi:10.1186/1757-7241-21-62
Cite this article as: Sørup *et al.*: Evaluation of emergency department performance – a systematic review on recommended performance and quality-in-care measures. *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine* 2013 **21**:62.

PAPER 2

Healthcare performance turned into decision support

Christian Michel Sørup and Peter Jacobsen
Technical University of Denmark DTU Management Engineering, Kongens Lyngby, Denmark

Abstract

Purpose – The purpose of this study is to first create an overview of relevant factors directly influencing employee absence in the healthcare sector. The overview is used to further investigate the factors identified using employee satisfaction survey scores exclusively. The result of the overall objective is a management framework that allows managers to gain insight into the current status of risk factors with high influence on employee absence levels.

Design/methodology/approach – The research consists of a quantitative literature study supported by formal and semi-formal interviews conducted at the case organisations. Employee satisfaction surveys were applied to analyse the development over time of selected factors correlated with concurrent employee absence rates. Checking for causal results, comparisons with the included published literature findings were also carried out.

Findings – Four major clustered factors, three of which constitute the term "social capital", showed a high degree of connection with employee absence rates. The factors are general satisfaction, fairness, reliance and co-operation. Integrating the four elements in a management framework will provide valuable and holistic information about the determinants with regard to current levels of employee absence. The framework will be a valuable support for leaders with the authority to alter the determinants of employee absence.

Research limitations/implications – Since a great part of the empirical material is supplied from the healthcare sector, the results obtained could be restricted to this sector. Inclusion of data from Arbejdsmarkedets Tillægspension (ATP) showed no deviation from the results in the healthcare sector.

Practical implications – The product of the study is a decision support tool for leaders to cope with levels of employee absence. The framework is holistic and can prove to be a valuable tool to take a bearing of where to focus future initiatives.

Originality/value – Gathering former observational studies in a complete overview embracing many relevant factors that influence sickness absence has not yet been attempted. Hospital management is provided with valuable information when given insight into the factors that control employee absence behaviour. Having this insight will enable the managers to promote a healthy working environment, thus lowering employee absence rates to a minimum.

Keywords Key performance indicators, Sickness absence, Performance measurement, Decision support, Health care, Performance management, Absenteeism, Health services

Paper type Research paper

Introduction

The health sector is currently under heavy political pressure. In recent years, rounds of layoffs and decreasing budget limits have prevailed in Danish public hospitals due to the economic recession that started in 2008 (Drachmann, 2010). Hence, a decreasing

Employee satisfaction survey scores and absence rates were supplied by Gentofte Hospital, Hillerød Hospital, Bispebjerg Hospital, Southern Jutland Hospital and ATP.

Journal of Health Organization and Management
Vol. 27 No. 1, 2013
pp. 64-84

1477-7266
DOI 10.1108/14777261311311807

amount of available resources will have to manage an increasing number of patients. At the same time, managers also have to do an increasing amount of administrative work alongside their daily operational tasks. These challenges haunt managers in hospitals, who therefore need proper information in order to maintain the correct treatment of patients and at the same time focus on maintaining a good working environment. Thus, it is of utmost importance that the remaining staff attend work as much as possible. How this is achieved could be through the retention of currently available human resources in the workplace. High employee attendance can be sought by directing the searchlight towards the reasons for employee absence and investigating how this topic can be conceptualised.

In an effort to conceptualise employee absence, the term "sickness absence" is elaborated. The term "sickness absence" itself is regarded by the author to be imprecise. As an example, understanding the term "sickness absence" literally does not correctly cover absence due to motivational reasons. Therefore, the term "employee absence" is introduced in this paper instead, covering absence due to both lack of motivation and being physically incapable of attending work.

Few attempts to conceptualise employee absence have been carried out to date; those that have been undertaken have focused on motivation (Nielsen, 2008; Kristensen *et al.*, 2006; Kaiser, 1998) or from an economic point-of-view (Barmby *et al.*, 2003). The most cited model based on individual motivation on a unit level is still that of Steers and Rhodes (1978).

Employee absence remains a complex phenomenon that can be caused by numerous factors (Duijts *et al.*, 2007). Asking employees in their natural working environment, having fewer employees to treat increasingly more patients will intuitively result in a poorer service level, increased waiting times and cancelled operations. Paradoxically, the direct opposite situation has proven to be the case in the short term. Shorter waiting times, lower mortality rates and sickness absence and higher patient satisfaction were obtained at Frederiksberg Hospital (Drachmann, 2011). This fact speaks of high complexity in coherence with sickness absence.

Common to most research is that a single or few factors is/are analysed in terms of sickness absence in a variety of sectors. These factors are, for example, the influence of work time (Laaksonen *et al.*, 2010), workload (Roelen *et al.*, 2007) and whether the employee feels that he or she is part of a social community (Böckerman and Ilmakunnas, 2008). The term "social capital", comprised of the elements "reliance", "fairness", and "co-operation", has been verified to impact on working conditions and employees' wellbeing and is receiving increasing attention from all kinds of organisations (Kristensen *et al.*, 2008b). Since high scores in social capital are equivalent to better working conditions and higher employee contentment, it is easy to hypothesise that social capital may also influence employee absence patterns.

Emphasis on a more holistic and integrative approach in absence management is suggested and desired (Dibben *et al.*, 2001). In order to make such an approach, the determinants of employee absence need to be recognised, but unfortunately they remain elusive (Kaiser, 1998). Mapping and correlating formerly analysed factors with direct influence on sickness absence will be an important step to understand which determinants underlie sickness absence. When having pinpointed the determinants, the stepping stone for future interventional studies in evaluating the effect of given initiatives is prepared.

For a more comprehensive assessment it is recommended to identify leading (determining future events) and lagging measures (the outcomes) (Evans, 2004). To identify the determinants of employee absence, many researchers carry out comprehensive observational studies involving broad pools of participants (Laaksonen *et al.*, 2010; Nielsen, 2008; Michie and Williams, 2003; Munch-Hansen *et al.*, 2008). In this study, the approach has been to extract information using only the scores from past employee satisfaction surveys.

The aim of this paper is to provide a clear and tangible overview of the factors directly influencing employee absence. This is done by dividing the factors as objectively as possible into two subcategories, i.e. voluntary and involuntary absence (Chadwick-Jones *et al.*, 1982). Identification of the factors leading to employee absence enables analysis of determinants through:

- calculating Pearson correlations (Johnson, 2005); and
- calculating social capital.

Pearson correlations investigate the possible coherence of singled out factors with absence rates, while social capital involves a different and more conceptual approach. Encircling plausible determinants is thus performed using the two approaches. The data material used for testing the literature findings on the case material is employee satisfaction survey scores from a selection of Danish public hospitals. Interestingly, one of the case organisations exhibits very low absence rates, while another is burdened by high absence rates. Common to both is that they cannot explain why the absence rates are as they are. This fact suggests a decisive difference in one or more measurements in determinants. A practical framework showing the degree of contentment in determinants controlling absence rate in the public healthcare sector is proposed.

Methodology

In the research community of the Operational Management Department at the Technical University of Denmark, the authors have previous experience with IDEF0, which originally was a technique applied to modelling manufacturing processes, but one that can be used for multiple purposes (Staccini *et al.*, 2004). In this study the aspect of layer representation had previously been applied to analyse the linkages between patient and employee satisfaction. Employee absence was, among many other factors, identified as an important factor connecting the two types of satisfaction. This allowed deeper analysis of the contents contained in sickness absence, thus advancing one level in IDEF0 terminology. Such an overview serves as an initial comprehension of the complexity of employee absence while systematising the literature findings in respective categories.

Analysing the causes of sickness absence

The present study is based on case research and focuses on sickness absence in detail. The study comprises two stages. In the first stage, a new, thorough and more specific literature study supplemented with structured interviews was performed. In the second stage, a quantitative comparison of selected factors and employee absence rates was made. The literature study served as a means of identifying the plausible coherence of factors both interrelated and directly influencing the level of employee absence. The fundamental literature was primarily based on all-peer reviewed journals published after the millennium. The digital library of the Technical University of Denmark was applied as a search engine. This digital library provides unified access to more than 50 million scientific articles as well as e-books and journals. The specific search words applied were "sickness absence" accompanied with more narrow terms such as, for example, "healthcare" and "social capital". Articles older than ten years were only included if they were of special importance and in agreement with recent publications.

Using an age criterion stems from the hypothesis that the healthcare environment has undergone many changes during the last few decades. Correlations older than ten years are, therefore, supposed to be of minor relevance.

Interviews were carried out in two rounds. Three interviews were conducted early in the study to identify which factors primarily were leading to employee absence. The interviews were of a semi-formal structure involving two senior doctors and two HR consultants with a duration of approximately one hour. Both interviews involved two interviewers, where one controlled the dialogue and the other took notes.

The first-round interviews served as a supplement to the literature study and sought to cover possibly neglected factors. Two second-round interviews were conducted four months later after having carried out the data analysis, undertaking a similar approach. The purpose of the second-round interviews was to seek approval of the results in the context of the case organisations.

Testing of the findings from the literature study and the interviews is based on a selection of employee satisfaction surveys obtained from ATP[1], the paediatric department at Hillerød Hospital[2] with an average of 6,500 annual discharges and the radiology department at Southern Jutland Hospital[3] (abbreviated SHS in the following) with an annual examination rate of 150,000 patients. ATP is a private pension scheme managing a series of public work-related administrative tasks. Inclusion of the ATP data was done with the desire to find possible cross-functional similarities. Furthermore, a completed project concerning psychological working environment and sickness absence (Mikkelsen *et al.*, 2008) and two recently completed projects by Merete Labriola (Labriola, 2006) and Martin Lindhardt Nielsen (Nielsen, 2010) respectively were also included. The results from these projects were used to provide further evidence on behalf of the factors investigated.

The ASUSI project was initiated in 2003 at the request of the Danish government. The ASUSI project (Mikkelsen *et al.*, 2008) is based on five years of research, shedding light on the socio-psychological causes of sickness absence. The project highlights privacy matters, such as marital status, social inheritance and personality. The ASUSI project is important because it contains data that are not included in standard employee satisfaction surveys. Both Merete Labriola's and Martin Lindhardt Nielsen's projects also concern the psychosocial working environment, but differ in their selected approach.

A final theory for inclusion is the term "social capital", here consisting of the three interacting elements of fairness, reliance and co-operation skills (Kristensen *et al.*, 2008a). Social capital is commonly used as a means to measure potentially profitable resources in an organisation. Quantification of social capital can be carried out by gathering scores from four pre-formulated questions (Kristensen *et al.*, 2008b). Connectivity between a high degree of social capital and high satisfaction levels has been obtained across sectors (Nielsen, 2010; Thomsen, 2010). Since connectivity between satisfaction levels and social capital has been confirmed, it is evident to test whether a similar connectivity between absence level and social capital prevails.

Linkages described in the literature

A new and more specific comprehensive literature study was performed, giving insight into which factors have been positively correlated with sickness absence in the healthcare sector. Common to most of the investigated literature, an observational approach was used, using surveys and interviews in order to clarify hypotheses of possible relations (Roelen *et al.*, 2007; Schreuder *et al.*, 2011). Unfortunately, these hypotheses usually give rise to new hypotheses that are left to be explored further.

Interventional research is thus highly necessary, but how the stated hypotheses are tested in practice is not suggested. The present study results in a possible approach.

The factors identified were systematically grouped into two main categories:

(1) involuntary absence; and

(2) voluntary absence (Driver and Watson, 1989)

The difference between these two categories is that involuntary absence can be described by the employee not being able to attend work, while voluntary absence can be described by the employee not being willing (but able) to attend work. Quantitative testing of the factors identified is made more obvious based on the amount of focus given to them in the literature. The sorting of the various factors was done as objectively as possible. Subjectivity cannot be avoided, since many of the factors could possibly be placed elsewhere in the overview map. The authors have tried their best to place the factors on either side of the voluntary/involuntary half. Validation of the suggested overview chart was done by a specialist doctor in work medicine at the Centre for Work Retention. Elaborating on the literature and interview findings, the overview map of employee absence is shown in Figure 1.

Clarification of clustered elements

Notice that the linkages shown in Figure 1 indicate solely leading factors to absence levels. Interrelated connections between factors are not shown, since these would only contribute to enhanced complexity without adding further value. The reason for not showing inter-connectivity is the desire to give a simplistic overview of a complex phenomenon, without pruning down the level of detail.

Involuntary absence

With primary focus on the involuntary reasons for employee absence, the factors can be clustered in a total of five different topics:

(1) working conditions;

(2) lifestyle;

(3) mental disorders;

(4) family causes; and

(5) common illnesses.

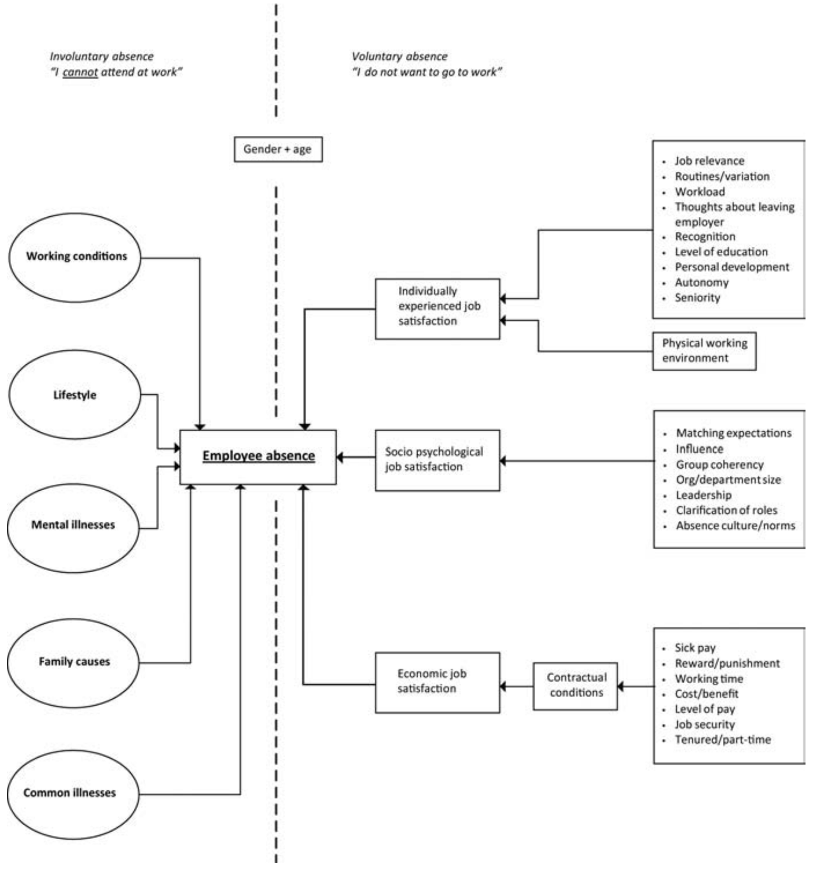


Figure 1.
Overview of factors with a direct influence on employee absence level

With regard to working conditions, Laaksonen *et al.* 2010 found consistent associations between sickness absence and the following factors:

- heavy workloads;
- hazardous exposure; and
- ergonomic working posture.

Ala-Mursala *et al.* (2002) showed that poor health caused by lifestyle, involving high alcohol consumption, smoking and obesity, had a negative impact on the level of sickness absence in a Finnish cohort study involving municipal workers. Another factor from the lifestyle topic contributing to sickness absence is excessive use of medical products (Duijts *et al.*, 2007).

Jensen *et al.* (2009) argue that mental disorders such as anxiety, stress and depression all impact on the psychosocial working environment. When undergoing such mental changes, the risk of being absent from work rises. Identification of employees with known sleep disturbances was found to be of great importance in the prevention of health problems, both mentally and physically (Vahtera *et al.*, 2006). With regard to family causes of employee absence, two main concerns are highlighted. Winkelmann (1999) argues that marital status and having a child predict one additional day of absence per year.

The last cluster of elements was not mentioned in any of the papers in the literature study, namely common illnesses such as headaches, having a cold or the flu (Mejlstrup, 2011). Even though this issue is not treated in the literature examined, common illnesses are regarded as being the main reason for being incapable of attending work. An interesting issue that arose is the individual comprehension of when one is too sick to attend work. This fact brings a related topic of "sickness presenteeism" to the attention, a topic that justifies a new study. Deciding to stay home can be considered to rely on several parallel factors, thus adding further to the difficulty of differentiating determinants in employee absence.

Voluntary and motivationally based absence

Voluntary, and thus motivationally based, absence is organised into three separate clustered topics. These are factors concerning individual, sociopsychological and economic job satisfaction (Løkke Nielsen, 2008; Kaiser, 1998; Kristensen *et al.*, 2006).

Individually experienced factors are the same as getting a personal yield for a day's work effort. Less absence from work was predicted when having a high degree of co-determination (autonomy) in a job, giving personal meaning. Variation in routines also has an impact on absence level and is tightly connected to having undergone a higher education (Notenbomer *et al.*, 2006). Balancing the workload is very dependent on personality features, both psychologically and physically (Bovier and Perneger, 2003). Absence levels will rise if the workload increases. A means to boost the threshold value for work effort is through recognition (Dieleman *et al.*, 2003). Bovier and Perneger (2003) additionally claim that greater seniority predicts lower absence levels.

Roelen *et al.* (2007) argue that being able to develop current personal skills will enhance job satisfaction, thus lowering the total frequency of employee absence periods. The last element making up the individually experienced job motivation cluster is the physical surroundings. When exposed to a poor working environment, an increase in functional inability seems inevitable regardless of ranking (Väänänen *et al.*, 2005).

The socio-psychological factors are defined as being a social phenomenon. The individual is affected by their co-workers' attitudes, moral code and ethics (Rentsch and Steel, 2003). Group coherence, the total number of employees in a department and influence on team decisions all affect employee absence (Duijts *et al.*, 2007).

The management style must be customised to fit separate teams, since guidance and control is needed in variable amounts (Böckerman and Ilmakunnas, 2008). Monitoring of teams becomes increasingly easier when expectations are matched (Labriola *et al.*, 2006). The clarification of team roles is of high importance when it comes to employee absence levels, as stated by Salminen *et al.* (2003). Last but not least, the absence culture and ethical norms must also be considered and can only be influenced by company policies or restrictions to a certain extent (Johns and Nicholson, 1985).

All of the contractual relationships point towards economic job satisfaction. Böckerman and Ilmakunnas (2008) point out that the employee will be influenced by the contractual safety (i.e. job security), level of pay and the economic benefits of going to work versus staying at home. When speaking of benefits, managers' use of either penalty or reward regulates employee attendance as well. Being tenured or working only temporarily influences the level of absence (Koopmans *et al.*, 2011).

A natural continuation of the explaination of absence due to working hours is whether people work part-time or full-time (Harrison and Martocchio, 1998). Furthermore, temporary regulations on sick pay in Germany proved to have a major impact on attendance (Kaufmann *et al.*, 2010; Puhani and Sonderhof, 2010; Ziebarth and Karlsson, 2010).

Finally, age and gender are been placed in the middle, since this factor inflicts on either side of the voluntary/involuntary cross line. Age and gender are personal characteristics that add to the heterogeneity of employee absence; as an example, personal comprehension of when "sufficient" illness has struck is highly independent.

It is important to note that none of the papers state that the given factors are of governing causes but only contribute to the level of employee absence.

Results

The derived results comprise two different approaches. First, the factors identified are correlated in turn with absence rates. Second, the social capital is calculated for selected departments.

Links identified in the literature correlated with gathered case material

The linkages identified in Figure 1 are correlated in turn with the employee satisfaction survey results, searching for similar tendencies in absence level development. The employee satisfaction surveys differ in structure, sample size and response rate. In Table I, the sample size and response rate for the departments involved is presented.

The order in which the factors are correlated is dependent on the amount of focus given in the literature studied. Pearson correlation coefficients are calculated, scouting for either highly negative or highly positive values. Causality between results is highlighted by comparing these with the literature reviewed and the results obtained in the ASUSI, Labriola and Lindhardt Nielsen projects.

Added to the table are also non-comparable though high correlation coefficients derived from a single department. In Figure 2, a schematic overview of the results derived with accompanying shaded space threshold values is presented.

Notice that the boundaries for the apprehended intervals have been defined with very sharp scales. Therefore, the reader is suggested to focus primarily on the correlation coefficients shown in Figure 2. Only the newest of published satisfaction survey scores were chosen for the analysis. The surveys chosen are only useful if the absence rates appertaining to them are stated. Subjectivity cannot be avoided when selecting

	Hospital 1			Hospital 2		
Year of survey	2006	2008	2010	2007	2009	2010
Sample size	160	180	150	98	125	72
Response rate (per cent)	64	77.8	93.8	71.2	74	58

Table I. Sample sizes and response rates for included employee satisfaction surveys

		ρ		
Category	**Driver**	**SHS**	**Hill**	**Causality**
Mental illness	No bullying		0.973	
	Fairness		-0.821	x
Working conditions	Gen. satisfaction with work. cond.	0.997		x
Economic job satisfaction	Job security	0.715		x
Sociopsychological job satis.	Cross-functional teamwork 2)	0.925		
	Cross-functional teamwork 1)	-0.990		
	Clarification of team roles	0.550	0.779	
	Management style	0.392		
Individual job satisfaction	Job relevance	0.768	0.982	
	Workload 1)	0.991	0.866	
	Development of personal skills	0.984	0.615	
	Thoughts of leaving current employ.	-0.774	-0.696	x
	Autonomy	0.946	-0.978	
	Workload 2)	0.824		
	Routines/variation	-0.567		
	General satisfaction 2)		-0.887	
	General satisfaction 1)		0.569	

ρ intervals	**Corresponding hachures**
[1;0,75]	
[0,75;0,25]	
[0,25;-0,25]	
[-0,25;-0,75]	
[-0,75;-1]	

Note: Below the table are shown the defined intervals for the shadings shown in Figure 8 and 9

Figure 2.
Comparable results between the radiology department at Southern Jutland Hospital and the paediatric department at Hillerød Hospital

indicative formulations for certain factors. For example, "fairness" is quantified by collecting the responses to the question "Are conflicts solved in a just manner?".

The most significant results obtained, using Pearson correlation and checking for causality, are fairness, job security, general satisfaction with working conditions and having thoughts about leaving one's current job (see the column headed "Causality" in Figure 2).

The correlation coefficient ρ is an expression of the similarity in development between two graphs. The interval for ρ will always be $[-1;1]$. A positive value of ρ indicates the same operational sign for the two graphs, whereas a negative value of ρ indicates opposite operational signs.

The equation used for deriving the correlation coefficients is as follows:

$$\rho = \frac{1}{n-1}\sum_{i=0}^{n}\left(\frac{x_i-\bar{x}}{s_x}\right)\left(\frac{y_i-\bar{y}}{s_y}\right),$$

where n is the total number of data sets correlated, x_i and y_i are the specific values of the observations made in each data set, x and y are the average observational values, and s indicates the standard deviation (Johnson, 2005, p. 374)).

Figure 3 shows the development between fairness and employee absence. The Pearson correlation coefficient returns −0.821, and is thus highly negative. The result indicates an opposite going direction between the two graphs, as can be seen in Figure 3.

The slope going from 2008 to 2010 becomes the dominant determinant in determining the final correlation coefficient. The x-axis corresponds to the years in which the employee satisfaction surveys were carried out. Two different y-axes are depicted. The left y-axis is the interval for fairness, while the right y-axis is for absence level.

Observing the development in job security contentment, a high positive Pearson correlation coefficient is achieved. Job security took a steep dive during the observed period, while the absence level seems to have followed the same tendency slightly (see Figure 4).

Being content with the working environment proved to be highly positively correlated with the employee absence level. This is a find similar to the results obtained in the research of Labriola (2006) and the ASUSI study (Mikkelsen *et al.*, 2008).

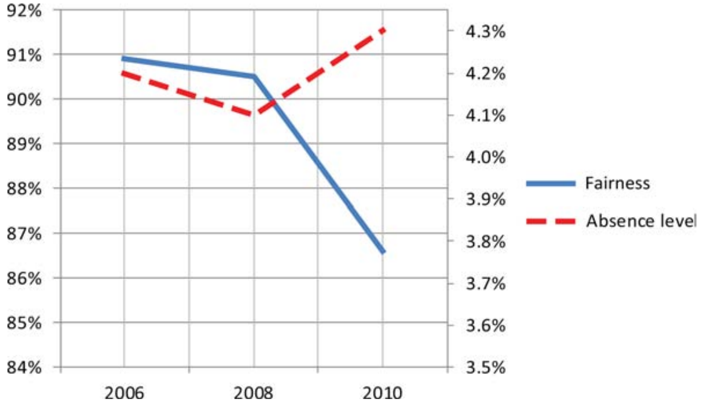


Figure 3.
Fairness versus absence level

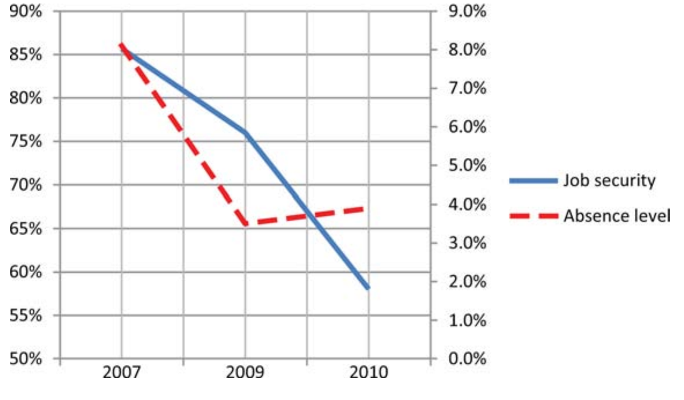


Figure 4.
Job security versus absence level

The question posed and used as underlying representation for working conditions embraced both satisfaction with ergonomic posture, exposure to hazardous elements, and also with having pleasant colleagues and good facilities. The graphical representation of the development of working conditions versus absence level demonstrates a very high positive correlation, as shown in Figure 5.

When an employee thinks of leaving the workplace it will result in a higher level of absence (Figures 6 and 7). The graphical representations are made for the two departments due to differences in the corresponding years in which employee surveys were conducted. The scores are extracted only from positively formulated questions, i.e. the percentage of employees who want to remain in their present position for the coming years.

Turning away from correlation analysis, calculating the present level of social capital can possibly be linked to the current level of absence. A total quantitative measure of social capital can be derived from the scores accumulated from four specific

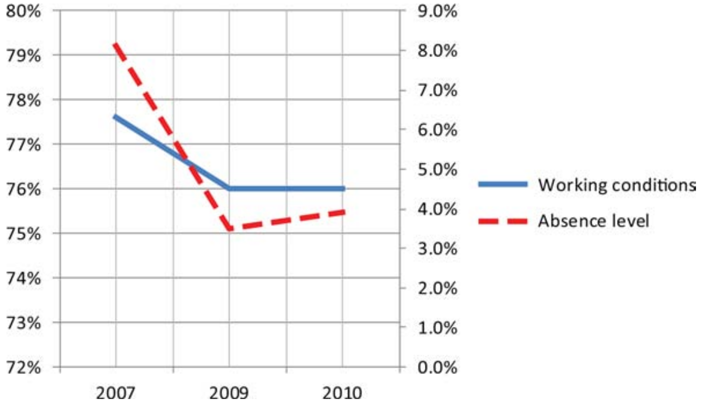


Figure 5.
Working conditions versus absence level

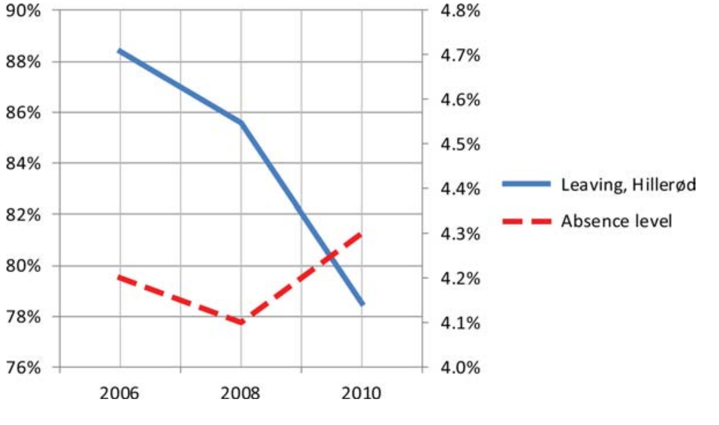


Figure 6.
Thoughts about leaving current employer, Southern Jutland Hospital versus absence level

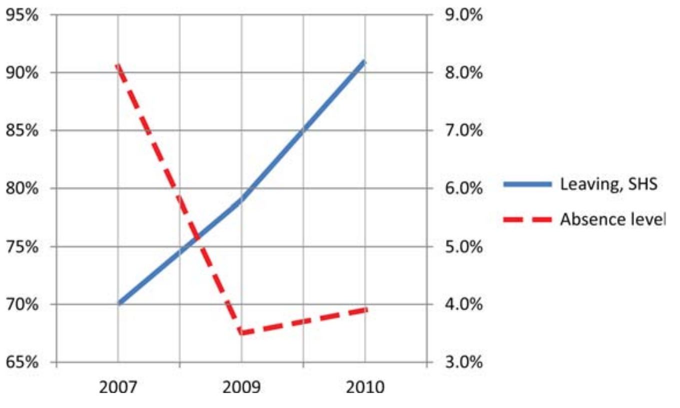


Figure 7. Thoughts about leaving current employer, Hillerød Hospital versus absence level

questions (Kristensen *et al.*, 2008a). The total score, which is indicated as social capital, is calculated by multiplying the distribution of answers (in percentages) with a factor of 0 to 4. Highly positive answers are multiplied by a factor of 4, and so forth.

For example, imagine a question containing five different answering options is distributed as:

- Very good, 12 per cent;
- Good, 28 per cent;
- Mediocre, 37 per cent;
- Poor, 20 per cent; and
- Very poor, 3 per cent.

The calculation of social capital is then carried out using the following approach:

$$(0.12*4) + (0.28*3) + (0.37*2) + (0.2*1) + (0.03*0) = 2.3\,\text{points}.$$

The scores achieved by the analysed organisations match the statement of coherency between the investigated factor and the absence level. Table II presents the calculated scores for the questions applied to quantify the social capital. The social capital levels are compared to the absence levels in Table III.

Hospital	Department	The management bears trust in the employees?	Confidence in the management's announcements?	Are the tasks distributed in a just manner?	Are conflicts solved justifiably?	Accumulated score, social capital
SHS	Radiology	2.5	2.2	2.1	1.8	8.6
Gentofte	All	2.6	2.7	2.4	2.4	10.1
Hillerød	Paediatric	2.8	2.4	2.2	2.4	9.8
Bispebjerg	Anaesthetics	2.5	2.2	2.1	1.8	8.6

Table II. Scores obtained based on separate hospital departments

The results presented in Table III suggest coherence between a high level of social capital and a low level of absence. Continuous monitoring of the highly correlated factors is recommended in order to minimise absence levels in future.

Practical application of results

Knowledge of significant and controlling factors can be transformed into practical use by creating a management framework. Accentuation and continuous monitoring of the factors will give managers an instant status update on whether or not the analysed department is in the danger zone of achieving high levels of absence. Hence, focus must be directed towards those wards where low determinant scores are present.

The framework uses carefully formulated questions in order to attain a numerical score representative for the separate factors (see Figure 8). The number of questions used to evaluate a given factor ranges from one to five.

Starting from the left hand side in Figure 8, the superior categories are given. Next, the significant clustered factors are stated. The "Note" column consists of the questions used explicitly to quantify the clustered factors. The questions can be incorporated as standard into future employee satisfaction surveys, providing a foundation for assessing the development of the department's social capital. In this way, departmental budgets would not be burdened by another item of expenditure. The scores are indicated here on a scale from 1 to 5. It is important to be cautious when stating the questions because the scores have to correspond. In order to present the scores in an intuitive and easily comprehensible manner, underlying macros transform the average scores (seen in the far right-hand column of Figure 8) into a shading code of dark, grey or light. Dark areas of interest require urgent attention. Light dotted areas of interest (such as, for example, reliance in Figure 9) need further monitoring to see whether the development turns better or worse. The striped areas of interest are well-functioning areas that give no reason for alert. If, however, a given score takes on a value near the threshold value between two intervals, the definitions do not take this into account.

A high average score will be interpreted as being satisfactory, thus attaining stripes (for example fairness in Figure 9). The questions must therefore be formulated positively. The two columns on the right hand side of Figure 9 show first the numerical score stated for the given question(s) and second the total average for the clustered factor. More questions can be added. In order to achieve a more varied picture of the present situation, a modification of the inserted questions can be made easily since the framework is provided with the required flexibility to do so.

Table III. Departments' absence rates compared to social capital scores

Hospital	Department	Absence rate (average, per cent)	Accumulated score, social capital
SHS	Radiology	3.9	11.1
Gentofte	All	4.9	10.1
Hillerød	Paediatric	4.3	9.8
Bispebjerg	Anaesthetics	5.2	8.6

22.02.2012		Project name:								
		Made by:	Clear scores	Low degree	- Some extent -			High degree		
Category	Factor	Colour	Note	1	2	3	4	5	Score	Average
Element 1, social capital	Fairness		Are conflicts solved in a just manner?					x	5	
			Are you acknowledged for a job well done?					x	5	
			Do the management always treat suggestions seriously?					x	5	
			Do work tasks get fairly distributed between employees?				x		4	4.8
Element 2, social capital	Reliance		As employee, are you able to express your opinion?			x			3	
			Do you have confidence in the way the top management manages the work place?			x			3	
			Can you trust the announcements made by the management?		x				2	
			Do the managers bear confidence in their employees?			x			3	2.8
Element 3, social capital	Cooperation		Is there a great working relations hip between your colleagues?					x	5	
			Do you experience a good cross sectional team work?				x		4	
			Do you have a good professional relations hip in your department?					x	5	
			Do you get aid and support from your colleagues when needed?			x			3	4.3
General satisfaction	Job security		Do you feel safe regarding your work-related future at the current employer?	x					1	
			Do you feel secure in your current employment?		x				2	1.5
	Working conditions		Are you content in terms of ergonomics? (lifting, working postures etc.)?		x				2	
			Are you content with the ergonomic tools used to lift heavy loads?		x				2	
			Are you content regarding the safety procedures (i.e. handling of needles, chemicals etc.)?				x		4	
			Are you satisfied with your working conditions in general?			x			3	2.8
	New job?		Do you plan on staying in your present job?	x					1	1.0

Figure 8.
Example of a checklist sheet

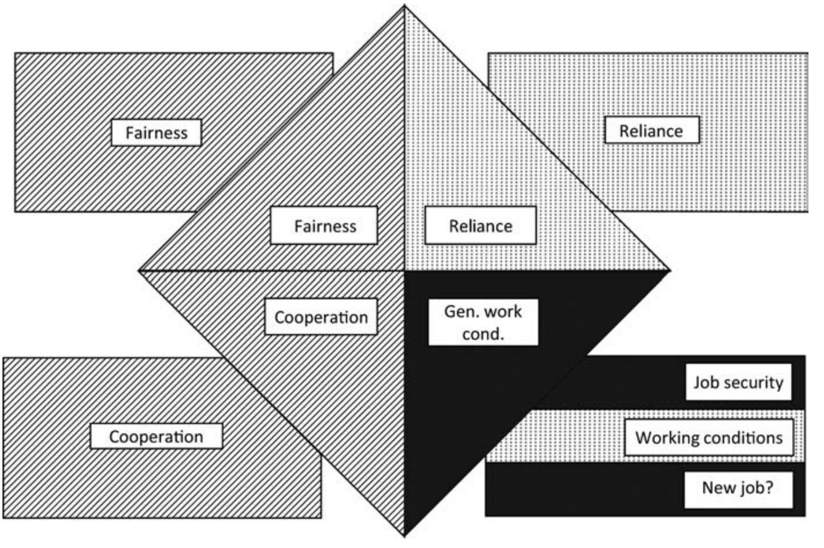


Figure 9.
Dominant factors represented by shaded spaces indicating areas of focus

The average scores are transformed automatically into a graphical representation that highlights the state of the department/organisation in terms of employee absence risks (see Figure 9).

The diamond form in the middle of Figure 9 reveals the overall clustered factors' status. The squares in the four corners are the questions making up the given clustered factor. The shaded spaces of the clustered factors shown in the diamond are determined by the average scores of the values attributed to the questions in the squares.

When the manager presents the scores to the employees, the graphical depiction requires a minimum of explanation. The framework does not provide managers with solutions as to how problems can be solved. This remains the manager's task.

Considering the broad initiatives launched in the case organisations (Resting, 2010; SHS, 2010), the management framework narrows the scope, specifically pinning down the optimal target in the quest to bring down absence rates.

Discussion

There are many potential possibilities in performance measurement using questionnaires. Through the use of questionnaires applied at two healthcare departments, this study suggests a total of four overall elements to monitor when managing the complex phenomenon of reasons for employee absence.

Concerning social capital, the scores from the four different questions are in correspondence with the previously published absence rates (Kristensen *et al.*, 2008a). Only the scores from one of the hospitals differ slightly from the overall tendency, being an opposite development between social capital and absence level. The reason for this difference can be a variation in the formulation of questions used as indicators for the scores.

At ATP the long-term focus on social capital has paid off in terms of high scores in employee satisfaction with the benefit of low absence rates (Kristensen *et al.*, 2008b). The low absence rate at ATP must be interpreted by bearing in mind that employees have the possibility of working at home. Therefore, the customer centre has been the object of analysis, since the call centre requires employees to be present.

Questionnaires are regarded as a potential method to investigate unwieldy issues such as possible reasons for employee absence (Traberg, 2010). Extraction of information through the employee satisfaction survey scores obtained can prove accurate, but there can be no guarantee that the management framework suggested in this study is applicable as a stand-alone tool. To support future findings, it is recommended that in-depth knowledge is gained about culture in a hospital department by applying a combination of social science and semiotics (Latour, 2005). The reason for recommending semiotics is due to its pragmatic approach comprehending three terms, i.e. humans, structure and technology. Although semiotics is a non-scientific approach, uniting these terms allows for in depth insight into the culture prevailing at a Danish public hospital. This knowledge can become explicit through interviews, workshops, and observations on site.

The limitations of questionnaires are described in much of the research literature, focusing amongst other things on representative sample size (Skalland, 2011), interpretation and understanding of the questions asked (Johns, 1994) and confidentiality issues in the health sector. Furthermore, the responses collected may prove biased by the mood of the respondent. Employees showing a high degree of satisfaction are anticipated to reply positively throughout a satisfaction survey. The opposite result will be the case for employees with a low degree of satisfaction.

However, the questionnaires analysed were approved in terms of their validity by external consultancy agencies, which speaks in favour of their statistical significance.

Since it is not obligatory for public hospitals to run employee satisfaction surveys, such surveys are carried out rather sporadically and only once a year. Higher quality data would allow for the use of multivariate statistical analysis, adding additional validation to the results obtained.

This fact is an obstacle in terms of data analysis and comparisons across departments. In the search to determine true findings, several hospitals were therefore contacted and asked for the results of employee satisfaction surveys that they had undertaken previously.

When doing comparisons of survey scores, it is important to interpret the results with caution because of differences in the structure of the surveys. The various consultancy companies providing the surveys present different standards and use a wide variety of formulations in questions and presentations of the scores obtained between customers. In years to come, more data will be readily available to further test this study's suggested management framework for its precision.

The results obtained from the interviews were basically gathered by applying the research strategy of case studies (Yin, 1989). One of the strengths of this particular strategy is that the researcher will gain in-depth knowledge about the subject of interest in its natural surroundings. Generalisation of the results must be done with care.

Regarding future application of the suggested framework introduced in this paper, the factors identified through multiple case studies originate from the Danish healthcare sector. The application of this framework to other sectors has not yet been tested. It may be that there are other more accurate factors dominating the level of absence in the private sector.

Moreover, one cannot neglect the variation between individuals making it difficult to pin down precise determinants in absenteeism that are valid for all. Generalisation of the results must be done with caution, since only a handful of implied parties have been included. To strive for a more solid basis for generalisation, the creation of a broader international case study that would involve a larger number of public hospitals is a way to obtain this. Having a broader span of data would allow for more in-depth statistical analysis, thus accepting or rejecting the current indicators for absenteeism.

The framework is ready for use and can be merged with already existing questionnaires. Monitoring the risk of possibly absent employees will open up the possibility of taking preventive action if necessary.

Conclusion

Comprehension and specification of the determinants that control employee absence rates was the main focus of this study. The approach started by conducting interviews and scouting literature and afterwards analysing the scores from a wide selection of employee satisfaction surveys carried out in the public healthcare sector. Having performed the empirical study, provision of a tangible overview of factors directly influencing absence rates, analogous to an IDEF0 chart, was made. The links identified were correlated with concurrent absence rates to test for possible coherence in development over time.

The final product of the study was a management framework which can provide healthcare managers with valuable and holistic information about the current status of the determinants that control employee absence rates. The management framework provides a practical and straightforward approach that transforms the examined theory into a practical solution, which represents the contribution of this study to sickness absence research. Additionally, the framework possesses a beneficial merging with satisfaction survey standards as well as an easily comprehensible and visual representation of scores. The manager is not only provided with insight into which factors generally contribute to absence rates, but can also quantify the effect of corrective actions. To support future findings from the use of the management framework, initiatives to improve employee satisfaction must be suggested on the basis of interviews and workshops with the employees concerned.

Since the suggested framework is not static, it is an iterative process to evaluate the current determinants to test their validity.

Notes

1. See www.atp.dk/X5/wps/wcm/connect/atp/atp.com
2. See www.hillerodhospital.dk/menu/Afdelinger/Boerneafdelingen/
3. See www.sygehussonderjylland.dk/wm221871

References

Ala-Mursala, L., Vahtera, J., Kivimäki, M., Kevin, M.V. and Pentti, J. (2002), "Employee control over working times: associations with subjective health and sickness absence", *Journal Epidemiology & Community Health*, Vol. 56, pp. 272-8.

Barmby, T., Ecolani, M. and Treble, J. (2003), "Sickness absence in the UK: 1994-2002", *Proceedings of the Economic Council of Sweden Conference*, pp. 1-15.

Böckerman, P. and Ilmakunnas, P. (2008), "Interaction of working conditions, job satisfaction, and sickness absence: evidence from a representative sample of employees", *Social Science and Medicine*, Vol. 67, pp. 520-8.

Bovier, P.A. and Perneger, T.V. (2003), "Predictors of work satisfaction among physicians", *European Journal of Public Health*, Vol. 13, pp. 299-305.

Chadwick-Jones, J., Nicholson, N. and Brown, C. (1982), *Social Psychology of Absenteeism*, Praeger, New York, NY.

Dibben, P., James, P. and Cunningham, I. (2001), "Developments: absence management in the public sector: an integrative model", *Public Money and Management*, Vol. 21 No. 4, pp. 55-60.

Dieleman, M., Cuong, P.V., Anh, L.V. and Martineau, T. (2003), "Identifying factors for job motivation of rural health workers in North Vietnam", *Human Resources for Health*, Vol. 1, pp. 1-10.

Drachmann, H. (2010), available at: www.politiken.dk/indland/ECE917324/hospitaler-fyrer-lige-efter-jobfest/

Drachmann, H. (2011), available at: www.e-pages.dk/politiken/6954/47

Driver, R.W. and Watson, C.J. (1989), "Construct validity of voluntary and involuntary absenteeism", *Journal of Business and Psychology*, Vol. 4 No. 1, pp. 109-18.

Duijts, S.F.A., Kant, I., Swaen, G.M., van den Brandt, P.A. and Zeegers, M.P. (2007), "A meta-analysis of observational studies identifies predictors of sickness absence", *Journal of Clinical Epidemiology*, Vol. 60, pp. 1105-15.

Evans, J.R. (2004), "An exploratory study of performance measurement systems and relationships with performance results", *Journal of Operational Management*, Vol. 22 No. 3, pp. 219-32.

Harrison, D.A. and Martocchio, J.J. (1998), "Time for absenteeism: a 20-year review of origins, offshoots and outcomes", *Journal of Management*, Vol. 24 No. 3, pp. 305-50.

Jensen, H.K., Wieclaw, J., Munch-Hansen, T., Thulstrup, A.M. and Bonde, J.P. (2009), "Does dissatisfaction with psychosocial work climate predict depressive anxiety and substance abuse disorders? A prospective study of Danish public service employees", *Epidimiology & Community Health*, Vol. 64, pp. 796-801.

Johns, G. (1994), "How often were you absent? A review of the use of self-reported absence data", *Journal of Applied Psychology*, Vol. 79 No. 4, pp. 574-91.

Johns, G. and Nicholson, N. (1985), "The absence culture and the psychological contract – who's in control of absence?", *Academy of Management Review*, Vol. 10 No. 3, pp. 397-407.

Johnson, R.A. (2005), *Miller & Freund's Probability and Statistics for Engineers*, 7th ed., Pearson Prentice Hall, Englewood Cliffs, NJ.

Kaiser, C.P. (1998), "What do we know about employee absence behaviour? An interdisciplinary interpretation", *Journal of Socio-Economics*, Vol. 27 No. 1, pp. 79-96.

Kaufmann, T., Wäschle, R., Bauer, M. and Schlüpfer, G. (2010), "Management von Kurzfehlzeiten im Krankenhaus", *Anaesthetist*, Vol. 59, pp. 433-42.

Koopmans, P.C., Bültman, U., Roelen, C.A., Hoedeman, R., van der Klink, J.J. and Groothoff, J.W. (2011), "Recurrence of sickness absence due to common mental disorders", *International Archives of Occupational and Environmental Health*, Vol. 84, pp. 193-201.

Kristensen, T.S., Hasle, P. and Pejtersen, J. (2008a), "Virksomhedens sociale kapital – en ny indfaldsvinkel til det psykiske arbejdsmiljø?", *Tidsskrift for Arbejdsliv*, Vol. 10 No. 2, pp. 30-45.

Kristensen, T.S., Hasle, P., Thoft, E. and Olesen, K.G. (2008b), *Virksomhedens sociale kapital (hvidbog)*, 1st ed., Det Nationale Forskningscenter for Arbejdsmiljø, Copenhagen.

Kristensen, K., Juhl, H.J., Eskildsen, J., Nielsen, J., Frederiksen, N. and Bisgaard, C. (2006), "Determinants of absenteeism in a large Danish bank", *International Journal of Resource Management*, Vol. 17 No. 9, pp. 1645-58.

Laaksonen, M., Pitkäniemi, J., Rahkonen, O. and Lahelma, E. (2010), "Work arrangements, physical working conditions, and psychosocial working conditions as risk factors for sickness absence: Bayesian analysis of prospective data", *Annals of Epidemiology*, Vol. 20, pp. 332-8.

Labriola, M. (2006), "Work environment factors associated with long-term sickness absence and return to work", National Institute of Occupational Health, Copenhagen.

Labriola, M., Lund, T. and Burr, H. (2006), "Prospective study of physical and psychosocial risk factors for sickness absence", *Occupational Medicine*, Vol. 56, pp. 469-74.

Latour, B. (2005), *Reassembling the Social: An Introduction to Actor-network-theory*, 1st ed., Oxford University Press, Oxford.

Mejlstrup, K.L. (2011), ATP, personal interview, 2 February.

Michie, S. and Williams, S. (2003), "Reducing work related psychological ill health and sickness absence: a systematic literature view", *Occupational & Environmental Medicine*, Vol. 60, pp. 3-9.

Mikkelsen, S. *et al.* (2008), "ASUSI Project Report", available at: http://www.asusi.dk/publikationer/asusirapport.pdf

Munch-Hansen, T., Wieclaw, J., Agerbo, E., Westergaard-Nielsen, N., Rosenkilde, M. and Bonde, J.P. (2008), "Sickness absence and workplace levels of satisfaction with psychosocial work conditions at public service workplaces", *American Journal of Industrial Medicine*, Vol. 52, pp. 153-61.

Nielsen, A.-K.L. (2008), "Determinants of absenteeism in public organizations: a unit-level analysis of work absence in a large Danish municipality", *International Journal of Human Resource Management*, Vol. 19, pp. 1330-48.

Nielsen, K.K. (2010), available at: www.etsundtarbejdsliv.dk/socialkapital/bar_sosu_projekt/bispebjerg_hospital.aspx (accessed 24 August 2011).

Notenbomer, A., Roelen, C.A.M. and Groothoff, J.W. (2006), "Job satisfaction and short-term sickness absence among Dutch workers", *Occupational Medicine*, Vol. 56, pp. 279-81.

Puhani, P.A. and Sonderhof, K. (2010), "The effects of a sick pay reform on absence and on health-related outcomes", *Journal of Health Economics*, Vol. 29, pp. 285-302.

Rentsch, J.R. and Steel, R.P. (2003), "What does unit-level absence mean? Issues for future unit-level absence research", *Human Resource Management Review*, Vol. 13, pp. 185-202.

Resting, D.N. (2010), "Nye fokusområder: flow, feedback og patientkommunikation", available at: www.bispebjerghospital.dk/NR/rdonlyres/F480A368-A183-4A54-A409-870FCFE92B88/0/Saertryk2_juni2010_5.pdf (accessed 14 May 2011).

Roelen, C.A.M., Koopmans, P.C., Notenbomer, A. and Groothoff, J.W. (2007), "Job satisfaction and sickness absence: a questionnaire survey", *Occupational Medicine*, Vol. 58, pp. 567-71.

Salminen, S., Kivimäki, M., Elovainio, M. and Vahtera, J. (2003), "Stress factors prediction injuries of hospital personnel", *American Journal of Industrial Medicine*, Vol. 44, pp. 32-6.

Schreuder, J.A.H., Roelen, C.A.M., van Zweeden, N.F. and Jongsma, D. (2011), "Leadership styles of nurse managers and registered sickness absence among their nursing staff", *Health Care Management Review*, Vol. 36 No. 1, pp. 58-66.

SHS (2010), available at: www.sygehussonderjylland.dk/dwn109912 (accessed 8 April 2011).

Skalland, B. (2011), "An alternative to the response rate for measuring a survey's realization of the target publication", *Public Opinion Quarterly*, Vol. 75 No. 1, pp. 89-98.

Staccini, P., Joubert, M., Quaranta, J.-F. and Fieschi, M. (2004), "Mapping care processes within a hospital: from theory to a web-based proposal merging enterprise modelling and ISO normative principles", *International Journal of Medical Informatics*, Vol. 74 Nos 2-4, pp. 335-44.

Steers, R.M. and Rhodes, S.R. (1978), "Major influences on employee attendance: a process model", *Journal of Applied Psychology*, Vol. 63 No. 4, pp. 391-407.

Thomsen, J.B. (2010), available at: www.e-pages.dk/hip/305/5 (accessed 29 June 2011).

Traberg, A. (2010), "Operational benchmarking of Japanese and Danish hospitals", *Proceedings of the 17th International Annual EurOMA conference, Managing Operations in Service Economies, 6 June*, pp. 1-9.

Väänänen, A., Pahkin, K., Huuhtanen, P., Kivimäki, M., Vahtera, J., Theorell, T. and Kalimo, R. (2005), "Are intrinsic motivational factors of work associated with functional incapacity similarly regardless of the country?", *Epidemiology and Community Health*, Vol. 59, pp. 858-63.

Vahtera, J., Pentti, J., Helenius, H. and Kivimäki, M. (2006), "Sleep disturbances as a predictor of long-term increase in sickness absence among employees after family death or illness", *Life Event – Sleep and Health*, Vol. 29 No. 5, pp. 673-82.

Winkelmann, R. (1999), "Wages, firm size and absenteeism", *Applied Economics Letters*, Vol. 6 No. 6, pp. 337-41.

Yin, R.K. (1989), *Case Study Research: Design and Methods*, Sage Publications, Thousand Oaks, CA.

Ziebarth, N.R. and Karlsson, M. (2010), "A natural experiment in sick pay cuts, sickness absence, and labor costs", *Journal of Public Economics*, Vol. 1, pp. 1108-22.

Further reading

Aronsson, G. and Gustafsson, K. (2005), "Sickness presenteeism: prevalence, attendance pressure factors, and an outline of a model for research", *Journal of Occupational and Environmental Medicine*, Vol. 47, pp. 958-66.

Collins, K., Jones, M.L., McDonnell, A., Read, S., Jones, R. and Cameron, A. (2000), "Do new roles contribute to job satisfaction and retention of staff in nursing and professions allied to medicine?", *Journal of Nursing Management*, Vol. 8, pp. 3-12.

Fried, Y., Melamed, S. and Ben-David, H.A. (2002), "The joint effects of noise, job complexity, and gender on employee sickness absence: an exploratory study across 21 organizations – the Cordis Study", *Journal of Occupational and Organizational Psychology*, Vol. 75, pp. 131-44.

Heponiemi, T., Elovainio, M., Mandervacka, K., Aalto, A.M., Kivimaäki, M. and Keskimäki, I. (2007), "Relationship between unemployment and health among health care professionals: health selection or health effect", *Journal of Psychosomatic Research*, Vol. 63, pp. 425-31.

Ichino, A. and Riphahn, R.T. (2005), "The effect of employment protection on worker effort: absenteeism during and after probation", *Journal of the European Economic Association*, Vol. 3 No. 1, pp. 120-43.

Kushnir, T. and Cohen, A.H. (2008), "Positive and negative work characteristics associated with burnout among primary care pediatricians", *Pediatrics International*, Vol. 50, pp. 546-51.

Moreau, M., Valente, F., Mak, R., Pelfrene, E., de Smet, P., De Backer, G. and Kornmitzer, M. (2003), "Occupational stress and incidence of sick leave in the Belgium workforce: the Belstress Study", *Journal of Epidemiology & Community Health*, Vol. 58, pp. 507-16.

Munir, F., Burr, H., Hansen, J.V., Rugulies, R. and Nielsen, K. (2011), "Do positive phychosocial work factors against 2-year incidence of long-term sickness absence among employees with and those without depressive symptoms? A prospective study", *Journal of Psychosomatic Research*, Vol. 70, pp. 3-9.

Nicol, A.-M. and Botterill, J.S. (2004), "On-call work and health: a review", *Environmental Health: A Global Access Science Source*, Vol. 3, pp. 3-15.

Nielsen, M.L. (2010), "Psychosocial work environment and sickness absence: a two-year follow-up on the IPAW study", Faculty of Health Sciences, University of Copenhagen, Copenhagen.

Nieuwenhuijsen, E.R. (2003), "Health behaviour change among office workers: an exploratory study to prevent repetitive strain injuries", *Work*, Vol. 23 No. 1, pp. 215-24.

Perry, E.L., Kulik, C.T. and Zhou, J. (1999), "A closer look at the effects of subordinate supervisor age differences", *Journal of Organizational Behaviour*, Vol. 20, pp. 341-57.

Scoppa, V. (2010), "Worker absenteeism and incentives: evidence from Italy", *Managerial and Decision Economics*, Vol. 31, pp. 503-15.

Svensson, A.L., Marott, J.L., Suadicani, P. and Ebbehoej, N.E. (2011), "Sickness absence in student nursing assistants following a preventive intervention programme", *Occupational Medicine*, Vol. 61, pp. 57-61.

Trinkoff, A.M., Le, R., Geiger-Brown, J., Lipscomb, J. and Lang, G. (Eds) (2006), "Longitudinal relationship of work hours, mandatory overtime, and on-call to muscoskeletal in nurses", *American Journal of Industrial Medicine*, Vol. 49, pp. 964-71.

Vahtera, J., Poikolainen, K., Kivimäki, M., Ala-Mursuka, L. and Pentti, J. (2002), "Alcohol intake and sickness absence: a curvilinear relation", *American Journal of Epidemiology*, Vol. 156 No. 10, pp. 969-76.

Vahtera, J., Kivimäki, M., Pentti, J., Linna, A., Virtanen, M., Virtanen, P. and Ferrie, J.E. (2004), "Organizational downsizing, sickness absence, and mortality: 10-town prospective cohort study", *British Medical Journal*, Vol. 328, p. 555.

Virtanen, M., Ferrie, J.E., Gimeno, D., Vahtera, J., Elovainio, M., Singh-Manoux, A., Marmot, M.G. and Kivimäki, M. (2009), "Long working hours and sleep disturbances: the Whitehall 2 Prospective Cohort Study", *Sleep*, Vol. 32 No. 6, pp. 737-45.

Corresponding author
Christian Michel Sørup can be contacted at: cmiso@man.dtu.dk

PAPER 3

Health Systems (2014) 3, 105-116

www.palgrave-journals.com/hs/

ORIGINAL ARTICLE

Patient safety and satisfaction drivers in emergency departments re-visited – an empirical analysis using structural equation modeling

Christian M. Sørup[1] and Peter Jacobsen[1]

[1] *Department of Management Engineering, Technical University of Denmark, Kongens Lyngby, 2800, Denmark*

Correspondence: Christian M. Sørup, Department of Management Engineering, Technical University of Denmark, Produktionstorvet 424, room 229, Kongens Lyngby, 2800, Denmark. Tel: +0045 45254768

Abstract

How can emergency department (ED) decision makers contribute to increase patient satisfaction rates? This question has been thoroughly investigated in many hospital departments but not so much in the ED, which has led to a number of untested hypotheses. Maximising value-added activities seen from a patient's perspective has become an essential outcome in health care, meaning that the untested hypotheses are in need of quantitative testing. This study proposes an integrated framework in which four latent constructs reflecting principal aspects of patient care are tested. The four constructs are entitled *safety and satisfaction*, *waiting time*, *information delivery*, and *infrastructure* accordingly. As an empirical foundation, a recently published comprehensive survey in 11 Danish EDs is analysed in depth using structural equation modeling (SEM). Consulting the proposed framework, ED decision makers are provided with information of where to launch high-impact initiatives to enhance current satisfaction levels.

Health Systems (2014) 3(2), 105–116. doi:10.1057/hs.2013.16; published online 6 December 2013

Keywords: emergency department; patient satisfaction; structural equation modeling; decision support

Introduction

Many aspects of the emergency department (ED) make it an excellent example of a complex health system. Several actors, such as doctors, nurses and secretaries, are required to collaborate under time pressure and limited space. Patient arrival patterns are difficult to forecast, since an ED can go from quiet to busy and back to quiet in a very short time. When patients enter the ED, their condition is typically unknown and is prone to possible changes over time in a non-deterministic manner. Continuous monitoring is thus an obligation in order to launch timely and appropriate action in case of unforeseen deterioration of a patient's health. Furthermore, every patient experiences a modifiable and continuous array of nursing and treatment actions throughout the ED system. All these uncertainties make the ED system subject to high degrees of variations, putting great demands on the organisation's employees, processes, logistics, technology, and human resources. When making the attempt to steer the ED towards continuous quality improvements, it is paramount to identify key areas of where to measure its obtained performance levels.

Received: 8 May 2013
Revised: 30 October 2013
Accepted: 31 October 2013

The World Health Organization (2010) has suggested three basic goals to include when assessment of a health system is to be performed. These are (1) improving health, (2) enhancing responsiveness to the expectations of the population, and (3) assuring financial equitability. All three goals are intertwined and are thus equally central to consider when establishing a quality improvement programme. In this study, the focus will be on how to address the expectations of the population (the ED patients) because the task of including all three aspects in a single study would become excessive.

Patient satisfaction has been considered a key outcome to measure in the assessment of EDs and has attracted much attention from both practitioners as well as academics especially throughout the last decade (Boudreaux & O'Hea, 2004; Boudreaux *et al*, 2004). What still remains a puzzle is which aspects to address to improve customer value. Health care organisations have sought inspiration in service industries but fall short because of the nature of the health care customer service model. A striking difference is evident when a patient exhibits high levels of satisfaction despite low-clinical care was delivered. Usually, patients arrive at the ED with a predefined set of unique expectations and requirements that must be met by the health care providers (Toma *et al*, 2009). These expectations tend to covary with level of acuity, since a patient requiring urgent care is bound to be less critical in terms of satisfaction compared with that of a patient suffering from minor injuries (provided that the urgent patient is treated adequately) (Smith *et al*, 2007). Nevertheless, there are common traits to be addressed in the attempt of promoting patient perceived experiences. Most often, the communication between provider and patient is mentioned as the primary driver for satisfaction, since a patient who is continuously monitored, approached, and assured of timely treatment will exhibit high levels of general satisfaction (Boudreaux & O'Hea, 2004; Welch, 2010).

Additional highly relevant drivers for patient satisfaction includes waiting times (Dinh *et al*, 2012), perceived safety (Fenton *et al*, 2012), and the health care providers' technical skills (Toma *et al*, 2009). These traits are justified in several studies but lack solid arguments for ranking these in accordance with value adding actions (Taylor & Benger, 2004). This study proposes an integrated framework in which four constructs reflecting key aspects of patient care are tested. ED health care decision makers are allowed insight into where resources are best utilised to improve patient satisfaction. The important assumption made in this study is that a patient's perceived safety heavily influences patient satisfaction and is thus an indirect indicator of satisfaction levels. This assumption is well established in literature. Rathert *et al* (2011) investigated patient safety perceptions as the mediating variable between service quality and patient satisfaction. Their analysis showed that in two of the three cases patient-perceived safety fully mediated the relationship between service quality and satisfaction, the third case partly mediated. If service quality is poor, rather than not meeting patient's predetermined expectations, this could influence a patient's perception more heavily compared with plain disappointment. Sorra *et al* (2012) examined the direct correlation between general inpatient satisfaction and the employees' impression of patient safety and found this relationship to be substantial. Hence, there is some evidence that suggests patient safety and satisfaction to be closely linked likely because of both being manifestations of a principal underlying hospital culture committed to patient welfare.

Hypotheses to be tested

Common to most patients presented in the ED is a tendency to overemphasise their need for urgent treatment. Often the triage system is misinterpreted by patients to give unfair advantages to those patients that arrive later than one self. Such an instance provides the source for a patient to perceive the wait as being unacceptably long and feel neglected. In a study by Bursch *et al* (1993), actual *vs* perceived waiting time was investigated and found to be the essential parameter that contributes to patient satisfaction rates. Their findings suggest matching patients' expectations, so that presumptions of timely treatment are aligned with current workloads. Perceived waiting times is also given the highest influence by Kennedy *et al* (2008), who used 'patient who leave the ED without being seen' as control variable.

H1: *Patients' satisfaction with waiting times is positively connected to perceived safety and satisfaction.*

Closely linked to a patient's satisfaction with waiting times is a continuous flow of information of when he or she can anticipate value-adding activities (Boudreaux *et al*, 2004; Taylor & Benger, 2004). If a patient experiences unexplained or unclear wait, the time will feel longer and may have a negative effect on general satisfaction and security (Maister, 1985).

A study from Barcelona conducted in 2008 proved that undertaking routine rounds of information delivery on the patient's progress in the ED decreased the number of patients who left before being seen; an outcome measure commonly interpreted as an indirect measure of patient satisfaction (Vidal *et al*, 2008). Information was delivered by oral tradition in the Barcelona study, but can be delivered through different techniques. Currently, the impact of delivering information through videotapes, closed circuit television, and pamphlets are being investigated but has so far not obtained the same positive result in patient satisfaction (Welch, 2010). Information delivery is hypothesised to effect satisfaction both on (1) perceived safety and satisfaction and (2) waiting times. Hence, two hypotheses are proposed:

H2: *Delivery of precise information on waiting times is positively connected to patients' perceived safety and satisfaction.*

H3: *Delivery of accurate information on estimated waiting times is positively connected to patients' satisfaction with waiting times.*

A new construct investigated in this study is named *Infrastructure*. This construct encompasses two aspects being (1) how easy it is for the patient to navigate to the correct treatment facility and (2) find an available parking slot. Even though the construct may appear somewhat peculiar at a first glance, both aspects have recently been proposed to be important determinants for patient safety and satisfaction rates (Kington & Short, 2010). A reason for including proper signposting and guidance is because of widespread general public confusion of the term ED. For instance, in Denmark, the former urgent care unit and ED were two independently working units, each denominated differently. Through a recent restructuring of the preliminary hospital treatment facilities, it was decided on a political level to merge these two units into one joint unit (Holm-Petersen, 2010). Media disposition becomes important for informing the public when such a new concept is launched. To date, familiarity with the new ED as a concept still has room for improvement in Denmark (Rimdal & Soerensen, 2012).

The merging of the two former units and concurrent closures of smaller hospitals increases the amount of patients to be handled. Hence, to ease the experience of getting to the ED, parking slots should be offered in an adequate amount to avoid patients having to park far from the ED. How to find ones way to the ED can become a concern if it is not intuitively clear where to go and where to park. If this becomes the case, a patient may experience despair causing a feeling of insecurity to evolve.

H4: *Satisfactory infrastructure is positively connected to patients' perceived safety and satisfaction.*

Despite researchers' intensive focus to highlight ED patient satisfaction determinants, the variability in study methodologies has been questioned and could potentially lead to misunderstandings (Welch, 2010).

This study answers the call for a more confirmatory approach in patient satisfaction research, incorporating *a priori* hypotheses and testing these on a comprehensive empirical data sample.

A framework will be proposed for ED decision makers to target future initiatives to have the best impact in enhancing current patient satisfaction levels. A graphical presentation of the four analysed hypotheses along with direction is presented in Figure 1.

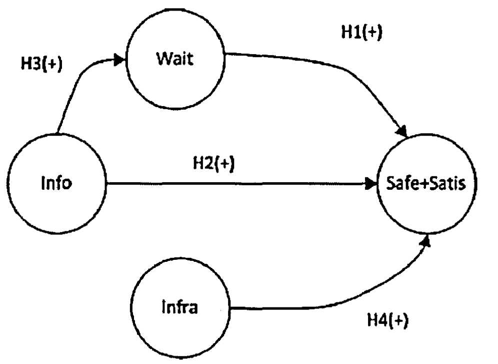


Figure 1 Theoretical model.

Methods

Sample and procedure for data gathering

The Unit of Patient-Perceived Quality (UPPQ), a decentralised unit referring to the Capital Region of Denmark, performs various patient surveys to map the general public opinion on the encounter between patients and hospitals or ambulatories on a national, regional, and departmental basis. Results from the surveys are primarily applied for quality improvement purposes, giving the hospitals an opportunity to benchmark their current performance to other hospitals. Patients are also given insight into high-performing hospitals, thus able to choose a preferred hospital accordingly. Unfortunately, patients in the Danish EDs have not been subject to such a patient survey until 2012.

The first comprehensive patient satisfaction survey in the Capital Region's EDs was conducted as a telephone survey in the period between 20 February and 4 March in 2012. Only patients who had visited one of the 11 EDs in the Capital Region of Denmark were eligible for the survey. Eligible respondents were drawn from the given ED's patient administrative database.

To identify which questions were to reflect all-round ED performance the best, multiple approaches were applied by UPPQ for triangulation purposes (Rimdal & Soerensen, 2012).

In establishing the telephone survey, some prior investigation was carried out. Initially, a literature study including both national reports and international emergency medicine literature was carried out to identify possible patient satisfaction elements. Afterwards, UPPQ established a user panel where two representatives, a doctor and a nurse, from each of the 11 EDs were summoned. Eight themes identified in the literature study that reflect typical aspects of patient experiences in an ED were subject to elaboration. The themes were: (1) guidance of directions, signposts and parking facilities, (2) reception, (3) waiting times (both actual and perceived), (4) relation and communication between patient and staff, (5) examination, treatment and care in accordance to safety and personal involvement, (6) information dispersion, (7) physical surroundings, and (8) an overall evaluation (Rimdal & Soerensen, 2012). Each theme was carefully elaborated upon and suggestions to which key questions should be posed were recorded. Subsequently, UPPQ conducted semi-structured interviews on location with ED patients. These interviews contained questions about purpose of visit and which conditions should be fulfilled for a positive experience and lasted in average 15 min each. 33 quantitative questions reflecting the eight themes reached consensus. Most questions utilised either a 3- or 5-point Likert scale with a few questions open for additional comments. One exception was the actual waiting times where each time interval was categorised on a scale from 1 to 6. In order to include actual waiting times in this study, the authors had to reverse the scale. The reversed scale can be found in Appendix A2.

Before the main data collection, the survey was piloted with a subsample of patients to ensure the content and construct validity. Results from the pilot study are available from the authors upon request. The gross empirical sample consists of 1940 ED patient responses. The distribution of responses was approximately 200 responses from each ED. Few EDs were represented by a lower number of patients as the attendance in the survey period was correspondingly low. Only patients directly discharged to their respective homes were included in the survey, which excludes ED patients referred to other hospital departments. A response analysis was conducted with gender as control variable to test whether the applied sample is representative for all ED patients. Males and females were found not to be statistically different in their representation (Rimdal & Soerensen, 2012). As the authors did not themselves conduct the telephone survey, this study draws on secondary empirical data material. Not all respondents proved eligible in this study because of possible bias. Responses made by either a patient's guardian, relative, or parents were excluded for further analysis (30%). It is deemed possible that these responses may reflect personal opinions rather than the patient's experiences solely. The total sample that has been used in this study contains a total of 685 patient responses.

Patient satisfaction data analysis

Three statistical techniques were applied to answer the research question posed in this study. According to the order used, these are first an exploratory factor analysis (EFA), then a confirmatory factor analysis (CFA), and finally structural equation modeling (SEM). The techniques will be briefly described as follows.

Exploratory factor analysis

In this study, a prior analysis of the underlying structure behind the selected variables was carried out through an EFA. This technique is particularly suitable as it emphasises the presence of potential cross-loading factors and illuminates unknown latent constructs, the latter being constructs that cannot be directly measured. In doing the EFA, an optimal number of latent constructs can be determined along with individual factor loadings. Factor loadings are numerical values that reveal direction and strength of a single item on its latent construct. High-factor loadings equal great coherence to a given construct. A prior EFA ensures that the following analysis will reveal the best possible results with the empirical data material at hand.

Attention to the theoretical foundation must remain in focus as the calculations do not take causality into account. The more detailed description of the study's EFA can be found in Appendix A1.

Confirmatory factor analysis

In the CFA, the number of latent constructs is specified along with the relations between the observed variables and latent constructs. EFA is different from CFA since the latter is a hypothesis testing procedure. The CFA model, which is commonly mentioned as a *measurement model*, is specified in advance and the empirical data used for testing the model consists of a correlation matrix containing the correlations between all included observed variables. Factor loadings and error variances, that is the parameters of the model, are estimated in order for the factor model to generate a correlation matrix, which is as close an approximation as possible to the sample correlation matrix. There are different methods of model estimation but the most frequently used is *maximum likelihood*, which has also been applied in this study. Discrepancies between the two correlation matrices are used to evaluate how well the hypothesised model fits the empirical data. During the past 10 years, which global goodness-of-fit measures to report and at what threshold level have been subject to heavy debate (Schreiber, 2008). This debate can be explained by the questionable nature of several studies' results that draws on the lack of consensus on various statistical measures in SEM (Shook *et al*, 2004). It is important to note that these measures do not necessarily tell if the model is 'true' but rather suggest if the model is statistically plausible.

This study uses threshold values recommended by Kline (2011). Those factor models that fit the data well are deemed plausible, whereas those that do not must be either re-specified or rejected.

The widely acknowledged and commonly used two-step approach by Anderson & Gerbing (1988) was followed. The two-step approach involves 'a separate estimation and respecification of the measurement model prior to the simultaneous estimation of the measurement and structural submodels' (Anderson & Gerbing, 1988, p. 417). Unlike the one-step approach where measurement and structural models are estimated simultaneously, the two-step approach has the strength of being purely confirmatory and the hypothesised model is tested independently. With the measurement model confirmed by sufficient model fit, a structural model served as means to quantify path coefficients, while testing the relationships between the constructs. Whereas all latent constructs are allowed to covary in the CFA, in SEM only the hypothesised paths are tested.

Structural equation modeling

To analyse relations among the latent constructs simultaneously results in a more precise estimation of constructs while dodging biases related to single-indicator models. An appropriate and powerful technique to do so is SEM. SEM is a composite technique, which includes three statistical techniques, namely, *path analysis*, *factor analysis*, and *multiple regression* that allows for modeling of complex 'causal' relationships between observed and latent variables. Compared with other statistical procedures, a more detailed set of assumptions about causal relations can be accommodated by SEM.

It is important to note that SEM is an *a priori* technique, implying that the technique is not suitable for exploratory purposes.

In this study, SEM has been applied using MPlus version 6.12 (Muthen & Muthen, 2007) to test the hypothesised relationships between the included latent constructs: (1) waiting times, (2) information delivery, (3) safety, and (4) infrastructure. In practice, the measurement model (CFA), in which all latent constructs were allowed to covary, is modified to only include the hypothesised links. SEM has been chosen as the preferred analytical tool because of its usefulness in testing the developed theoretical model as it allows for concurrent estimation of many relationships between observed/manifest and latent variables. Moreover, SEM takes measurement error into account. One nested structural model (competing model) was assessed to determine which model represented the best fit, thus reflecting the underlying data set the most while still being grounded in theory.

Results

Survey responses

In total, 1.940 patients reported their experiences in the telephone survey. Of these, 53% were males and 47% were females. The age distribution in the telephone survey is presented in Table 1.

A sample of patients from all EDs in the capital region of Denmark where drawn and the distribution of patients is presented in Table 2.

Extraction of the 685 patient responses was done based on a listwise deletion of the eligible 1358 patient responses, that is, only patients who had answered all 14 questionnaire items were usable. No response rate has been calculated for the telephone survey because of the difficulties connected to this technique. The difficulties are owing to the patients who could not be dialled because of not having registered a telephone number or patients who did not answer the telephone call. Therefore, more patients could potentially have been included in the telephone survey. The upper limit of 200 patients per ED was set because of time limitations.

Exploratory factor analysis

A final set of mutually uncorrelated factors was extracted through the EFA using varimax rotation. The rotated component matrix is presented in Table 3. As a requirement,

Table 1 Age distribution among respondents

Age	0–4	5–9	10–19	20–39	40–59	60–79	80+
Percentage	9	8	20	19	21	17	6

Table 2 Number of telephone interviews sorted by ED

Unit	*Amount of patients surveyed*
Amager hospital	198
Bispebjerg hospital	158
Bornholm hospital	67
Frederiksberg hospital	184
Frederikssund hospital	149
Gentofte hospital	206
Glostrup hospital	208
Elsinore hospital	137
Herlev hospital	204
Hospital of Northern Zealand	208
Hvidovre hospital	194
Amount in total	1.940

factor loadings must be >0.7 to exhibit sufficient affiliation to a given construct (Farrell, 2010).

The two items denoted 'Facilities' were removed from further analysis because of insufficient factor loadings. By applying the EFA, it can be concluded that the involved items defines four constructs to satisfactory. Subsequently, the four constructs are linked to existing literature.

On the basis of prior research suggesting important ED patient satisfaction drivers, Welch (2010) promoted five key constructs and several constructs of minor importance in influencing patient satisfaction rates. In this study, patient satisfaction and safety is examined based on four latent constructs developed through the prior EFA while rooted in existing literature.

Following Welch (2010), a patient's expectation to be treated with minimum waiting time has received a lot of focus in the past decade. *Waiting times* are operationalised in terms of both actual waiting times and the patients' perceived waiting times. The previous is measured for (1) the entire stay in the ED and (2) the wait from entrance to initial encounter with a health care professional.

Adopted from Boudreaux *et al* (2004), *information delivery* is measured since well-informed patients are likely to feel more secure and thus exhibit greater degrees of satisfaction (Krishel & Baraff, 1993; Thompson *et al*, 1996). Unexplained and uncertain waits feel protracted and has a negative impact on the patient's perceived wait, which suggest a hypothesised causal link between information delivery and perceived waiting times. *Safety and Satisfaction* is another included construct that encompasses primarily perceived safety but also contains a single element concerning an all-round evaluation of the entire ED stay.

The safety and satisfaction construct draws upon different aspects of how secure the patient feels about the examination and treatment given, how secure the patient feels in returning home, what symptoms to be aware of after discharge and who to contact in case of exacerbation when returned home. One additional control and outcome measure is included in the safety construct, namely, the total impression of the ED visit. As safety can be regarded as a more critical measure, it is found reasonable

Table 3 EFA factor loadings ($n=685$)

		Construct			
		1	2	3	4
sp7	Intuitive signposting				0.883
sp8	Parking possibilities				0.867
sp14	Wait times, perceived (arrival to treatment)		0.871		
sp15	Information, reason for wait times			0.887	
sp16	Information, progress in wait times			0.894	
sp17	Wait times, perceived (total stay)		0.767		
sp20	Safety, correct treatment	0.739			
sp25	Safety, symptom awareness	0.803			
sp26	Safety, contact person	0.818			
sp27	Safety, discharge	0.799			
sp28	Facilities, overall assessment	0.315		0.331	
sp29	Facilities, cleanliness	0.519			
sp31	Total impression of ED visit	0.705	0.373		
sp13a	Wait times, actual		0.850		

Table 4 Descriptive statistics

Construct	*Cronbach's α*	*Items*
Safety and satisfaction (5-point Likert scale)	0.825	Do you feel very secure, secure, insecure, or very insecure that you have received the correct examination and treatment in the ED? How do you evaluate the information about which symptoms you should be aware of after you returned home? How do you evaluate the information about who to contact in case you experienced any symptoms after you returned home? Did you feel very secure, secure, insecure, or very insecure in returning home from the ED? What is your total impression of your visit in the ED?
Waiting times Rescaled (1–6) Reverse scaled (3-point Likert scale), (5-point Likert scale)	0.811	Waiting times distributed in triage levels How will you describe the duration of wait time from your reception till your initial examination? How do you evaluate the length of the total wait time during your entire visit at the ED?
Infrastructure (newly developed) (5-point Likert scale)	0.687	How do you evalute the posting of signs to the ED? How do you evaluate the parking possibilites when you arrived at the ED?
Information delivery (5-point Likert scale)	0.677	Where you informed of why there was wait time from your reception to your initial examination? Where you continuously informed of the development in wait time from your reception to your initial examination – for instance, by the personnel or shown on an information board/screen?

by the authors to place an overall evaluation measure in this construct.

Finally, the last measure denoted that *infrastructure* is a newly developed construct and refers to (1) the provision of precise and understandable guidance to the ED and (2) adequate number of patient parking slots near the ED. This issue should receive special attention if new technical terms for the ED are applied or the layout is changed.

Table 4 summarises all applied questionnaire items accompanied with Cronbach's α as estimate for each construct's internal consistency, which means how closely a set of item relates as a group. In this study, a Cronbach's α cut-off value of 0.6 was deemed sufficient as advocated by (Hair *et al*, 2010, p. 125). Specification of scales is stated for each measure.

Confirmatory factor analysis

The CFA was used to evaluate the properties of the scaled questionnaire items for each construct extracted from the obtained patient satisfaction survey. Means, standard deviations, and correlations for the observed variables are presented in Table 5.

All four constructs displayed acceptable levels of reliability as indicated by (1) composite reliability ranging from 0.68 to 0.92; and (2) average variance extracted (AVE)

Table 5 Correlation analysis including means and standard deviations

Numbers	*Items*	*Mean*	*Standard deviation*	1	2	3	4	5	6	7	8	9	10	11	12
1	Guidance assessment	0.26	0.20	1.00											
2	Parking possibilites assessment	0.26	0.19	0.52	1.00										
3	Waiting times, actual	0.43	0.21	0.08	0.08	1.00									
4	Waiting times, perceived initial	0.33	0.21	0.12	0.09	0.76	1.00								
5	Waiting times, perceived total	0.23	0.13	0.02	0.05	−0.07	−0.01	1.00							
6	Information, waiting times reason	0.24	0.12	0.08	0.05	0.01	0.05	0.51	1.00						
7	Information, waiting times development	0.06	0.12	0.14	0.12	0.45	0.57	0.02	0.11	1.00					
8	Safety, correct treatment	0.16	0.17	0.05	0.11	0.23	0.29	0.04	0.07	0.31	1.00				
9	Safety, symptom awareness	0.23	0.17	0.13	0.15	0.16	0.22	0.04	0.11	0.19	0.43	1.00			
10	Safety, contact person	0.22	0.16	0.14	0.16	0.13	0.18	0.05	0.08	0.14	0.37	0.72	1.00		
11	Safety, discharge	0.16	0.17	0.09	0.09	0.15	0.23	0.02	0.05	0.23	0.57	0.48	0.45	1.00	
12	Satisfaction, total impression of ED visit	0.20	0.17	0.16	0.14	0.33	0.43	0.07	0.14	0.43	0.54	0.44	0.40	0.48	1.00

ranging from 0.78 to 0.93 (Fornell & Larcker, 1981). The AVE number indicates the average amount of variation that a given latent construct can explain in the associated observed variables and should equal more than 0.50 (Hair *et al*, 2010). As an example, latent construct A correlates with observed variables x_1 and x_2. The square of the correlation gives the amount of variation in the separate observed variables that the latent variable accounts for (called the shared variance). Averaging the variance across the observed variables related to a latent construct ultimately gives the AVE (Farrell, 2010). The baseline measurement model is presented in Figure 2.

As follows, recommended fit indices were extracted to assess how well the *a priori* model fitted the sample data.

The measurement model's χ^2 test of model fit was found to be significant ($\chi^2=152.126$, df=48, $P<0.001$). Great model fit is recommended to provide an insignificant result above a P-value of 0.05 and is often mentioned as measuring 'badness-of-fit' (Barrett, 2007). However, the χ^2 measure, in its nature as a statistical significance test, is sensible to sample size, which nearly always results in a rejection of the model when larger samples are applied (Jöreskog & Sörbom (1993); Kline, 2011). Again, the data sample size was $n=685$ for this study. Complete reliance on the χ^2 measure is unwarranted and criticised in literature, drawing attention to other comparative fit indices. These comparative fit indices measure incremental improvements of the model fit by comparison of the hypothesised model and a restricted baseline model. Root Mean Square Error of Approximation (RMSEA) is one measure that takes the error of approximation in the population into account, while being sensitive to the number of estimated parameters in the model (Muthen & Muthen, 2007). RMSEA favours parsimony as it will prefer the model with minimum estimated parameters.

Cut-off values have lowered steadily during the past decade and now, a value below 0.08 is considered a good fit. Further, the Comparative Fit Index (CFI) considers sample size and values above 0.90 indicates good model fit (Hu & Bentler, 1999). If the survey instrument contains items with different scales, the Standardised Root Mean Square Residual (SRMR) can be calculated as an estimate of the difference between the sampled and predicted correlation matrix. An SRMR value of 0 indicates perfect fit but values below 0.08 are deemed a good fit (Hu & Bentler, 1999). The measurement models goodness-of-fit indices were satisfactory. RMSEA equalled 0.056 with a 90% confidence interval of 0.046 (low) to 0.066 (high). CFI was 0.963 and SRMR was 0.049.

Structural equation modeling

Evaluation of the causal links between the latent constructs is examined in the structural model. The baseline theoretical model (Model 1) that fitted the data sufficiently was tested (RMSEA=0.058, CFI=0.960, SRMR=0.058, $\chi^2=161.840$, df=49, $P<0.001$). Afterwards, the baseline model was compared with another nested structural model grounded in theory. Two models can be said to be nested if one is a subset of the other (Kline, 2011, p. 214). The two models (baseline and competing) are depicted in Figure 3.

The fully mediated model (Model 2) was compared with the baseline model by removing the direct path from information delivery to safety and satisfaction. Such a change suggests that satisfaction with infrastructure and safety and satisfaction is mediated by both information delivery and waiting times and as such is a potentially more parsimonious solution. Unfortunately, eliminating the path between information delivery and perceived safety seems to have damaged the model slightly because of an increase in RMSEA (ΔRMSEA=+0.01), a decrease in CFI (ΔCFI=−0.02), an increase in SRMR (ΔSRMR=+0.02), and an insignificant increase in χ^2 ($\Delta\chi^2=+5.96$, $P<0.02$).

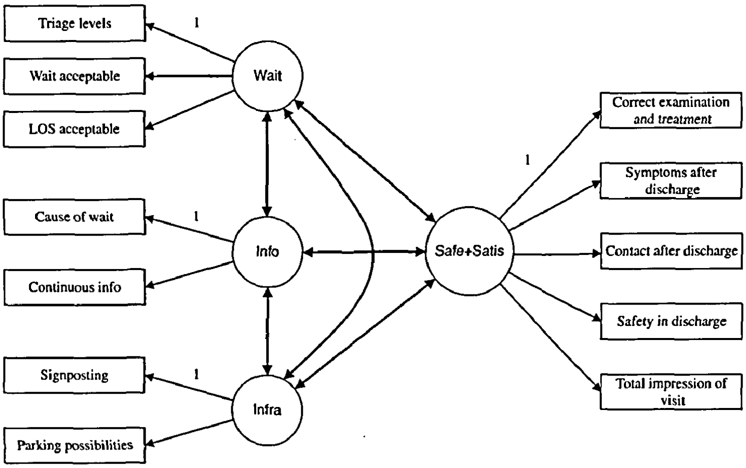


Figure 2 Measurement model (CFA). Error terms omitted. LOS = Length of Stay.

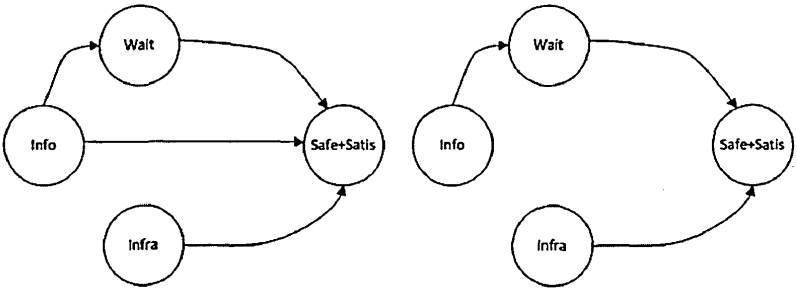


Figure 3 Baseline and competing SEM model.

Reviewing the fit indices, Model 2 obtains reduced values that indicate minor misspecification compared with the baseline model, even though the χ^2 test is only near significant (at significance level $\alpha = 0.01$). The nested models are presented in Table 6.

The latent constructs infrastructure and information delivery comprise less than the recommended three observed variables. This can result in an unidentified model which is not possible to solve. Despite having two items define infrastructure and information delivery accordingly, model identification did not prove to be a problem since no error messages was stated when the model was run (error messages notify the user of identification issues). Mplus can conduct a non-identification check through the use of the singularity check of the sums of squares and cross-products of first-order derivatives. Such a check will notify the user in case of non-identification issues.

Requirements for the sample size depend on a row of factors. Instances of such factors are amounts of missing data, distribution of variables, reliability of variables, and size of the hypothesised model (Muthen & Muthen, 2007). This study contains a sufficiently large sample given the relatively simple model making the authors confident of the robustness of the results. In addition, the data sample was prepared before Anderson and Gerbing's suggested two-stage approach with no missing data, high-reliability coefficients, and near normal distributed items.

Transformation of data was necessary because of a high all-round level of patient satisfaction reported in the empirical material (Rimdal & Soerensen, 2012), making the response distribution skewed to the right. All items were subject to a natural logarithmic data transformation to enhance a near normal distribution character.

Hypotheses tested

By application of Model 1, that is, the best of the tested structural models, it was possible to test and acquire estimates of all stated hypotheses. Both standardised and non-standardised path coefficients are reported in Table 7.

Beginning with the relationship between safety and satisfaction and waiting time (H1), a positive and significant

Table 6 Fit indices for structural models

Number	Model	CFI	ΔCFI	RMSEA	ΔRMSEA	χ^2	df	$\Delta\chi^2$	Δdf	P
0	Measurement model	0.963		0.056		152.126	48			
1	Base line model	0.960		0.058		161.840	49			
	Model 1–0 difference		−0.003		0.002			9.71	1	$P<0.01$
2	Delete Safe+Satis ON Info	0.958		0.059		167.800	50			
	Model 2–1 difference		−0.002		0.001			5.96	1	$P<0.05$
	Model 2–0 difference		−0.005		0.003			15.67	2	$P<0.01$

Table 7 SEM results of tested hypotheses

Hypotheses	Description of path	Hypothesised direction	Standardised path coefficients	Non-standardised path coefficients
H1	Safe+Satis – Wait	+	0.425**	0.338**
H2	Safe+Satis – Info	+	0.097*	0.192*
H3	Wait – Info	+	0.051	0.127
H4	Safe+Satis – Infra	+	0.161**	0.156**

Note: $N=685$, $P<0.05$*, $P<0.01$**.

path coefficient was confirmed ($\gamma=+0.338$, $P<0.01$). H2, which anticipated safety and satisfaction to be positively connected with information delivery, was found to be supported to a significant degree ($\gamma=+0.192$, $P<0.05$). Interestingly, the hypothesised relationship between waiting times and information delivery (H3) did not prove statistically significant despite correct direction ($\gamma=+0.127$, $P<0.1$). Finally, safety and satisfaction connected to infrastructure (H4) did receive support ($\gamma=+0.156$, $P<0.01$). All regression coefficients extracted from the best fitted structural equation model can be seen in Figure 4.

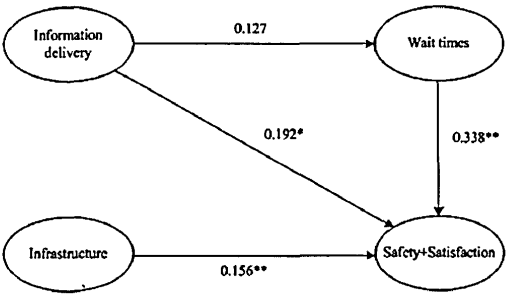


Figure 4 SEM results of tested hypotheses, non-standardised path coefficients only
Note: $N=685$, $P<0.05$*, $P<0.01$**.

Discussion

ED patient satisfaction can be viewed as a key component of a larger more complex ED performance evaluation system built from performance metrics of special interest to several stakeholders. This study adds to existing knowledge by introducing a more advanced statistical method for testing hypotheses proposed in recent peer-reviewed scientific literature. In this study, some of the key underlying mechanisms of what drives ED patient satisfaction has been illuminated and tested. This new knowledge can prove beneficial for multiple ED stakeholders as where to put effort in order to enhance patient satisfaction levels.

Proven in this study, minimising waiting times will increase safety and satisfaction levels the most; a result consistent with much other literature findings. For instance, Pines *et al* (2008) connect patient's length of stay (LOS) to the likelihood of recommending the ED to others and perceived overall hospital-care assessment. The key issue is to limit the time interval from the time where the patient arrives to the time for initial assessment, while the patient is given the feeling that he or she is continuously being cared for. Especially being cared for in a timely manner is highlighted by multiple authors as being the prevailing determinant in patient satisfaction (Welch, 2010). A recent study by Dinh *et al* (2012) points out the coherence between actual and perceived waiting times. Indeed, acceptable waiting times is a subjective matter, since one patient may find a certain wait too long, whereas another patient, suffering from similar symptoms and within the same triage level, would find the wait perfectly acceptable.

In continuation of waiting times, dispersion of continuous information acting as mediator was confirmed to be the second most important measure in connection to perceived safety and satisfaction. Past studies have shown that improved information does lead to enhanced satisfaction levels and can be obtained by fairly easy changes, for instance, handing out a business card upon first encounter with the patient (Guss *et al*, 2013). However, not all interventions have positive outcomes making information delivery techniques a prospective research area for the future.

The obtained SEM results were presented to the UPPQ at Frederiksberg Hospital in February 2013. Of the 20 tenured employees in the unit, 15 participated in a 1 hour workshop, reflecting on the SEM outcomes and pinpointing potential future research areas. The SEM results were confirmed to be aligned with their experience based on their former studies in other hospital departments, thus adding to external validity (Karlsson, 2009).

Patient satisfaction data has recently been used in similar SEM studies. SEM has been used to model and test hypotheses, for instance, concerning the health care provider's interpersonal and technical skills (Berg *et al*, 2012), relationships between different decision-making-related constructs (Hölzel *et al*, 2013), and various background variables connected to all-round patient satisfaction (Sahin *et al*, 2007). As such, the studies mentioned each investigate an isolated part in order to better understand the complexity behind patient satisfaction. There is a potential to combine these studies in a larger model to conduct a simultaneous estimation of more relevant hypotheses. If such a model is pursued, a better understanding of what the patients want during their ED stay will be obtained.

As an alternative to SEM, partial least squares regression (PLS) could be applied in cases where the theory is not well developed and where observed variables are likely not to support a sufficiently specified measurement model (McDonald, 1996). PLS applies limited information methods to provide some statistical estimates, which implies minimal requirement of the data. This comes in contrast to SEM, where overall optimisation in parameter estimates is sought through a complete information estimation technique (for instance, maximum likelihood). Examples of PLS use in health care includes measuring out-patient satisfaction (Wicks & Chin, 2008) and on best practice in implementing a Balanced Scorecard (Lovaglio, 2011). In this study, PLS would not be deemed sufficient to fully exploit the potential of the empirical data set.

This study provides valuable information on which patient satisfaction elements to monitor. As the hypotheses could not be disregarded, the constructs examined could be implemented in already existing quality-monitoring systems. In this way, potential initiatives could more easily be assessed if they have the wanted impact on patient satisfaction rates and changes in procedures and policies could be made accordingly.

Limitations

Some practical constraints do limit the interpretation of the results obtained. On the other hand, these limitations also make fellow researchers aware of potential pitfalls. First, the obtained data set does not allow for preclusion of alternative causal explanations. Reliability would be enhanced if the study was replicated by the use of a longitudinal study design. Second, all included latent variables have been extracted and documented in published peer-reviewed literature. However, this does not rule out the existence of other equivalently important constructs, for instance, staff communication skills and technical competencies (Berg *et al*, 2012). Much emphasised, and deemed to be the strongest predictors of ED patient satisfaction, is the communication between patients and health care professionals (Boudreaux *et al*, 2004). Interpersonal interaction with ED clinicians or nurses can be improved by changes in attitude to promote expressive quality, for instance, by apologising for prolonged waiting times, have good eye contact, appear interested, and concerned when the patients describe their problems, and explain test results and plausible implications.

One potential shortcoming of this study is the fact that the telephone survey, which provided the empirical foundation for the analysis, was not designed for a subsequent SEM approach in mind. For future research with a wish to apply SEM, it would be imperative to formulate survey questions in constructs of three or preferably more questions (observed variables) each (Kline, 2011, p, 126). This is in compliance with Anderson *et al* (1998) who recommend aggregating patient satisfaction scores based on multiple measures rather than on single measures. With sufficient observed variables compared with model complexity, the model will become over identified, meaning that it will be theoretically possible for the simulation software to derive a set of parameter estimates based on the input data.

Although, the telephone survey was not intended for such an analysis, Likert scales were primarily utilised making it possible to apply SEM with reason. To advance current knowledge, future studies could extent the presented model with more latent constructs including control measures, such as 'likelihood to recommend' and 'overall satisfaction'. By inclusion of such measures, a separate construct denoted 'satisfaction' could rightfully be established thus disentangling safety from satisfaction in future studies. Normal distributed input data is usually a general requirement for SEM but is most often not the case in practice. Therefore, a data transformation was necessary to modify the shape of the data into an adequate distribution. Given the skewed distribution of satisfaction survey responses towards a positive opinion, an altered Likert scale should be applied to make the answers more normal distributed. For a 5-step Likert scale, this could be defined as 'poor, fair, good, very good, excellent' in order to give a more nuanced view on satisfaction (Trout *et al*, 2000).

The authors acknowledge that other aspects on patient satisfaction survey design improvements, such as respondent biases, survey timing, standardisation issues, cross-survey comparisons, and language barriers are evident. However, these instances are addressed in much other literature. The analysed model was chosen for parsimonious reasons to reduce model complexity.

Conclusion

This study's key contribution is an empirical investigation of four stated hypotheses in contemporary patient satisfaction literature concerning what patients prefer the most when in need of services from an ED system. In a recent comprehensive sample of patient satisfaction scores stemming from 11 Danish EDs, three of four hypotheses were confirmed with statistical significance by the use of SEM. Even though the path coefficients obtained are of minor magnitude, most emphasised is the connection between perceived safety and satisfaction and waiting times. If this link in particular is addressed in practice, ED patient satisfaction feedback is likely to improve the most. ED decision makers are endowed with insight to launch initiatives with potentially higher impact, serving to refine an important cog in a highly complex health system.

Future research includes extending the presented SEM model with more latent constructs that potentially could shed light on other valued patient satisfaction drivers. In addition, an interventional longitudinal study design is deemed appropriate to see if specific initiatives targeted to this study's presented results have the wanted effect on ED patient satisfaction levels.

Acknowledgements

The authors would like to thank the UPPQ for allowing insight into the individual patient responses gathered from the telephone survey and for participation in a workshop. Moreover, we thank Professor Bo. B. Nielsen of Copenhagen Business School, Denmark for his guidance in establishing the CFA.

References

Anderson FD, Maloney JP and Beard LW (1998) A descriptive, correlational study of patient satisfaction, provider satisfaction, and provider workload at an army medical center. *Military Medicine* 163(2), 90–94.

Anderson JC and Gerbing DW (1988) Structural equation modeling in practice: a review and recommended two-step approach. *Psychological Bulletin* 103(3), 411–423.

Barrett P (2007) Structural equation modeling: adjudging model fit. *Personality and Individual Differences* 42(5), 815–824.

Berg GM, Crowe RE, Nyberg S and Burdsal C (2012) Trauma patient satisfaction with physician assistants: testing a structural equation model. *Journal of the American Academy of Physician Assistants* 25(5), 42–51.

Boudreaux ED, D'Autremont S, Wood K and Jones GN (2004) Predictors of emergency department patient satisfaction: stability over 17 months. *Academic Emergency Medicine* 11(1), 51–58.

Boudreaux ED and O'Hea EL (2004) Patient satisfaction in the emergency department: a review of the literature and implications for practice. *The Journal of Emergency Medicine* 26(1), 13–26.

Bursch B, Beezy J and Shaw R (1993) Emergency department satisfaction: what matters most? *Annals of Emergency Medicine* 22(3), 586–591.

Dinh MM, Enright NM, Walker A, Parameswaran A and Chu M (2012) Determinants of patient satisfaction in an Australian emergency department fast-track setting. *Emergency Medicine Journal* 30(10), 824–827.

Farrell AM (2010) Insufficient discriminant validity: a comment on Bove, Pervan, Beatty, and Shiu (2009). *Journal of Business Research* 63(3), 324–327.

Fenton JJ, Jerant AF, Bertakis KD and Franks P (2012) The cost of satisfaction: a national study of patient satisfaction, health care utilization, expenditures, and mortality. *Archives of Internal Medicine* 172(5), 405–411.

Fornell C and Larcker DF (1981) Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research* 18(1), 39–50.

Guss DA, Leland H and Castillo EM (2013) The impact of post-discharge patient call back on patient satisfaction in two academic emergency departments. *The Journal of Emergency Medicine* 44(1), 236–241.

Hair Jr. JF, Black WC, Babin BJ, Anderson RE and Tatham RL (2010) *Multivariate Data Analysis*, 7th edn, Pearson Higher Education, Upper Saddle River, NJ.

Holm-Petersen C (2010) Fremtidens arbejdstilrettelæggelse for speciallæger på fælles akut modtagelser, http://projekt.dsi.dk/uploads/upload_4c908f0440a57.pdf, accessed 3 June 2013.

Hu LT and Bentler PM (1999) Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria *vs* new alternatives. *Structural Equation Modeling* 6(1), 1–55.

Hölzel LP, Kriston L and Härter M (2013) Patient preference for involvement, experienced involvement, decisional conflict, and satisfaction with physician: a structural equation model test. *BMC Health Services Research* 13(231), 1–10.

Jöreskog K and Sörbom D (1993) *LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language*. Scientific Software International, Chicago, IL.

Kaiser HF (1958) The varimax criterion for analytic rotation in factor analysis. *Psychometrika* 23(3), 187–200.

Karlsson C (2009) *Researching Operations Management*. Routledge, New York.

Kennedy M, MacBean CE, Brand C, Sundararajan V and Taylor D MCD (2008) Review article: leaving the emergency department without being seen. *Emergency Medicine Australasia* 20(4), 306–313.

Kington M and Short AE (2010) What do consumers want to know in the emergency department? *International Journal of Nursing Practice* 16(4), 406–411.

Kline RB (2011) *Principles and Practice of Structural Equation Modeling*. Guilford Press, New York.

Krishel S and Baraff LJ (1993) Effect of emergency department information on patient satisfaction. *Annals of Emergency Medicine* 22(3), 568–572.

Lovaglio PG (2011) Model building and estimation strategies for implementing the balanced scorecard in health sector. *Quality & Quantity* 45(1), 199–212.

Maister D (1985) *The psychology of waiting lines*, http://davidmaister.com/wp-content/themes/davidmaister/pdf/PsycholgyofWaitingLines751.pdf, accessed 5 April 2013.

McDonald RP (1996) Path analysis with composite variables. *Multivariate Behavioral Research* 31(2), 239–270.

Miller I, Freund J and Johnson R (2005) *Miller and Freund's Probability and Statistics for Engineers*. Pearson Higher Education, New Jersey.

Muthen LK and Muthen BO (2007) *Mplus User's Guide*, 5th edn, Los Angeles, CA.

Nunnally JC and Bernstein IH (1994) *Psychometric Theory*. McGraw-Hill, Michigan.

Pines JM, Iyer S, Disbot M, Hollander JE, Shofer FS and Datner EM (2008) The effect of emergency department crowding on patient satisfaction for admitted patients. *Academic Emergency Medicine* 15(9), 825–831.

Rathert C, May DR and Williams ES (2011) Beyond service quality: the mediating role of patient safety perceptions in the patient experience–satisfaction relationship. *Health Care Management Review* 36(4), 359–368.

Rimdal TO and Soerensen PV (2012) Patienters oplevelser i Region Hovedstadens akutmodtagelser og akutklinikker, http://www.patientoplevelser.dk/log/medie/Rapporter/2012/akut2012.pdf, accessed 9 August 2013.

Rimdal TO and Soerensen PV (2012) Bilagstabeller: patienters oplevelser i Region Hovedstadens akutmodtagelser og akutklinikker, http://www.patientoplevelser.dk/akut2012/bilagstabel.pdf, accessed 5 August 2013.

Sahin B, Yilmaz F and Lee K (2007) Factors affecting inpatient satisfaction: structural equation modeling. *Journal of Medical Systems* 31(1), 9–16.

Schreiber JB (2008) Core reporting practices in structural equation modeling. *Research in Social and Administrative Pharmacy* 4(2), 83–97.

Shook CL, Ketchen Jr. DJ, Hult GT and Kacmar KM (2004) An assessment of the use of structural equation modeling in strategic management research. *Strategic Management Journal* 25(4), 397–404.

SMITH GD *et al* (2007) Diagnosis and satisfaction scores in emergency department patients who return a mailed survey. *Journal of Emergency Medicine* 32(2), 131–135.

SORRA J, KHANNA K, DYER N, MARDON R and FAMOLARO T (2012) Exploring relationships between patient safety culture and patient's assessment of hospital care. *Journal of Patient Safety* 8: 131–139.

TABACHNICK BG and FIDELL LS (2013) *Using Multivariate Statistics*, 6th edn, Allyn & Bacon, Boston.

TAYLOR C and BENGER JR (2004) Patient satisfaction in emergency medicine. *Emergency Medicine Journal* 21(5), 528–532.

THOMPSON DA, YARNOLD PR, WILLIAMS DR and ADAMS SL (1996) Effects of actual waiting time, information delivery, and expressive quality on patient satisfaction in the emergency department. *Annals of Emergency Medicine* 28(6), 657–665.

TOMA G, TRINER W and MCNUTT L (2009) Patient satisfaction as a function of emergency department previsit expectations. *Annals of Emergency Medicine* 54(3), 360–367, e6.

TROUT A, MAGNUSSON AR and HEDGES JR (2000) Patient satisfaction investigations and the emergency department: what does the literature say? *Academic Emergency Medicine* 7(6), 695–709.

VIDAL O *et al* (2008) Factors associated with patient satisfaction and care of over 65 year olds in emergency departments. *Cirurgía Española* 83(5), 260–265.

WELCH SJ (2010) Twenty years of patient satisfaction research applied to the emergency department: a qualitative review. *American Journal of Medical Quality* 25(1), 64–72.

WICKS AM and CHIN WW (2008) Measuring the three process segments of a customer's service experience for an out-patient surgery center. *International Journal of Health Care Quality Assurance* 21(1), 24–38.

World Health Organization. (2010) Monitoring the building blocks of health systems: a handbook of indicators and their measurement strategies, http://www.who.int/healthinfo/systems/WHO_MBHSS_2010_full_web.pdf, accessed 10 May 2013.

Appendix A1

Selected measures from the survey instrument were examined for an initial unrotated solution by calculating the 'Kaiser-Meyer-Olkin (KMO)' measure of sampling accuracy and 'Bartlett's Test of Sphericity' (Miller *et al*, 2005). These two tests advise the researcher whether or not factor analysis is an appropriate technique for categorising the underlying data set. The method of extraction used was the Principal Components Analysis with varimax rotation. A gross list of 14 items listed in the third column of Table 4 were analysed on a sample of 685 eligible patients. KMO equalled 0.767 that is in range of the acceptable boundaries {0.6; 1} (Tabachnick & Fidell, 2013). Bartlett's Test of Sphericity tests the null hypothesis that the correlation matrix is an identity matrix, thus rejection is wanted. This test displayed a satisfactory *P*-value ($P<0.001$), since this value must be significant at $\alpha=0.05$ (Nunnally & Bernstein, 1994). Determination of the optimal number of clusters can be estimated by Kaiser's criterion, which relies on eigenvalues>1 (Kaiser, 1958). Kaiser's criterion revealed a total of four clusters. It is important to acknowledge that Kaiser's criterion is criticised for its unilateral threshold value for the eigenvalues. This means that an item with eigenvalue 1.01 is included, whereas another item with eigenvalue 0.99 is excluded. Hence, the scree plot was consulted for confirmation purposes. Again, four clusters were found to be the optimum number of clusters.

Appendix A2

Table A2 Categorisation of actual waiting times

Minutes	*Corresponding category*
0	1
1–15	2
16–60	3
61–180	4
181–240	5
>241	6

PAPER 4

Health Care Management Science manuscript No.
(will be inserted by the editor)

Balancing patient flow and returning patients: a system dynamics study on emergency department crowding factors

Christian Michel Sørup · Daniel Sepulveda Estay · Peter Jacobsen · Philip Dean Anderson

Received: date / Accepted: date

Abstract Emergency department (ED) crowding has been an increasing problem worldwide. Previous research has identified factors that contribute to ED crowding. However, the relationships between these remain incompletely understood. This study's objective was to analyse the effects of initiating a local protocol to alleviate crowding situations at the expense of increasing returning patients through the development of a system dynamics (SD) simulation model. The SD study is from an academic care hospital in Boston, MA. Data sources include direct observations, semi-structured interviews, archival data from October 2013, and peer-reviewed literature from the domains of emergency medicine and management science. The SD model shows interrelations between inpatient capacity restraints and return visits due to potential premature discharges. The model reflects the vulnerability of the ED system when exposed to unpredicted increases in demand. Default trigger values for the protocol are tested to determine a balance between increased patient flows and the number of returning patients. Baseline simulation runs for generic variables assessment showed high leverage potential in bed assignment- and transfer times.

A thorough understanding of the complex non-linear behaviour of causes and effects of ED crowding is enabled through the use of SD. The vulnerability of the system lies in the crucial interaction between the physical constraints and the expedited patient flows through protocol activation. This study is an example of how hospital managers can benefit from virtual scenario testing within a safe simulation environment to immediately visualize the impacts of policy adjustments.

Christian Michel Sørup
Niels Koppels Allé 426D
Tel.: +45-45254768
E-mail: cmiso@dtu.dk

Peter Jacobsen
Niels Koppels Allé 426D

Keywords System dynamics · computer simulation · crowding · hospital discharge · emergency department

1 Introduction

Emergency department (ED) crowding has for decades been acknowledged as an international problem compromising health outcomes [16]. Hospitals have been cutting the number of inpatient beds for budget reasons while the demand for emergency care services is on the rise [4]. The American College of Emergency Physicians (ACEP) updated their position statement in February 2013 along with various policy recommendations aimed at mitigating crowding situations [1]. Despite the recommendations, crowded EDs were still a growing problem. In 2006, the Institute of Medicine published their report entitled *Hospital-Based Emergency Care at the Breaking Point* with a warning that the current system is not sustainable [9]. In response, The Joint Commission stated that every hospital needs a plan for dealing with crowding in their ED [7]. However, congested EDs remain an increasing problem in which causative factors remain to be fully elucidated [6]. While this paper refer to crowding in the ED, the authors do recognise that crowding is a symptom of a system-wide problem associated with potential deterioration of health outcomes for patients regardless department [24].

Simulation studies are increasingly acknowledged as a means to better understand the complex interconnections causing ED crowding [15]. Many of these are discrete event studies while system dynamics studies are rarer. Examples of such include discharging strategies [25] and the competition for inpatient beds between ED- and elective patients [12]. Because of the problem's dynamic complexity, time delays through non-linear behaviour and feedback mechanisms, ED crowding seems apt for a system dynamics analysis. The problem is a classic example of "policy resistance" when attempted interventions show poor results or even contradict intended effects. According to Sterman, the problem persists due to an imperfect perception of what constitutes the problem [21]. This study's focus is twofold. First, a local protocol ('Code Help') for responding to ED crowding at a major tertiary level one teaching hospital in Boston, MA is explored for its system-wide effect on patient flow. When predetermined criteria are met, the policy can be activated so that inpatient wards are bound to receive ED boarders by either reallocating current inpatients or discharge faster than normally. An ED boarder is defined as a patient who remains in the ED after having been admitted to an inpatient ward but has not yet been transferred there [1]. Congestion can dissolve but comes at the expense of a higher risk of premature discharges [19, 5, 10]. Second, the model considers the effects of 1) inpatient capacity [23], 2) bed assignment time [18], 3) transfer time [2], and 4) number of incoming elective patients [13] as potential levers for ED crowding.

2 Method and Materials

2.1 Setting

The system dynamics model shows the admission process at the case hospital. Most patients go through the ED in order to get a preliminary diagnosis and are either fully treated in the ED or, if further treatment is required, admitted to an inpatient ward. 35 % of the ED patients are admitted. Occasionally, the patient flow is impaired partly due to ED boarding. This paper uses the notion of ED patients meaning essentially all acute patients. This is true for the most hospitals today but some hospitals may still receive acute patients directly at inpatient wards.

2.2 Model construction

Three main elements constitute the model. These are 1) the ED admission process, 2) two feedback mechanisms controlling the available bed capacity and the effects of initiating the protocol, and 3) the competing requirement for inpatient beds by elective patients.

For the admission process, all patients that enter the ED undergo an initial ESI triage assessment [20]. Patients then receive medical evaluation and a preliminary diagnosis is established.

In the model, treatment is either finished in the ED or specialized treatment is needed elsewhere in the hospital. Patients requiring admission need to have an inpatient bed assigned. At the case study hospital, this happens via an electronic request initiated from the ED where the receiving ward can choose either to accept the patient or request further discussion with the ED team. This solution was implemented as part of a Lean initiative to avoid cumbersome telephonic consultation. Next, patient admission is restricted by the bed occupancy rate to avoid assigning more beds than available. In case of ED crowding, the protocol can be activated if three criteria are met:

1. ED boarders are 10 or more patients,
2. ED census is 54 or more patients, and
3. Criteria 1. and 2. are met for two consecutive hours.

Patients are either relocated or discharged faster from the inpatient wards, thereby increasing capacity for incoming patients. Hence, the patient flow is increased and congestion eased. The expense is an increased risk of discharging patients prematurely which could result in more patients returning unexpectedly [19, 5, 10].The model was built using daily historical data from October 2013 at the study site. Inputs to the model include day of the week averages for patient arrivals (shown in Figure 1), ED census registered every ten minutes, ED patients that await inpatient bed assignment, and ED boarders. Vensim DSS version 6.3 was utilized to construct the model. Complete model documentation can be acquired with the authors. The model is simulated for a one week period.

2.3 Model validation

If the model is deemed sufficiently valid then the stakeholders are more likely to be assured that the system is replicated to satisfaction. Tenured physicians were iteratively consulted to qualitatively agree on the correctness of the process, feedback mechanisms, and assumptions modelled. Quantitative testing of the model was done through comparison of model output to historical data of ED census from October 2013. As advocated by Sterman, replication of historical data does not provide sufficient evidence for a robust model [22]. Therefore, the developed SD model was subject to a number of tests to ensure feasible behaviour under extreme conditions.

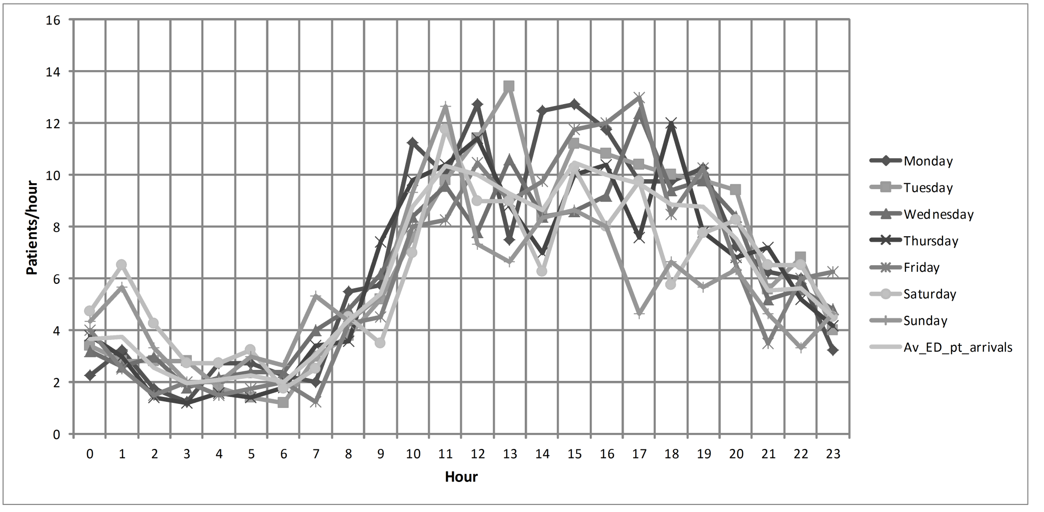


Fig. 1 Hourly ED patient arrival pattern following Lane et al., 1996

2.4 Sensitivity testing

After model validation, several simulations were run with various parameter settings. Adjusting the trigger threshold values to see the effects on ED crowding level and returning patients was our primary interest. In order that the protocol would become activated temporarily, the baseline model was stressed through a sudden surge of patients arriving within three hours. The admission fraction had to be increased in the period of the surge after a small delay corresponding to patients having undergone treatment. Trigger values were changed separately to determine the subsequent effects on the outcome variables. First, the ED census trigger was allowed to vary from 40 to 68 (default = 54). Next, the ED boarder trigger would vary from 7 to 13 (default = 10). The ED census trigger criterion is met on a regular basis even in the baseline simulation. Based on the semi-structured interviews conducted with staff physicians, the main indicator of ED crowding was the accumulation of ED boarders. It was expected that altering this trigger would have the largest effect on the system. The simulation model developed offer ED decision makers insights into other potential policy changes. The parameters analysed are inpatient capacity, bed assignment time, transfer time, and number of incoming elective patients.

3 Results

3.1 Model structure

A simplified presentation of the main system dynamics model is shown in Figure 2.

The main flow presents the ED patient journey with special emphasis on the admitted patients. The model includes two sub views (not shown). The first sub view is an exogenous mechanism that can generate an inflow of patients in a user specified pulse, step, ramp, or sine wave function. This inflow is directly connected to the patient arrivals inflow. The second sub view introduces a memory function representing the third criteria of the protocol, being thresholds for both ED census- and boarders respectively are met for two consecutive hours.

The balancing feedback loop restricts the amount of patients to be admitted. The auxiliary variable 'occupancy rate' returns a value of zero to one according to how many patients are already admitted from the ED and elsewhere in the hospital plus the ED patients that have an inpatient bed assigned. With a higher occupancy rate it will become increasingly difficult to admit new ED patients.

The reinforcing feedback loop explicitly models the return flow of admitted patients that have been discharged. Here, a fraction of discharged patients need repeated treatment upon discharge within 72 hours. This fraction increases when the protocol is activated because this implies a discharging of patients faster than normal. The upper right corner in Figure 2 models the

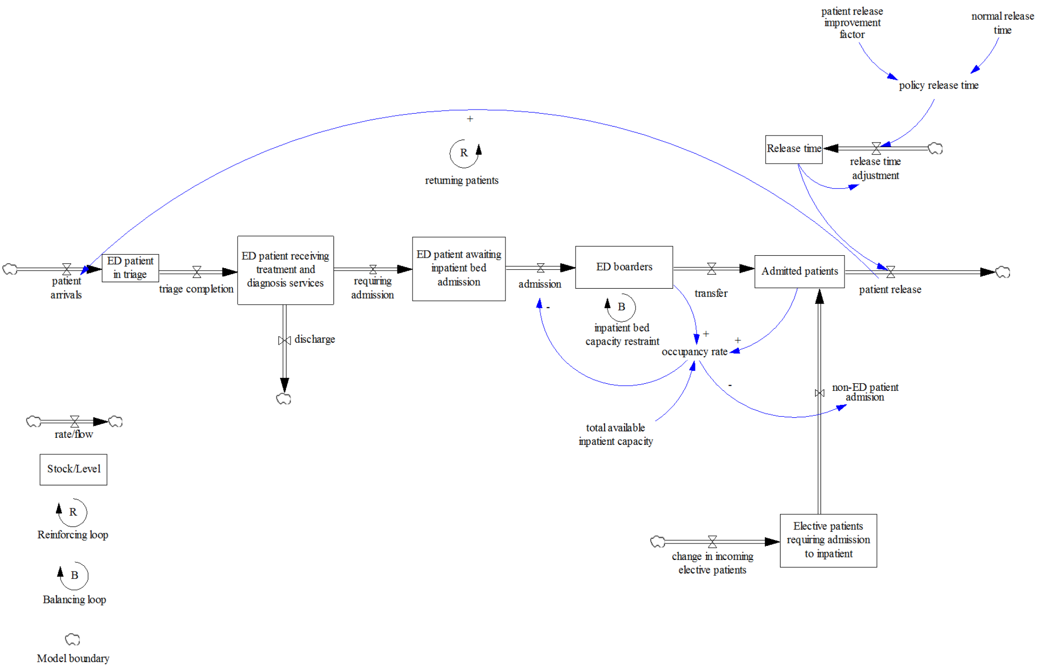


Fig. 2 Reduced presentation of the system dynamics model with auxiliary variables omitted. Patients arrive to the ED based on an average repeated diurnal pattern and stay in the ED until either discharged or admitted. Patients destined for admittance first need an inpatient bed assigned, becoming ED boarders when the inpatient bed is assigned and are then transferred out of the ED. The feedback mechanisms, denoted R and B respectively, and the inflow of elective patients are also presented.

release time, equivalent to a patient's total time as inpatient. Activation of the protocol impacts the policy release time and the protocol's effect will show after a 30 minutes delay. The before mentioned fraction of patients returning to the ED is based on estimates found in literature [19, 5].

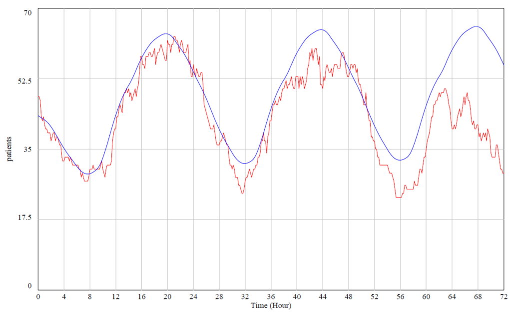


Fig. 3 Comparison of ED census data and simulation output for first 72 hours (October 1st to 3rd, 2013). The smooth graph shows the simulation output while the jagged graph shows imported historical data.

3.2 Model validation

Quantitative validation was done by comparing the simulation output for ED census to the acquired historical data reflecting the actual ED census throughout October 2013. Generally, there is a good fit between the two graphs (see Figure 3). Extreme tests were performed as to determine system behaviour when stressed. Confidence in the model's robustness was established between authors and physicians involved in the development of the model.

3.3 Sensitivity testing

As the model can replicate actual behaviour adequately, the study proceeds to testing how the ED census- and ED boarders trigger values can change the model's behaviour. Such changes correspond to an earlier or later activation of the protocol. Effects of parameter changes are measured on the occupancy rate, admitted patients, ED census, and ED boarders. Testing the settings for each trigger is conducted in isolation keeping all other parameters constant. The baseline model investigates other relevant generic parameters that are potential levers for alleviating ED congestion. All tests and their effects are summarised in Tables 1 and 2 [17].

Two scenarios are used to compare the effects of changing parameter values. These are the baseline model, i.e. 'business as usual' corresponding to what was experienced in October 2013 and the stressed model, where an exogenous surge of patients trigger the protocol temporarily. The model shows some sensitivity towards a change in the trigger value for ED boarders. For instance, changing the threshold from a default value of 10 to 7 yields a 5.69 % increase in the number of ED boarders, due to the number of returning patients while occupancy is lowered by 2.54 % (see Figure 4). Varying the ED census trigger had no effect on the system.

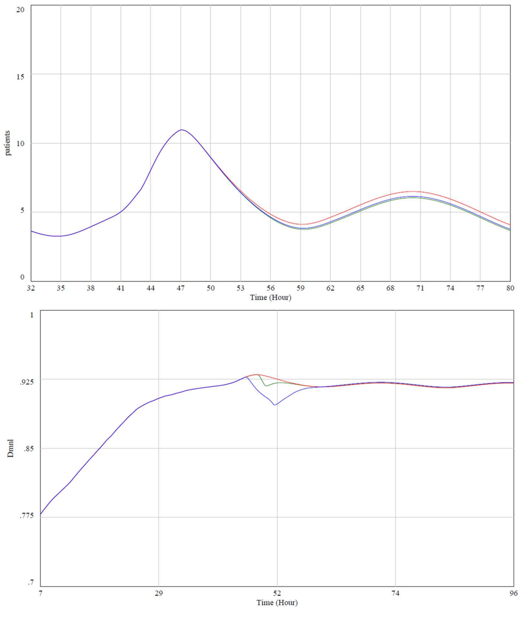


Fig. 4 The effect of changing the trigger threshold for ED boarders, measured on total number of ED boarders (upper graph) and occupancy rate (lower graph).

Switching the attention to the baseline model, four parameters were subject to analysis. These were the total inpatient capacity, bed assignment time, transfer time, and mean incoming elective patients. In Figure 5, the largest effect can be seen when changing the transfer time (53.28 % and 44.79 % respectively).

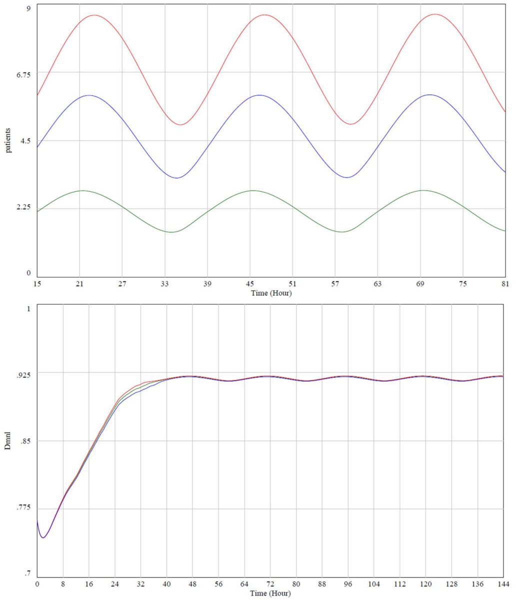


Fig. 5 The effect of changing transfer times on ED boarders (upper graph) and occupancy rate (lower graph). Lowering the transfer time both decrease the number of ED boarders and the occupancy rate. Similarly, increasing the transfer time yields approximately the same effect in opposite direction.

Adjusting the inpatient capacity showed very little impact on the ED census (between 0.18 % and 1.06 %). The results obtained on both transfer time and inpatient capacity are consistent with recently published studies applying other simulation techniques [11, 8]. Being able to assign an inpatient bed faster shows some leverage potential, implying enhanced communication between the receiving inpatient ward and the ED. Lastly, changing the number of incoming elective patients had little impact on the ED census- and boarders, while occupancy rates and the number of admitted patients were impacted.

4 Discussion

This study presents a system dynamics model capable of analysing the effects of a local protocol to alleviate ED crowding alongside other more generic parameters that have been identified in the literature as relevant

Table 1 Stressed model sensitivity test results for protocol trigger values. The percentages of deviation to the default value are in parentheses. The values have been extracted to the time where the three graphs (base, maximum, and minimum) show the biggest variance. Most leverage potential is found in the largest percentages deviation. Note no effect on varied ED census trigger due to the threshold being frequently exceeded

	trigger, ED boarders			trigger, ED census		
	Base	7	13	Base	48	60
Occupancy	0.9205	0.8971 (2.54)	0.926 (0.59)			
time, max diff	51					
ED census	61.02	64.34 (5.47)	60.06 (1.57)			
time, max diff	67.5					
ED boarders	6.153	6.503 (5.69)	6.052 (1.64)			
time, max diff	70.125					
Admitted patients	453.7	442 (2.58)	456.6 (0.64)			
time, max diff	51					

Table 2 Baseline sensitivity test results for selected parameters associated with ED crowding

	tot avail inpatient beds			bed assign time			transfer time			mean elective patients		
	Base	400	900	Base	1.8	4	Base	0.5	2.5	Base	100	200
Occupancy	0.9167	0.9185 (0.20)	0.7891 (13.92)	0.833	0.89 (0.79)	0.865 (2.04)	0.909	0.905 (0.43)	0.912 (0.37)	0.836	0.689 (17.59)	0.917 (9.75)
time, max diff	56			24			32			16		
ED census	59.96	59.35 (1.02)	60.07 (0.18)	59.53	56.72 (4.72)	62.03 (4.20)	59.53	56.65 (4.84)	61.82 (3.85)	59.96	59.14 (1.37)	59.98 (0.03)
time, max diff	139.5			19.625			19.5			139.5		
ED boarders	6.031	5.967 (1.06)	6.043 (0.20)	5.869	5.403 (7.94)	6.238 (6.29)	5.959	2.784 (53.28)	8.628 (44.79)	6.031	5.945 (1.42)	6.033 (0.03)
time, max diff	142.5			24			23.125			142.25		
Admit patients	455.6	364.7 (19.95)	479.3 (5.20)	438.6	440.3 (0.39)	437 (0.36)	455.7	458.6 (0.64)	453.2 (0.55)	414.7	340.8 (17.82)	455 (9.72)
time, max diff	100			24			48			16		

levers for crowding. Three main findings can be extracted from the study.

First, the developed system dynamics was validated both qualitatively, through iterative consensus-seeking consultations with emergency physicians on the process being modelled, and quantitatively, through satisfactory replication of census data.

Next, the developed model was stressed so that the protocol was activated in order to analyse the effects of changing the trigger values for both ED census- and boarders. Having an ED census beyond the threshold limit is no rare incidence at the case hospital. The ED is routinely busy during the afternoon where patient arrivals have been high for several hours. Patients typically arrive faster than those either discharged directly or leaving for an inpatient ward. However, the congestion is time limited since the patient arrivals will decrease, giving the ED a chance to recover. Therefore, the ED boarder trigger is the parameter that can impair the patient flow. If this threshold is lowered to 7 from 10, ED census increases 5.47 % and ED boarders increase 5.69 % while the occupancy rate and number of admitted patients decrease by approximately 2.5 %. A higher ED census is explained by the increased amount of returning patients due to premature discharge from an inpatient ward.

Finally, simulation of the baseline model yielded a number of other relevant findings. Fast transportation of ED boarders is identified as the main lever for clearing ED congestion [20].

The current trends towards decreasing available inpatient beds should be evaluated in light of these effects. Bed assignment time was also revealed to be a potential focus area for improving patient flow. When decreased to an average of 1 hour and 48 minutes (from a base of 2 hours and 55 minutes), ED census is decreased by 4.72 % while the occupancy rate is elevated by 2.89 %. Inpatient bed capacity did not provide great leverage while changing the number of elective patients primarily had an influence on admitted patients because of idle capacity.

The SD model can be extended in several ways. A benefit of system dynamics is its capability of modelling human behaviour. Social aspects of care and human factors can therefore be included, for instance the performance effects of coping with increased workloads [14]. Some of the constants included in the model can be altered into dependent variables with further analyses. One such example would be the relation between faster discharging times on the probability of increased ED patient returns. Additionally, the ED patients that do return for more treatment may be at a higher risk of getting admitted once more. Another possibility would be to focus on the transition between departments by explicitly modelling the receiving inpatient ward. Adopting a broader system focus would be an important first step in finding sustainable efficiency improvements for the hospital as compared to looking at the ED in isolation [3].

5 Limitations

The simulation model developed is only an approximation of reality as perceived by a limited number of ED physicians' opinions and is based on a limited data sample. Reality is much more complex than the model presented explaining the difference in displayed behaviour shown in Figure 3. Patients are treated as one large quantity despite being categorized according to their respective levels of acuity. Such differentiation could in the model be obtained by the introduction of subscripts.

The SD model introduces an even inflow of elective patients around the clock which represents an idealized scenario desirable from an ED perspective. The reality is that most hospitals schedule elective admissions in batches typically in the mornings on Monday to Thursday. Elective patients getting admitted after surgery are in need of inpatient beds at the same time that the daily wave of patients admitted from the ED are looking for beds. Furthermore, the hospital may have policies as to admit either ED patients or elective patients in favour of the other in case of restricted inpatient capacity; something the model does not account for.

The results generated from the developed SD model are dependent on the parameter representing the likelihood of a prematurely discharged inpatient returning to the ED. This value was estimated based on current literature but may in reality be a variable that depend on how early the patient is discharged. Since data stems from a single hospital, the model resembles one particular system. Generalizability potential is thus reduced. However, there are conclusions that may be of interest to other hospitals with similar admission processes. The model's validity would be improved if similar data sets were acquired at other emergency care hospitals. Finally, any interventional pilot studies that may follow the use of the proposed model would have to be assessed and compared to that of the model's output.

6 Conclusion

The main objective of this study was to investigate the effects of altering the trigger thresholds for a local protocol used for alleviating crowding situations at a large tertiary care center and teaching hospital in Boston, MA. A side objective was to change the focus to more general parameters hypothesized as alternative levers for ED crowding control. Through the development of a system dynamics model, it was shown that, for the protocol, changing the ED census trigger had no effect on patient flow while the ED boarder trigger could speed up the patient flow when lowered on the expense of a slight increase in ED census. More generally, faster bed assignment- and transfer times had a positive effect on the patient flow and are thus a recommended target area for potential process improvements. This study underlines the capabilities of simulation as a feasible option in the search of sophisticated management decision support tools for complex dynamic systems.

References

1. American College of Emergency Physicians (2013) Crowding. URL http://www.acep.org/Clinical—Practice-Management/Crowding/
2. Beniuk K, Boyle AA, Clarkson PJ (2012) Emergency department crowding: prioritising quantified crowding measures using a Delphi study. Emergency Medicine Journal 29(11):868–871
3. Brailsford SC, Lattimer Va, Tarnaras P, Turnbull JC (2004) Emergency and on-demand health care: modelling a large complex system. Journal of the Operational Research Society 55(1):34–42
4. Cameron P (2006) Hospital overcrowding: a threat to patient safety? The Medical Journal of Australia 184(5):203–4
5. Crawford Ea, Parikh PJ, Kong N, Thakar CV (2014) Analyzing discharge strategies during acute care: a discrete-event simulation study. Medical decision making : an international journal of the Society for Medical Decision Making 34(2):231–41
6. Derlet RW, Richards JR (2000) Overcrowding in the Nation' s Emergency Departments : Complex Causes and Disturbing Effects. Annals of Emergency Medicine 35(January):63–68
7. Fee C, Weber EJ, Maak CA, Bacchetti P (2007) Effect of emergency department crowding on pneumonia admission care components. Annals of Emergency Medicine 50(5):501–509
8. Hurwitz JE, Lee JA, Lopiano KK, Mckinley SA, Keesling J (2014) A flexible simulation platform to quantify and manage emergency department crowding. BMC Medical Informatics and Decision Making 14(50):1–11
9. Institute of Medicine (2006) Hospital-Based Emergency Care At the Breaking Point. The National Academies Press, Washington DC

10. Jack BW, Chetty VK, Anthony D, Greenwald JL, Sanchez GM, Johnson AE, Forsythe SR, O'Donnell JK, Paasche-Orlow MK, Manasseh C, Martin S, Culpepper L (2009) A Reengineered Hospital Discharge Program to Decrease Rehospitalization. Annals of Internal Medicine 150(3):178–187
11. Khare RK, Powell ES, Reinhardt G, Lucenti M (2009) Adding more beds to the emergency department or reducing admitted patient boarding times: which has a more significant influence on emergency department congestion? Annals of emergency medicine 53(5):575–85
12. Lane DC, Monefeldt C, Rosenhead JV (1996) Looking in the wrong place for healthcare improvements: A system dynamics study of an accident and emergency department. Journal of Operational Research Society 51(5):518–531
13. Morris ZS, Boyle A, Beniuk K, Robinson S (2012) Emergency department crowding: towards an agenda for evidence-based intervention. Emergency Medicine Journal 29(6):460–466
14. Morrison JB, Wears RL (2011) Emergency Department Crowding : Vicious Cycles in the ED. In: 29th International Conference of the System Dynamics Society, June, pp 1–36
15. Paul Sa, Reddy MC, DeFlitch CJ (2010) A Systematic Review of Simulation Studies Investigating Emergency Department Overcrowding. Simulation 86(March):559–571
16. Pines JM, Asplin BR (2011) Conference Proceeding: Improving the quality and efficiency of emergency care across the continuum: a systems approach. Academic Emergency Medicine 18(6):655–61
17. Rahmandad H, Sterman JD (2012) Reporting guidelines for simulation-based research in social sciences. System Dynamics Review 28(4):396–411
18. Rechel B, Wright S, Barlow J, McKee M (2010) Hospital capacity planning: from measuring stocks to modelling flows. Bulletin of the World Health Organization 88(8):632–6
19. Rising KL, White LF, Fernandez WG, Boutwell AE (2013) Emergency department visits after hospital discharge: a missing part of the equation. Annals of Emergency Medicine 62(2):145–50
20. Shelton R (2010) ESI: A better triage system? Nursing Critical Care 5(6):34–37
21. Sterman JD (2000) Business Dynamics - Systems Thinking and Modeling for a Complex World, 1st edn. Irwin/McGraw-Hill, Boston, MA
22. Sterman JD (2002) All models are wrong: reflections on becoming a systems scientist. System Dynamics Review 18(4):501–531
23. Trzeciak S, Rivers EP (2003) Emergency department overcrowding in the United States: an emerging threat to patient safety and public health. Emergency Medicine Journal 20(5):402–405
24. Viccellio P, Zito Ja, Sayage V, Chohan J, Garra G, Santora C, Singer AJ (2013) Patients overwhelmingly prefer inpatient boarding to emergency department boarding. Journal of Emergency Medicine 45(6):942–946
25. Wong HJ, Wu RC, Caesar M, Abrams H, Morra D (2010) Smoothing inpatient discharges decreases emergency department congestion: a system dynamics simulation model. Emergency medicine journal : EMJ 27(8):593–8

The goal of this PhD project was to develop a performance measurement model for EDs. The new model comprises only the most important performance measures that provide an estimate for overall ED performance levels. Furthermore, a thorough analysis of the interdependencies between the included performance measures was conducted in order to gain deeper knowledge of the ED as a system. The model enables monitoring of how well the ED performs over time, including how performance is impacted by the various initiatives. In the end, the developed model will be an important management tool to meet the management's vision of providing the best possible care for the patient meanwhile achieving the highest possible utilisation of resources.

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